International Journal of Latest Trends in Computing
E-ISSN: 2045-5364

Volume 2, Issue 1, March 2011

IJLTC Board Members

TABLE OF CONTENTS

University Students' Usage of e-Student Services

Ahmet Naci Coklar[1] and Semseddin Gunduz[2]

[1,2] Selcuk University, Faculty of Education, Computer and Instructional Technologies Department
42090 Konya,Turkey

Corresponding Adresses
ahmetcoklar@selcuk.edu.tr, sgunduz@selcuk.edu.tr

***Abstract*:** The aim of this study is to find out the frequency students-as individuals in information age use the internet. With this aim in mind, a data collection device based on e-student services provided by Turkey Information Council was developed. The measurement device developed was applied to 390 students from different departments at Selçuk University in 2009-2010 education year. As a result of the study, it was found out that students used social and education related e-student services frequently and used health and economy related e-students services sometimes. It was also found out that male students used e-student services more than female students and students who stayed with their families used e-students services more than those who stayed in a dormitory. It was also concluded that the internet usage frequency has effect on students' use of e-students services.

***Keywords*:** e-Students, the internet usage, university student**.**

1. Introduction

The internet and internet based technologies make their effects felt more and more in daily life. Especially in recent years, with ADSL technology becoming widespread, the internet which entered many places like home, workplace and school has changed life styles, as well. According to data from TurkStat, while the internet which was started to be used in Turkey, which has a population of over 70 million, in the 1990s was being used by 19.7% of the population in 2007 t, this rate raised up to 34.0% as of the end of 2009 [1]. In a period of two years, increase and spread of internet usage was remarkable. When new technologies like 3G -which makes wireless internet connection possible -are considered, it can be said that the internet will affect many more lives.

1.1 The Services Offered On the Internet

The internet is the most important global network and information source in information age. When the operations on the internet are thought, the first things that come to the mind are communication (mail and chat), information source (search engines) and entertainment (video and music). In addition to these services, there are many services of the internet used in daily life. Eastman [2] classified the services the internet offers as general services which range from communication services, information source services, shopping services, group services, online banking, job application to hotel reservation. The data provided by TurkStat [3] about the use and usage purposes of the internet in Turkey are seen in Table 1.

Table 1. Results of the Internet Usage in Households and by individuals.

Purposes	**Rate of Population (%)**
Sending / receiving e-mails	72.4
Reading/downloading online newspapers/news magazines	70.0
Posting messages to chat sites, newsgroups or on-line discussion forum	57.8
Playing or downloading games, images or music or film	56.3
Finding information about goods or services	52.9
Telephoning over the Internet	49.8
Seeking health-related information (e.g. injury, disease, nutrition, improving health, etc.)	45.1
Listening to web radios/watching web television	43.3
Consulting the Internet with the purpose of learning	31.7
Looking for information about education, training or course offers	25.7
Uploading self-created content (text, images, photos, videos, music etc.) to any website to be shared	24.1
Using services related to travel and accommodation	22.5
Downloading software (other than games software)	15.2
Internet Banking	14.0
Looking for a job or sending a job application	13.2
Using peer-to-peer file sharing for exchanging movies, music, video files	5.8
Selling goods or services (e.g. via auctions)	2.0

As it can be seen in Table 1, the most common purpose for using the internet are the operations like e-mail services, reading newspapers and magazines playing games, downloading films or music, searching information. On the other hand, it can be said that services like internet banking, job hunting, buying and selling operations are also used, to a little extent tough.

1.2. E-student Services Offered on the Internet

The internet offers services appropriate for people of all ages and occupation. People from all walks of life, employees, public workers, workers, civil servants can use the internet

for their own purposes. However, student stands out as individuals who made use of the internet most. According to TurkStat [4] as of 2009, in terms of using the internet students with their 88.2% are on the top, which is followed by employers with 66.1% and by regular employee and casual employee with 56.8%.

Students as the individual group who use the internet most make use of a wide range of services from sharing information to health services. In the report of e-society study by Turkey Information Council (TIC) in 2004 carried out in Middle East Technical University, internet based services students, employers and retired people can use were determined [5]. According to the report, e-student services mean that daily life activities being done more easily and practically on the internet and are composed of four sub-dimensions. In Table 2, the sub-dimensions which make up e-student services and the comparison of the operations in these sub-dimensions with their traditional counterparts are seen.

Table 2. e-Student Applications with their Sub-dimensions and the Comparison of e-Student Applications with Traditional Methods

e-Student Sub-dimensions	e-Student What and How	Traditional Student What and How
SOCIAL		
Communication with Family	e-mail, video conference, messenger etc. Following events via the internet devices	In city/ abroad communication via mail , telephone
Cultural Relations	, e-Ticket	By going or watching the media
Communication with Friends	e-Mail, Chat , video conference etc.	Face to face or on the telephone communication
HEALTH		
Communication with the doctor, following diseases and getting information	e-mail, msn etc.	Face to face or on the telephone communication
Taking Appointment/ examination results	Via search engines and sites and entering the site or e-mail	Restricted Search via Book or acquainted doctor on telephone or by going
EDUCATION		
Individual Development	e-Education	By going, searching
School relations	Registration, following courses, notes/homework/exam dates,	by going
Communication with teachers	msn, e-mail, video conference etc.	On the Telephone and face to face
Searching for education opportunities abroad	Web and e-application	Very difficult ways, limited facilities
ECONOMY		
Banking operations	Via Web, e-banking	By going
Searching for Job opportunities and application	Via Web, e-mail	Newspaper/media opportunities, by going
Bill payments	e-banking	by going
Shopping	e-banking, e-trade sites	

e-Student applications are composed of four different sub-dimensions -that is operational groups- as social, health, education and economy related operations (Table 2). Students' communication with families and friends via e-mail, video conference, using the internet opportunities for online education, meeting their need for library by accessing to sources from home, following cultural events, performing banking operations like paying bills, getting appointments and analysis results from health institutions are among e-Student applications.

1.3. The Significance of the Study

The internet has witnessed an important development in the last twenty years and life style has been shaped again. Changing life style has introduced new concepts like e-individual, e-society, e-citizen and e-Student. In the basis of e-Student services lies the philosophy that students are to be able to do their daily activities easily on the internet. In this study, the extent to which students use e-Student services in education, health, economy and social sub-dimensions is examined. The study is significant in that it emphasizes the effect of the internet on daily life and propounds students' internet usage proficiencies and the states of e-Student services offered on the internet. Besides, it is thought that the study is significant in that it will show direction to the development of e-Student services.

1.4. The Aim of the Study

The aim of this study is to determine the usage of e-Student services offered on the internet by students. Besides, it attempts to reveal the relation between students' e-Student services usage states and individual characteristics. With this aim in mind, answers for the following questions were sought:

1. What are the Students states in terms of using e-Student services?
2. What are the students states in terms of using the e-Student services in
   a. Social
   b. Health
   c. Education
   d. Economy
   sub-dimensions?
3. Does the extend students use e-Student services vary significantly in terms of
   a. Gender
   b. The place they stay
   c. The frequency of using the internet
   variables?

2. Method

In this section, there are information about the model of the study, population and sample, means of data collection and analysis of data.

2.1 Study Model

In this study in which the relation between students' status of using e-Student services and individual characteristics is sought to be determined, single and relational survey models were used. In the study, students' usage of e-Student services and personal information was described with single survey model. By using relational survey method, the relation between students' status of using e-Student services and gender, residence place and the frequency of using the internet were examined.

2.2 Population and Sample

The population of the study is composed of the students attending Faculty of Education at Selcuk University and taking "Computer I" course. As the population is very large, sampling method was used and simple two stage sampling method was applied. In simple two-stage sampling method, the aim is to have sub-sample of the same size from each one of the clusters which have been chosen with simple random method in the first stage of selection process [6]. In the first stage, Primary Education Department was chosen by lot, in the second stage, three classes randomly selected from each of Science, Class, Social Sciences and Primary Mathematic Teaching Department were included in the sample. It was learnt from Registrar Office of Education Faculty at Selcuk University that the number of the students in these classes to be 478. The measurement device was applied to 407 students and valid data from 390 students was included in the study. Demographic information about participants is given in Table 3.

Table 3. Demographic Information about the Participants

		f	*%*
Gender	Male	173	44.4
	Female	217	55.6
	Total	390	100
Department	Primary school	87	22.3
	Mathematics		
	Science	76	19.5
	Class	119	30.5
	Social Sciences	108	22.7
	Total	390	100
Frequency of Internet Usage	Everyday	106	27.2
	Once in 2-3 days	117	30.0
	Once a week	55	14.1
	Less Frequently	112	28.7
	Total	390	100
Residence	Family	96	24.6
	Dormitory	117	30.0
	Friends	177	45.4
	Total	390	100

When Table 3 is examined, it is seen that participants are composed of 390 students, 173 (44,4%) of whom are male, 217 (55,6%) of whom are female. A hundred and six of the students (27,2%) use the internet everyday, 117 (30,0%) use it once in 2-3 days, 55 (14,1%) use it once a week, and the rest 112 (28,7%) use it less frequently. Ninety-six of the students (24,6%) stay with their family, 117 (30,0%) at dormitory and the rest 177 (45,4%) stay with their flat-mates.

2.3. Data Collection Devices

Data collection device is composed of two parts as individual information form and e-Student survey:

Personal Information Form: It is a form which includes question items about gender, department, residency, internet usage frequency of the students in the study group that will help us know them and that can be related with their usage of e-Student services.

E-Student Survey: A measurement device of 20 items was developed by researchers to determine students' state of using e-Student services. In the measurement device, students were asked their frequency of use of e-Student service for each item. While the survey was being prepared, e-Student result report by Turkey Information Council (TIC) was used as base. In the mentioned study, e-Student was examined in 4 dimensions as social, health, education and economy. In the social dimension there are 5, in health dimension there are 3, in education dimension there are 7 and in economy dimension there are 5 items. E-Student survey was prepared by the researchers based on these items. In order to ensure content validity of the survey, expert views were used. As a result of feedback from 3 experts in the field from Computer and Instructional Technologies Education Department, 2 items in the survey was re-written. In order to determine the reliability of the survey, pilot application was administrated

to 30 students. As a result of application, cronbach alpha of inner consistency coefficient value was calculated to be .81.

2.4. Analysis of Data

In transferring the items in data collection device on the computer, the responds in 5-item likert scale were scored as "I never Use, 1", "I use it rarely, 2", "I use it sometimes, 3", " I use it frequently, 4" and "I use it very frequently, 5". To assess arithmetic mean interval, 5 columns and 4 interval logic was used. The value of the interval 4/5= 0.8. Accordingly, the frequency of usage was assessed according to Table 4.

In order to determine students' states in terms of usage of e-Student services, descriptive statistic data; namely, percentage, frequency, arithmetic mean and standard deviation was used. To determine whether the usage frequency of e-Student services change in terms of gender, Independent Samples t Test was used. In order to determine whether the frequency of the use of e-Student services vary according to where they live One Way ANOVA test was used.

Table 4. Assessment Criteria for e-Student Services

Data Interval	Usage Frequency Assessment Criteria
1.00-1.80	Never
1.81-2.60	Rarely
2.61-3.40	Sometimes
3.41-4.20	Frequently
4.21-5.00	Very Frequently

In all analysis carried out, the level of significance was accepted as .05. In statistical analysis of the study, SPSS 15.0 (Statistical Package for the Social Sciences) package software was used.

3. Findings

In this section, findings with regard to students' state of using e-Student services and whether their states of usage vary according different variables are given under different headings.

3.1. Students States of Using e-Student Services

In line with the aim of the study, how often students use e-Student services were analyzed and the findings are given in Table 5.

Table 5. The frequency students use e-Student Services

E-Student Services	*Never*		*Rarely*		*Sometimes*		*Freq.*		*Very Freq.*		*MEAN*	
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*	*f*	*%*	$\overline{X}$	*ss*
1 Communication with people (friends family etc.)	38	9.7	43	11.0	78	20.0	106	27.2	125	32.1	3.61	1.30
2. Communication with people by written chats (msn, Skype, yahoo Msn)	26	6.7	35	9.0	48	12.3	106	27.2	175	44.9	3.95	1.24
3. Chat, spoken and visual communication with people chat (msn, Skype, etc…)	33	8.5	46	11.8	88	22.6	111	28.5	112	28.7	3.57	1.25
4.Social networks to communicate with people (facebook, twitter, blog etc)	33	8.5	67	17.2	48	12.3	73	18.7	169	43.3	3.71	1.39
5Use the internet to follow cultural events (cinema, concert, sport etc.)	34	8.7	43	11.0	81	20.8	107	27.4	125	32.1	3.63	1.27
Mean of social dimension											***3.69***	***1.04***
6.Use the internet to get appointment from health institutions	69	17.7	171	43.8	74	19.0	46	11.8	30	7.7	2.48	1.14
7.Use the internet to get information about health institutions (features, contact details etc.)	43	11.0	92	23.6	118	30.3	73	18.7	64	16.4	3.06	1.23
8.Use the internet to do search about health problems (disease)	27	6.9	47	12.1	101	25.9	113	29.0	102	26.2	3.55	1.20
Mean of health dimension											***3.03***	***0.98***
9.Use the internet to register to a course	30	7.7	32	8.2	53	13.6	95	24.4	180	46.2	3.93	1.27
10.USE the internet to receive course on the web	57	14.6	117	30.0	81	20.8	60	15.4	75	19.2	2.95	1.35
11.Use the internet to do research and homework	7	1.8	3	0.8	22	5.6	103	26.4	255	65.4	4.53	.79
12. Use the internet to follow announcements about course (date of exams, exam results, homework announcement)	12	3.1	10	2.6	38	9.7	74	19.0	256	65.6	4.42	.98

13.Use the internet to get in communication with teacher	58	14.9	162	41.5	76	19.5	57	14.6	37	9.5	2.62	1.18
14.USE the internet to search supplementary education opportunities (courses, scientific events etc)	42	10.8	70	17.9	87	22.3	103	26.4	88	22.4	3.32	1.29
15. Use the internet to use library services (to search and access sources)	16	4.1	42	10.8	73	18.7	103	26.4	156	40.0	3.87	1.17
Mean of education dimension											***3.66***	***0.74***
16. USE the internet to search financial support (scholarship etc.)	31	7.9	69	17.7	87	22.3	107	27.4	96	24.6	3.43	1.25
17.Use the internet for banking (sending money paying bills etc.)	108	27.7	196	50.3	37	9.5	27	6.9	22	5.6	2.13	1.06
18.use the internet to hunt job	64	16.4	141	36.2	72	18.5	53	13.6	60	15.4	2.75	1.30
19.Use the internet for price search	51	13.1	95	24.4	79	20.3	76	19.5	89	22.8	3.15	1.36
20.use the internet for shopping	99	25.4	197	50.5	43	11.0	33	8.5	18	4.6	2.16	1.04
Mean of education dimension											***2.72***	***0.86***
E-STUDENT SERVICES GENERAL MEAN											***3.28***	***0.73***

When Table 5 is examined, it is seen that the mean of the frequency students make use of e-Student services is =3.28. According to this result, it can be stated that students use e-Student services sometimes. When they were examined in terms of sub-dimensions, it was found out that students use e-Student services "frequently" for social purposes (=3.69) and for education purposes (=3.66) most and for health-related purpose (=3.03) and for economic purposes (=2.72) "sometimes".

When the items in the social dimension are examined, it is seen that students use e-student services in the items "frequently". In social dimension of e-Student services, students use chat programs most frequently ($\overline{X}$ =3.95), which allows simultaneous written communication with people.

As for the items in health dimension, it is remarkable that the frequency the items are used vary. While it is stated that the internet is used "frequently" ($\overline{X}$ =3.55) to search for information about health problems, it is stated that they use the internet to get information about health institutions "sometimes" ($\overline{X}$ =3.06). As for getting appointment which is one of the e-Student services, students stated that they use the internet "rarely" ($\overline{X}$ =2.48).

It is seen that the items in e-Student services with education purpose are used "very frequently", "frequently" and "sometimes". Students stated that they use the internet "very frequently" to do home work and research and that they use the internet "frequently" to register for courses and to make use of library services. Besides, it was also found out that they sometimes used the internet to search for supplementary education services, to communicate with the teacher and taking a course on the web.

The frequencies the items in e-Student services for economic purposes sub-dimension which has the least frequently used items also vary. Among all items using the internet for internet banking services ($\overline{X}$ =2.13) and shopping ($\overline{X}$ =2.16) is used "rarely" and in the least used e-Student services. On the other hand, it was found out that students use the internet "frequently" to do search for financial support opportunities ($\overline{X}$ =3.43), and to search job opportunities ($\overline{X}$ =2.75) and to do price search ($\overline{X}$ =3.15) "sometimes".

3.2. Findings with Regard to Student's use of e-Student Services in Terms of Gender

The study examined whether students' use of e-Student services changed in terms of gender. T-test was used in order to determine whether students' scores of using e-Student services varied in terms of gender. The findings are shown in Table 6.

Table 6. Data from Students' Use of e-Student Services according to Gender

Group	n	Mean	Std Deviation	t	p
Male	173	3.46	.66	4.57	.01*
Female	217	3.13	.74		

As it can be seen in Table 6, male students use e-Student services "frequently" ($\overline{X}$ =3.46). On the other hand, women students use e-Student services "sometimes" ($\overline{X}$ =3.13). In order to test whether the difference is statistically significant, t values obtained as a result of t test (t= 4.57, p<.05) was found to be significant.

In other words, male and female students' use of e-Student services show difference. Male students make use of the internet for e-Student services more than female students.

3.3. Findings with Regard to Student use of e-Student Services in terms of Place of Residence

The study also examined whether students use of services showed difference according to the place of residence. The findings with regard to the students' use of e-Student services in terms of the place of residence are given in Table 7.

Table 7. Means and Standard Deviation Values of the Scores of the students in terms of their place of residence

Place of Residence	f	$\overline{X}$	Sd
A- With His/her Family	96	3.43	.68
B- Dormitory	117	3.14	.73
C- Shares a flat with Friends	177	3.29	.75
Total	390	3.28	.73

According to the data in Table 7, the students who live with their families use e-Student services "frequently" ($\overline{X}$ =3.43). The students who stay at dormitory ($\overline{X}$ =3.14) and the students who shares a flat with their friends ($\overline{X}$ =3.29) use these services "sometimes". In order to test whether arithmetic means between groups differed significantly, one-way variance analysis was carried out. The results of this analysis are shown in Table 8.

Table 8. the Students e-Student Service Usage Scores in terms of their Place of Residence

Source	df	Sum of Squares	Mean Squares	f	p	Differs
Between groups	2	4.52	2.26	4.29	.01*	A-B*
Within group	387	204.09	.53			
Total	389	208.61				

As it can be seen in Table 8, according to F value calculated from the scores of the groups in relation to their use of e-Student services (F= 4.29, p<.05), it can be said that the difference between groups is significant. Scheffe test was administrated in order to determine from which group or groups this difference stemmed from. According to the results of Scheffe test, the students who stay with their families use e-Student services more frequently than those who live in the dormitory. However, there is not a significant difference between the students who share a flat with friends and who stay at other places (at home with family or at dormitory) in terms of using e-student services.

3.4. Findings with Regard to Students' scores on Use of e-Student Services in terms of Internet Usage Frequency

Internet is one of the technologies most frequently used technology. The study also examined whether the frequency students use e-Student services showed difference in terms of the time students spend on the internet. In Table 9, data about the use e-student services by student in terms of the frequency they use the internet.

Table 9. Mean and standard deviation of e-Student services usage scores according to the frequency students use the internet

The Internet Usage Frequency	f	$\overline{X}$	Sd
A- Everyday	106	3.68	.57
B- Once in 2-3 days	117	3.52	.53
C- Once	55	3.33	.55
D- Less Frequent	112	2.62	.68
Total	390	3.28	.73

It is seen that students use of e-Student services varied according to their internet use frequency: those who use everyday ($\overline{X}$ =3.68) and once in 2-3 days ($\overline{X}$ =3.52) used e-student services "frequently", and those who use it one a week ($\overline{X}$ =3.33) and less frequently (2.62) use e-student services "sometimes" (Table 9).In order to test whether the difference between group arithmetic means are statistically significant, one way variance analysis was carried out. The results of this analysis are given in Table 10.

Table 10. One-way Variance Analysis Results of e-student services usage scores in terms of the frequency they use the internet

Source	df	Sum of Squares	Mean Squares	f	p	Differs
Between groups	3	73.52	24.50	70.02	.01*	A-C, A-D, B-D, C-D
Within group	386	135.09	.35			
Total	389	208.61				

As it can be seen in Table 10, students use of e-student services shows difference according to the frequency they use the internet (F= 70.02, p<.05). Scheffe test was used to determine the source of difference. According to result of the analysis, students who use the internet every day use e-student services more frequently than those who use the internet once a week and less frequently. At the same time, the students who use the internet once in 2-3 days and once a week use e-student services more frequently compared to students who use the internet less.

4. Results and Discussions

E-Student services mean students' using the internet to improve their quality of life. In this study in which the frequency e-student services used is examined including the sub-dimension, it was found out that university students used these services "sometimes". While students use e-student services for health and economic purposes "sometimes", they use social and education related e-student services "frequently".

The difference students show in terms of the frequency they use e-student services can be attributed to the limitations of the services offered on line. The findings of the studies on the

purposes students the internet support this implication. In other studies on this subject, it is reported that students use the internet more frequently for socialization (interaction-communication etc.) and education (access to information-doing homework etc) purposes compared to health and economy related purposes [7], [8], [9], [10], [11],[12].

As students are in education process, it is natural that instruction/learning focused internet usage comes to forth. On the other hand, it can be said that the limitedness of the services offered on the internet is decisive. Not all health institutions in Turkey have e-health services like getting appointment on-line, seeing analysis results and finding out information about doctors and diseases on-line. Financial conditions of students could have played role in the frequency they use e-student services. The use of e-student services for economic purposes is restricted as students -who are yet economically dependent on their families- have a limited budget. So, it is not surprising that they do not use credit card and internet banking and the students who live in dormitories do not have to pay bills and use supermarkets instead of the internet as they are more practical for their small scale shopping.

In the scope of the study, it was found out that male students use e-student services more frequently compared to females. The fact that male students use e-student services more frequently than female students can be accounted for with the fact that male students are more knowledgeable, experienced and self-efficient in terms of using the internet. In their study Schumacher and Morahan-Martin (2001) concluded that that male students have more computers and have had more experiences, they are more competent in using the computer and the internet. The result of the survey carried out by Turkish Statistical Institute– [1] indicated that the internet usage rate for males was 48.6%, while it is 28% for females. This result indicates that males use the internet more than females and thus are more experienced. Similarly, in a study by Peng, Tsai and Wu [14] on university students it was reported that male students are more self-efficient than female students.

Among the variables that affect e-student services, the place of residence and the frequency they use the internet attracts attention, too. It was concluded that the more frequently students used the internet the more frequently they used e-student services and those who live their family used e-student services more frequently. This result can be accounted for with the findings of Moon [15] that being knowledgeable about internet usage would increase internet usage. Therefore, it is sensible that individuals who use the internet more frequently and have more opportunities to access to the internet use e-student services more frequently. In a survey conducted by TurkStat [1] in Turkey in 2009, it was reported that while homes with 57,6 % ranked the first among the places accessed to the internet, internet cafes which were previously preferred more turned out to be preferred at 24,1 %. The finding that the frequency e-student services used at homes was higher than that in dormitories can be attributed to such problems as rooms' being shared by several students, time restriction on the internet use and using common computers all of which restrict individual internet use. Nie and Ebring [16] indicate that web users use more versatile and efficient internet sources as the internet used more. The difference in the frequency e-students service-which include health, social, economic and education sub-dimensions- are used could also have resulted from individuals' internet usage periods.

References

[1] TurkStat. "Information And Communication Technology (ICT) Usage Survey On Households And Individuals", Retrieved: 23 02, 2010. from http://www.turkstat.gov.tr/PreHaberBultenleri.do?id=4104, 2009.

[2] Eastman Internet. "Advantages of internet". Eastman Internet Reports. Retrieved: 12.10.2006 from http://www.eastman.com.au/e-commerce/advantages.html, 2003.

[3] TurkStat. "Internet activities of individuals who have accessed the Internet in the last 3 months by private purposes", Retrieved: 25 02, 2010. from http://www.tuik.gov.tr/PreIstatistikTablo.do?istab_id=44, 2009.

[4] TurkStat. "Individuals using the computer and accessing the Internet in the last 3 months by age, education level and employment situation", Retrieved: 25 02, 2010. from http://www.tuik.gov.tr/PreIstatistikTablo.do?istab_id=42, 2009.

[5] Turkey Information Council – TIC. "e-Toplum Çalışma Grubu Sonuç Raporu (e-Society Study Group Final Report)". *II. Turkey Information Council*. 30-31 May 2004, Ankara, Turkey: Middle East Technical University, 2004.

[6] Ozmen, A. *Uygulamalı Araştırmalarda Örnekleme Yöntemleri*. Eskişehir: Anadolu Üniversitesi Yayınları, pp. 197, 2000.

[7] Atav, E., Akkoyunlu, B. & Sağlam, N. "Prospective teachers' internet access facilities and their internet usage- (öğretmen adaylarının internete erişim olanakları ve kullanım amaçları)" *Hacettepe University Journal of Education*. 30, pp:37-44, 2006.

[8] Atıcı, B. & Dikici, A. "The relationship between the internet use and educational level of individuals who connect to internet from internet cafes". *Fırat University Journal of Social Science*. 13(2), pp:129-146, 2003.

[9] Jackson, L.A., Zhao, Y., Qiu, W., Kolenic III, A., Fitzgerald, H.E., Harold, R. & Eye, A.V. "Culture, gender and information technology use: A comparison of Chinese and US children", *Computers in Human Behavior*. 24(6), pp: 2817-2829, 2008.

[10] Joiner, R., Brosnan, M., Duffield, J., Gavin, J. & Maras, P. "Including the special issue: avoiding simplicity, confronting complexity: advances in designing powerful electronic learning environments", *Computers in Human Behavior*. 23(3), pp: 1408-1420, 2007.

[11] Li, Nai & Kirkup, G. "Gender and cultural differences in Internet use: A study of China and the UK", *Computers & Education.* 48(2), pp: 301-317, 2007.

[12] Torkzadeh, G. & Van Dyke, T.P. "Effects of training on Internet self-efficacy and computer user attitudes", *Computers in Human Behavior.* 18(5), pp: 479-494, 2002.

[13] Schumacher, P. & Morahan-Martin, J. "Gender, internet and computer attitudes and experiences". *Computers in Human Behavior,* 17 (1), pp:95–110, 2001.

[14] Peng, H., Tsai, C.C. & Wu, Y.T. "University students' self-efficacy and their attitudes towards the internet: the role of students' perceptions". *Educational Stuies.* 32(1), pp:73-86, 2006.

[15] Moon, B-J. "Consumer adoption of the internet as an information search and product purchase channel: some research hypotheses", *International Journal of Internet Marketing and Advertising,* 1(1), pp: 104-118, 2004.

[16] Nie, N.H. & Erbring, L. "Internet and Society:A Preliminary Report". Stanford Institute of Quantitative Study Soc., 2009.

Author Biographies

Ahmet Naci Çoklar is an assistant professor at the Computer and Instructional Technologies Education Department of Education Faculty, Selcuk University, Konya Turkey. His research interest is technology standards in education, e-learning and the effects of internet on people.

Semseddin Gunduz is an assistant professor at the Computer and Instructional Technologies Education Department of Education Faculty, Selcuk University, Konya, Turkey. He classifies his academic interest as development strategies of e-learning, computer and ethics, human-computer interaction.

Survey of Fault Recovery in Wireless Sensor Networks

Neda Beikmahdavi
Department of Computer Engineering
IAU – Ayatollah Amoli Branch
Amol,Iran

Abolfazl Akbari
Department of Computer Engineering
IAU – Ayatollah Amoli Branch
Amol, Iran

Fatemeh Akbari
Department of Computer Engineering
IAU – Ayatollah Amoli Branch
Amol, Iran

***Abstract*—** *In the past few years wireless sensor networks have received a greater interest in application such as disaster management, border protection, combat field reconnaissance, and security surveillance. Sensor nodes are expected to operate autonomously in unattended environments and potentially in large numbers. Failures are inevitable in wireless sensor networks due to inhospitable environment and unattended deployment. The data communication and various network operations cause energy depletion in sensor nodes and therefore, it is common for sensor nodes to exhaust its energy completely and stop operating. This may cause connectivity and data loss. Therefore, it is necessary that network failures are detected in advance and appropriate measures are taken to sustain network operation. In this paper we survey cellular architecture and cluster-based to sustain network operation in the event of failure cause of energy-drained nodes. The failure detection and recovery technique recovers the cluster structure in less than one-fourth of the time taken by the Gupta algorithm and is also proven to be 70% more energy-efficient than the same. The cluster-based failure detection and recovery scheme proves to be an efficient and quick solution to robust and scalable sensor network for long and sustained operation. In cellular architecture the network is partitioned into a virtual grid of cells to perform fault detection and recovery locally with minimum energy consumption. Fault detection and recovery in a distributed manner allows the failure report to be forwarded across cells. Also this algorithm has been compared with some existing related work and proven to be more energy efficient.*

Keywords- *Cluster head, Sensor Networks, clustering, fault detection, fault recovery & virtual grid*

I. INTRODUCTION

Energy-constrained sensor networks require clustering algorithms for tackling scalability, energy efficiency and efficient resource management. Clustering prolongs the network lifetime by supporting localized decision-making and communication of locally aggregated data within the clusters thereby conserving energy [1]. The amount of energy consumed in a radio transmission is proportional to the square of the transmission range. Since the distance from sensor node to sensor node is shorter than sensor node to the base station, it is not energy efficient for all sensor nodes to send their data directly to a distant base station. Therefore cluster-based data gathering mechanisms effectively save energy [1].

There are many clustering algorithms proposed in the literature [1–5]. Failures are inevitable in sensor networks due to the inhospitable environment and unattended deployment. The failures arise out of energy loss in the sensor nodes, faulty data reporting, faulty hardware and damage due to climatic conditions. Failures occurring due to energy depletion are continuous, and as the time progresses, these failures may increase. Even if the nodes are deployed uniformly at the onset of the network, as time progresses, nodes will become inactive randomly due to varying traffic characteristics, resulting in a non-uniform network topology. This often results in scenarios where a certain segment of the network becomes energy constrained. The problem that can occur due to sensor node failure is loss in connectivity and in some cases network partitioning. There may also be some delay due to the loss in connection and the resulting data may not reach in time. In clustered networks, it causes holes in the cluster topology and disconnects the clusters, thereby causing data loss and connectivity loss. Therefore to overcome sensor node failure and to guarantee system reliability, failing nodes should be identified and appropriate measures to recover network or cluster connectivity must be taken to accommodate for the failing node. They are limited at different levels. Existing approaches are based on hardware faults and consider hardware components malfunctioning only. Some assume that system software's are already fault tolerant as in [6, 7]. Some are solely focused on fault detection and do not provide any recovery mechanism [8]. Sensor network faults cannot be approached similarly as in traditional wired or wireless networks due to the following reasons [9]:

1. Traditional wired network protocol are not concerned with the energy consumptions as they are constantly powered and wireless ad hoc network are also rechargeable regularly.

2. Traditional network protocols aim to achieve point-to-point reliability, where as wireless sensor networks are more concerned with reliable event detection.

3. Faults occur in wireless sensor networks more frequently than traditional networks, where client machine, servers and routers are assumed to operate normally.

Therefore, it is important to identify failed nodes to guarantee network connectivity and avoid network partitioning. We aimed to maintain the cell structure in the event of failures caused by energy-drained nodes. In our scheme, the whole network is divided into a virtual grid where each cell consists of a group of nodes. A cell manager and a secondary manager are chosen in each cell to perform fault management tasks. A secondary manager is a back up cell manager, which will take control of the cell when cell manager fails to operate. These cells combine to form various groups and each group chooses one of their cell managers to be a group manager. The failure detection and recovery is performed after the formation of virtual grid. The energy drained nodes are detected and recovered in their respective cells without affecting overall structure of the network. We considered the case of node notifying their cell managers, when their residual energy is below the threshold value. The virtual grid based failure detection and recovery scheme is compared to Cluster-based failure detection and recovery scheme [10]. It can be result that failure detection and recovery in virtual grid based algorithm is more energy efficient and quicker than that of Cluster-based. In [10], it has been found that Cluster-based algorithm is more energy efficient in comparison with crash fault detection [11] and fault tolerant clustering approach proposed by Gupta and Younis [12]. Therefore, we conclude that our proposed algorithm is also more efficient than Gupta and Crash fault detection algorithm in term of fault detection and recovery.

II. RELATED WORK

In this section, we review the related works in the area of fault detection and recovery in wireless sensor networks (WSNs). Many techniques have been proposed for fault detection, fault tolerance and repair in sensor networks [13–16]. A survey on fault detection in the context of fault management can be found in [17]. Fault tolerance in Internet such as network availability and performance has been discussed in [14]. Hierarchical and cluster-based approaches for fault detection and repair have also been dealt by researchers in [12]. Some authors use routing techniques to identify the failed or misbehaving nodes [18–20]. In [21], a failure-detection scheme that using management architecture for WSNs called MANNA is proposed and evaluated. However, this approach requires an external manager to perform the centralized diagnosis and the communication between nodes and the manager is too expensive for WSNs. Several localized threshold-based decision schemes were proposed by Krishnamachari and Iyengar [24] to detect both faulty sensors and event regions. Luo et al. [22] did not explicitly attempt to detect faulty sensors, instead the algorithms they proposed improve the event-detection accuracy in the presence of faulty sensors. There have been several research efforts on fault repair in sensor networks. However, most existing approaches require knowledge of accurate location information. Some algorithms employ mobile sensor nodes to replace the faulty sensors and rectify coverage and connectivity holes. However, movement of the sensor nodes is by itself energy consuming and also to move to an exact place to replace the faulty node and establish connectivity is also tedious and energy-consuming. Mei et al. [23] proposed a method to use mobile robots to assist sensor replacements for the failed sensor nodes. They study the algorithms for detecting, reporting sensor failures and coordinating the energy-efficient movement of the mobile robots. A replacement protocol for failures in hybrid sensor networks is proposed in [25]. In this paper, the mobile sensors are used to recover from faults or to improve the coverage and connectivity of the network. The faulty sensors locate redundant sensors and initiate request for replacement. In [26], holes are detected using Voronoi diagrams and a bidding protocol is proposed to assist the movement of the sensor nodes for healing the holes. Three distributed self-deployment protocols involving movement of sensor nodes to rectify the holes is proposed in [27].

Ganeriwal et al. [28] proposed an algorithm called coverage fidelity maintenance algorithm (Co-Fi), which uses the mobility of sensor nodes to repair the coverage loss. To repair a faulty sensor, Wang et al. [29] proposed an algorithm to locate the closest redundant sensor using cascaded movement and replace the faulty sensor node. In [30], the authors proposed a policy-based framework for fault repair in sensor network and a centralized algorithm for faulty sensor replacement.

III. SURVEY VIRTUAL GRID BASED FAILURE DETECTION AND RECOVERY ALGORITHM

A. Cellular formation

The sensor network nodes configure themselves into a virtual grid structure, in which the network nodes are partitioned into several cells each with a radius that is tightly bounded with respect to a given value R. Detail of this cellular architecture has been revealed in [31]. A cell can be considered as a special kind of clustering. However it is more systematic and scalable. Cells can merge together to produce large cells that would be managed using the same process. Division of network into virtual grid helps in achieving self configuration, in which it must actively measure network states in order to react to the network dynamics. A grid-based architecture is feasible in a network in which nodes are relatively regularly deployed. We assume that communicated data is fault free and that all semantic-related generic faults are detected and removed by the application itself. Furthermore, we assume that there will be no alterations or creations of messages over the transmission links. One node in each cell is distinguished as the cell manager, to represent this cell in the network. All cell managers in the network form an upper level grid and the remaining nodes form a lower level grid. Fig 1 depicts the organization of the nodes in a virtual grid. After the division of the network into small virtual cells as shown in Fig. 1, a cell manager is appointed in each cell. Initially, node with the highest co-ordinates in a cell becomes cell manager and node with the second highest co-ordinates becomes secondary cell manager.

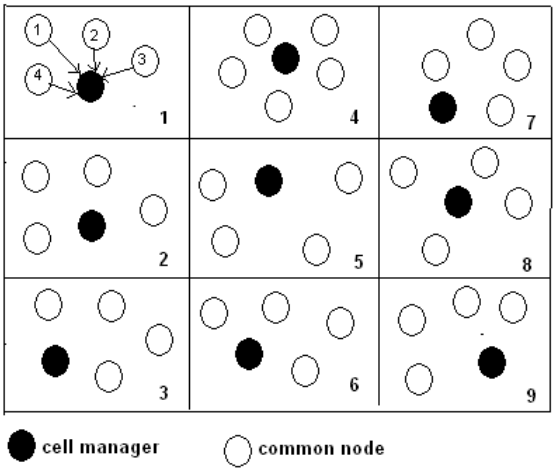


Fig. 1. virtual grid of nodes

Later on, selection of cell manager and secondary cell manager will be based on available residual energy. The node with the maximum residual energy will be chosen as cell manager or secondary manager. The cell manager receives data directly from its cell members and passes it to other neighboring cell managers. There is a one-hop communication between cell manager and its common members as shown in fig 1. After the selection of cell managers and secondary cell manager, cells combine to form various virtual groups. Each group of cells then selects a group manager with mutual co ordination. A group manager is a cell manager which performs its normal tasks for its own cell but at the same time act as a group manager for a group of cells. This is another level of virtual grid, on top of cell managers. The main goal of introducing group manager is to perform high level management tasks and predict future faults.

B. In cell failure detection and recovery

In this section, we discussed the mechanism to detect energy depletion failures in the network and how it is reported to relevant nodes to initiate recovery. Identification of faulty nodes can be achieved by two mechanisms i.e. through regular energy messages to cell manager and nodes themselves notify the managing nodes of its residual energy (if its below the required threshold value). In regular energy messages to managing nodes, common nodes in each cell send their energy status as a part of update_msg to their cell manager. The update_msg consists of node ID, energy and location information. A faulty node will be identified, if the cell manager does not hear from it. In this paper, we focused on the first mechanism as fault identification through regular energy messages has been discussed in [19]. A node is termed as failing when its energy drops below the threshold value. When a common node is failing due to energy depletion, it sends a message to its cell manager that it is going to a low computational mode due to energy below the threshold value. Thus, no recovery steps are required in the failure of common node. Cell manager and secondary cell manager are known to their cell members. If cell manager energy drops below the threshold value, it then sends a message to its cell member including secondary cell manager. Which is an indication for secondary cell manager to stand up as a new cell manager and the existing cell manager becomes common node and goes to a low computational mode. Common nodes will automatically start treating the secondary cell manager as their new cell manager and the new cell manager upon receiving updates from its cell members; choose a new secondary cell manager. Recovery from cell manager failure involved in invoking a backup node to stand up as a new cell manager. The failure recovery mechanisms are performed locally by each cell. In Fig.1, let us assume that cell 1's cell manager is failing due to energy depletion and node 3 is chosen as secondary cell manager. Cell manager will send a message to node 1, 2, 3 and 4 and this will initiate the recovery mechanism by invoking node 3 to stand up as a new cell manager.

C. *Overall cell failure detection and recovery*

Each cell maintains its health status in terms of energy. It can be High, Medium or Low. These health statuses are then sent out to their associate group managers. Upon receiving these health statuses, group manager predict and avoid future faults. For example; if a cell has health status high than group manager always recommend that cell for any operation or routing but if the health status is medium than group manager will occasionally recommend it for any operation. Health status Low means that the cell has un-sufficient energy and should be avoiding for any operation. Therefore, a group manager can easily avoid using cells with low health status. Consider Fig.1, let cell 4 manager be a group manager and it receives health status updates from cell 1, 2 and 3. Cell 2 sends a health status low to its group manager, which alert group manager about the energy situation of cell 2.

IV. SURVEY CLUSTER-BASED FAILURE DETECTION AND RECOVERY ALGORITHM

The nodes are organized into clusters and network operates for some time. The data communication and network operation causes energy depletion in the sensor nodes. The schemes for failure detection and cluster recovery are activated in the event of failures due to energy-drained nodes. In this paper, the maintenance and recovery of the cluster structure in the event of node failures is termed as failure recovery. We now further elaborate on the failure detection and failure recovery mechanisms.

A. Cluster formation

The sensor nodes are dispersed over a terrain and are assumed to be active nodes during clustering. The cluster heads are selected based on a weight, which is a function of number of neighbours and residual energy. Every cluster head starts the formation of the cluster by selecting its first hop members. The first hop members then select the second hop members using an expanding ring-search technique. The nodes select a maximum of D number of nodes as their immediate hop. The admission of nodes takes place till the number of members in a cluster reaches a maximum of S or when all the nodes are clustered. This algorithm has been dealt elaborately in [32].

Limiting the cluster size contributes to notable energy savings. Our investigations show that the energy savings are more prominent for higher values of transmission range. The detailed simulations proving the energy efficiency of the clustering algorithm are dealt in [32].

B. Failure detection

In this section, we discuss the method to detect energy failures in the nodes and report the same to the respective members of the clusters. This detection is essential for the cluster members as they have to invoke the mechanism for the repair and recovery of those failures so as to keep the cluster connected.

Every node has a record of its balance energy. The nodes in each cluster send their energy status as a part of the hello_msg, to their first hop members including their parent. The hello_msg consists of the location (x and y coordinates), energy and node ID. This hello_msg conveys the current energy status of the node. When the node is failing, it sends the failure report message fail_report_msg to its parent and children. A node is termed as failing when its energy level drops below the threshold value, Eth. The threshold value, Eth, is the energy required to transmit D number of l-bit messages across a distance equal to the transmission range.

In Fig. 2, let us assume that node 7 is failing,and then it sends a fail_report_msg to node 3, its parent and node 10 its child. Here we deal with failures related to energy exhaustion, and therefore we assume that the failing node can send the failure report to its immediate hop members before it dies completely.

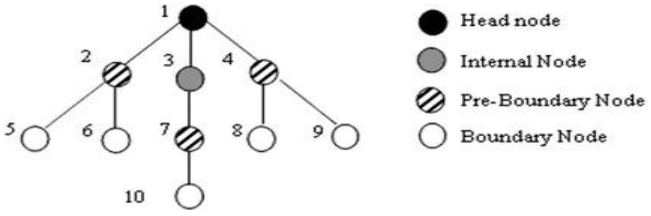


Figure 2. Cluster topology

This information of the failure report is an indication to start the failure recovery process. The parent and children of the failing node are sufficient to invoke the failure-recovery mechanism. Therefore energy is saved by not allowing all the nodes in the cluster to detect a failure. This is the method by which all the nodes in the cluster know about the failure of its first hop members and the corrective action is taken by only those nodes that have the information.

C. Failure recovery

In this section, we discuss the mechanisms for failure recovery. The failure recovery here refers to the connectivity recovery after the node has failed. The node failures discussed here is confined to failure due to energy exhaustion. The failure-recovery mechanisms are performed locally by each cluster. When a node fails, the failing node's parent and children take appropriate action to connect the cluster and bridge the gap formed by the failing node. We have proposed four types of failure mechanisms depending on the type of node in the cluster. The nodes in the cluster are classified into four types, boundary node, pre-boundary node, internal node and the head node. The descriptions of each node are explained in Table 1 and illustrated in Fig. 1. Fig. 1 gives an illustration of the organization of the nodes in the clusters formed by our proposed method in [32]. The nodes in the clusters are organized in a tree-like manner with a parent and children. Let us consider the cluster to have a size of 10 and supportable degree (number of neighbours that each node can have as the next hop) as 3. Every node has a different mechanism for failure recovery. We now discuss the various failure recovery algorithms for a boundary node, pre-boundary node, internal node and a head node. We first explain the routines that are commonly implemented by the four types of nodes.

Type of node	Description	Figure description (Fig. 2)
Boundary node	a node which has no children	nodes 5, 6, 8, 9, 10
Pre-boundary node	a node whose children are all boundary nodes	nodes 2, 4, 7
Internal node	a node which has at least one pre-boundary or internal node as child.	node 3
head node	Cluster head for the cluster	node 1

Table 1. Different types of nodes

D. Common routines followed by recovery Algorithms

- ***Failure reporting*:**

A node is considered failing if its energy falls below the threshold energy. A failure report message, fail_report_msg, is sent by the failing node to its parent and children. This helps the children to realize that they need to search for another suitable parent for further operation. Once the parent receives the failure report, it ignores the failing node for further data transactions and considers it a non-active member.

- ***Procedure for finding a suitable healthy parent:***

A join_request_mesg is sent by the healthy child of the failing node to its neighbours. All the neighbours within the transmission range respond with a join_reply_mesg / join_reject_mesg message. The healthy child of the failing node then selects a suitable parent by checking whether the neighbour is not one among the children of the failing node

and whether the neighbour is also not a failing node. If the healthy child is a boundary node, it first searches for a parent within a cluster, if not successful, it then searches for a parent outside the cluster. While searching for a parent, it checks whether the supportable degree of the neighbour is within the limit D, if the parent is of the same cluster. If the parent is from different cluster, the supportable degree of the neighbour must be within the limit D and the cluster size limit also must be within S. If a healthy child is an internal node, it searches for a suitable parent inside the cluster only. If a suitable parent is found, then the healthy child node attaches itself to the chosen parent. The cluster_info_mesg is exchanged if the chosen parent is from a different cluster. The cluster parameters of the child are updated to that of the new chosen parent through update_mesg and data transmissions then follow the new paths. The failing node is then left with the original parent and its children are all allocated different parents to keep their data transmissions uninterrupted.

- **Boundary node failure-recovery algorithm**

First, the failure reporting takes place as explained in the Section 4.4.1. The failing boundary node is ignored since it does not affect the connectivity of the other nodes in the cluster.

- **Pre-boundary node failure-recovery algorithm**

Failure reporting is done by the failing pre-boundary node. If all the children of the failing pre-boundary node are failing as well, then the whole scenario is ignored as in the case of boundary nodes. If any one or more of the children of the failing pre-boundary node are failing, then the failing child alone is ignored alone. The healthy children then find another suitable healthy parent. If a suitable parent is not found and if the healthy child is a boundary node, it is left with the failing node itself.

- **Internal node failure-recovery algorithm**

Failure reporting is done by the failing internal node as explained in Section 4.4.1. If the child of the failing internal node is failing as well, then the treatment depends on the type of the node. If it is a boundary node, it is left as it is. If it is a pre-boundary node or an internal node, then that will be treated accordingly in one of the procedures for the failure recovery at a later stage, we do not perform recurring failure recovery in one procedure. We allow it to be taken care in the next stage as an internal node or a pre-boundary-node recovery procedure. If a suitable parent is not found, the child starts a cluster of its own. A cluster split happens with that child as the cluster head for the new cluster. The cluster members would be all the children below the new cluster head. Now the failing internal node is left with the original parent and its children are all allocated different parents to keep their data transmissions uninterrupted.

- ***Head node failure-recovery algorithm***

. If the child of the failing cluster-head node is failing as well, then the treatment depends on the type of the node. The failing child is ignored completely if it is a boundary node. If the failing child is not a boundary node, then it will be ignored in this stage of head-node recovery. However, this node will be treated accordingly in one of the procedures for the failure recovery at a later stage as an internal or a pre-boundary node. Soon after the cluster head fails, another new cluster head is elected to manage the cluster.

- **Procedure for choosing another suitable**

Cluster head for the cluster: The children of the failing cluster-head node exchange their energy status. The children who are failing are not considered for the new cluster-head election, and they send tentative_CH_mesg. The healthy child with the maximum residual energy is selected as the new cluster head and it sends a final_CH_mesg. After the new cluster head is selected, the other children of the failing cluster head are attached to this new cluster head and the new cluster head becomes the parent for these children. The failing cluster head also makes the new cluster head as its parent. Since the supportable degree limits need to be maintained, the children of the new cluster head find a suitable parent inside the cluster. This re-allocation helps maintain the cluster size limits and also the supportable degree limits.

VI. CONCLUSION

In this paper we survey a localized cellular based scheme and Cluster-based Scheme for fault detection and recovery in wireless sensor network of. Clustering has been used to address various issues i.e. routing, energy efficiency, management and huge-scale control. Therefore clustering can be formed in several ways. Nodes generally form a cluster in two stages: (1) a header is selected among the nodes through election algorithm, randomized election, degree of connectivity or pre-definition, and (2) the headers and the nodes interact to form a group or a cluster [33]. Cluster heads are responsible for coordinating the nodes in their clusters and generally are more resourceful than its cluster members. Cluster heads are the traffic bottlenecks; their failure may cause several problems. Also, if a cluster head failed to operate then no messages of its cluster will be forwarded to the base station and selection of the new cluster head is energy consuming. Virtual Grid based architecture also divides the network into small virtual cells and each cell consists of a group of nodes, managed by a cell manager. In clustering, the most intuitive way to recover from a cluster head failure is to re-cluster the network. However, re-cluster is not only a resource burden on the sensor nodes but often very disruptive to the ongoing operation. Therefore, we introduced a backup node for recovery from cell manager failure. It does not affect network operation and consume no energy in order to recover from cell manager failure. Heterogeneous network comprises of nodes with different energy levels. Some nodes are less energy constrained than others. In such type of networks the less energy constrained nodes are chosen as cluster head of the cluster. Usually, these less energy constrained node are uniformly distributed with multi-hop communication. Nodes close to these cluster heads are under sever load as traffic

routed from different areas of the network to the cluster head is via the neighbours of the cluster head. This results rapid dying of the nodes in the vicinity of the cluster heads, creating connectivity loss and in some cases network partitioning. Our approach addresses this challenge by employing a load balancing strategy so that all nodes operate together for as long as possible. We consider that all the nodes in the network are equal in resources and no node should be more resourceful than any other node. The optimal role assignment and reconfiguration scheme support the network management system to utilize the network nodes in the most efficient manner. Our approach does not rely on specific nodes with extra resources but assign tasks due to their optimal capabilities. Nodes are ranked according to their available energy. Therefore, the selection of cell manager and group manager is based on the available energy. The basis idea of this design is to encourage nodes to be more self manageable and extend the network life time for as long as possible. Also, distributed management system has lower communication costs and provides better reliability and energy efficiency. Virtual Grid based divides the whole network into a virtual grid and enables the network to perform local detection and distribute the management tasks across the network. This approach helps sensor nodes to take more management responsibility and decision-making in order to success the vision of self managed WSNs. Also, this increases network life time. The cellular architecture is for management purpose only so they can be merged into clusters for routing or any other purpose if needed. This scheme outperforms the Cluster-based algorithm with respect to fault detection and recovery in term of energy efficiency and time. The results obtained clearly that virtual grid based algorithm perform failure detection and recovery much faster than cluster-based algorithm and consumed significantly low energy.

VII. REFERENCES

[1] YOUNIS O., FAHMY S.: 'HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks', IEEE Trans. Mobile Comput., 2004, 3, (4),pp. 366–379

[2] CHATTERJEE M., DAS S.K., TUGUT D.: 'WCA: a weighted clustering algorithm for mobile ad-hoc Networks', J. Cluster Comput., 2002, 5, (2), pp. 193–204

[3] HEINZELMAN W., CHANDRAKASAN A., BALAKRISHNAN H.: 'Energyefficient routing protocols for wireless microsensor networks'. Proc. 33rd Hawaii Int. Conf. System Sciences, Hawaii, USA, January 2000, p. 8020

[4] LI C., YE M., CHEN G., WU J.: 'An energy-efficient unequal clustering mechanism for wireless sensor networks'. Proc. 2nd IEEE Int. Conf. Mobile Ad-hoc and Sensor Systems, Washington, DC, USA, November 2005, p.

[5] BANDYOPADHYAY S., COYLE E.: 'An energy-efficient hierarchical clustering algorithm for wireless sensor networks'. Proc. IEEE INFOCOM, San Francisco, USA,March 2003, vol. 3, pp. 1713–1723

[6] J. Chen, S. Kher and A. Somani, "Distributed Fault Detection of Wireless Sensor Networks", in DIWANS'06. 2006. Los Angeles, USA: ACM Pres.

[7] F. Koushanfar, M. Potkonjak, A. Sangiovanni- Vincentelli, "Fault Tolerance in Wireless Ad-hoc Sensor Networks", Proceedings of IEEE Sensors 2002, June, 2002

[8] W. L. Lee, A. Datta, and R. Cardell-Oliver, "Network Management in Wireless Sensor Networks", to appear in Handbook on Mobile Ad Hoc and Pervasive Communications, edited by M. K. Denko and L. T. Yang, American Scientific Publishers.

[9] L. Paradis and Q. Han, "A Survey of Fault Management in Wireless Sensor Networks", Journal of Network and Systems Management, vol. 15, no. 2, pp. 171-190, 2007.

[10] G. Venkataraman, S. Emmanuel and S.Thambipillai, "Energy-efficient cluster-based scheme for failure management in sensor networks" IET Commun, Volume 2, Issue 4, April 2008 Page(s):528 – 537

[11] S. Chessa and P. Santi, "Crash faults identification in wireless sensor networks", Comput. Commun., 2002, 25, (14), pp. 1273-1282.

[12] G. Gupta and M. Younis; Fault tolerant clustering of wireless sensor networks; WCNC'03, pp. 1579.1584.

[13] ZHOU Z., DAS S., GUPTA H.: 'Fault tolerant connected sensor cover with variable sensing and transmission ranges'. Proc. IEEE Sensor and Ad Hoc Communications and Networks, Santa Clara, USA, September 2005, pp. 594–604

[14] DING M., CHEN D., XING K., CHENG X.: 'Localized fault-tolerant event boundary detection in sensor networks'. Proc. [15] MENG X., NANDAGOPAL T., LI E., LU S.: 'Contour maps: monitoring and diagnosis in sensor networks', Proc. Int. J. Comput. Telecommun. Netw., 2006, 50, (15), pp. 2820–2838

[16] HARTE S., RAHMAN A., RAZEEB K.M.: 'Fault tolerance in sensor networks using self diagnosing sensor nodes'. Proc. IEE Int. Workshop on Intelligent Environments, UK, June 2005,

[17] OATES T.: 'Fault identification in computer network: a review and a new approach'. Technical Report UM-CS- 1995-113, University of Massachusetts Amherst, 1995

[18] STADDON J., BALFANZ D., DURFEE G.: 'Efficient tracing of failed nodes in sensor networks'. Proc. 1st ACM Int.Workshop on Wireless Sensor Networks and

[19] TANACHAIWIWAT S., DAVE P., BHINDWALE R., HELMY A.: 'Secure locations: routing on trust and isolating compromised sensors in location-aware sensor networks'. Proc. 1st Int. Conf. Embedded Networked Sensor Systems, Los Angeles, USA, November 2003, pp. 324–325

[20] PERRIG A., SZEWCZYK R., WEN V., CULLER D.E., TYGAR J.D.: 'SPINS: security protocols for sensor networks', Proc. Wirel. Netw., 2002, 8, (5), pp. 521–534

[21] RUIZ L.B., SIQUEIRA I.G., OLIVEIRA L.B., WONG H.C., NOGUEIRA J.M.S., LOUREIRO A.A.F.: 'Fault management in event-driven wireless sensor networks'. Proc. 7th ACM Int. Symp. Modeling, Analysis and Simulation of Wireless and Mobile Systems, Venice, Italy, October 2004,pp. 149–156

[22] LUO X., DONG M., HUANG Y.: 'On distributed fault-tolerant detection in wireless sensor networks', Proc. IEEE Trans. Comput., 2006, 55, (1), pp. 58–70

[23] MEI Y., XIAN C., DAS S., HU Y.C., LU Y.H.: 'Repairing sensor networks using mobile robots'. Proc. ICDCS Int. Workshop on Wireless Ad Hoc and Sensor Networks, Lisboa, Portugal, July 2006

[24] KRISHNAMACHARI B., IYENGAR S.: 'Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor network', IEEE Trans. Comput., 2004, 53, (3), pp. 241–250

[25] LE T., AHMED N., JHA S.: 'Location-free fault repair in hybrid sensor networks'. Proc. first ACM Int. Conf. Integrated Internet Ad Hoc and Sensor Networks, Nice, France, May 2006, vol. 138, Article no: 23

[26] WANG G., CAO G., LA PORTA T.F.: 'A bidding protocol for deploying mobile sensors'. Proc. 11th IEEE Int. Conf. Network Protocols, Atlanta, USA, November 2003, pp. 315–324

[27] WANG G., CAO G., LA PORTA T.F.: 'Movement-assisted sensor deployment', IEEE Trans. Mobile Comput., 2006, 5, (6), pp. 640– 652

[28] GANERIWAL S., KANSAL A., SRIVASTAVA M.B.: 'Self aware actuation for fault repair in sensor networks'. Proc. IEEE Int. Conf. Robotics and Automation, New Orleans, USA, April 2004, vol. 5, pp. 5244–5249

[29] WANG G., CAO G., PORTA T., ZHANG W.: 'Sensor relocation in mobile sensor networks'. Proc. IEEE INFOCOM, Miami, USA, March 2005, vol. 4, pp. 2302–2312

[30] LE T., AHMED N., PARAMESWARAN N., JHA S.: 'Fault repair framework for mobile sensor networks'. Proc. First Int. Conf. Communication System Software and Middleware, New Delhi, January 2006, pp. 1–8

[31] M. Asim, H. Mokhtar and M. Merabti, "A Fault Management Architecture for Wireless Sensor Networks", in IWCMC'08. 2008. Crete Island, Greece: IEEE.

[32] VENKATARAMAN G., EMMANUEL S., THAMBIPILLAI S.: 'DASCA: a degree and size based clustering approach for wireless sensor networks'. Proc. IEEE, Int. Symp. Wireless Communication Systems, Siena, Italy, September 2005, pp. 508–512

[33] J.L Chen , H.F Lu and C.A Lee, "Autonomic selforganization architecture for wireless sensor communications", International Journal of Network Management, v.17 n.3, p.197-208, June 2007

Computational Study of H1N1 Viral Segments Inserted Within the Regions of SARS Genome

Dr.DSVGK Kaladhar[1], and A.Krishna Chaitanya[2]

[1,2]GITAM University, Visakhapatnam

Corresponding Addresses
dr.dowluru@email.com

Abstract: SARS viruses have been transforming both unnecessary and necessary viral gene segments of H1N1 genes which are emerged recently due to change of climatic factors. In this point of view, eight viral sequences of H1N1 were downloaded from NCBI Tax Browser and scanned against complete genome of H1N1 for the presence of possible viral inserts in SARS genome. The results from the computational analysis provided that H1N1 virus resulted in viral segments inserted in the intron and exon regions of SARS genome. The alignments showed 13 proteins related to SARS, 7 proteins related to Influenza, 8 proteins related to other species and 2 unpredicted segments.

***Keywords*:** SARS, H1N1 influenza, alignment, gene prediction.

1. INTRODUCTION

Most virus detection methods are providing information in the detection of specific viral targets that cause emerging viral infections [1]. These protein targets may cause diseases within a short time or may prolong and occurs suddenly after long time. The etiologic agent causing severe acute respiratory syndrome (SARS) and H1N1 influenza, continue to emerge based on the hybridization or mutational signals and have become major health problems worldwide. Immunologic assays and the PCR method are some traditional viral detection methods using in present decades [2].

Scientists know very little about SARS, a virus belongs to the *Coronaviridae* family[3], are still not certain of the precise origin and evolution of the disease, which is the key to predicting its nature of occurrence. Influenza A viruses belong to the family Orthomyxoviridae[4], and the genomes are plastic, owing to point mutations and reassortment events that contribute to the emergence of new variants [5].

Particularly in childhood, acute lower respiratory tract infections have become a major worldwide health problem. These infections are ranked first among the conditions which are contributing to the global burden of disease [6]. Avian influenza and H1N1 influenza in South East Asia provides a real possibility for the emergence of a novel influenza virus pathogenic in humans.

Some of the viruses can also cause fever in relation to neurologic and gastrointestinal symptoms [7] which may also leads due accumulation of dopamine leads to diseases such as Alzheimer and Parkinson at the ages ranging from 36 to 87. There were also be reductions in mortality regardless of age, sex, race and risk factors, adapt to become a virus readily transmissible between humans [8]. The etiological agent of SARS was identified as a group II coronavirus, which was named SARS-CoV [9], [10].

2. METHODOLOGY

2.1 Phylogeny

Multiple nucleic acid sequences are aligned for three principal purposes: to identify common motifs in sequences with a conserved biological function, constructed newly formed organism and charecterise proteins in the newly mutated virus that may provide insight into its biological functions. Phylogenetic analysis of swine flu genome segments with genomes of SARS virus, Myxoma virus, Koiherpes virus and Vaccinia virus has been constructed during present experimentation. A multiple alignment was made using CLUSTAL W

Steps for phylogenetic construction

1. Multiple alignment
2. Distance calculation
3. Tree construction

2.2 Characterization of mutated swine flu segments with SARS genome.

SARS genome (GenBank Accession Number DQ071615) has been modified to the new viral segment by substituting the eight genomes segments of swine flu virus (NCBI Accession Numbers NC 002016 to 23) by 7th January 2010 in NCBI server. The newly modified genome was submitted to FGENES V0, a viral gene prediction server. The produced translated genes has been separately submitted to BLASTP for knowing characteristics of the protein sequences.

3. RESULTS AND DISCUSSION

The results of the present research have shown that Swine flu

virus is related to Vaccinia and SARS Genomes (**Figure 1**). 30 genes are predicted due to changes of SARS genome with Swine flu segments. The new organism can transfer genes from other organisms such as *Eggerthella lenta, Thioalkalivibrio sp, Oreochromis mossambicus, Pelobacter carbinolicus, Drosophila melanogaster, Neisseria meningitides* and *Sulfitobacter* (**Table 1**).

The changing sizes of the coronavirus RNA genome uses a special mechanisms by these viruses to produce an extensive set of subgenomic RNAs is linked to a replicative machinery and enzymatic activities. There is a huge variation in climate and even some of the hidden objects are also transforming from other planets onto the earth. Due to these objects, the mutagenic changes has been occurring in the genomic changes in biological organisms, which in turn have the capability to regenerate using complex organisms as hosts.

FIGURE 1 CLADOGRAM

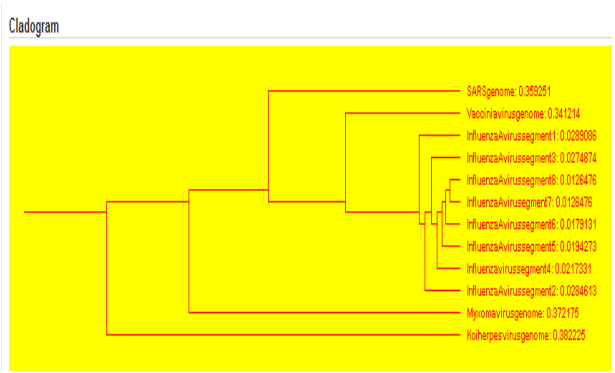


TABLE 1: CHARECTERIZATION OF MUTATED SWINEFLU SEGMENTS WITH SARS VIRUS

GENE	STR AND	START TO STOP AND LENGTH(IN BP)	CHARACTERIZATION AND ORGANISM
1	+	291-1136/ 846	non-structural polyprotein pp1a [Bat SARS]
2	+	808-1209/ 402	uncharacterized protein 1c [SARS]
3	+	2132-2347/ 216	non-structural polyprotein pp1a [Bat SARS]
4	+	2777-3106/ 330	non-structural polyprotein [Bat SARS]
5	+	3959-4195/ 237	nsp3 [Bat SARS]
6	-	5466-5702/ 237	Dehydrogenase/ reductase [Eggerthella lenta]
7	-	6163-6369/ 207	No significant similarity found
8	+	6657-7004/ 348	orf1a polyprotein [SARS]
9	-	8303-8662/ 360	ATP-binding protein [Thioalkalivibrio sp]
10	+	10227-10463/ 237	P1a [Bat SARS]
11	+	11425-11670/ 246	orf1ab [SARS]
12	-	11750-11983/ 234	ELAV-like [Oreochromis mossambicus]
13	-	14136-14465/ 330	hypothetical protein [Pelobacter carbinolicus]
14	+	15409-15612/ 204	hemagglutinin [Influenza A virus]
15	-	19210-19428/ 219	cytochrome c oxidase, cbb3-type, subunit II [Thioalkalivibrio sp]
16	+	19575-19832/ 258	P1b [Bat SARS]
17	+	21000-21503/ 504	polymerase PB1 [Influenza A virus]
18	+	21666-22043/ 378	polymerase [Influenza A virus]
19	-	22151-22501/ 351	GH14316p [Drosophila melanogaster]
20	+	22494-22709/ 216	nucleoprotein [Influenza A virus]
21	+	23100-23318/ 219	nucleoprotein [influenza A virus]
22	-	23427-23672/ 246	hypothetical protein [Neisseria meningitides]
23	-	25644-26003/ 360	deacetylase / probable acetylpolyamine aminohydrolase [Sulfitobacter]
24	+	27050-27316/ 267	matrix protein 1 [Influenza A virus]
25	+	27920-28198/ 279	S2 protein [SARS]
26	+	28337-28549/ 213	S2 protein [SARS]
27	+	28584-28880/ 297	hypothetical protein [Bat SARS]
28	+	29009-29227/ 219	ORF3a [Bat SARS]
29	-	31337-31558/ 222	No significant similarity found
30	+	32232-32465/ 234	nonstructural protein 1 [Influenza A virus]

4. CONCLUSION

Biological research provides deeper insights in analyzing protein-protein system relationship to system-system relationship. The living nanosystems such as viruses will provide impact on control of human cellular system. Combining the wetlab and computer science will increase the efficiency of biological research which can provide the mutagenic activity of two viral particles causes fall of human populations. Hence complete control of gene exchanges between viruses can make positive effects in control of new

emerging viruses such as SARS, swine flu etc..

REFERENCES

[1] Cheng-Chung Chou, Te-Tsui Lee, Chun-Houh Chen, Hsiang-Yun Hsiao, Yi-Ling Lin , Mei-Shang Ho, Pan-Chyr Yang, Konan Peck, "Design of microarray probes for virus identification and detection of emerging viruses at the genus level", BMC Bioinformatics 2006, vol 7, pp. 232, 2006.

[2] Elnifro EM, Ashshi AM, Cooper RJ, Klapper PE, "Multiplex PCR: optimization and application in diagnostic virology", Clin Microbiol Rev.,vol 13, No. 4, pp. 559-570, 2000.

[3] Stefano Ricagno,Marie-Pierre Egloff, Rachel Ulferts, Bruno Coutard, Didier Nurizzo, Valérie Campanacci, Christian Cambillau, John Ziebuhr, Bruno Canard, "Crystal structure and mechanistic determinants of SARS coronavirus nonstructural protein 15 define an endoribonuclease family", PNAS, vol 103, No.32, pp. 11892-11897, 2006.

[4] Jessica A. Belser, Carolyn B. Bridges, Jacqueline M. Katz, Terrence M. Tumpey, "Past, Present, and Possible Future Human Infection with Influenza Virus A Subtype H7", Emerging Infectious Diseases, Vol. 15, No. 6, pp. 859-865, 2009.

[5] Gabriele Neumann, Takeshi Noda, Yoshihiro Kawaoka, "Emergence and pandemic potential of swine-origin H1N1 influenza virus", Nature, Vol 459,pp. 931-939, 2009.

[6] Maria Zambon, "Influenza, respiratory syncytial virus and SARS"**,** Medicine, vol 33, No. 5, pp. 130-134, 2005.

[7] D. Weiss, D. Carr, J. Kellachan, C. Tan, M. Phillips, E. Bresnitz, M. Layton, West Nile, "Virus Outbreak Response Working Group, Clinical findings of West Nile virus infection in hospitalized patients, New York and New Jersey, 2000", Emerg Infect Dis. vol 7, No. 4, pp. 654–658, 2001.

[8] J S M Peiris, Y Guan, K Y Yuen, "Severe acute respiratory syndrome", Nature Medicine, vol **10**, pp. S88 - S97, 2004.

[9] Ron A. M. Fouchier, Nico G. Hartwig , Theo M. Bestebroer , Berend Niemeyer , Jan C. de Jong , James H. Simon, Albert D. M. E, "Osterhaus . A previously undescribed coronavirus associated with respiratory disease in humans", PNAS, vol 101 No. 16, pp. 6212-6216, 2004.

[10] Naomi Takasuka, Hideki Fujii, Yoshimasa Takahashi, Masataka Kasai, Shigeru Morikawa, Shigeyuki Itamura, Koji Ishii, Masahiro Sakaguchi, Kazuo Ohnishi, Masamichi Ohshima, Shu-ichi Hashimoto, Takato Odagiri, Masato Tashiro, Hiroshi Yoshikura, Toshitada Takemori, Yasuko Tsunetsugu-Yokota, "A subcutaneously injected UV-inactivated SARS coronavirus vaccine elicits systemic humoral immunity in mice", Int. Immunol., vol 16, No. 10, pp. 1423-1430, 2004.

Author Biographies

Dr.DSVGK Kaladhar: Dr. Kaladhar was born in Vijayawada in 1974. He has got his doctoral degree in Biotechnology and M,Sc in Microbiology from ANU, Guntur. He is presently working as Asst. Professor in Bioinformatics, GITAM University, Visakhapatnam. His field of study is neural networks, Image analysis, Bioinformatics and Biology

Mr. A.Krishna Chaitanya: Mr. Chaitanya was born in Rajahmundry in 1984. He has acquired Master of Science in Biochemistry from AU, Visakhapatnam. He is presently working as Asst.Professor in GITAM University, Visakhapatnam. His specializations are Molecular modeling, Bioinformatics and Biochemistry.

Self -Managing Fault in Wireless Sensor Networks

Abolfazl Akbari
Department of computer Engineering
Islamic Azad University Ayatollah Amoli Branch
Amol, Iran
a.akbari@iauamol.ac.ir

Neda Beikmahdavi
Department of computer Engineering
Islamic Azad University Ayatollah Amoli Branch
Amol, Iran
n.beikmahdavi@iauamol.ac.ir

Abstract— *Wireless sensor networks are characterized by dense deployment of energy constrained nodes. in the past few years wireless sensor networks have received a greater interest in application such as disaster management, border protection, combat field reconnaissance and security surveillance. Sensor nodes are expected to operate autonomously in unattended environments and potentially in large numbers. Failures are inevitable in wireless sensor networks due to inhospitable environment and unattended deployment. The data communication and various network operations cause energy depletion in sensor nodes and therefore, it is common for sensor nodes to exhaust its energy completely and stop operating. This may cause connectivity and data loss. Therefore, it is necessary that network failures are detected in advance and appropriate measures are taken to sustain network operation. In this paper we proposed a new mechanism to sustain network operation in the event of failure cause of energy-drained nodes. The proposed technique relies on the cluster members to recover the connectivity. The proposed recovery algorithm has been compared with some existing related work and proven to be more energy efficient[20].*

Key words: *Sensor Networks, clustering, fault detection, fault recovery.*

1. Introduction

Energy-constrained sensor networks require clustering algorithms for tackling scalability, energy efficiency and efficient resource management. Clustering prolongs the network lifetime by supporting localized decision-making and communication of locally aggregated data within the clusters thereby conserving energy. The amount of energy consumed in a radio transmission is proportional to the square of the transmission range. Since the distance from sensor node to sensor node is shorter than from sensor node to the base station, it is not energy efficient for all sensor nodes to send their data directly to a distant base station. Therefore cluster-based data gathering mechanisms effectively save energy . In clustered networks, it creates holes in the network topology and disconnects the clusters, thereby causing data loss and connectivity loss. Good numbers of fault tolerance solutions are available but they are limited at different levels. Existing approaches are based on hardware faults and consider hardware components malfunctioning only. Some assume that system software's are already fault tolerant as in [7, 8]. Some are solely focused on fault detection and do not provide any recovery mechanism [9]. Sensor network faults cannot be approached similarly as in traditional wired or wireless networks due to the following reasons [11]:

1. Traditional wired network protocol are not concerned with the energy consumptions as they are constantly powered and wireless ad hoc network are also rechargeable regularly.

2. Traditional network protocols aim to achieve point to point reliability, where as wireless sensor networks are more concerned with reliable event detection.

3. Faults occur more frequently in wireless sensor networks than traditional networks, where client machine, servers and routers are assumed to operate normally.

Therefore, it is important to identify failed nodes to guarantee network connectivity and avoid network partitioning. The New Algorithm recovery scheme is compared to Venkataraman algorithm proposed in [10]. The Venkataraman algorithm is the latest approach towards fault detection and recovery in wireless sensor networks and proven to be more efficient than some existing related work. It solely focused on nodes notifying its neighboring nodes to initiate the recovery mechanism. It can be observed from the simulation results that failure detection and recovery in our proposed algorithm is more energy efficient and quicker than that of Venkataraman algorithm. In [10], it has been found that Venkataraman algorithm is more energy efficient in comparison with Gupta and Younis [15].

Therefore, we conclude that our proposed algorithm is also more efficient than Gupta and Crash fault detection algorithm in term of fault detection and recovery.

This paper is organized as follows: Section 2 provides a brief review of related work in the literature. . In section 3, we provided a detail description of our clustering algorithm. In section 4, we provided a detail description of our proposed solution. The performance evaluation of our proposed algorithm can be found in Section 5, Finally, section 6 concludes the paper.

2. RELATED WORK

In this section we will give an overview about existing fault detection and recovery approaches in wireless sensor networks. A survey on fault tolerance in wireless sensor networks can be found in [12]. A detailed description on fault detection and recovery is available at [11]. WinMS [13] provides a centralized fault management approach. It uses the central manager with global view of the network to continually analyses network states and executes corrective and preventive management actions according to management policies predefined by human managers. The central manager detects and localized fault by analyzing anomalies in sensor network models. The central manager analyses the collected topology map and the energy map information to detect faults and link qualities. WinMS is a centralized approach and approach is suitable for certain application. However, it is composed of various limitations. It is not scalable and cannot be used for large networks. Also, due to centralize mechanism all the traffic is directed to and from the central point. This creates communication overhead and quick energy depletions. Neighboring co-ordination is another approach to detect faulty nodes. . For Example, the algorithm proposed for faulty sensor identification in [16] is based on neighboring co-ordination. In this scheme, the reading of a sensor is compared with its neighboring' median reading, if the resulting difference is large or large but negative then the sensor is very likely to be faulty. Chihfan et. al [17] developed a Self monitoring fault detection model on the bases of accuracy. This scheme does not support network dynamics and required to be pre configured. In [14], fault tolerance management architecture has been proposed called MANNA (Management architecture for wireless sensor networks). This approach is used for fault diagnosis using management architecture, termed as MANNA. This scheme creates a manager located externally to the wireless sensor network and has a global vision of the network and can perform complex operations that would not be possible inside the network. However, this scheme performs centralized diagnosis and requires an external manager. Also, the communication between nodes and the manager is too expensive for WSNs. In Crash fault detection scheme [18], an initiator starts fault detection mechanism by gathering information of its neighbors to access the neighborhood and this process continue until all the faulty nodes are identified. Gathering neighboring nodes information consumes significant energy and time consuming. It does not perform recovery in terms of failure. Gupta algorithm [15] proposed a method to recover from a gateway fault. It incorporates two types of nodes: gateway nodes which are less energy constrained nodes (cluster headers) and sensor nodes which are energy constrained. The less energy constrained gateway nodes maintain the state of sensors as well as multi-hop route for collecting sensors. The disadvantage is that since the gateway nodes are less energy constraint and static than the rest of the network nodes and they are also fixed for the life of the network. Therefore sensor nodes close to the gateway node die quickly while creating holes near gateway nodes and decrease network connectivity.

Also, when a gateway node die, the cluster is dissolved and all its nodes are reallocated to other healthy gateways. This consume more time as all the cluster members are involved in the recovery process. Venkataraman algorithm [10], proposed a failure detection and recovery mechanism due to energy exhaustion. It focused on node notifying its neighboring nodes before it completely shut down due to energy exhaustion. They proposed four types of failure mechanism depending on the type of node in the cluster. The nodes in the cluster are classified into four types, boundary node, pre-boundary node, internal node and the cluster head. Boundary nodes do not require any recovery but pre-boundary node, internal node and the cluster head have to take appropriate actions to connect the cluster. Usually, if node energy becomes below a threshold value, it will send a fail_report_msg to its parent and children. This will initiate the failure recovery procedure so that failing node parent and children remain connected to the cluster.

3. Cluster formation

The sensor nodes are dispersed over a terrain and are assumed to be active nodes during clustering.

- ***Problem Definition***

The clustering strategy limits the admissible degree, *D* and

the number of nodes in each cluster, *S*. The clustering aims to associate every node with one cluster. Every node does not violate the admissible degree constraint, *D* and every cluster does not violate the size constraint, *S* while forming the cluster. The number of clusters(*C*) in the network is restricted to a minimum of N/S, $N < C < N/S$, where *N* is the number of nodes in the terrain.

- ***Sensor Network model***

A set of sensors are deployed in a square terrain. The nodes possesses the following properties

i. The sensor nodes are stationary.
ii. The sensor nodes have a sensing range and a transmission range. The sensing range can be related to the transmission range, Rt> 2rs.
iii. Two nodes communicate with each other directly if they are within the transmission range
iv. The sensor nodes are assumed to be homogeneous i.e. they have the same processing power and initial energy.
v. The sensor nodes are assumed to use different power levels to communicate within and across clusters.
vi. The sensor nodes are assumed to know their location and the limits *S* and *D*.

- ***Description of the Clustering Algorithm***

Initially a set of sensor nodes are dispersed in the terrain. We assume that sensor nodes know their location and the limits *S* and *D*. Algorithms for estimating geographic or logical

coordinates have been explored at length in the sensor network research [21,22].

In our algorithm, the first step is to calculate *Eth* and *Eic* for every node i, N <I< 1. Eth is the energy spent to communicate with the farthest next hop neighbor. *Eic* is the total energy spent on each link of its next hop neighbors. Every node *i* has an initial energy, E_{init}. A flag bit called "*covered flag"* is used to denote whether the node is a member of any cluster or not. It is set to 0 for each node initially.

- ***Calculation of Eth and Eic***

i. Nodes send a message *hello_msg* along with their coordinates which are received by nodes within the transmission range. For example in figure (1) nodes *a*,*b*,*c*,*d*,*w*,*x*,*y* are within transmission range of *v*.

ii. After receiving the *hello_msg*, the node *v* calculates the distance between itself and nodes *a*,*b*,*c*,*d*,*w*,*x*,*y* using the coordinates from *hello_msg*. It stores the distance d_i and the locations in the *dist_table*.

iii. Nodes within the sensing range are the neighbours of a node. In fig (1) nodes *w*,*x*,*y*,*b* are neighbours of *v*.

iv. Among the nodes within the sensing range, it chooses the first *D* closest neighbours as its potential candidates for next hop. Assuming *D*=3, in figure (1), the closest neighbours of *v* are *w*,*x*,*y*.

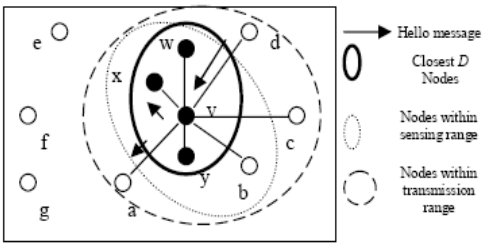


Fig 1. Topology

v. Among the potential candidates, the farthest node's distance, d_{max} is taken for the calculation of E_{th} .
vi. Suppose a node needs power E to transmit a message to another node who is at a distance 'd' away, we use the formula $E = E=kd^{c}$ [7,8], where k and c are constants for a specific wireless system. Usually 2<c<4. In our algorithm we assume k=1, c=2. For a node v, d *1*d^2max = E thv, since there are D members to which a node sends message.
vii. *Eic* is the total energy spent on each of link of the *D* closest neighbours. For a node *v*,

$$E_{ic\,v} = \sum_{i=1}^{D} d_{iv}^2$$

Where k= 1. d_{iv} is the distance between node *i* and node *v*. After the calculation of threshold energy *Eth*, nodes become eligible for cluster head position based on their energies. A node *v* becomes eligible for the cluster head position if it's $E_{init} > Eth_v$. *And* node with the second $E_{init} > Eth_v$ becomes secondary Cluster heed.. When no nodes satisfy this condition or when there is insufficient number of cluster heads, the admissible degree *D* is reduced by one and then *Eth* is recalculated. The lowest value that *D* can reach is one. In a case where the condition $E_{init} > Eth_v$ is never satisfied at all, clustering is not possible because no node can support nodes other than itself. There may also be situations where all the nodes or more number of nodes are eligible for being cluster heads. A method has been devised by which the excess cluster heads are made to relinquish their position.

- ***Relinquishing of the cluster head position***

i. Every cluster head sends a message *cluster_head_status msg* and *Eic* to its neighbours (within sensing range).
ii. Every cluster head keeps a list of its neighbor cluster heads along with its *Eic*
iii. The nodes which receive *Eic* lesser than itself relinquishes its position as a cluster head.
iv. The cluster heads which are active send their messages to the cluster_head_manager outside the network. Thecluster_head_manager has the information of the desired cluster head count.
v. If the number of cluster heads are still much more than expected, then another round of cluster head relinquishing starts. This time the area covered would be greater than sensing range.
vi. The area covered for cluster head relinquishing keeps increasing till the desired count is reached.

- ***Choosing cluster members***

i. The cluster head select the closest *D* neighbours as next hop and sends them the message *cluster_join_msg*. The *cluster_join_msg* consists of cluster ID, S_a, *D*, *S*, *covered flag*. S_a is (*S*-1)/number of next hop members
ii. Energy is expended when messages are sent. This energy, *Eic* is calculated and reduced from the cluster head's energy.
iii. The cluster head's residual energy $E_r = E_{init} - Eic$. E_{init} is the initial energy when the cluster is formed by the cluster head.
iv. After receiving the *cluster_join_msg*, the nodes send a message, *cluster_join_confirm_msg* to the cluster head if they are uncovered, else they send a message, *cluster_join_reject_msg*.
v. After a *cluster_join_confirm_msg*, they set their *covered_flag* to 1.
vi. The next hop nodes now select *D*-1 members as their next hop members. *D*-1 members are selected because they are already associated with the node which selected them. For example, in figure (1) where *D* =3, a node *v* selects node *w*, node *x* and node *y*. In the next stage, node y selects only node *a* and node *b* because it is already connected to node *v* making the *D* = 3.

vii. After selecting the next hop members, the residual energy is calculated, E_r (new) = E_r (old) - *Eic*
viii. This proceeds until S is reached or until all nodes have their *covered_flag* set to 1.

- ***Tracking of the size***

The size *S* of the cluster is tracked by each and every node. The cluster head accounts for itself and equally distributes *S*-1 among its next hop neighbors by sending a message to each one of them. The neighbours that receive the message account for themselves and distribute the remaining among all their neighbours except the parent. The messages propagate until they reach a stage where the size is exhausted. If the size is not satisfied, then the algorithms terminates if all the nodes have been covered. After the cluster formation, the cluster is ready for operation. The nodes communicate with each other for the period of network operation time.

4. Cluster Heed Failure recovery Algorithm

we employ a back up secondary cluster heed which will replace the cluster heed in case of failure. no further messages are required to send to other cluster members to inform them about the new cluster heed. Cluster heed and secondary cluster heed are known to their cluster members. If cluster heed energy drops below the threshold value, it then sends a message to its cluster member including secondary cluster heed. Which is an indication for secondary cluster heed to standup as a new cluster heed and the existing cell manager becomes common node and goes to a low computational mode. Common nodes will automatically start treating the secondary cluster heed as their new cluster heed and the new cluster heed upon receiving updates from its cluster members; choose a new secondary cluster heed. Recovery from cluster heeds failure involved in invoking a backup node to standup as a new cluster heed.

5. PERFORMANCE EVALUATION

The energy model used is a simple model shown in [19] for the radio hardware energy dissipation where the transmitter dissipates energy to run the radio electronics and the power amplifier, and the receiver dissipates energy to run the radio electronics. In the simple radio model [19], the radio dissipates E_{elec} = 50 nJ/bit to run the transmitter or receiver circuitry and E_{amp} = 100 (pJ/bit)/m^2 for the transmit amplifier to achieve an acceptable signal-to-noise ratio. We use MATLAB Software as the simulation platform, a high performance discrete-event Java-based simulation engine that runs over a standard Java virtual machine.

The simulation parameters are explained in Table 1.
We compared our work with that of Venkataraman algorithm [10], which is based on recovery due to energy exhaustion.

Table 1.Simulation parameters

Simulation parameters	Value
terrain dimensions	1 km^2
total number of nodes in terrain, *N*	100−1200
transmission range	100−450 m
cluster size limit, *S*	10−50
supportable degree, *D*	3−10

5.1. Characteristics of the Clusters

Fig.1 depicts the percentage of cluster heads observed with varying cluster ranges. The cluster range was varied from 200 to 400. The size limit, S in our algorithm was set to 50 with admissible degree, D set to 3. The percentage of cluster heads was observed and noted for about 10 runs of the clustering algorithm. The percentage of cluster heads does not vary over various runs of the algorithm. This is because for a total number of N nodes in terrain, the limit S is set to 50 leading to N/50 cluster heads or clusters. Due to this limitation the results do not show a decrease or an increase in the cluster heads. Though the percentage of cluster heads is not changing, the role of cluster heads is exchanged among the nodes in the network.

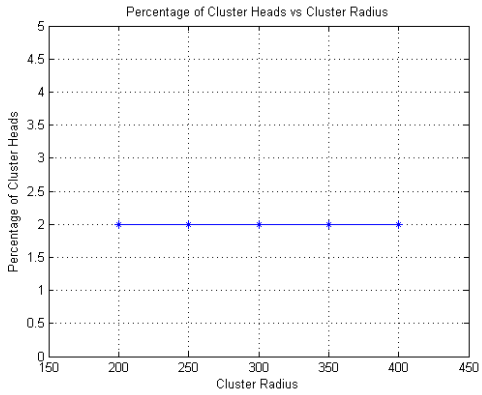
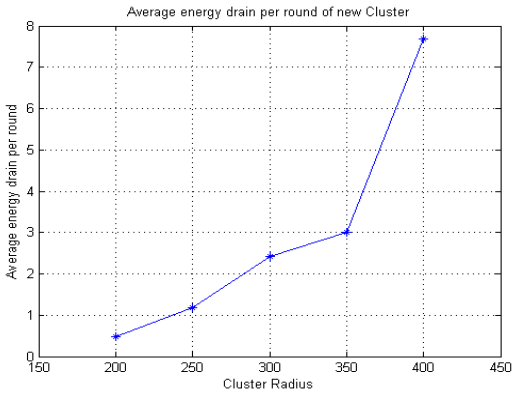


Fig2. Percentage of cluster heads with varying cluster radius

5.2. Energy characteristics in Clusters

Fig.3 depicts the energy drain during the cluster formation. Energy drain refers to the loss of energy in each node after every round of cluster formation and operation. The energy loss is according to the relation in the first order radio model. Total energy loss would be the energy loss due to transmission added to the loss due to receiving. energy consumption depends on parameters used in first order radio model, distance and the number of bits, k. This energy consumption is also dependent on how many nodes would the concerned node be transmitting to and receiving from. In this clustering algorithm, the distance is sensing range, which is



about 50 % of the transmission range. Also the number of nodes each node would handle is D. These two factors make the loss of energy at every stage uniform.

Fig 3. Ratio of average balance energy drain per round with varying cluster radius

In Venkataraman algorithm, nodes in the cluster are classified into four types: boundary node, pre-boundary node, internal node and the cluster head. Boundary nodes does not require any recovery but pre-boundary node, internal node and the cluster head have to take appropriate actions to connect the cluster. Usually, if node energy becomes below a threshold value, it will send a fail_report_msg to its parent and children. This will initiate the failure recovery procedure so that failing node parent and children remain connected to the cluster. A join_request_mesg is sent by the healthy child of the failing node to its neighbors. All the neighbors with in the transmission range respond with a join_reply_mesg /join_reject_mesg messages. The healthy child of the failing node then selects a suitable parent by checking whether the neighbor is not one among the children of the failing node and wEther the neighbor is also not a failing node. In our proposed mechanism, common nodes does not require any recovery but goes to low computational mode after informing their cell managers. In Venkataraman algorithm, cluster head failure cause its children to exchange energy messages. The children who are failing are not considered for the new cluster-head election. The healthy child with the maximum residual energy is selected as the new cluster head and sends a final_CH_mesg to its members. After the new cluster head is selected, the other children of the failing cluster head are attached to the new cluster head and the new cluster head becomes the parent for these children. This cluster head failure recovery procedure consumes more energy as it exchange energy messages to select the new cluster head. Also, if the child of the failing cluster head node is failing as well, then it also require appropriate steps to get connected to the cluster. These can abrupt network operation and is time consuming.

In our proposed algorithm, we employ a back up secondary cluster heed which will replace the cluster heed in case of failure.

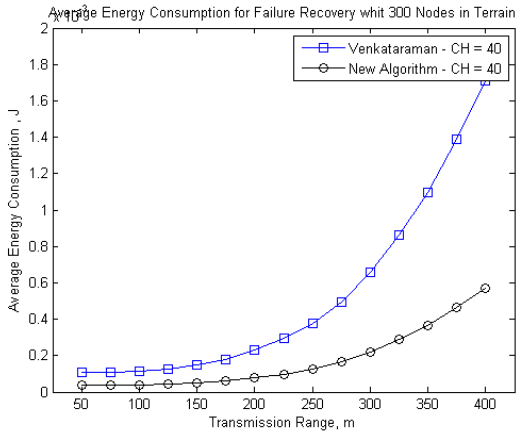


Fig 4. Average energy loss for cluster-head recovery

no further messages are required to send to other cluster members to inform them about the new cluster heed Figs. 4 and 5 compare the average energy loss for the failure recovery of the three algorithms. It can be observed from Fig. 4 that when the transmission range increases, the greedy algorithm expends the maximum energy when compared with the Gupta algorithm and the proposed algorithm. However, in Fig. 5, it can be observed that the Gupta algorithm spends the maximum energy among the other algorithms when the number of nodes in the terrain increases.

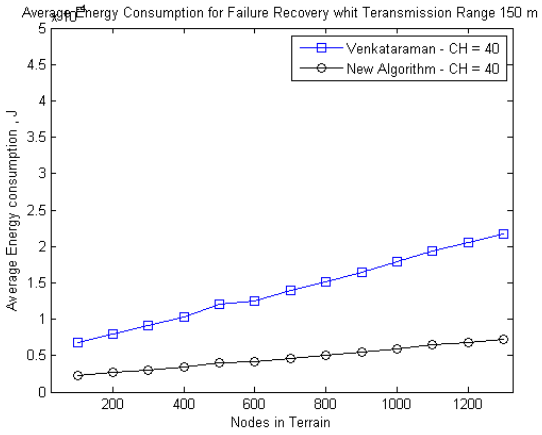


Fig 5. Average energy loss for cluster-head recovery

6. Conclusion

In this paper, we have proposed a cluster-based recovery algorithm, which is energy-efficient and responsive to network topology changes due to sensor node failures.. The proposed cluster-head failure-recovery mechanism recovers the connectivity of the cluster in almost less than of the time taken by the fault-tolerant clustering proposed by Venkataraman . The Venkataraman algorithm is the latest approach towards fault detection and recovery in wireless sensor networks and proven to be more efficient than some existing related work. Venkataraman

algorithm is more energy efficient in comparison with Gupta and Algorithm Greedy Therefore, we conclude that our proposed algorithm is also more efficient than Gupta and Greedy[20] algorithm in term of fault recovery.

The faster response time of our algorithm ensures uninterrupted operation of the sensor networks and the energy efficiency contributes to a healthy lifetime for the prolonged operation of the sensor network.

References

[1] A. Bharathidasas, and V. Anand, "Sensor networks: An overview", Technical report, Dept. of Computer Science, University of California at Davis, 2002

[2] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: Scalable coordination in sensor networks", in Proceedings of ACM Mobicom, Seattle, Washington, USA, August 1999, pp. 263-- 270, ACM.

[3] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "A Survey on Sensor Networks", IEEE Communications Magazine, pp. 102--114, August 2002.

[4] D. Estrin, L. Girod, G. Pottie, M. Srivastava, "Instrumenting the world with wireless sensor networks", In Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 2001.

[5] E. S. Biagioni and G. Sasaki, "Wireless sensor placement for reliable and efficient data collection", in the 36th International Conference on Systems Sciences, Hawaii, January 2003.

[6] G. Gupta and M. Younis, "Load-Balanced Clustering in Wireless Sensor Networks", in the Proceedings of International Conference on Communication (ICC 2003), Anchorage, AK, May 2003.

[7] J. Chen, S. Kher and A. Somani, "Distributed Fault Detection of Wireless Sensor Networks", in DIWANS'06. 2006. Los Angeles, USA: ACM Pres.

[8] F. Koushanfar, M. Potkonjak, A. Sangiovanni- Vincentelli, "Fault Tolerance in Wireless Ad-hoc Sensor Networks", Proceedings of IEEE Sensors 2002, June, 2002.

[9] W. L. Lee, A. Datta, and R. Cardell-Oliver, "Network Management in Wireless Sensor Networks", to appear in Handbook on Mobile Ad Hoc and Pervasive Communications, edited by M. K. Denko and L. T. Yang, American Scientific Publishers.

[10] G. Venkataraman, S. Emmanuel and S.Thambipillai, "Energy-efficient cluster-based scheme for failure management in sensor networks" IET Commun, Volume 2, Issue 4, April 2008 Page(s):528 – 537

[11] L. Paradis and Q. Han, "A Survey of Fault Management in Wireless Sensor Networks", Journal of Network and Systems Management, vol. 15, no. 2, pp. 171-190, 2007.

[12] L. M. S. D. Souza, H. Vogt and M. Beigl, "A survey on fault tolerance in wireless sensor networks", 2007.

[13] W. L Lee, A.D., R. Cordell-Oliver, WinMS: Wireless Sensor Network-Management System, An Adaptive Policy-Based Management for Wireless Sensor Networks. 2006.

[14] L. B. Ruiz, I. G.Siqueira, L. B. Oliveira, H. C. Wong, J.M. S. Nigeria, and A. A. F. Loureiro. "Fault management in event-driven wireless sensor networks", MSWiM'04, October 4-6, 2004, Venezia, Italy

[15] G. Gupta and M. Younis; Fault tolerant clustering of wireless sensor networks; WCNC'03, pp. 1579.1584.

[16] M. Ding, D. Chen, K. Xing, and X. Cheng, "Localized fault-tolerant event boundary detection in sensor networks", in Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05), vol. 2, pp. 902–913, Miami, Fla, USA, March 2005

[17] C. Hsin and M.Liu, "Self-monitoring of Wireless Sensor Networks", Computer Communications, 2005. 29: p. 462-478

[18] S. Chessa and P. Santi, "Crash faults identification in wireless sensor networks", Comput. Commun., 2002, 25, (14), pp. 1273-1282.

[19] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proc. Hawaii Int'l Conf. System Sciences 2000.

[20] GUPTA G., YOUNIS M.: 'Fault-tolerant clustering of wireless sensor networks'. Proc. IEEE WCNC, New Orleans, USA,March 2003, vol. 3, p. 1579 – 1584

[21] N. Bulusu, J. Heidemann and D. Estrin, "GPS-less Low Cost Outdoor Localization For Very Small Devices", IEEE Personal Communications, Special Issue on "Smart Spaces and Environments", Vol. 7, No. 5, pp. 28-34, October 2000.

[22] Radhika Nagpal, "Organizing a Global Coordinate System from Local Information on an Amorphous Computer", MIT AI Memo 1666, August 1999

Cloud Computing: Its security & Privacy Aspects

Dr. Deepshikha Jamwal[1], Abhishek Singh Sambyal[2], Prof.G.S Sambyal[3]

Department of Computer Science & IT

University of Jammu

jamwal.shivani@gmail.com,abhishek.sambyal@gmail.com

ABSTRACT
Cloud computing has changed the way we perceive the picture of the computers. We can get high number-crunching performance, virtually infinite storage capacities, GB's of RAM at very low cost. But as with all other technologies, cloud computing is not on all white side. Cloud computing has its disadvantages too. Privacy is an important issue for cloud computing, both in terms of legal compliance and user trust, and needs to be considered at every phase of design. We must ensure that the risks to privacy are mitigated and that data is not excessive, inaccurate or out of date, or used in unacceptable or unexpected ways beyond the control of data subjects. Also, we must ensure that there are no conflicts regarding the ownership of data and responsibility in case of damage of data. In this paper the authors try to find out the privacy & security challenges and issues that arises when targeting the cloud as their production environment to offer services assessed, and key design principles to address these as suggested by author.

KEYWORDS
Cloud computing, Cloud Computing Security.

1. Introduction

Security is one of the concerns about cloud computing that is delaying its adoption. As only 5% turn over to cloud computing. One of the biggest security concerns about is that when you move your information over the cloud you will loose control of it. The cloud gives you to access data, but you have no way of ensuring no one else has access the data. In a cloud-based software environment, physical security is stronger because the loss of a client system doesn't compromise data or software [1]. Cloud computing seems offer some incredible benefits for communication: the availability of an incredible array of software application, access to lightning-quick processing power, unlimited storage, ad the ability to easily. Cloud computing takes hold as 69% of all internet users have either stored data online or used a web-based software application. "Washington, DC – Some 69% of online Americans use webmail services, store data online, or use software programs such as word processing applications whose functionality is located on the web. In doing so, these users are making use of "cloud computing," an emerging architecture by which data and applications reside in cyberspace, allowing users to access them through any web-connected device [2].

1.1 Applications of cloud computing in different fields.

- A number of companies, including Goole, Microsoft, Amazon, and IBM, have built enormous datacenter-based computing capacity all over the world to support their Web services offerings (search, instant messaging, Web-based retail, etc). With this computing infrastructure in place, these already poised to other new cloud-based, software applications.

- Large enterprise software solutions, such as ERP (Enterprise Resource Planning) application, have traditionally only been affordable to very big enterprises with big IT budgets. However companies that sell these solutions are finding they can reach small to medium business by making their very expensive, very complex applications available as Internet-based software services. These new market segments have encouraged them to expand their SaaS offerings.

- New kinds of hardware, such as Mobile Internet Devices (MIDs), lighter and more portable notebook computers (net book), even high-end smart phones and internet accessibility, make it easier for end-user to log their cloud-based applications any time, any place. This means the market for SaaS is also being driven at the user end by new Internet accessible drivers.

1.2 Needs of Security in Cloud Computing

1. Bribery, extortion and other con games have found new life online. Today, botnets threaten to take vendors down; scammers seduce the unsuspecting on dating sites; and new viruses encrypt your hard drive's contents, then demand money in return for the keys [3].

2. In June 2008, Amazon's U.S. site went off the air, and later some of its other properties were unavailable. It was a denial-of-service attack aimed at the company's load-balancing infrastructure.

3. Wanting to survive an attack is yet another reason for start-ups to deploy atop cloud computing offerings from the likes of Amazon, Google, Joyent, XCalibre, Bungee, Enki and Heroku. [4].

4. Cloud operators need to find economies of scale in their security models that rival the efficiencies of hackers.

2. Security Issues of Cloud Computing

Here are the few real issues that are vitally important to grapple with –

- **ENCRYPTING DATA-AT-REST –** Alex Stamos, a prominent BlackHat researchers raised this challenge for cloud computing providers/ Stamos' point is not that many cloud providers don't offer encryption (true) but also that there are inherent challenges in producing sufficiently random encryption keys. Fascination stuff you can read here if you're interested.
- **DON'T MISS THE BIGGER POINT THROUGH –** while much attention is given to network security (firewall/IDS), no one is talking about the need for securing data-at-rest in a cloud environment…and we really should. Depending in the kind of data you're dealing with, it may not be necessary to require encryption for all corporate data. However, the need to except critical data-at-rest like access information (username, password), privacy related and corporate sensitive information is a must. As a consumer of cloud based services or applications, you need to ask the right question and demand the right answers from your provider
- **SECURITY POLICY CONTROL**. Here's an area where cloud computing security has to grow up. For cloud computing to go main stream beyond SMB market, it must offer IT organizations the ability to enforce corporate policy. This policy control ranges from the simple daily policy issues (like enforcing rules to ensure "strong" passwords) to the more complex (like conducting security related forensics) and everything in between. Many cloud providers fail to offer the kind of granular policy control that many organizations require.

- **THE HUMAN FACTOR** - what about the people who keep the cloud running? Does your cloud provider conduct background checks on all the individuals who have access to your data? By the way, do you know who specifically has access to your data? How about policies that govern who can and cannot see your data? The point is that most security breaches are not about arcane attacks but really about social engineering and negligent or lax policies. This is why I firmly believe every consumer of cloud based applications and services should demand their providers to be SAS 70 audited. If nothing else, this ensures transparency and consistency of policy enforcement [5].

2.1 Some Other Issues

There are some issues defined below that need to be addressed before enterprises consider switching to the cloud computing model. They are as follows:

1. **PRIVILEGED USER ACCESS** - information transmitted from the client through the Internet poses a certain degree of risk, because of issues of data ownership; enterprises should spend time getting to know.

2. **REGULATORY COMPLIANCE** - clients are accountable for the security of their solution, as they can choose between providers that allow to be audited by third party organizations that check levels of security and providers that don't.

3. **DATA LOCATION** - depending on contracts, some clients might never know what country or what jurisdiction their data is located.

4. **DATA SEGREGATION** - encrypted information from multiple companies may be stored on the same hard disk, so a mechanism to separate data should be deployed by the provider.

5. RECOVERY - every provider should have a disaster recovery protocol to protect user data

6. INVESTIGATIVE SUPPORT - if a client suspects faulty activity from the provider, it may not have many legal ways pursued an investigation.

7. LONG-TERM VIABILITY - refers to the ability to retract a contract and all data if the current provider is bought out by another firm [6].

Given that not all of the above need to be improved depending on the application at hand, it is still paramount that consensus is reached on the issues regarding standardization.

3. Challenges faced to overcome the security issues.

Challenges that cloud computing currently faces in being deployed on a large enterprise scale:

1. **SELF-HEALING** - in case of application/network/data storage failure, there will always be a backup running without major delays, making the resource switch appear seamless to the user.

2. **SLA-DRIVEN** - cloud is administrated by service level agreements that allow several instances of one application to be replicated on multiple servers if need arises; dependent on a priority scheme, the cloud may minimize or shut down a lower level application.

3. **MULTI-TENANCY** - the cloud permits multiple clients to use the same hardware at the same time, without them knowing it, possibly causing conflicts of interest among customers.

4. **SERVICE-ORIENTED** - cloud allows one client to use multiple applications in creating its own.

5. **VIRTUALIZED** - applications are not hardware specific; various programs may run on one machine using virtualization or many machines may run one program.

6. **LINEARLY SCALABLE** - cloud should handle an increase in data processing linearly; if "n" times more users need a resource, the time to complete the request with "n" more resources should be roughly the same.

7. **DATA MANAGEMENT** - distribution, partitioning, security and synchronization of data [7].

4. Methodology

The methodology used by the authors consists of two types i.e. primary and secondary method.

The primary method consists of questionnaire containing questions like percentage usage of cloud computing by an organisation, bandwidth consumed by originations etc.

The second method is the personal interviews, visiting organisations, consulting journals, research papers, various published articles in reputed journals & magazines.

5. Analysis

As per the authors study reveals that online users who take advantage of "cloud" applications say they like the convenience of having access to data and applications from any Web-connected device. At the same time, however, they express high levels of concerns about storing personal data online when presented with scenarios about possible uses of their data by companies providing cloud services. "These findings give consumers, the technology community, and policymakers a chance to discuss the tradeoffs between convenience and privacy and figure out where there are needs for education to improve public understanding."

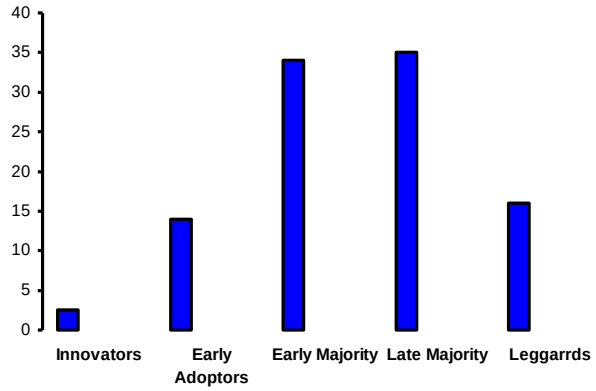


Figure 1: Percentage Usage of Cloud Computing

This above analysis shows there are very less % of innovators i.e. 2.5%, there are 3.5% of Early adopters, 34% of Early Majority, again 34% of Late Majority, and 16% of Laggards.

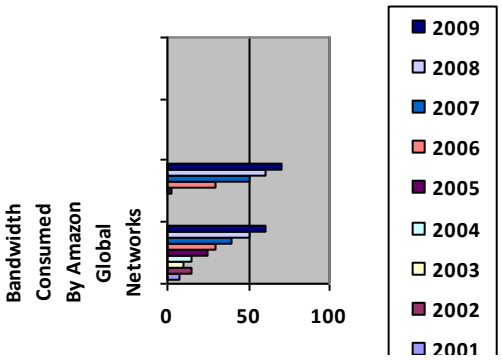


Figure 2. Bandwidth usage by different networks services.

In this above survey, there is a comparison between the Bandwidth Consumed by Amazon Global Networks and the Bandwidth Consumed by Amazon Web Services from 2001 to 2009.

6. Conclusion

After the thorough study done by the authors the cloud computing data comes from a survey of 2,251 adults. Some 1,553 respondents in the survey were internet users and the margin of error is plus or minus 3 percentage points for results based on internet users. The main purpose of this paper is to investigate what benefits the cloud computing

technology can bring to an organization. "Even as large numbers of users turn to 'cloud computing' applications, many may lack a full understanding of possible consequences of storing personal data online. In this paper the author has presented the technologies for determining various applications or workloads which are most amenable to cloud deployment.

7. Future Findings

1. To find the standardized platform for the companies to shift from one cloud providers to another.
2. To find out how market leaders using best practice in cloud computing and how it may benefit to web 2.0.
3. To find out what risk migrations strategies need to be adopted.

References

[1] Dikaiakos et.al, "***Cloud Computing: Distributed Internet Computing for IT and Scientific Research",*** IEEE, Volume 13, Issue 5, Sept.-Oct. 2009, Page: 10 - 13.

[2] Liang-Jie Zhang et.al, "***CCOA: Cloud Computing Open Architecture***", IEEE, 6-10 July 2009, Page(s):607 – 616.

[3] Kandukuri et.al, ***"Cloud Security Issues***", IEEE, , 21-25 Sept. 2009, Page(s):517 – 520.

[4] Kaufman et.al, ***"Data Security in the World of Cloud Computing***", IEEE, Volume 7, Issue 4, , July-Aug. 2009, Page(s):61 – 64.

[5] Xiong, Kaiqi et.al, "***Service Performance and Analysis in Cloud Computing***", IEEE, 6-10 July 2009, Page(s):693 – 700.

[6] Lin, G et.al, "***Cloud Computing and IT as a Service: Opportunities and Challenges***", IEEE, 23-26 Sept. 2008, Page(s):5 – 5.

[7] Cong Wang et.al, ***"Ensuring data storage security in Cloud Computing***", IEEE, 13-15 July 2009, Page(s):1 – 9.

High Performance Analysis of Frequency Reuse Schemes in Cellular Mobile Environment

A.K.M Fazlul Haque[1], A.F.M. Shahen Shah, Md. Abdul Hannan, Nusrat Jahan, jasim Uddin Ahmed, Md. Abu Saleh
Department of Electronics and Telecommunication Engineering
Daffodil International University
Email: [1]akm_haque@yahoo.com, [1]akmfhaque@daffodilvarsity.edu.bd

***Abstract*---** *Frequency planning is the prime important issues for the Radio Frequency (RF) engineers. The principle goal of the analysis done in various frequency reuse factor is for proper frequency planning. In this paper, the performances for frequency reuse schemes have been analyzed. Using MATLAB R2008a, the performance of frequency reuse factors for both OMNI and sectoring cell has been evaluated. In order to fulfill the traffic demand and all the factors related with the performance of respective model for different environment have also been proposed. With considering the traffic variation in different geographical areas i.e. rural, urban, highways etc., the respective model not only provide the satisfactory level of traffic but also provide the standard interference level.*

Keywords- *Frequency reuse schemes, omni, sectoring, traffic, grade of service, trunking efficiency, co-channel interference.*

1. Introduction

Frequency planning means to optimize spectrum usage, enhance channel capacity, and reduce channel interference [1]. Frequency planning is not a major concern in rural areas rather it becoming typical in urban areas. Due to the increasing traffic demand and higher interference level in urban areas, a RF engineer should provide a good solution which fulfils the demand for operators as well as customers. Since there are a limited number of channels, these channel groups are reused in regular distance intervals. This is an important engineering task that requires a good compromise between capacity and performance. The mechanism that governs this process is called frequency planning [1-5]. Through the paper has been proposed frequency planning for urban, rural based on the analysis for both omni and sectoring cell. A cell is a geographical area covered by the RF signal. The cell uses omni directional antenna is known as omni cell. An omni directional antenna is an antenna system which radiates power uniformly in all direction (360 degree). Sectorization scheme is achieved by dividing a cell into number of sectors. It may 3sectors (120 degree), 4 sectors (90 degree), 6 sectors (60 degree) etc. each sector is treated as a logical omni cell, where directional antenna are used in each sector. Here the directional antenna may be semi directional or highly directional. In mobile cellular system, the frequency resource should be used repeatedly over some geographic area. There are some works on frequency reuse scheme [6-9]. Ki Tae Kim [6] et al proposed a universal frequency reuse (UFR) system that can reuse effectively a given frequency resource in a mobile cellular environment. Z. Wang et al [7] proposed a new frequency reuse scheme for coherent orthogonal frequency division multiplexing (OFDM) cellular systems using high level modulation schemes, whereby both data and pilot parts can have different frequency reuse factors, is described. Syed Hussain Ali and Victor C. M. Leung [10] proposed dynamic frequency allocation in fractional frequency reused OFDMA networks. M. Zhang et al [9] proposed an alternate row antenna rotation (ARAR), for 4x12 GSM frequency reuse. In their work, the cell edge coverage is mutual enhanced by adjacent sectors thus no dead spot exists and narrow-beam antenna can be used. Theodore S. Rappaport and Robert A. Brickhouse [11] provide a simulation of cellular system growth and its effect on urban in-building parasitic frequency reuse. Romeo Giuliano and Cristiano Monti [12] analyzed the Fractional Frequency Reuse (FFR) scheme in rural areas evaluating the increase of the overall system capacity. Most of the works proposed their individual model and they analyzed the performances based on their given model only. So, it is needed to analyze the performance of all existing frequency reused schemes where it is used in various environments in the world. In this paper, the concept of FRF 7 OMNI cell in rural areas, FRF 4 – 120 degree sector cell in urban areas have been proposed based on the existing performance criteria and found the satisfactory level of traffic. Here the performance of each model in different environment has been analyzed and by this way it can be understood which model is fast convergence or efficiently work in specific environment. Applying the experimented efficient model, RF engineer

can optimize spectrum usage, enhance channel capacity and reduce channel interference.

2. Backgrounds

Four factors for performance analysis have been considered. They are Channel Capacity or Traffic, Grade of Service (GoS), Trunking efficiency, and co channel interference or carrier to interference ratio (CCI or C/I). Channel Capacity is measured by the available voice channels per cell, translated into Erlangs, known as Traffic. Grade of Service (GoS) is defined as the probability of call failure. It is measured in percentage. Thus the GoS lies between 0 and 1 i.e. 0<GoS<1. Let N be the total number of channel, T be the offered Traffic in erlangs, then the probability that all the channel being busy will be given by following Poisson's Distribution [1]

Where $P(N;T) = \frac{(T^N * e^{-T})}{N!}$ P(N;T) is the blocking rate or GoS. Trunking efficiency also known as channel utilization efficiency is determined by the amount of traffic per channel. Co channel interference or carrier to interference ratio (C/I) arises from multiple use of same frequency. The carrier to interference ratio (C/I) is defined by

$$\frac{C}{I} = 10\log\left[\frac{1}{j} * \left(\frac{D}{R}\right)^{\gamma}\right] ----(1)$$

Where, j = number of co channel interferer (j= 6 in omni and 2 in sectoring cell (1st tier)), γ = propagation constant (γ=4 in cellular mobile environment), D = frequency reuse distance, R = radius of the cell. The above equation may be written as [6]

$$\frac{C}{I} = 10\log\left[\frac{1}{j} * \left(\sqrt{3K}\right)^{\gamma}\right] ----(2)$$

Where,

$\frac{D}{R} = \sqrt{3K}$; K= frequency reuse factor

3. Simulations and Results

The analysis has been done for both OMNI and sectoring cell for urban and rural areas using MATLAB R2008a . The simulated values for frequency reuse factors are shown in table 1 for OMNI cell and table 2 for sectoring cell and the performances are also described in Figure 2, figure 3 and figure 4. The algorithm of the evaluation process has also been introduced in appendix.

Table 1. Performance evaluation of frequency reuses factors for OMNI cell

FRF	No. of voice channel/cell	Traffic (1% GoS) Erlang	Traffic (2% GoS) Erlang	Traffic (3% GoS) Erlang	Trunking efficiency (2% GoS)	C/I (dB)
1	375	350.75	360.8	368.4	96.2%	1.78
3	125	107.75	112.3	115.6	89.8%	11.3
4	94	78.4	82.2	84.9	87.5%	13.8
7	54	41.5	44	45.8	81.5%	18.7
9	42	30.8	32.8	34.3	78%	20.9
12	32	22	23.7	24.9	74%	23.3

Table 2: Performance evaluation of frequency reuse factors for Sectoring cell

FRF	No. of sectors	No. of channel	Traffic (1%GoS) Erlang	Traffic (2%GoS) Erlang	Traffic (3%GoS) Erlang	Trunking efficiency (1%GoS)	Trunking efficiency (2%GoS)	Trunking efficiency (3%GoS)	C/I (dB)
1	3	125	107.75	112.3	115.6	86%	89%	92%	6.53
	4	94	78.4	82.2	84.9	83%	87%	90%	
	6	63	49.7	52.5	54.5	79%	83%	87%	
3	3	42	30.8	32.8	34.3	73%	78%	82%	16
	4	32	22	23.7	24.9	69%	74%	78%	
	6	20	12	13.2	14	60%	66%	70%	
4	3	32	22	23.7	24.9	69%	75%	78%	18.7
	4	24	15.3	16.6	17.6	64%	69%	73%	
	6	16	8.88	9.83	10.5	56%	61%	66%	
7	3	18	10.8	11.5	12.2	58%	64%	68%	23.4
	4	14	7.35	8.20	8.80	53%	59%	63%	
	6	9	3.78	4.34	4.75	42%	48%	53%	

A. Frequency Plan for Rural areas

This is the area where the population density is low. High gain antenna with wide opening angle is suitable in this kind of coverage. Since the population density (Traffic) is very low in this environment, capacity and traffic is not a great issue here which is shown in figure 1. But simply it is said about coverage. Also multi path propagation is not a factor in this environment. Therefore FRF 7 OMNI cell concept is proposed. In this environment, there is no need to sector the cell due to the absence of traffic issue. So omni cell concept is appeared the better here not only satisfy traffic demand but also cover the large area by increasing its power.

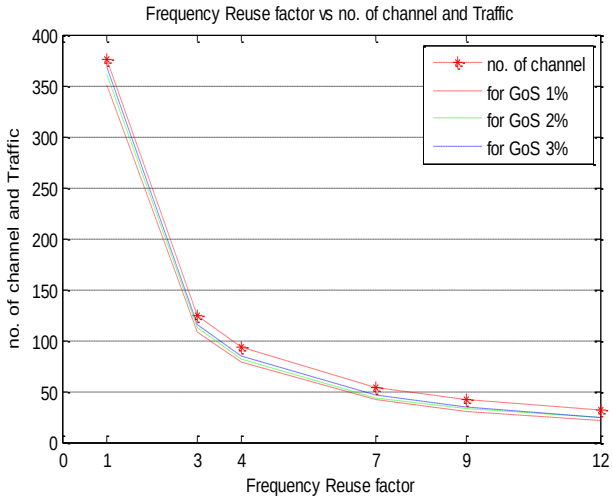


Fig.1A

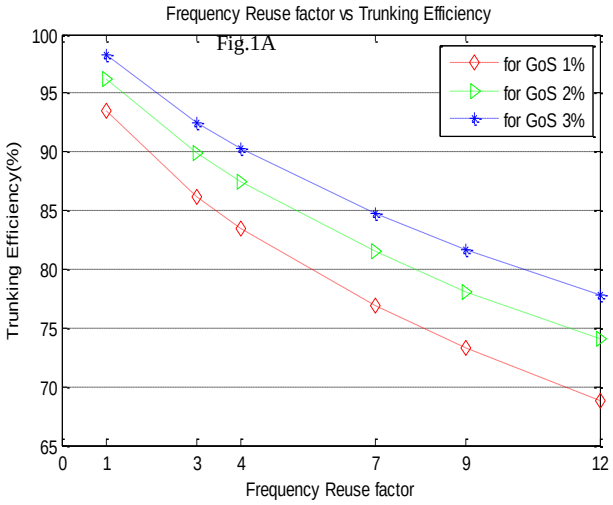


Fig.1B

Fig. 1(1A+1B) Graphical representation of Traffic and Trunking efficiency 1, 2, and 3% GoS

Since the numbers of available channels are limited, therefore, the frequency is needed to be reused properly to get the desired demand. It is seen that traffic depends on number of channels and trunking efficiency depends on traffic, and finally, number of channel depends on frequency reuse factors (FRF) which is shown in figure 2. It is obvious that FRF

number is increased and corresponding channel is decreased. Therefore, Traffic also is decreased as shown in Fig. 1 and 2.

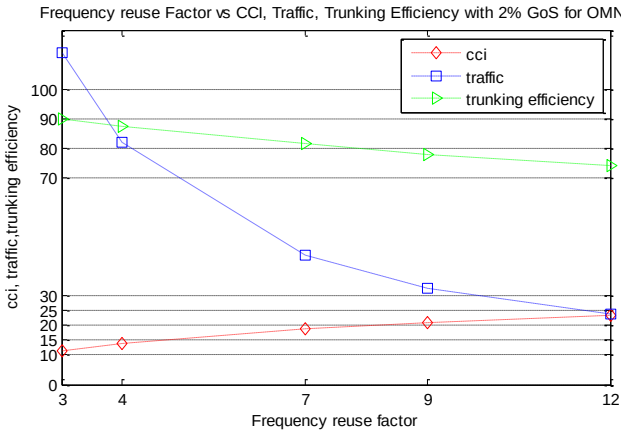


Fig. 2 CCI (C/I), Traffic, Trunking efficiency with 2% Gos

B. Frequency plan for urban areas

Frequency planning is the most important concern for radio engineers especially in urban environment. Because in this environment the RF signal faced lots of impairments. The most important factor is multipath propagation due to random buildings. The received signals in this environment are a result of direct rays, reflected rays, and shadowing or any combination of these signal components. Therefore interference is great issue in this environment. Another important issue is about its population density. The urban environment is highly populated i.e. high traffic density. Therefore more voice channel must be ensuring to provide Quality of Service (QoS). Considering all above factors and evaluated performance, FRF 4, 120 degree sectorization (3 sectors) has been proposed.

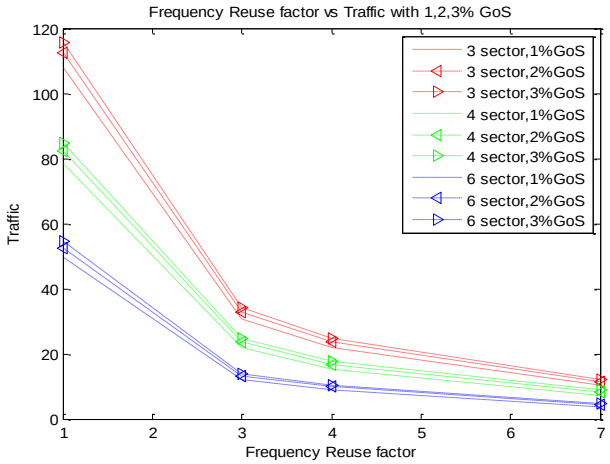


Fig.3A

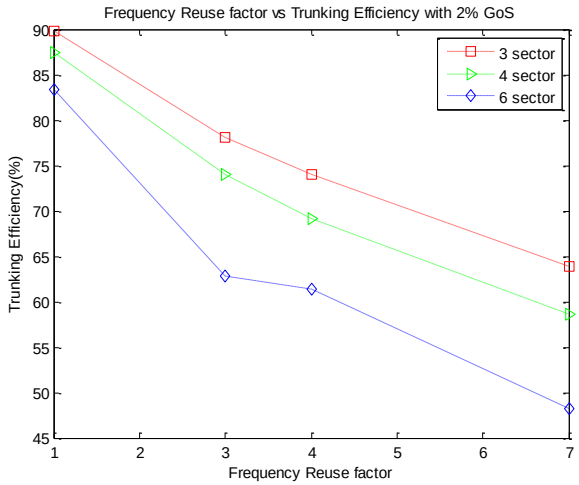


Fig.3B

Fig.3 (3A+3B) Traffic and Trunking efficiency for 3, 4, 6 sectoring cell

Now there is question may arise that why it doesn't choose OMNI cell due to FRF 4 for omni provides greater number of channel, traffic, and also trunking efficiency as well which is shown in figure 4. According to the reference of Electronic Industry Association (EIA) standard about C/I, the standard defines that C/I ≈ 17 dB [1]. But FRF 4 OMNI gives C/I =13.8dB. FRF 4 in OMNI cell should not be used. Because in this case interference is so high that hampers desired Quality of Service (QoS).

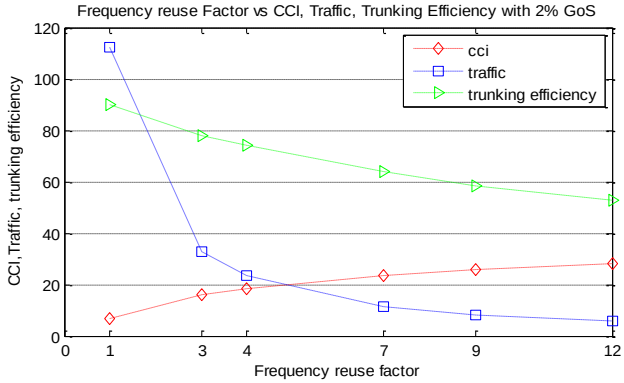


Fig. 4 C/I, Traffic, and Trunking efficiency for 3 sector cell with 2% GoS

From the Fig. 3, it is seen that the number of channel has been reduced in sectoring cell. As stated earlier that with the decreasing of number of channel, corresponding traffic and trunking efficiency is decreased. It is worth mentioning that the above concept is valid only for omni cell. If one cell in sectoring is considered then the concept is true as shown in figure 3. In reality, in case of

sectoring cell, the frequency is reused frequently, and enhanced the number of channel.

4. Conclusion

In this paper, the performances of frequency reuse schemes in mobile cellular environment have been simulated and evaluated. Based on simulation, two respective models (sectoring and omni) for two propagation environment (urban and rural) have been proposed. The demand of the traffic and all factors related with performance respective model for different environment has also been proposed. Finally it is experimented in different geographical areas, especially in rural and urban and found the satisfactory level of traffic.

References

[1] Faruque, Saleh, "Cellular Mobile System Engineering", Artech House Publishers, 1996.

[2] Lee, William C.Y, "Mobile Cellular Telecommunication", McGraw – Hill Book Company, 1995.

[3] Samad, Md. Abdus, "Communication Engineering Fundamental", UGC Publication, Bangladesh, 2009.

[4] Carpenter, Tom, "Certified Wireless Network Administator", McGraw – Hill Book Company, 2007.

[5] Faruque, Saleh, "High Capacity Cell Planning Based on Fractional Frequency Reuse with Optimum Trunking Efficiency", Vehicular Technology Conference, 1998. VTC 98. 48th IEEE, Page(s): 1458 - 1460 vol.2.

[6] Ki Tae KimO, Seong Keun Oh, "A Universal Frequency Reuse System in a Mobile Cellular Environment" Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th, Page(s): 2855 – 2859.

[7] Z. Wang, and R A. Stiring-Gallacher, "Frequency reuse scheme for cellular OFDM systems," Electron. Lett., vol. 38, no. 8, pp. 387-388, Apr. 2002.

[8] T. S. Rappaport, Wireless communications principles and practice, PH PTR, 2nd ed., 2002.

[9] M. Zhang, and V. H. MacDonald, "Generalized cell planning technique applied for 4x12 frequency-reuse," in Proc. PIMRC2003, vol. 2, pp. 1884-1888, Sep. 2003.

[10] Syed Hussain Ali and Victor C. M. Leung, "Dynamic Frequency Allocation in Fractional Frequency Reused OFDMA Networks", IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 8, NO. 8, AUGUST 2009, Page(s): 60 – 65.

[11] Theodore S. Rappaport and Robert A. Brickhouse, "A Simulation of Cellular System Growth and its Effect on Urban In-Building Parasitic Frequency Reuse", IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 48, NO. 1, JANUARY 1999, pp. 286 – 294.

[12] Romeo Giuliano and Cristiano Monti, "WIMAX FRACTIONAL FREQUENCY REUSE FOR RURAL ENVIRONMENTS", IEEE Wireless Communications, June 2008, Vol 15, Issue 3. pp. 60-65.

Appendix

Algorithm sectoring

```
Begin
Set number of co channel interferer (js)= 2;
Set grade of service (gos) = 0.02;
Set frequency reuse factor (frf) =[1 3 4 7];
Set total channel (tch)=[375 375 375 375];
Set number of channel per cel (cec) = ceil(tch./frf);
```

```
Set number of channel per sector with 3 sector (cps3) = ceil(cec/3);
Set number of channel per sector with 4 sector (cps4) = ceil(cec/4);
Set number of channel per sector with 6 sector (cps6) = ceil(cec/6);
```

// Initialization
Traffic with 1% gos for 3 sector (trs13) = [107.75 30.8 22 10.4];
Traffic with 2% gos for 3 sector (trs23) = [112.3 32.8 23.7 11.5];
Traffic with 3% gos for 3 sector (trs33) = [115.6 34.3 24.9 12.2];
Traffic with 1% gos for 4 sector (trs14) = [78.4 22 15.3 7.35];
Traffic with 2% gos for 4 sector (trs24) = [82.2 23.7 16.6 8.20];
Traffic with 3% gos for 4 sector (trs34) = [84.9 24.9 17.6 8.80];
Traffic with 1% gos for 6 sector (trs16) = [49.7 12 8.88 3.78];
Traffic with 2% gos for 6 sector (trs26) = [52.5 13.2 9.83 4.34];
Traffic with 3% gos for 6 sector (trs36) = [54.5 14 10.5 4.75];

//Processing
Trunking efficiency with 1% gos for 3 sector = (trs13./cps3).*100;
Trunking efficiency with 2% gos for 3 sector = (trs23./cps3).*100;
Trunking efficiency with 3% gos for 3 sector = (trs33./cps3).*100;
Trunking efficiency with 1% gos for 4 sector = (trs14./cps4).*100;
Trunking efficiency with 2% gos for 4 sector = (trs24./cps4).*100;
Trunking efficiency with 3% gos for 4 sector = (trs34./cps4).*100;
Trunking efficiency with 1% gos for 6 sector = (trs16./cps6).*100;
Trunking efficiency with 2% gos for 6 sector = (trs26./cps6).*100;
Trunking efficiency with 3% gos for 6 sector = (trs36./cps6).*100;

// Output
Co-channel interference for sectoring = 10*(log10((1/js)*(3*frf).^2));
End

Algorithm Omni
Begin

```
Set number of co channel interferer (j) = 6;
Set grade of service = 0.02;
Set frequency reuse factor (frf) = [1 3 4 7 9 12];
Set total channel (tch) = [375 375 375 375 375 375];
Set number of channel per cel (cec) = ceil(tch./frf);

//Initialization
```

Traffic for 1% gos (tr1) = `[350.75 107.75 78.4 41.5 30.8 22];`
Traffic for 2% gos (tr2) = `[360.8 112.3 82.2 44 32.8 23.7];`
Traffic for 3% gos (tr3) = `[368.4 115.6 84.9 45.8 34.3 24.9];`

```
//Processing
```

Trunking efficiency for 1% = `(tr1./cec).*100;`
Trunking efficiency for 2% = `(tr2./cec).*100;`
Trunking efficiency for 3% = `(tr3./cec).*100;`

```
//Output
```

Co-channel interference `=10*(log10((1/j)*(3*frf).^2));`
End

Handgrip Recognition on Joystick

Zong Chen

School of Computer Sciences and Engineering
Fairleigh Dickinson University
1000 River Road, Teaneck, NJ 07666
zchen@fdu.edu

Abstract: Dynamic handgrip recognition is a new biometric authentication method based on the human grasping behavior. Handgrip pattern recognition seeks to analyze the dynamics inherent in grasping behavior such as how the pressure varying during the grasping process. A real-time biometric system based on handgrip pattern is proposed for aircraft pilot identification in this paper. An experiment was initialized and the results proved that handgrip pattern recognition performance.

Keywords: Biometric authentication, handgrip recognition, pilot identification.

1. Introduction

In response to the growing threats of global terrorism, aircraft designers are looking to develop technology in planes designed to foil hijacking began after the 11 September terrorist attacks. An aircraft anti-hijacking system includes components for producing informational signals reflecting conditions on board. A transceiver communicates the aircraft with at least one remote guidance facility. In case of hijacking, an emergency signal would be sent to a remote ground facility. Responsive to the emergency signal, the ground facility deactivates on-board control of selected aircraft flight systems and the autopilot system, and directs the autopilot to fly the aircraft to a safe landing.

The core component of the anti-hijacking system is the pilot identification device. Many technologies have been studied for pilot identification. The most well known technology is fingerprint identification. Humans have used fingerprint for personal identification for centuries and the validity of fingerprint identification has been well established. However, fingerprint identification can not prevent 9/11-style hijacking.

Starting 2008, Israel requires pilots who fly to its airports to use the Security Code System (SCS). The Security Code System (SCS), also known as Code Positive, is a pilot identification system for civil aircraft flying in Israeli airspace. Developed by Israel's Elbit Systems, Code Positive consists of a smart card reader installed on every aircraft participating in the program, and a smart card that is assigned to each pilot. When the aircraft approaches Israeli airspace, the pilot must verify his or her identity using the card with the control center. However, the smart card could be taken from pilot by terrorist and be used to defraud the security system.

Dynamic handgrip recognition is a novel behavior based biometric technology. The behavioral biometric technology is based on the fact that human actions are predictable in the performance of repetitive, routine tasks [1]. Dynamic keystroke recognition and handwritten signature verification are such examples. In 1895, observation of telegraph operators showed that each operator had a distinctive pattern of keying messages over telegraph lines [2]. When a person types, the latencies between successive keystrokes, keystroke duration, finger placement and applied pressure on the keys can be used to construct a unique signature for that individual [3]. For regularly typed strings, such as username can be quite consistent. Researches have also shown that a unique handwritten signature template can be built for an individual. Each time he signs, he traces out his template, not exactly identical but with small stochastic variations [4]. These variations include the speed of writing, rotation, scale, and shear. Society has relied on the written signature to verify the identity of an individual for hundreds of years. The complexity of the human hand and its environment make written signatures highly characteristic and difficult to forge precisely.

Dynamic handgrip recognition seeks to analyze the dynamics inherent in human grasping process. How hard does people grasp the object? How does the pressure vary from the beginning to the end of the grasping process? How and when does the pressure cross a threshold? There are a number of such parameters inherent within the dynamic process. By comparing those features, an authorized user can be recognized. Handgrip pattern recognition had proven its accuracy in [5] and [6]. A pilot identification system based on handgrip pattern recognition technology is proposed in this paper. An experiment was initialized and the results proved that handgrip pattern recognition performance.

2. Data Acquisition From Joystick

A piezoelectric ceramic PZT disk has been configured as a pressure sensor. The PZT disk can translate the pressure variation on it into voltage change. There are eleven piezoelectric sensors embedded into the joystick handle and wired to operational amplifiers that are located in the joystick base. Since most users are using right hand to grasp the joystick, eight sensors are put at the front side of the joystick handle and three are put at the other side as shown in Figure 1. The push button trigger switch provides the trigger signal (data collection start point), and is connected through channel 12.

Figure 1. Sensor arrangement on joystick

The handgrip pattern recognition focuses on the pressure variation over the time not the absolute amplitude of the handgrip signal, because the absolute amplitude of the handgrip signal may have big variations in different cases. The hand pressure will be much weaker when a person is sick or injured. However, the relative pressure waveform will not change much in such cases. Normalization needs to be done before to create patterns for each subject.

To normalize the data, all DC values from all samples were removed first. Then, the maximum pressure value on all sensors from each sample was calculated and set to +/- 1. The other values were normalized in the same ratio so that the pressure waveform kept same.

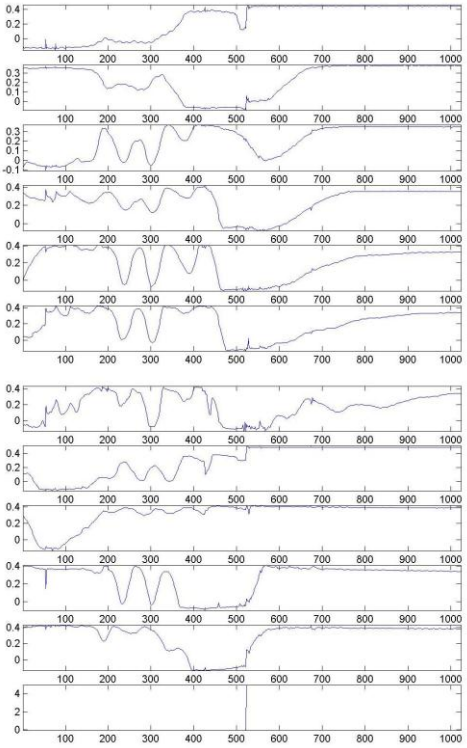


Figure 2. Data collected from joystick.

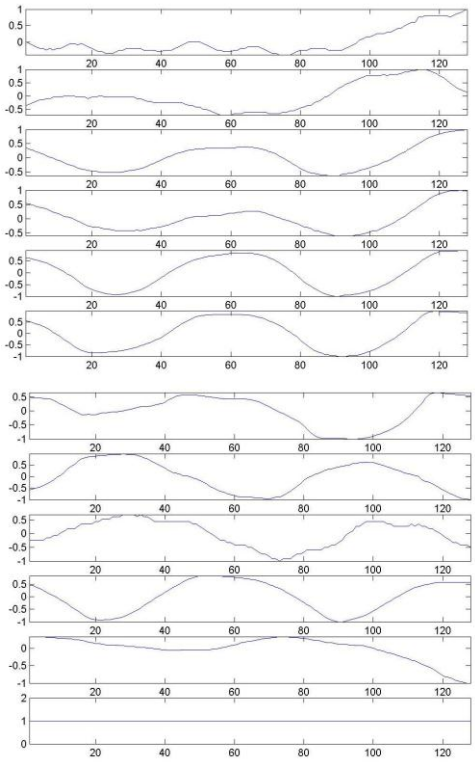


Figure 3. Details in the 128ms before trigger pulled.

A DAQ PC-card (DAQ/112B from IoTech Company) is used to collect pressure data at the sample rate 1000 samples/second. Figure 2 shows what happened during 1 second (1000 ms) of data collecting process.

Time constraint of the pressure signal needs to be considered. Only the pressure variation during a short period when the joystick trigger pulled is considered as the valid signal. Generally, people can click a button in 100ms or remove his/her hand from a button in 200ms when this person receives a "go" or "stop" signal. Therefore, 100ms is considered as the psychological moment. It means that if an action is done in 100ms, people will consider this action done instantaneously. Therefore, the handgrip signal during the period of 100ms when the trigger pulled is the signal needs to be studied.

Note in the next sections, the data during 128ms before the trigger being pulled instead of 100ms will be used because 128 is the nearest integer of 2^n to 100, and it is more convenient to use 2^n to do wavelet transform or other transform in the future. The detail of what happens during the 128ms before the trigger being pulled is shown in Figure 3.

3. Similarity Measurement

In order to depict the similarity between two time series, it needs to define a similarity measurement during the matching process. Given two time series *x* and *y*, a standard approach is to compute the Manhattan distance between time series *x* and *y*. The Manhattan distance function on one channel is defined as

$$D(x,y)_{ch} = \sum_{t=1}^{T} |x(t) - y(t)| \quad (1)$$

And the total Manhattan distance over all channels is defined as

$$D(x,y) = \sum_{ch=1}^{m}\sum_{t=1}^{T} |x(ch,t) - y(ch,t)| \quad (2)$$

Given the template data $\{y_1(ch, t), y_2(ch, t), \ldots, y_n(ch, t)\}$, an pattern for *y* can be created as

$$\bar{y}(ch,t) = \frac{1}{n}\sum_{i=1}^{n} y_i(ch,t) \quad (3)$$

If the distance $D(x,\bar{y})$ smaller than the threshold, then *x* is considered to similar to *y*.

Similarity measurement can be calculated in transform domain too. Wavelet transforms have advantages for representing functions that have discontinuities and sharp peaks, and for accurately deconstructing and reconstructing finite, non-periodic and/or non-stationary signals. There are a wide variety of popular wavelet algorithms, including Haar wavelets, Daubechies wavelets, Mexican Hat wavelets and Morlet wavelets. Due to the approximation quality and the simplicity, the Haar wavelet, which requires the least computation, becomes a good choice for the real-time work.

The original handgrip data would be transformed to Haar wavelet coefficients. Then, use the same distance measurement defined above to measure the wavelet distance between *x* and *y*. The input *x* would be identified as a qualified user only when both the distance in time domain and distance in wavelet domain between *x* and *y* are smaller than the thresholds.

4. Real-time Handgrip Recognition

The patterns (average waveforms in time domain and wavelet domain) for each subject need to be built up in training process. Then the in-group average distance and standard deviation need be calculated for each subject. Smaller in-group average distance and standard deviation means more consistent grasping behavior of this subject, and therefore smaller threshold could be selected.

The in-group average distance and standard deviation for subject *y* can be calculated as

$$D_{group} = \frac{1}{n}\sum_{i=1}^{n} D(y_i, \bar{y}) \quad (4)$$

$$D_{std} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (D(y_i, \bar{y}) - D_{group})^2} \quad (5)$$

By Chebyshev's inequality, an observation is rarely more than a few standard deviations away from the mean. Especially, at least 75% of the values are within 2 standard deviations from the mean. Therefore, the threshold could be determined by

$$Threshold = D_{group} + 2D_{std} \quad (6)$$

The handgrip pattern recognition algorithm is presented as following:

1. Calculate the distances between the test data *x* and the stored patterns in time domain and wavelet domain.

2. Find out the which pattern the test data x closest to by selecting the smallest distance from $D(x, \bar{y}_j)$, $j = 1, 2, ..., N$, where there are N patterns totally.
3. Assume y_k has been selected as the closest pattern, then a threshold can be setup as (6) for y_k.
4. If $D(x, \bar{y}_k)$ <threshold, then x is identified as y_k. Otherwise, x is identified as unauthorized user.

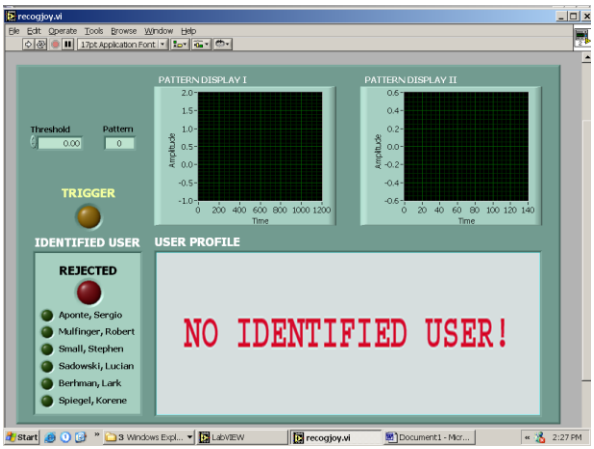


Figure 4. Initial interface

Figure 4 shows the initial interface of the joystick handgrip recognition system. In Figure 4, there are two control panels on the top left. The number in the "Threshold" control panel determines how sensitive the system is, and can be modified by the user. If the threshold is set to 0 (the initial state), the program will use the default threshold of 2. A bigger threshold will increase the possibility that an authorized user is recognized correctly, but it will also increase the possibility that an unauthorized user is mistakenly recognized as an authorized user. A smaller threshold will decrease the possibility that an unauthorized user is mistakenly recognized as an authorized user, but this in turn will also decrease the possibility that authorized user is correctly identified. The default threshold is an optimized tradeoff between these two possibilities.

The "Pattern" control panel displays the recognition result as a number. The number 0 means that an unauthorized user is trying to use this device. The numbers 1-6 will each correspond to a specific authorized user.

Every time a user grasps the joystick and pulls the trigger, this program will try to recognize who is using the device and will give the recognition result. Figure 5 shows an example of an unauthorized user trying to use this joystick.

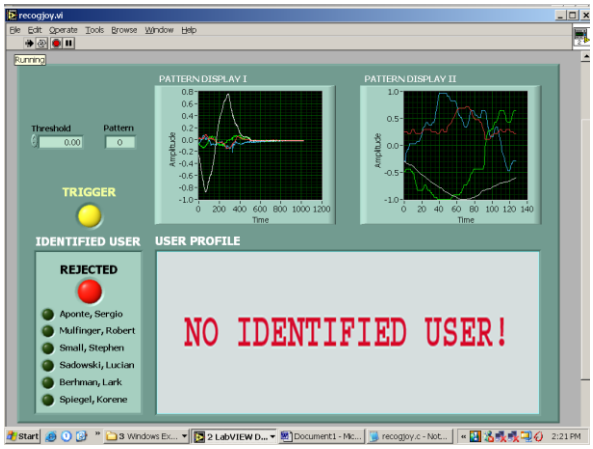


Figure 5. Unauthorized user being identified

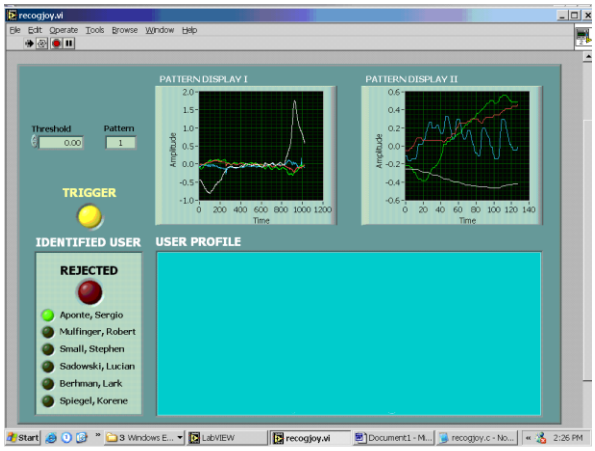


Figure 6. Authorized user being identified

The yellow button is the trigger button which indicates whether the trigger is being pulled or not. Two pattern display panels on the top right show the pressure signal. The left one shows the whole pressure waveform during the trigger-pulling process. The right one zooms the time window onto a much shorter time period during the trigger pulling process, so that more details can be shown.

If an unauthorized user is recognized, a red light will be turned on to show that this person is rejected from using this device. The user profile window will display the message "NO IDENTIFIED USER!" as a warning. If an authorized user is recognized, a green light will be turned on to show which person is recognized. Figure 6 shows an example of an authorized user being identified. The user profile window will display a photograph of this person (the photograph has be hidden to protect privacy in this paper), along with their name and other information.

This system has been tested in a small group. In the first stage, six subjects were asked to grasp the joystick for 30 times to collect handgrip data and build patterns by calculating average waveform and the in-group average distance with standard deviation for each subject.

After building patterns, these 6 subjects with other 4 not-in-group subjects were asked to grasp the joystick and let the system identify who was using the joystick. Each subject repeated this process for 20 times. Among total 200 tests, this system gave answer correctly for 172 times, approximately 86% success rate.

6. Conclusions

In this paper, a real-time identification system based on handgrip pressure variation is proposed to recognize the authorized users of joystick. A prototype system was developed for this research. The results proved that handgrip recognition is a promising technology for pilot identification. Handgrip recognition has other possible applications too, such as car driver authorization and security door authorization by calculating the handgrip pressure variation on transmission shifter or door handler.

7. Acknowledgements

The author would like to thank Dr. Michael Recce and Dr. Tim Chang for many helpful comments. The author would also like to thank New Jersey Institute of Technology and Picatinny Arsenal for the project sponsorship.

References

[1] D. Umphress, and G. Williams, "Identity Verification through Keyboard Characteristics", *International Journal Man-Machine Studies*, Vol. 23, pp. 263-273, 1985.

[2] W. L. Bryan, and N. Harter, "Studies in the Physiology and Psychology of the Telegraphic Language", *The Psychology of Skill: Three Studies*. New York Times Corporation, 1973.

[3] F. Monrose, and A. Rubin, "Authentication via Keystroke Dynamics", *Proceedings of the 4th ACM conference on Computer and communications security*, pp. 48-56, 1997.

[4] T. Hastie, E. Kishon, M. Clark, and J. Fan, "A Model for Signature Verification", *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 191-196, 1991.

[5] Z. Chen and M. Recce, "Verification of Fingertip Pattern on Gun Handle", *Proceedings of the 2005 International Conference on Biometric Authentication*, pp. 421-424, Las Vegas, NV, June, 2005

[6] Z. Chen and M. Recce, "Dynamic Handgrip Recognition", *Proceedings of the 2005 International Conference on Biometric Authentication*, pp. 434-439, Las Vegas, NV, June, 2005

An Approach for Loudness Analysis of Voice Signal (LAVS) Using MatLab

Shiv Kumar[1], Aditya Shastri[2] and RK Singh[3]

[1]Technocrat Institute of Technology, Bhopal (M.P.)-462021, India
[2]Vice Chancellor, Banasthali University, Banasthali, Tonk (Rajasthan), India
[3]Vice Chancellor, ITM University, Gurgaon (Haryana), India
Email: shivksahu@rediffmail.com, adityashastri@yahoo.com, vcrks@rediffmail.com

Abstract: In this paper titled: "An Approach for Loudness Analysis of Voice Signal (LAVS) Using MatLab" is an approach that is used to identify sound pressure level (loudness) of voice signal which contains speech. For this Reply Gain and Equal Loudness filter are used. An underlying purpose of this paper is to investigate if it seems to be possible to use a normalization algorithm to calculate loudness of audio files containing speech which adjust the overall level of the files to match each other by assigning a single gain to every file.

***Keywords*:** Algorithm for LAVS, Equal Loudness Filter, Replay Gain Filter, RMS, VRMS.

1. Introduction

For companies working with broadcasting, it is important to be able to keep an equal perceived sound level, or equal loudness, between various programs and program parts. Many people with different backgrounds, knowledge and preferences make the programs, and in the current situation this would lead to major level differences if program material for a whole day would be broadcasted without mixing. These differences would be annoying for the listeners, which would have to increase or decrease the volume every time the program content changes.

One of the abilities of the auditory system is to be able to order sounds from weak to strong. Loudness is defined as that attribute of the auditory system. One problem concerning loudness is that it is a subjective entity, which means that it can not be measured directly [10].

Loudness level can be measured for any sound, but most widely known are the "Fletcher-Munson-curves" which is shown in Fig.-1. This is based on the work of Fletcher and Munson at Bell labs in the 30s, or perhaps refinements made more recently by [13]. These were made by asking people to judge when pure tones of two different frequencies were the same loudness. This is a very difficult judgment to make, and the curves are the average results from many subjects, so they should be considered general indicators rather than a prescription as to what a single individual might hear. The numbers on each curve identify it in terms of phons, a unit of loudness that compensates for frequency effects. To find the phon value of an intensity measurement, find the db reading and frequency on the graph graph, then see which curve it lands on. The interesting aspects of these curves are that it is difficult to hear low frequency of soft sounds, and that the ear is extra sensitive between 1 and 6 kilohertz [6-7].

With loudness on a logarithmic vertical axis and loudness level on the horizontal axis, it can be seen that if the loudness level is raised by 9 phon, the loudness is doubled [6-7]. This rule is linear over approximately 40 phon as shown in Fig.-2.

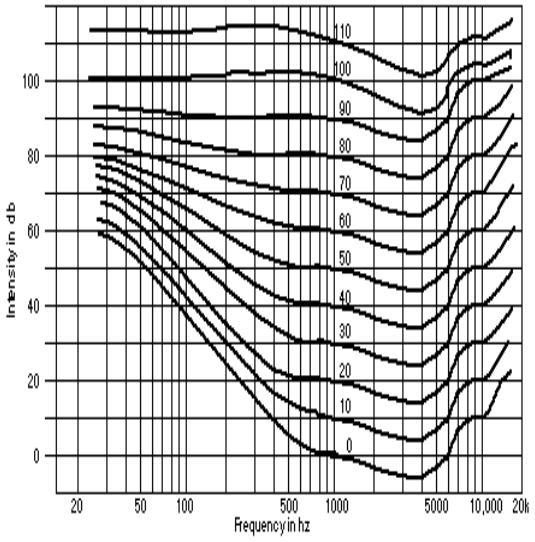


Figure1. Fletcher-Munson-curves

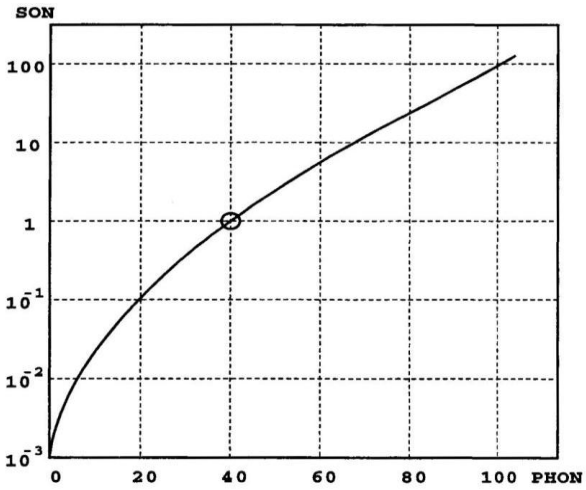


Figure2. Loudness as a function of loudness level

1.1 Critical Band Effect for Loudness

Critical band is a way to divide the frequency range of the ear in to different regions. This makes it possible to calculate the loudness on wide band sounds by adding the loudness in each of the critical bands. The frequency range is usually divided in to 25 regions with the unit Bark. Every Bark has a center frequency and a bandwidth. The size of the bandwidth is dependent on the center frequency. The bandwidth

increases with higher center frequencies. One way to determine the critical bandwidth at a certain center frequency is to let test subjects listen to narrow band noise with a bandwidth that is gradually increased with a total sound level that is constant. Up to a certain bandwidth the loudness will be perceived as constant. When the loudness is perceived as higher than before, the critical bandwidth has been exceeded and consequently the critical band around a specific center frequency can be determined [14].

1.2 Spectral Effect for Loudness

Different sounds have different spectrum. Certain sounds only have one frequency, and other sounds have wide band spectrums. For example, the loudness of 60 dB UEN(Uniform Exiting Noise) is perceived approximately 3.5 times as loud than the loudness of a 1 kHz sinusoid with the same physical sound pressure level. Therefore, the loudness is dependent on the width of the frequency content. Fig.-3 illustrates the loudness difference between the 1 kHz sinusoid and the noise, which starts at the vertical dashed line. The line coincides with the bandwidth of the critical band at 1 kHz for the auditory system. If the bandwidth of the noise fits within a critical band, the loudness increase mentioned above does not appear. This is confirmed by similar measurement at various center frequencies [8].

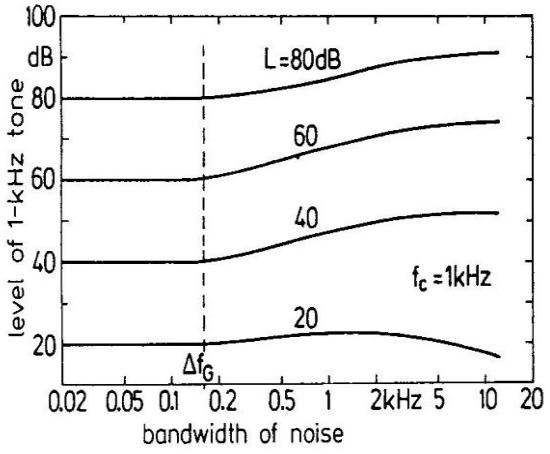


Figure3. The sound level of a 1-kHz tone that is perceived as loud as band pass filtered noise at different bandwidths. The total intensity of the noise is held constant.

A background noise can mask another sound, which also affects the loudness of the sound. Fig.-4 shows the loudness of a 1-kHz sinusoid as a function of its sound pressure level. The dashed line corresponds to the loudness without the noise, and the two lines correspond to the loudness with the masking noise. In this case pink noise with a sound pressure level of 40 dB, and 60 dB per 1/3 octave band is present.

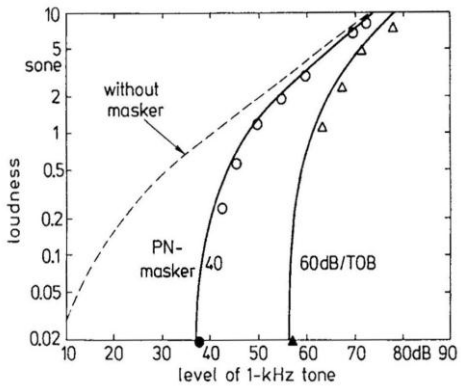


Figure 4. The loudness of a 1-kHz tone as a function of its sound pressure level. The dashed line corresponds to the loudness without the noise and the two lines correspond to the loudness when a pink noise source with a sound pressure level of 40 dB and 60 dB per 1/3 octave band is present.

1.3 Temporal Effect for Loudness

Most sounds vary over time, for example, speech and music. This is why the temporal aspects of sounds also are important for how humans perceive sound levels [8]. For example, a sinusoid with a length of 10 ms is perceived as weaker than a similar sinusoid with the same sound pressure level, but with the length 100 ms. The length of a sound is affecting the loudness, and this is illustrated in Fig.-5. This feature of the auditory system is called temporal integration and holds up to around 100 ms, and after that, the loudness of a sound is constant [8, 10].

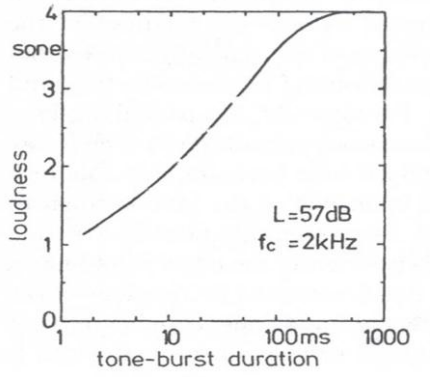


Figure 5. The loudness of a 2-kHz tone with a sound pressure level of 57 dB, as a function of the tone length in milliseconds.

Another temporal effect that affects the loudness of a sound is masking sounds that come before or after in time. Fig.-6 shows loudness measurements when a 2-KHz tone is played before noise (UEN). The tone is 5 ms long and has a sound level of 60 dB. Δt is the distance in time between the tone and the masking noise. The loudness will decrease when the distance between tone and masker decreases. The implication of this is that even if the auditory system perceives the tone it takes some time to create the impression

of loudness. If the auditory system is disturbed by another sound the impression of the earlier tone is interrupted [8].

Something that also affects the loudness of a sound is if a person is exposed to high sound levels a longer period of time. Eventually the hearing threshold will rise temporary. This effect can be measured right after a person has been exposed to the sound and is called temporary threshold shift [10].

1.4 Spatial Effect for Loudness

The filtering of the ear, depending of the direction of the sound, affects the perception of loudness. The fact that sound often reach the ears with a slight difference in time also affects the loudness. This creates a complex phenomenon, for example, when a person listens to music with two loudspeakers in stereo, and is called binaural loudness summation. The phenomenon can cause the perception of loudness with two loudspeakers to increase compared to one speaker when the both speaker systems are playing at the same physical sound level [9].

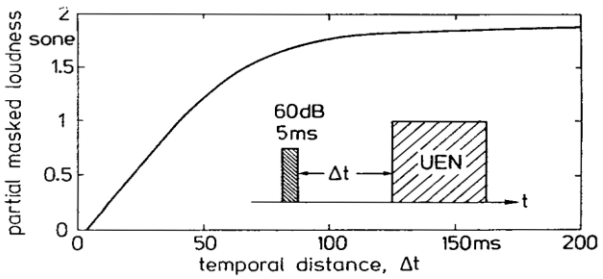


Figure 6. The loudness of a 2-kHz tone that is played before noise, as a function of the distance between tone and masker in milliseconds

1.5 Gain Normalization for Loudness

Normalization of audio files is a way to adjust the overall level of the files to match each other by assigning a single gain to every file. There are many different methods to normalize audio files. One way is to normalize the peak amplitude, i.e. find the largest sample value in each file and adjust the overall gain so that the largest sample in every audio file matches [12]. Another method is to measure the mean level in the files and adjust them so that all files get the same mean level.

If equal loudness is the goal with the normalization, a weighting filter often is applied before measuring the mean level in a file. This is done to give different frequencies different weightings dependent on how loud the human ear perceives them [9].

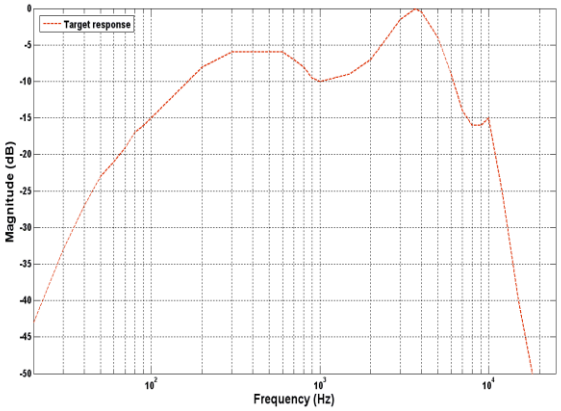


Figure 7. The target response for the Replay Gain filter, which is an inverted approximation of Fletcher-Munson curves.

1.6 Replay Gain

The main idea of Replay Gain is to calculate and store a value of the gain correction needed on an audio file to match the perceived sound level of other audio files where the algorithm have been applied. Replay Gain sets one gain correction value over a whole audio file, which for example could mean that the overall gain in a file could be adjusted with -10 dB [10].

There are two versions of Replay Gain:

- The first is called Radio Replay Gain and works as the explanation above.
- The second is called Audiophile Replay Gain that calculates the gain correction needed over a whole album, so that the intentional level differences within the album are left unchanged.

In this research work only Radio Replay Gain will be described and evaluated which is shown in Fig.-7.

2. Problem Identification and Literature Survey

As per Shaleena Jeeawoody [1], develop a more efficient car security system using the technology voice recognition. The function of this voice recognition car security system is to unlock only when it recognizes a password spoken by the password holder. As per results show that among the 15 words tested, no two voices overlapped. For a given word, the voice spectrum differs from one person to another. But if voice recognition is done with loudness analysis then car security system will develop better result which was not added.

According to Khalid Saeed [2], explained voice recognition where it is not defined that which type file is recognized and how? In this research work, .wav file is taken for voice detection and loudness analysis is one of the parts for them.

According to Maria Markaki[3], a novel feature set for the detection of singing voice in old and new musical recordings but what about if fake voice is recorded for voice detection. This problem is resolved by “automatic voice signal detection” and loudness analysis is one of them.

As per Vijay K Chaudhari[4-5], Voice Spectrum Analysis for fake voice is done manually with the help of various DSP parameters. But if Voice Spectrum Analysis is done with loudness analysis then it will develop better result which was not added.

3. Methodology

3.1 Block Diagram for LAVS

The proposed block diagram for LAVS is given in Fig.-8 in which multiple .wav files are called to calculate loudness. Called .wav files are filter by different filters. In this research work Equal Loudness, and Replay Gain (Radio) filters are used. In Equal Loudness filter, sampling frequency is specified first. If it is not specified then default sampling frequency (CD) is used. Then specify 80dB equal loudness curve, after that convert frequency and amplitude of the equal loudness curve into format suitable for yulewalk which is filter by IIR and 2nd order high pass filter at 150Hz to finish the process of Equal Loudness filter.

Filtered signal by Equal Loudness filter is now filtered by Radio Replay Gain Filter. The calculation process is divided into four steps, where the first is to filter the signal. The filter used is an inverted approximation of the Fletcher-Munson curves. Step two is to calculate the energy of the signal. The signal is divided into blocks of 50 milliseconds and the RMS (Root Mean Square) energy is calculated over each block. Each value is stored in an array. The energy of stereo files is calculated by adding the means of the two channels and divide by two before the square root is calculated. After the RMS energy calculation, all values are converted into decibel (dB). The third step is to choose one RMS value from all the 50-millisecond blocks. Replay Gain sorts all values into numerical order and picks the value that is stored 5% down in the array from the largest value. The last step is to compare the calculated value with a reference. Replay Gain uses pink noise with the RMS energy -20 dBFS (Decibels relative to full scale which is used in digital systems with a maximum available level). The pink noise signal is sent through the algorithm and the result is stored. Then the difference between the reference and the audio file Replay Gain values is calculated, and this value is called the Replay Gain. The obtained value is converted in decibel which is called loudness of the voice signal. 0-Leve DFD for LAVS is given in Fig.-9.

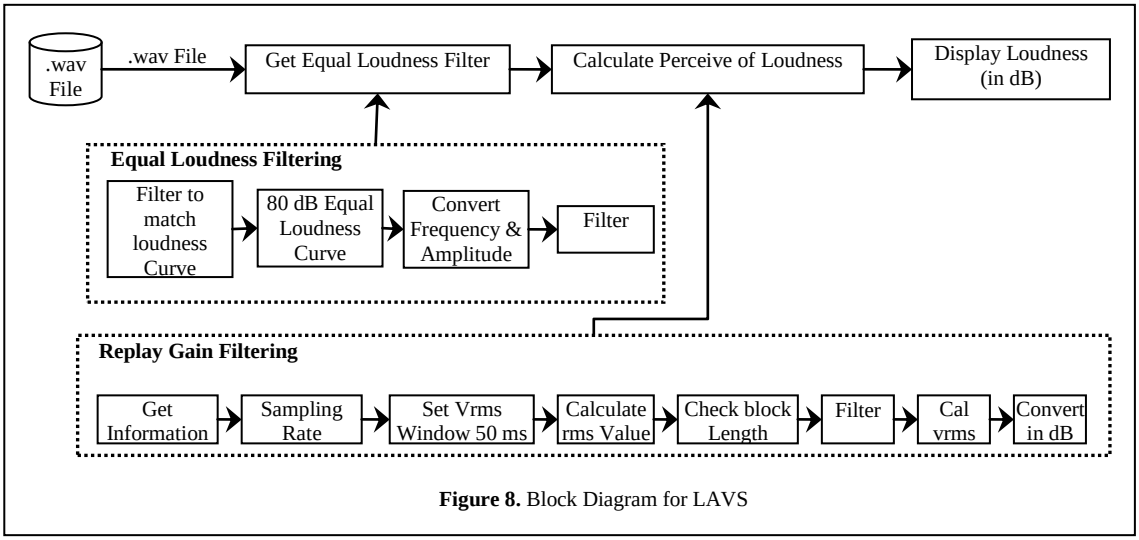


Figure 8. Block Diagram for LAVS

3.2 0-Level DFD for LAVS

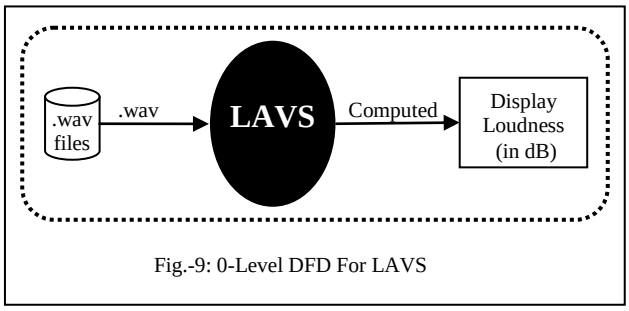


Fig.-9: 0-Level DFD For LAVS

3.3 Algorithm/ Method for LAVS

The proposed algorithm for LAVS is given in following steps in which multiples .wav files are selected in step-1, and from step-2 to step-6 is for Equal Loudness Filter, and Replay Gain filter is illustrated by step-7 to step-20. Loudness of voice signal is computed in step-21.

1. Select multiple .wav files.
2. Do equal loudness filter to match equal loudness curve.
3. Specify the 80 dB equal Loudness curve.
4. Convert frequency and amplitude of the loudness curve.
5. Do IIR filtering.
6. Add a 2nd order high pass filter at 150Hz to finish the process for Equal Loudness filter.
7. Get information of files.
8. Read sampling rate and No. of bits.
9. Generate appropriate equal loudness filter.
10. Set Vrms window to 50 ms.
11. Calculate RMS value to 95%.
12. Set amount of data (in seconds).
13. Determine how many RMS values to calculate per block of data.
14. Check the size of selected file.
15. Grab a selection of file.
16. Do filtering by Equal Loudness curve.
17. Calculate Vrms.

18. If signal is mono then just do the channel and go to step-19 else go to step-19.
19. If signal is stereo then take average Vrms of both channels.
20. Convert in dB.
21. Pick the 95% value.

4. RESULT

The given result is proposed as per proposed algorithm which is shown from Table 1- 4. There is taken 12 .wav file to compute the result. .wav files are recorded by both male and female in different environment with different sentence. The recorded .wav files for female category are dsp_lp2, dsp_c, dsp1_c, dsp_hp1, dsp_hp2, and dsp_hp22. The recorded .wav files for male category are sks_lp, sks_lp2, sks_lp3, sks_lp_c, sks_lp2_c, and sks_lp3_c which contain same sentence. Files dsp_c, and dsp1_c contain same sentence. Files dsp_lp1, dsp_lp2, dsp_hp1, dsp_hp22 contain same sentence.

Table-1: Loudness Analysis for Female Category Voice Signals

Selected Files (.wav)	Reference File Name	dsp_lp2.wav, Type: Stereo		
	Ref. Vrms	Loudness (in dB)	Elapsed Time (in Sec.)	File Type
dsp_lp2	-23.02220	0.00000	0.29600	Stereo
dsp_c	-23.02220	2.13300	0.46800	mono
dsp1_c	-23.02220	2.13300	0.42200	mono
dsp_hp1	-23.02220	-5.82910	0.62500	Stereo
dsp_hp2	-23.02220	-2.68210	0.35900	Stereo
dsp_hp22	-23.02220	-0.71543	0.35900	Stereo

Table-2: Loudness Analysis for Female Category Voice Signals

Selected Files (.wav)	Reference File Name	dsp_c.wav, Type: Mono		
	Ref. Vrms	Loudness (in dB)	Elapsed Time (in Sec.)	File Type
dsp_lp2	-25.15510	-2.13300	0.32800	Stereo
dsp_c	-25.15510	0.00000	0.35900	Mono
dsp1_c	-25.15510	0.00000	0.34400	Mono
dsp_hp1	-25.15510	-7.96200	0.34400	Stereo
dsp_hp2	-25.15510	-4.81510	0.31300	Stereo
dsp_hp22	-25.15510	-2.84840	0.29700	Stereo

Table-3: Loudness Analysis for Male Category Voice Signals

Selected Files (.wav)	Reference File Name	Sks_lp.wav, Type: Stereo		
	Ref. Vrms	Loudness (in dB)	Elapsed Time (in Sec.)	File Type
sks_lp	-26.76030	0.00000	0.34300	Stereo
sks_lp2	-26.76030	-0.84745	0.32800	Stereo
sks_lp3	-26.76030	-5.81840	0.35900	Stereo
sks_lp_c	-26.76030	-0.15865	0.28100	Mono
sks_lp2_c	-26.76030	-0.97951	0.25000	Mono
sks_lp3_c	-26.76030	-5.90750	0.23400	Mono

Table-4: Loudness Analysis for Male Category Voice Signals

Selected Files (.wav)	Reference File Name	Sks_lp_c.wav, Type: Mono		
	Ref. Vrms	Loudness (in dB)	Elapsed Time (in Sec.)	File Type
sks_lp	-26.60160	0.15865	0.31300	Stereo
sks_lp2	-26.60160	-0.68880	0.29700	Stereo
sks_lp3	-26.60160	-5.65970	0.34400	Stereo
sks_lp_c	-26.60160	0.00000	0.29700	Mono
sks_lp2_c	-26.60160	-0.82086	0.28100	Mono
sks_lp3_c	-26.60160	-5.74880	0.25000	Mono

Refernces:

[1] Shaleena Jeeawoody, "Voice Analysis and Recognition as a Car Theft Deterrent", California State Science Fair 2008, Project No: J1307, Ap2/08
http://www.USC.edu/CSSF/Current/Projects/J1307.pdf

[2] Khalid Saeed, "Sound and Voice Verification and Identification a Brief Review of Töeplitz Approach", Znalosti 2008, pp. 22-27, ISBN 978-80-227-2827-0.FIIT STU Bratislava, .stav informatiky a softvÈrovÈho inæinierstva, 2008.
http://znalosti2008.fiit.stuba.sk/download/articles/znalosti2008-Saeed.pdf

[3] Maria Markaki, Andre Holzapfel, Yannis Stylianou "Singing Voice Detection using Modulation Frequency Features", Computer Science Department, University of Crete, Greece.
http://www.sapa2008.org/papers/122.pdf

[4] Vijay K Chaudhari, Dr. R.K. Singh, Dr. Dinesh Varshney, Shiv Kumar, "A New Approach for Voice Spectrum Analysis (VSA) Suitable for Pervasive Computing Using Matlab", 2009 International Conference on Software Technology and Engineering, Chennai-India on July 24 - 26, 2009, ISBN: 978-9-8142-8724-1, Volume No.: 22, Page No.: 390, Published by the World Scientific, Chennai, India, © Copyright 2009 IACSIT.
http://www.iacsit.org/proceeding/icste09.toc.pdf
http://www.worldscibooks.com/etextbook/7533/7533_toc.pdf

[5] Vijay K Chaudhari, Dr. R.K. Singh, Dr. Dinesh Varshney, Shiv Kumar, "Voice Signal Compression and Spectrum Analysis (VSCSA):

Suitable for Pervasive Computing and Limited Storage Devices", Published in "IEEE Xplore Digital Library", Published in Proceeding of "IEEE International Advance Computing Conference (IACC-2009), on 6-7 March 2009, Page No.: 780-784, Location: Patiala-India, ISBN: 978-1-4244-2927-1, INSPEC Accession Number: 10520803, Digital Object Identifier: 10.1109/IADCC.2009.4809113, Current Version Published: 2009-03-31, © Copyright 2009 IEEE. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?isnumber=4808969&arnumber=4809113&count=300&index=143http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F4799789%2F4808969%2F04809113.pdf%3Farnumber%3D4809113&authDecision=-203

[6] Udk Berlin Sengpiel, "Normal equal-loudness-level contours - ISO 226:2003 Acoustics International Organization for Standardization (ISO) 2nd edition" http://www.sengpielaudio.com/Acoustics226-2003.pdf

[7] Audiholic Online A/V Magazine, "Human Hearing: Amplitude Sensitivity Part 1", webpage http://www.audioholics.com/education/acoustics-principles/human-hearing-amplitude-sensitivity-part-1

[8] Zwicker, E. & Fastl, H. (1999). Psychoacoustics: Facts and Models, P. 216-219. Berlin: Springer.

[9] Skovenborg, E. & Nielsen, S.H. (2004), "Evaluation of Different Loudness Models with Music and Speech material", In Proc. of the 117th AES Convention, P. 3-9 , San Francisco.

[10] Moore, B. (2003). An Introduction to the Psychology of Hearing, p. 146. Great Britain: Academic Press.

[11] Replay Gain, 2009-10, Webpage. www.replaygain.org

[12] Vickers, E. (2001), "Automatic Long-term Loudness and Dynamics Matching", In Proc. of the 111th AES Convention, P.-2, New York.

[13] David Robinson, July 2001, Webpage. http://www.David.Robinson.org

[14] Granqvist, S. & Liljencrants, J. (2004), P.1-9, *Kompendium i Elektroakustik*. Stockholm: Royal Institute of Technology.

Author Biographies:

Shiv Kumar received Diploma (Lather Technology Branch) from Govt. Lather Institute, Agra (U.P.)/ Board of Technical Education, Lucknow (U.P.)-India in year 2000. He worked as a Tanner in Ajaction Lather Punjab Ltd. (Punjab)-India in during 2000 to 2001. After that he completed B. Tech. (Information Technology Branch) from Bhagwant Institute of Technology, Muzaffarnagar (U.P.)/ Uttar Pradesh Technical University-Lucknow (U.P.)-India in year 2004, and M. Tech. (Honors, Information Technology Branch) from Technocrat Institute of Technology (TIT), Bhopal (M.P.)/ Rajeev Gandhi Technical University, Bhopal (M.P.)-India in year 2010. Now he is Pursuing Ph.D. (Computer Science & Engineering Branch) from Banasthali University, Tonk (Rajasthan). He worked as a lecturer in the department of C.S.E. / I.T. at Chouksey Engineering College (CEC), Bilaspur (C.G.)-India from Dec-2004 to April-2006, and at Shri Shankaracharya College of Engineering and Technology (SSCET), Bhilai (C.G.)-India in between May-2006 to August-2006. Now he is working as an Asst. Professor in the Department of Information Technology at Technocrat Institute of Technology (TIT), Bhopal (M.P.)-India since 07 Nov. 2006. His research area interest includes Voice Signal Compression, Tonality Computation of Voice Signal, Image Processing, Spectrum Analysis of Voice Signal, Voice Recognition, Pitch Computation of Voice Signal, Formant Analysis of Voice Signal, Loudness Analysis of Voice Signal, and Voice Signal Detection. He is having 18 papers/articles in International Journals (IJCEE, IJET, CiiT International Journal of Digital Signal Processing, CiiT International Journal of Image Processing, CiiT International Journal of Programmable Device Circuits and Systems, CiiT International Journal of Biometrics and Bioinformatics, IJCEA, IJCM, IJES, IEEE Xplore Digital Library, IEEE CS Digital Library, ACM Digital Library, and WSP), 07 papers in International Conferences, 8 papers in National Conferences, 05 International/National Seminar, and 01 Text Book (Concept of Operating System). He is reviewer for 05 International Journals (Journal of Electrical and Engineering and Research (JEEER)-Victoria (Island), Journal of Computer and Communication (JCC)-USA, Journal of Computer Technology and Application (JCTA)-USA, International Journal of Computer Theory and Engineering (IJCTE)-Singapore). He is senior member of IACSIT & IAENG. He has given great contribution in the field of Voice Signal Analysis. He believes in only and only work.

Prof. Aditya Shastri is presently Professor of Computer Science and Vice Chancellor of Banasthali University Rajasthan, India. His research interests are in Discrete Mathematics, Graph Theory and Mobile Computing. He has written more than 50 research papers in journals of repute and authored 5 textbooks. Prof. Shastri, Published more than 200 papers/articles in international journals/ Conference proceeding on Discrete Mathematics, Graph Theory, Ramsey Theory, Lattices and Partial Orders, Design and analysis of Algorithms**,** Parallel and Distributed Algorithms Theory of Computation, E-Commerce and Mobile Computing.

Prof. R.K.Singh received his Ph.D.degree from the Banaras Hindu University in 1969 for the Development of New Three-Body Force Shell Model for Ionic Solids. He pursued higher studies at the University of Edinburgh (U.K.) as a Commonwealth Fellow. As an excellent academic achiever with significant contributions in Science and Technology, Prof. Singh is the first recipient of Dr. K.N. Katju Award for Sciences, presented by HE President of India in 1984 for outstanding contributions in Condensed Matter Physics. Prof. Singh has published 641 Research Papers (out of which 277 are in International Journals and 364 are in Conf. Proceedings) including 10 exhaustive Critical Reviews in contemporary areas and 9 Books (edited). He has visited several countries (UK, USA, Canada, Italy, Japan, Korea, Hong Kong, Singapore, Malaysia, Cyprus, Eqypt, China, Mauritius, etc.) repeated number of times to deliver lectures as Visiting Professor and participate and chair and deliver invited talk in International Conferences/ Symposium. During his research endeavor, he has supervised 60 Ph.D. Scholars. The research interest of Prof. Singh spans over a wide spectrum of Condensed Matter Physics and Materials Science, which includes Nano-materials, Superconductivity, CMR materials, Multiferroic Manganites, Phonon Dynamics, Phase Transition, Orientationally Disordered Materials, Diluted Magnetic Semiconductors, Plastic Crystals and many other Electronic Ceramics. His pioneering work on many body interactions and a comprehensive review on the said topic in Physics Reports 85 (1982) 259-401 and published a Paper in Proceedings of Royal Society (London) A349 (1976) 289-309, which where were heralded by the worldwide scientific community with great appreciation. In addition to his outstanding work on the condensed matter physics, he has explored experimentally several properties of materials, particularly, high temperature superconductors. His innovative ideas are implemented through a well established Superconductivity Research Laboratory, established by him at Bhopal and also through several collaborative programmes. As an Acclaimed Institution Builder. Prof. Singh established the Department of Physics as its Founder Professor and Head at Barkatullah University, Bhopal from 1984-95 and got sanctioned and implemented Superconductivity, Special Assistance and Electronics Programmes from the University Grants Commission. As Hon. Director, he established a Centre for Science and Technology Development Studies at MP Council of Science and Technology. As Founder Director, he established the Computer Centre and Computer Science Department. He was also Head of the University Science Instrumentation Centre at Barkatullah University, Bhopal from 1984 -95. Prof. R.K. Singh has been the Vice Chancellor of Guru Ghasidas University, Bilaspur for two terms (from June 22, 1995 to Jan 22, 2002), Prof. Singh introduced 45 courses and established 8 new Institutes (namely, Institutes of Technology, Pharmacy, Rural Technology, Biotechnology, Computer Applications, Information Technology, Distance Education, and Chhattisgarh Institute of Medical Sciences for MBBS course), besides strengthening the existing Institutes and Departments in the University, and raised the annual budget from 3.00 Crores to 36.00 Crores from 1995-2002. Prof. R.K. Singh has been the Vice Chancellor of M. P. Bhoj (Open) University, Bhopal (from January 23, 2002 to January 4, 2005) with his wide experience and vision of establishing an Institute of Distance Education at Guru Ghasidas University. He participated in the conferences of Asian Association of Open Universities Organized by IGNOU (2001) and Korea (2002). He planned and introduced 50 new professional and job oriented courses of distance education, particularly for the students from reach to unreach. Gyanvani FM band channel was launched in Bhopal in September 2002 in collaboration with IGNOU and DEC for mass literacy mission. Besides, he got established the Institutes of Basic Sciences, Information Technology, General Education and the Centre for Information and Communication Technology (ICT) for imparting online/e-learning courses in Management/Computer/Special Education. Prof R K Singh has been the Senior most Vice Chancellor of Madhya Pradesh

Universities. He was elected the Member of the Executive Council of the Association of Commonwealth Universities for 2002-03 and 2003-04. He is the elected Vice President of the International Academy of Physical Sciences for 2004-06. He was elected as Vice President, Association of Indian Universities in November, 2004. Prof. Singh has been working as Dean, Academic Affairs and Director, University Centre for Distance Learning, Chaudhary Devi Lal University, Sirsa (Haryana), from Sept 2006-Nov 2007. Prof R.K. Singh has been the Pro-Vice Chancellor, MATS UNIVERSITY, RAIPUR (C.G.). Prof. Singh is Emeritus Professor of Physical Sciences at Indian Institute of Information Technology, Allahabad. Prof R.K. Singh is working as Vice Chancellor, ITM UNIVERSITY, GURGAON (HRY.) – 122017 from January 2010.

Radio Link Reliability in Indian Semi-Desert Terrain under Foggy Conditions

Naveen Kumar Chaudhary[1], DK Trivedi[2], Roopam Gupta[3]

[1] Military College of Telecommunication Engineering, Mhow (M.P), India,
[2] RITS, Bhopal (M.P), India,
[3] University Institute of Technology, RGPV, Bhopal, (M.P), India
{nachau2003@yahoo.com, trivedi_devesh@yahoo.co.in, roopamgupta@rgtu.net}

Abstract**:** *Radio links operating in semi-desert region are expected to operate under harsh climatic conditions. Fog is one of the factors which affect the reliability of the radio link in semi-desert terrain during winters. In this paper a case study has been carried out to ascertain signal attenuation in radio relay link due to fog in Indian semi-desert terrain. The empirical formula for fog considering visibility as a function of fog density is used to obtain the fog attenuation. The radio meteorological data was obtained during winter season to study phenomenon of fog attenuation on the radio relay link.*

Keywords**:** *Radio link, fog, visibility, fog attenuation, liquid water content, semi-desert terrain.*

1. Introduction

The radio links are widely used in civil as well as military application for engineering point to point communication. These radio line-of-sight link works on the principle of radio line-of-sight communication. The vehicle based mobile Radio terminals are extensively used for engineering communication network in military operations. The combat communicators plan these links for forward, lateral as well as rear communication support. These links not only facilitates commanders to command and control the battle field but also enables planning of combat logistics. The deployment of these radio mobile terminals is based on distance-based deployment rule. These links are expected to operate satisfactorily under local climatic condition, however it has been observed that reliability of the link is affected in foggy condition due to attenuation.

The climate of Indian semi-desert region is very extreme. The summer temperature reaches up to 50º Celsius and winter temperature dips just around 0º Celsius. The average annual rainfall is only 20 cms. Fog is a natural phenomenon which has been observed during winter season in northern region of India and Indian semi-desert region of Rajasthan. In [1], it is described that fog forms from the condensation of atmospheric water vapor into water droplets that remain suspended in the air. Fog is suspension of very small microscopic water droplets in the air. Fog forms during early morning hours or night when radiative cooling at the earth's surface cools the air near the ground to a temperature at or below its dew point in the presence of a shallow layer of relatively moist air near the surface. The characterization of fog is based on water content, optical visibility, temperature and drop size distribution [2]. In [3], the types of fog, strong advection fog, light advection fog, strong radiation fog, and light radiation fog is mentioned. Attenuation in the foggy days may cause significant anomalous attenuation for radio relay links in climatic regions such as semi-desert terrain. Radiation fog which is more pertinent to semi-desert climate during night is formed as a result of ground becoming cold at night and cooling the adjacent air mass until super saturation is reached. This necessitates analyzing the impact of the fog on radio relay line of sight communication link.

2. Method for Calculating Fog Attenuation

There are various methods to determine fog attenuation, however calculation methods are extensively used. In [4], a radiation fog model was discussed in detail considering description of interaction between microphysical structure of fog and atmospheric radioactive transfer. In [2], the various methods for calculation of fog attenuation pertaining to microwave and millimeter wave frequencies were discussed. In [5], model describing attenuation due to clouds and fog expressed in terms of water content was discussed. In [6], under foggy air conditions the propagation properties of millimeter wave and microwave frequencies were discussed. In [7], it is mentioned that attenuation due to fog is a complex function of the density, extent, index of refraction and wavelength. An empirical formula to calculate fog attenuation in the microwave and millimeter wavelength regions can be obtained by [7], [8].

$$F_A = L_D\,[0.0372\,\lambda_t + 180/\lambda_t - 0.022\,T - 1347]\ \text{dB Km}^{-1} \quad (1)$$

Where F_A is attenuation in (dB/km), L_D is liquid water content in ($g/m^{3)}$, λ_t is wavelength in mm, and T is temperature in degree Celsius (º C). The relation in Equation (1) is valid only if 3 mm$<\lambda_t<$3 cm and -8 º C < T < 25 º C.
The liquid water content L_D is given in terms of visibility in Km, when fog density data are not available but visibility data are available[7],[9].

$$L_D = (0.024/V)^{1.54}\ g^{-3} \quad (2)$$

Where V is the visibility in km and L_D is the liquid water content in g/m^3.
The definition of visibility is given in [7], which define visibility as the greatest distance at which it is just possible for an observer to see a prominent dark object against the sky at the horizon. The visibility is also defined as that distance from an observer at which a minimum contrast ratio C between a black target and a bright background is equal to C=0.02 [10].

3. Case Study for the Effect of Fog Attenuation on Radio Link

This section describes case study based on empirical formula to ascertain the degree of attenuation due to various condition of visibility. An existing digital line-of-sight radio link, located in semi-desert terrain of Rajasthan, India at latitude 29.3167 and longitude 73.9000, tuned to transmitting frequency 10.7 GHz and receiving frequency 11.2 GHz was used to analyze the effect of fog attenuation.

3.1 Radio-meteorological Data

Radio-meteorological data was obtained from the local metrological department and visibility and temperature data was further analyzed to ascertain the results. The formation of fog is predominantly observed in January and February month during winters in Indian semi-desert terrain.

The events when visibility reduced less than six hundred meters for the month of January and February for the year 2008 and 2009 were observed. The measured visibility data for the instances when visibility reduced lesser than 600 meters was analyzed and shown in table 1 and 2. The instances of visibility(V<600 meter) existed for less than one hour is not shown in table 1 and 2 as its value of percentage of occurrence was not very significant.

Table 1: Visibility condition in Suratgarh, Rajasthan,India for the year 2008 (January and February)

Visibility (Km)	January (Hours)	% of time in month (%)	February (Hours)	% of time in month (%)
0	15.5	2.08	7	1.00
0.1	1.5	0.201	-	-
0.5	4.5	0.646	2	0.287

Table 2: Visibility condition in Suratgarh, Rajasthan, India for the year 2009 (January and February)

Visibility (Km)	January (Hours)	Percentage of time in month (%)	February (Hours)	Percentage of time in month (%)
0	24.5	3.64	10.5	1.56
0.1	2.5	0.37	1	0.14
0.2	1	0.148	2	0.29
0.3	-	-	3.5	0.52
0.5	9	1.33	2	0.29

3.2 Calculated Fog Attenuation

The value of liquid water content (L_D) for different visibility conditions is given in table 3. The mean temperature for the different instances of time when visibility reduced less than 600 meter is shown in table 4. The attenuation due to different visibility condition (V<600 meter) corresponding to the mean value of temperature observed at different instances of time is shown in table 5. The plot of visibility verses mean temperature for visibility values less than 600 meters is shown in figure1. The plot of visibility versus attenuation for corresponding mean value temperature is shown in figure2.

TABLE 3: The value of fog water content (L_d)

Visibility(V) in Km	Fog water content L_d in (g/m^3)
0.1	0.11
0.2	0.038
0.3	0.020
0.5	0.010

Table 4: The mean temperature for different instances of time when visibility reduced less than 600 meter

Visibility (Km)	Mean Temperature(°C)			
	January 2008	February 2008	January 2009	February 2009
0.1	10.33	5.1	9.64	11.2
0.2	11.2	6.22	13.2	12.73
0.3	9.78	5.5	10.20	10.6
0.5	7.53	7.04	6.75	11.6

Table 5: Attenuation due to fog for corresponding mean value of temperature

V (Km)	L_d (g/m^3)	Attenuation (dB/Km) 2008		Attenuation (dB/Km) 2009	
		January	February	January	February
0.1	0.11	0.647	0.660	0.649	0.645
0.2	0.038	0.223	0.227	0.2211	0.2215
0.3	0.020	0.118	0.119	0.1178	0.1176
0.5	0.010	0.0595	0.0596	0.0596	0.0586

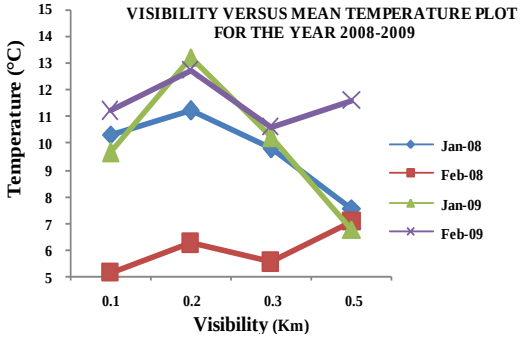


Figure1. Visibility versus mean temperature plot for the year 2008-2009

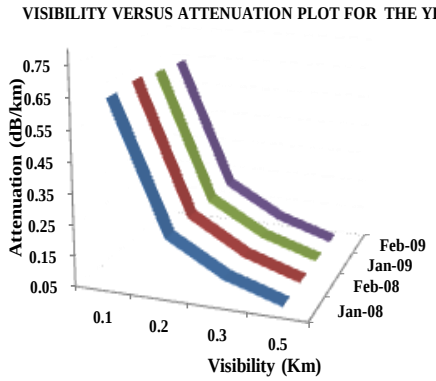


Figure2. Visibility versus attenuation plot for the year 2008-2009

4. Measurement Results

It is observed from table 5 that fog attenuation depends on liquid water content (L_D) and temperature (T) for different instances of visibility. The lowest average temperature 5.96 °C was observed in the month of February 2008 for all the instances of visibility less than 600 meters, this caused average attenuation of 0.2666 dB/Km, this was the highest attenuation value due to fog for the year 2008-2009. The highest average temperature was observed in February 2009 for different instances of visibility (V<600 meter), this caused average attenuation of 0.2606 dB/Km. The average attenuation profile of different instances of visibility (V<600 meter) for the year 2008-2009 is shown in table 6. The plot of average temperature versus attenuation for January, February 2008-2009 is shown in figure 3. It was also observed from the analyzed data of table 5 that for the different instances of visibility (V<600 meter) when temperature increased the attenuation decreased.

TABLE 6: Average attenuation profile for different instances of visibility (v<0.6 Km) for the year 2008-2009

Months	Average Temperature (°C)	Average Attenuation(dB/Km)
January 2008	9.71	0.2618
February 2008	5.96	0.2666
January 2009	9.94	0.26165
February 2009	11.53	0.2606

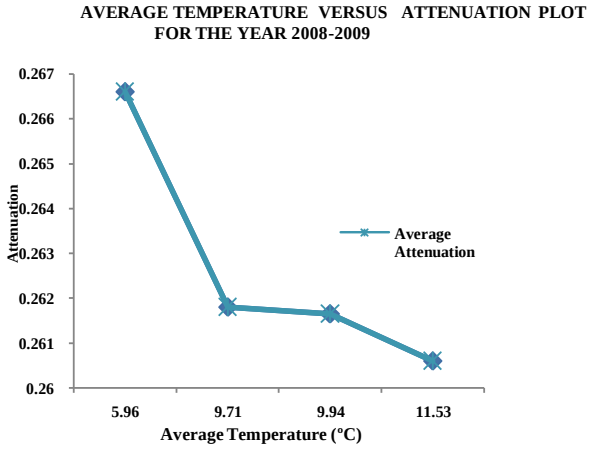


Figure3. Average temperature versus attenuation plot for the year 2008-009

5. Conclusion

Foggy condition affects the reliability of radio line-of-sight link in Indian semi-desert terrain. The results obtained from table 6, indicates that attenuation due to fog predominantly occurs in the month of January and February when the impact of fog is more prevalent in these areas. The results obtained in this paper can be utilized by the planners to cater for suitable fade margin during winter season to overcome the problem of fog attenuation. The study results can also be utilized to improve the performance of existing radio line-of-sight links operational in Indian semi-desert terrain during the months of January and February.

References

[1] Freeman, R. L., *Radio System Design for Telecommunications*, John Wiley & Sons, Inc., New York, 2009.

[2] S. Tamosiunas, M. Tamosiunaite, M. Zilinskas, and M. Tamosiuniene "The influence of fog on the propagation of the electromagnetic waves under Lithuanian climatic conditions", Progress in electromagnetic research symposium (PIERS) online (ISSN 1931-7360),Vol.5, No. 6, 2009

[3] Galati, G., I. Dalmasso, G. Pavan, and G. Brogi, "Fog detection using airport radar," *Proceedings of International Radar Symposium IRS 2006*, 209–212, Krakow, Poland, May 24–26, 2006.

[4] Bott, A. U. Sievers and W. Zdunkowski, "A radiation fog model with a detailed treatment of the interaction between radiative transfer and fog microphysic," J. Atmospheric Sciences, Vol. 47, No. 18, 2153-2166, 1990.

[5] Attenuation due to clouds and fog, ITU-R Recommendation PN 840-3, 1-7, 1999.

[6] Liebe, H. J. T. Manabe, G. A. Liebe, and G. A. Hufford, "Millimeter-wave attenuation and delay rates due to fog/cloud conditions," IEEE Trans. On Antennas and Propagation, Vol. 37, No. 12, 1617-1623 1989.

[7] Altshuler, E. E., "A simple expression for estimating attenuation by fog at millimeter wavelengths," *IEEE Trans. on Antennas and Propagation*, Vol. 32, No. 7, 757–758, 1984.

[8] O. E. Eyo, A. I. Menkiti, S. O. Udo "Microwave signal attenuation in Harmattan weather along Calabar-Akampkpa line-of-sight link" journal of Turk J phys, 27 (2003), page 153-160.

[9] Eldridge, R. G. "Haze and fog aerosol distributions," J. Atmospheric Sciences, Vol. 23, No. 5, 605-613, 1996.

[10] Zhao, Z. and Z. Wu, "Millimeter-wave attenuation due and clouds," Int. J. of Infrared and Millimeter Waves, Vol. 21, No. 10, 1607-1615, 2000.

Author Biographies

Naveen Kumar Chaudhary obtained his BTech degree in Telecommunication and Information Technology from Military College of Telecommunication Engineering (MCTE), Mhow, India and ME degree in Digital Communication from Rajiv Gandhi Technology University (RGTU), Bhopal, India. He is presently posted in Faculty of Communication Engineering, Military College of Telecommunication Engineering, Mhow, India. He is pursuing PhD in electronics and communication engineering from University Institute of Technology, Rajiv Gandhi Technology University, Bhopal, India. His research interest includes: Wireless communication, Tactical Radio line-of-sight communication, Tactical communication Networks and adhoc-mobile network. He is member of IETE.

Devesh Kumar Trivedi obtained his Masters degree M.sc. (Tech) in Electronics from BITS Pilani in 1967. After serving TIFR Bombay as Scientific officer for about five years, he joined IIT Kanpur as Research Engineer in Advanced Center for Electronics System (ACES) in 1973. Here he completed his Ph.D. in Electrical Engg. in 1981. Joined MACT as a faculty member in 1981 and served the institute for above 25 years, in the capacities of Professor, HOD, and Dean. His present areas of interest are Microwave Integrated Circuits (MICs), Computer Aided Design (CADs), Digital Signal Processing. He is member of IEEE and Life member of IETE.

Dr. Roopam Gupta is currently working as Reader & HOD in IT department, University Institute of Technology , Rajiv Gandhi Technology University, Bhopal, India. She received her Ph.D from Barkatullah University, Bhopal, India and M.Tech from MANIT, Bhopal, India. She has seventeen years of teaching experience. Her research interests are in the areas of Communication and Computer Networking. She is the recipient of MPCST,Young Scientist Award ,1999.She is Life Member of IETE and ISTE.

Computational Decision-Making via Game-theoretic Approach and ANN-Based Methods: An Example of Power Utility Conservation Under Non-green/Green Environments

Dr. Dolores De Groff

Department of Civil, Environmental, and Geomatics Engineering
Florida Atlantic University, Boca Raton, Fl. 33421
degroff@fau.edu

***Abstract*:** Green-environment is an art sought for "better-living" albeit growing pollution concerns. In search of a green environment, prudent decisions on energy conservation etc. are imperative. Even while considering non-green ambient, such smart decision-making toward logistical energy conservation is a primary interest. Presented in this paper are ways of decision-making via power utility conservation. Relevant examples are furnished to illustrate the underlying considerations and are discussed. Specifically, the non-green ambient is considered and the payoff resulting from each combination of strategies adopted (or courses of action taken) by technology-option participants (in the market) is evaluated. A neural-network based optimization is also illustrated rationally fusing the game-theoretic concepts and the artificial neural network (ANN) heuristics. Scope of the study for extension to green-environment pursuits in general, is identified.

***Keywords*:** Power Utility Conservation, Payoff-matrix, Competitive and Noncompetitive Markets, Coal, Cleaner Coal, Climate Change, Natural Gas, Game Theory, Artificial Neural Network

1. Introduction

Power-utility centric pollution is a nemesis of global economic growth sprouted over the last century. Typically fuels used are seen with contempt due to their after-the-use consequences. Coal has consistently produced about 40 percent of world's total electricity since early 1970s [1] and 50 percent of the electricity generated in the U.S. is from coal [2]. Spiraling energy demand means that the use of coal is on the rise. In particular, the United States, China, and India have immense coal reserves. China is currently constructing the equivalent of two, 500 megawatt, coal-fired power plants per week [2]. For the United States, China, and India, as well as for importers of coal in Europe and East Asia, economics and security of supply are significant incentives for the continuing use of coal. Between 1999 and 2006, coal use around the world grew by 30 percent [3]. Similar increases are predicted in the near future because coal is cheap and abundant. In any foreseeable scenario, coal, in significant quantities will remain indispensable. Low-cost electricity is crucial for the survival of some industries and is vital for household expenditures of average people. Indeed, organizations that advocate abandoning plans to build new coal-fire power stations are often met with replies of "it's absolutely needed for the economy." However, coal conversion generates a number of gaseous and solid pollutants [1]. The following is a brief outline on the related issues:

- Nitrogen oxides are responsible for the formation of acid rain, photochemical smog, and ground-level ozone. Coal-fired plants are responsible for only a small portion of Nitrogen oxide emissions with the bulk coming from automobile emissions.
- Sulphur dioxide (SO_2) is the leading contributor to acid rain formation and is an irritant to lungs. Coal combustion makes up the majority of SO_2 emissions from utilities.
- Coal is the dominant source of world's carbon dioxide (CO_2) emissions. If plans to build new coal-fired power plants go ahead (without technology to make coal cleaner) CO_2 emissions from coal will increase 60 percent by 2030 [3]. There is now almost 40 percent more CO_2 in the atmosphere than before the Industrial Revolution. Current CO_2 levels are higher than at any point in the last 650,000 years. CO_2 is considered a greenhouse gas because it gives rise to the greenhouse-like warming of Earth. A warming world could experience sea-level rise, coastal erosion, increasing intensity of natural disasters such as floods, droughts, severe storms and forest fires, as well as the spreading of diseases such as dengue fever and malaria. Thus the risk of adverse climate change from global warming forced in part by growing greenhouse gas emissions is serious. There is now wide acceptance that global warming is occurring, that the human contribution is important, and that the effects may impose significant costs on the world economy.
- In a coal-burning plant a portion of the ash is released to the atmosphere through the stack, with the very fine particulates having harmful health effects [1].
- Most heavy metals like lead, tin, and magnesium are collected in particulate control systems, but due to its low vaporization point (356^{o}C), much of the mercury found in coal escapes into the environment, making coal combustion the primary source for mercury emissions.

2. Non-green Environment Optimization

Given that coal use will increase in the future notwithstanding the pollutants that pervade the earth's atmosphere, it is imperative to identify a combination(s) of technology options that would facilitate the achievement of carbon emission reduction goals, as well as reduction of other pollutants. Regulations have already helped to reduce the emissions of acid rain and mercury. The utility industry has been able to meet emission challenges in the past.

Using coal abundantly and reducing pollution effects lead to a multivariable decision problem seeking a solution via technological options to reduce environmental impact of coal energy generation while ensuring that the proposed technology solutions meet conflicting performance and cost requirements. Prudent decisions of what combination(s) of technology options to follow can be made so as to optimize cost-effective requirements and justify engineering decisions. Such an analysis should be done to determine the optimal solution prior to implementation or development.

2.1 Concept of Making Coal Cleaner and/or More Efficient versus Natural Gas Alternative

A number of technologies are available or under development to make the process of converting coal into a transmittable form of energy a less polluting process. Some of the techniques that would be used to accomplish this include the following:

- Washing minerals and impurities from the coal: This process removes mineral impurities that would otherwise be released in the conversion of the coal, polluting the atmosphere. Coal cleaning mainly focuses on reducing sulphur, ash, and heavy metal in the coal [1]
- Gasification or integrated gasification combined cycle (IGCC): Gasification is essentially the process of converting the organic material of coal into gaseous form. It is the centerpiece of near-zero-emission power plants and could provide valuable byproducts such as hydrogen from coal. Hydrogen is suitable for use in transportation and fuel cells. An IGCC effort would gasify coal into fuel gas to fire a gas turbine and use the waste heat to generate steam to run a steam turbine
- Treating the flue gases with steam to remove sulfur dioxide: This is a postconversion technology to strip the flue gas of SO_2 not eliminated through preconversion and conversion techniques
- Carbon-capture and storage (CCS) technologies to capture the carbon dioxide from the flue gas: CCS is a means to capture carbon dioxide from any source, compress it to a dense liquid-like state, and inject and permanently store it underground. Potential hazards include leakage of sequestered CO_2 to the atmosphere, induced geological instability, or contamination of aquifers used for drinking water supplies. The process is very costly at present and far from ready to be deployed at a useful scale
- Dewatering lower rank coals (brown coals) to improve the calorific value, and thus the efficiency of the conversion into electricity: Lignite, often referred to as brown coal, has a high content of volatile matter which makes it easier to convert into gas and liquid petroleum products than higher ranking coals. However, brown coal has a high inherent moisture content which leads to excess transport and handling costs [4]. The bulk of the cost associated with processing is due to dewatering.

Currently, there are no coal-fired power plants in commercial production or construction, which capture all carbon dioxide emissions. There is an uncertainty as regard to the future costs of clean-coal technology. However, the coal industry continues to advertise a future in which coal can be used without environmental damage. Environmental groups believe these cleaner-coal technologies (when and if they are used) will create new waste pollution streams.

There are, however, currently existing and running "critical" and "supercritical" power plants, which employ higher steam temperatures. Thus the amount of coal needed to produce a watt of energy is decreasing. China is a leader in this cutting-edge field [5]. One set of ultra-supercritical steam boilers, built by Siemens and installed in China's Zhejiang province, claims 46% efficiency.

The most efficient coal plants currently operating in the U.S. attain efficiencies around 40% [5].

An alternative to coal, Natural gas is an abundant resource with broad worldwide distribution [6]. There are those today who say that natural gas's time has come. Natural gas combined cycle and cogeneration turbine technologies have resulted in lower cost electricity and dramatically reduced environmental impact. An economical, environmentally friendly, and efficient energy source, natural gas is the cleanest-burning conventional fuel producing 50% to 70% lower levels of greenhouse gas emissions than coal.

3. Decision-making in Power Conservation

A major issue under existing non-green ambient is to make a prudent decision on exercising efficient power conservation. Pertinent to such issues, the decision has to be made among possible alternatives with all the certainty as well as uncertainty considerations duly taken into account. That is, relevant decisions to be made are dictated by a prescribed set of decision criteria.

Hence, evolving such criteria and heuristically applying them to the problem is attempted in this paper with relevant examples of optimal decisions. Both game-theoretic and artificial neural network (ANN) theoretic methods are indicated. The purpose of this study is to extrapolate relevant efforts toward a better ambient sustainability towards a green ambient.

3. 1 Decision-making: A Review

Decision-making is a crucial aspect of any goal-oriented problem. It refers to making the best rational judgment with minimum risk. In power engineering it involves prudent

managerial steps of saving the energy by minimizing losses. Hence, relevant steps that are involved largely refer to critical aspects of making judgment on the proper choice of equipment and systems (specified by the associated technological merits and engineering advantages). In reality, it goes without saying that, decision-making involves both – an optimal choice of engineering methods and a pragmatic selection of technology that matches the environment in hand. The engineering aspects of decision-making are aimed at deciding whether it is worthy of trying a particular method of taking into account of all the oddities of implications. Typically, such efforts are based on what is known as the *present worth analysis*. It refers to a decision-process on outcomes of feasible alternatives (with reference to a project proposal) with selection criterion being the extent of "green efficiency". The associated task of prudent decision among the alternatives involved is to foresee results with (i) a maximized result for a given (fixed) input; (ii) a minimized input leading to a desired (fixed) output and (iii) a maximized difference between the output and the input under the condition that neither the input nor the output is fixed. The corresponding *present worth* (PW) of a contemplated effort is decided on the basis of the life-span of the ambient specified within a finite period of time. This finite time-period is known as the *analysis period*. Within this analysis period, the useful life could be identical in respect of all the possible alternatives indicated for a non-green project; or, it could be different for each alternative. Another situation may correspond to the analysis period being of infinite extent. Adjunct to PW considerations, the other criteria on decision-making involve other factors, namely, *success rate of the project* and *eco-friendliness* [7].

3.2 Are Statistical Decisions Risky?

Decision-making is often risky, difficult and complex. This is because such decisions may have to be made mostly with insufficient information. Lack of or sparse information is undesirable but often unavoidable. This condition of insufficient information depicts the "positive entropy" –a measure of uncertainty. Such an uncertain state will lead to admitting that, different potential outcomes may be recognized without reasonable projections of their probability of occurrence. This admission portrays the implicit risk in the decision made thereof on the outcomes. *Risk* is the extent of variability among the outcomes associated with a particular strategy pursued in a business. It exists whenever, there is a range of possible outcomes that go with the decision and the statistics of such incomes are known (in terms of the probabilities of the outcomes). In general, efforts that offer higher expected results involve greater risk. Further, the risk is directly linked to the utility. That is, the utility function that measures the extent of satisfaction will respond to the extent of risk involved in a specific strategy pursued. If the utility function is plotted against the risk (for example, dollar expended on a project), there are three possible scenarios:

(i) If the resulting plot is concave, the utility function is *risk-averse*

(ii) A convex graph means that the utility function is *risk-seeking*

(iii) A linear (straight) relation between the utility function versus the risk corresponds to a *risk-neutral* situation.

Further, whenever utility of an exercised effort is greater than the expected utility of the effort, the pursued strategy is risk-averse. When utility of effort exercised is equal to the expected utility of the effort, the pursued strategy is risk-neutral; and, when utility of effort exercised is less than the expected utility of the effort, the pursued strategy is risk-seeking

3.3. Traditional Risk Management in Non-green Environment of Coal-utilization

The coal industry has put more than $50 billion towards the development and deployment of cleaner coal technologies including CCS. However, the expenditure has been unsuccessful in that there is not a single commercial scale coal fired power station in the U.S. that captures and stores more than token amounts of CO_2.

Cleaner-coal technologies have high costs at present. Private investors are not willing to invest in such technologies as carbon sequestration. An explicit and rigorous regulatory process that has public and political support is prerequisite for implementation of technologies such as carbon sequestration on a large scale. There is no economic incentive for private firms to undertake cleaner-coal projects. Government assistance is needed and should be provided for demonstration projects as well as for the supporting research and development program.

However, this investment in cleaner-coal technology leaves less investment for the deployment of renewable energy sources. It is a risk weather or not CCS and other technologies will deliver. It is also unknown if governments will enact strict penalties in the future that punish industries who emit CO_2. Could the use of natural gas with at least 50% reduction over coal in CO_2 be a viable alternative?

3.4 Making a decision is however, "a storm of uncertainty with lightning of risks" [8]

Notwithstanding the prevalence of risk in the business economics, the decision made could be either competitive or noncompetitive. As the competitive business has the attribute of future states in which one outcome will presumably occur with its likelihood being uninfluenced by the decision-maker. This decision bears a neutral environment. Further, a competitive decision presupposes a smart competitor who can select a future state of the opponent that maximizes the competitor's returns at the opponent's expense.

Pertinent to noncompetitive and/or competitive market considerations, the decision that can be made among possible alternatives (with the prospects of certainty and uncertainty duly taken into account) is dictated by certain decision criteria. Relevant heuristics are summarized below:

3.5 Decision-making heuristics in Power-plants based markets

Power-generating steam plants are fueled primarily by coal, gas, oil, and uranium. Of these, coal is the most widely used fuel in the United States due to its abundance in the country (and the largest source of energy for the generation of electricity worldwide). In 2004, approximately 50% of electricity in the United States was generated from coal. According to some estimates, coal reserves in the U.S. are sufficient to meet U.S. energy needs for the next five-hundred years. Cheap and reliable power can be obtained from coal. However, the implementation of public policies that have been proposed to reduce carbon dioxide emissions and air pollution may reverse the trend of increasing coal use without technologies to reduce these pollutants.

Centralized power generation became possible when it was recognized that alternating current power lines can transport electricity at very low costs across great distances by taking advantage of the ability to raise and lower the voltage using power transformers.

There are methods that exist to increase the efficiency of power delivery in non-green environments where coal would be the fuel used to meet electric energy requirements. When coal is used for electricity generation, it is usually pulverized and then combusted (burned) in a furnace with a boiler. The furnace heat converts boiler water to steam, which is then used to spin turbines which turn generators and create electricity. The thermodynamic efficiency of this process has been improved over time. "Standard" steam turbines have topped out with some of the most advanced reaching about 35% thermodynamic efficiency for the entire process, although newer combined cycle plants can reach efficiencies as high as 58%. Increasing the combustion temperature can boost this efficiency even further.⁻ Old coal power plants, especially "grandfathered" plants, are significantly less efficient and produce higher levels of waste heat.

The emergence of the supercritical turbine concept envisions running a boiler at extremely high temperatures and pressures with projected efficiencies of 46%, with further theorized increases in temperature and pressure perhaps resulting in even higher efficiencies. Other efficient ways to use coal are combined cycle power plants, combined heat and power cogeneration, and an MHD topping cycle.

The current electric power industry structure is a marketplace that is best represented as an auction [7]. The electrical generators offer to supply quantities of power for a given price, and the customers bid to buy power to supply their required loads. Between these two is the dispatcher who can buy power from the generators and then resell it to the customers. The mode of operation for the dispatchers is "buy low and sell high" [8].

The dispatchers control the transmission system and thus decide who can provide power to the transmission system, which will eventually satisfy customer loads. The dispatcher will then determine how much power will be accepted by the system from each of the interconnected electrical power generators and thus determines the price that the customer will pay.

Producers compete for a larger assignment of market share. (The dispatcher determines the fraction of loading on each of the generators). In most electrical power systems there is either one large generator or a grouping of smaller generators at a single facility. The dispatcher considers the customer's desires (cost of generating the electrical power, reliability, and environmental impact) and the total load required to be served in determining the market share assigned to each of the generating buses (and thus to each producer).

The existence of alternatives (coal, natural gas, etc.) as above gives multiple technology options to the power-utility companies. Under such options, crucial decisions are to be made. Relevant decision under certainty-risk-uncertainty structure of the technology choice can be illustrated *via* specific matrix formats of certainty, risk and uncertainty pertaining to the alternatives *versus* the associated probabilities. It implies considerations specified in terms of what is known as a *payoff matrix* described below.

4. Pay-off Matrix: Game-theoretic Heuristics

In competitive technology situations, two or more decision-makers are involved. The theory of equality applies to presuming that each decision-maker of a competing technology vendor is as good, intelligent and rational as the decision-makers of other competing units. However, each decision-player is confronted with the odds of other decision–makers' ploys. Thus, the decision-makers are *players*, the conflicts they face are *games* and the rationale of their competition is referred to as the *game theory* [11]. To illustrate the buried concepts of the game theory, it is necessary first to define and explain certain terms of relevance**:**

Strategies:	Alternative courses of action taken by game-players
Pure strategy:	A player adopting the same alternative at every play
Game of strategy:	A game with the best course of action for a player being dependent on what that player's adversaries can do
Optimum strategy:	Using one alternative or a mix of alternatives (that is, using different alternatives) for successive plays
Payoff matrix:	In two-person games, for example, the rows of the payoff matrix contain the outcomes for one player, and the columns contents of the columns carry the outcomes for the other player
Saddle-point:	A saddle-point is identified by an outcome, depicting both the smallest number in its row and the largest number in its column in the payoff matrix. Considering a hypothetical payoff matrix indicated in Table 1, its saddle-point can be identified as shown. Note that the table is constructed with reference to

three alternatives A(I), A(II), and A(III) and corresponding identified states of nature N1, N2, N3, and N4.

Table 1: Illustration of a saddle-point in a hypothetical payoff matrix

	N1	N2	N3	N4
A(I)	(9)	10	11	12
A(II)	1	2	3	4
A(III)	5	6	7	8

(The saddle point (9) is the largest in its column and smallest in its row).

Value of the game: It is the return from playing one game depicting the amount that a player nets from the ensuing outcome

Mixed strategy: It is the policy pursued by the players in the absence of the saddle- point. That is, different alternatives are used for a fixed proportion of the plays, but the alternative employed for each play is a random choice from those available

Nash equilibrium: It corresponds to a set of strategies such that none of the players can improve their payoff, given the strategies of the other players. The success of using game theory largely stays with applying judicious logic consistent with practical implications. The procedure involved includes: Assigning meaningful payoffs, solving the associated matrix, (which is usually large), handling multiple players and accounting for the possibilities of collusion, conciliation, irrationality of players, and nonconformance of traditional game theory principles.

In essence, the game theory applies to competitive decisions under uncertainty. It covers the following versions of game: (i) Zero-sum/two-person game where two opponents are presumed to have the same knowledge of outcomes; and, the winnings of one equates to the losses of the other. (ii) Zero-sum/two-person optimal pure (or mixed strategy) that allows selecting a single course of action (or randomly mixed actions) constrained by a set of proportions maintained, leads to the calculation of the value of the game assuring minimum return; (iii) multiple player games and (iv) nonzero-sum games. The last two games are more involved to formulate and solve. However, they carry promising applications in the technoeconomic world.

The following example pertinent to power utility service is indicated to illustrate the game-theory applications and solving approaches. The example refers to a two player/zero-sum case .

4.1 Example: A game-theoretic approach to electric power utility pricing issues

A producer of electrical power using natural gas as the source fuel (or, simply, producer of natural gas, PNG) seeks to increase its market share in an area where it faces competition from an incumbent producer (PC) using coal as the source fuel. The PNG finds that the PC's promotional strategy (to influence the dispatcher to allocate more market share) varies from reduced price/service charges (C1) of electricity during low demand periods, and promoting the promise of no disturbances in power quality and reliability (C2) by designing and deploying its own neighborhood electricity circuit controlled by on-off switches and protected by circuit breakers. In order to competitively influence the dispatcher to allocate more market share, the PNG also comes out with a plan that comprises of: (i) The new service at a rate (I1); (ii) promise of the same rate (I1) above for a certain period of time regardless of the fluctuation of natural gas prices (I2); or (iii) offering the ability to use natual gas with 50% less CO_2 emissions than coal to attract the dispatcher who wants to satisfy those customers who have environmental concerns (I3).

The PNG and the PC are equally competitive in their promotional efforts but differ in certain aspects of their technology-specified expertise. Taking these facts into account, the marketing department of the PNG has arrived at a payoff matrix depicting the percentage gain (or loss) in net revenue for the different outcomes under the service plans of PNG and PC as shown in Table 2. Suppose it is required to determine the proportion of PNG's efforts (pertinent to I2 and I3) that can be pursued so as to get an advantageous edge on (PNG's) revenue by providing the service plans under discussion.

Table 2: Payoff matrix on the service plans of the PNG and PC in Example

		PC	
		C1	**C2**
PNG	**I1**	**– 3 %**	**– 8 %**
	I2	**+ 2 %**	**– 6 %**
	I3	**– 1 %**	**+ 5 %**

4.1.1 Solution to Example

Graphical method: This method corresponds to entering the data from the payoff matrix with payoffs on the ordinates and mixed strategy on the abscissa. The vertical scale should be chosen such that, it must accommodate the entire range of payoffs involved; and, the horizontal scale extends from 0 to 1.

The subsequent steps are as follows: (1) Determination of the dominance and reduce the pay off matrix to (2 × N) size. (2) Graphical representation of the payoff. (The scales of the graph always represent the player with only two alternatives. The third alternative is removed as a result of dominance criterion). Illustrated in Figure 1 is the graphical solution under consideration. (3) Determination of the fractional (proportioned) use of alternatives by PNG in a mixed strategy; and (4) evaluation of value of the game to the PNG.

The proportioned (mixed) use of I2 and I3 by PNG can be

determined by knowing the coordinates of the point of intersection of the corresponding lines drawn on the graph (Figure 1). The coordinates can be either directly read off from the graph or by solving the simultaneous equations of the lines. The equations for the lines can be determined from the geometry as: For I3, [y + 6x = 5]; and, for I2, [y – 8x = – 6]. Hence solving them, the point of intersection is (11/14, 3/14).

Suppose θ is the fractional use of I2 and (1 – θ) is the fractional use of I3. Now, consider the reduced payoff matrix shown in Table 3. In terms of θ, it follows that, [(+ 2) × θ + (1 – θ) × (– 1) ≡ 11/14 ≡ (– 6 × θ + (1 – θ) × 5], solving which yields θ = 6/14. Hence, (1 – θ) = 8/14 as indicated on the reduced payoff matrix (Table 3).

This leads to the conclusion that, the PNG should adopt 43 % effort on I2 strategy (promise of same rate above for a period of time regardless of fluctuation of natural gas prices) and 57 % of effort on I3 strategy (offering the ability to use natual gas with 50% less CO_2 emissions than coal to attract the dispatcher who wants to satisfy those customers who have environmental concerns).

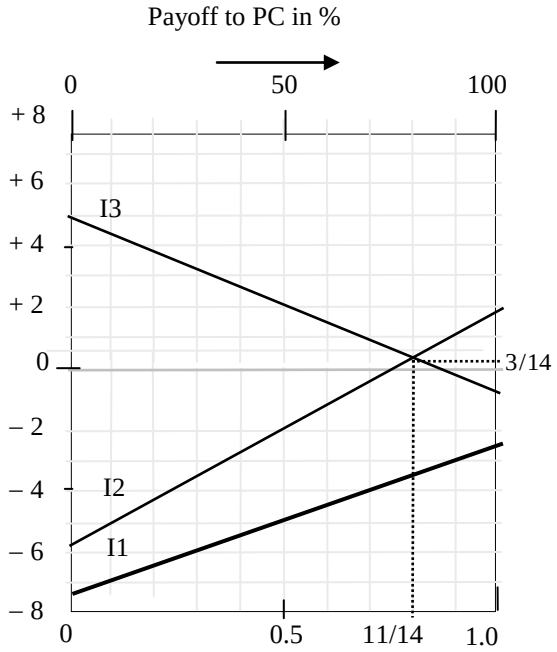


Fig.1. Graphical solution to the problem. (I1-alternative is not taken into the solution as it has less dominance than the other alternatives)

Table 3: Reduced payoff matrix of Table 2

		PC		
		C1	**C2**	
PNG	**I2**	**+ 2 %**	**– 6 %**	**6/14**
	I3	**– 1 %**	**+ 5 %**	**8/14**
		11/14	**3/14**	

If such a mixed proportion of efforts is exerted by the PNG, then what is its value of the game? This value refers to maximum advantage (gain) to the PNG, should the PC follow optimum strategy. It is an *expected value*, EV (for PNG) and can be deduced (using the reduced payoff matrix indicated above) as follows:

EV (PNG) = [6/14 × 11/14 × (+ 2]
+ [6/14 × 3/14 × (– 6)]
+ [8/14 × 11/14 × (– 1]
+ [8/14 × 3/14 × (+ 5)] ≅ 0.285 %

The above calculation takes into account the mutually independent strategies of the PNG and the PC and how their relative (proportionate) efforts weigh the percentages of payoff under possible alternatives are adopted.

Thus, by resorting to the mixed proportion of extending I2 and I3 strategies, the PNG is likely to gain a small revenue growth of about 0.285 % in the presence of the PC adopting optimal strategy. The game-theoretic matrix indicated above with the corresponding result on EV can be framed into an artificial neural network (ANN) so as to automatically optimize the outcome (EV) given a set of matrix elements are introduced as the ANN inputs. This strategy is illustrated via an illustrative example using the numerical results of Table 3.

4.2 ANN-based optimization of the nonzero-sum/two-person game problem

For the purpose as indicated above, a feed-forward ANN architecture can be constructed as illustrated in Figure 2. It is a multilayer perceptron that uses backpropagation [12]. It consists of nine units in the input layer there are two/one hidden layers with one unit output layer. It uses eight elements of the matrix of Table 3 plus EV as the 9th input. Further the computed EV as the supervising value (T), an error (ε) is generated, which is back-propagated to modify the values of the connection weights. This is iteratively done until the error tends to zero.

This network is first trained with a set of training vectors (generated from Table 3) and using the converged set of weights from the training phase, if any new set of inputs are addressed to the ANN, it would yield the corresponding (optimized) out put of EV. The operation of the test ANN is illustrated with the numerical example presented with the results as in Table 3.

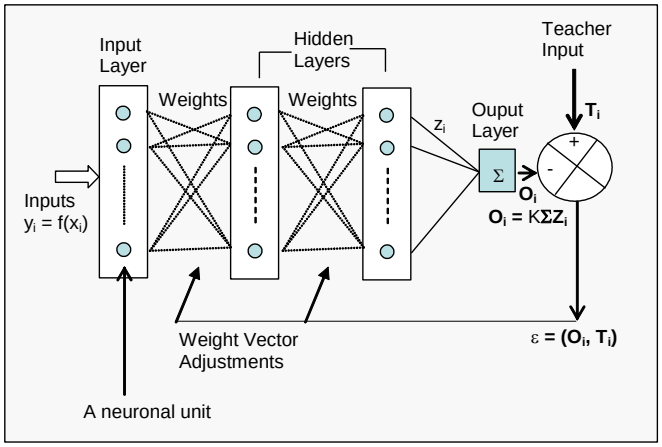


Fig2. Description of the ANN Architecture

4.2.1 Algorithm

1. Input data generation for training phase

First the test matrix (table 3) data are normalized and kept all values positive. For this purpose, each element value (taken in ratio) of the matrix is added with 1 and dived by 2. The resulting matrix with normalized elements is as follows:

Table 4: Normalized payoff matrix of Table 3

		PC		
		C1	**C2**	
PNG	**I2**	**IP3 = + 0.510**	**IP1= +0.470**	**IP6 = + 0.715**
	I3	**IP2= +0.495**	**IP4 = 0.525**	**IP7 = + 0.785**
		IP8 = +0.895	**IP5 = 0.605**	

And the corresponding normalized EV or IP9 = 0.501425. Here, the elements are designated as IP1, IP2, …, IP8 in the ascending order of their values; and, together with IP9, they are used as the inputs to the ANN . Now, construct a function $f(x_i = IP)$, i = 1, 2, …, 9)) versus IP2, IP2, …, IP9 using the matrix data as above. It is of the form: $a_n x^n + a_{n-1}x^{n-1} + \ldots . a_0$.

2 Training the neural network

a. Initially assume a set of uniformly distributed random weights (–1 to + 1). For 9 inputs and 9 neurons one requires a total weight matrix of 9 × 9. Assume zero bias input.

b. The output of a neuron is calculated using the following algorithmic steps:

- $Y = \Sigma W_i X_i$ where i ranges from 0 to 9.
- Multiply and calculate for each of the 9 neurons. The result will then be an output vector of 9 × 1 matrix with each entry corresponding to the output of one neuron.
- Apply the nonlinear (hyperbolic tangent) activation function to each of the outputs thus generating another vector of 9 × 1 size
- Sum all elements in this vector and compare the result with the teacher value
- Calculate the error using ENTF
- Adjust the weights proportional to the value of ENTF and learning coefficient (0.001).
- Repeat steps above till the error ENTF reduces to less than 0.001.
- Store the final weight values in an array
- Repeat steps above for all the 100 sets of inputs. 100 sets of weight values (9 × 9) arrays are obtained after this step
- Average the weight values to obtain one set of 9 × 9 array. Use these weights for testing
- Plot the change of error with respect to the number of iterations for any one input.

c. Start the testing phase

- Randomly generate 9 sets of 9 points on the given function
- Randomize and normalize the test inputs.
- Calculate the outputs of each neuron by using the weights obtained after the training phase.
- Apply the activation function to each of the outputs and sum the outputs.
- Calculate the ENTF using the desired value being the teacher value.
- Display the output for each test as well as the error in each set.

d. End

3. Details:

Function obtained, namely f(x) for the matrix of Table 4 is:

$f(x) = 0.491x^6 + 16.825x^5 - 47.958x^4 + 47.326x^3 – 19.449x^2 + 3.4945x + 0.2715;$

$f_S(x) = \tanh(x)$; ENTF used

The following figure shows the functional variation for one set of inputs, randomized and normalized. This set of inputs is used for training.

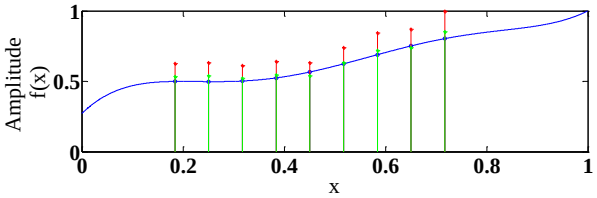


Fig. 3. Function (blue), constructed with one set of inputs after randomizing (green), one set of inputs after randomizing plus normalizing (red)

Fig. 4 shows the learning curve; that is, the variation of ENTF with iterations for one set of inputs.

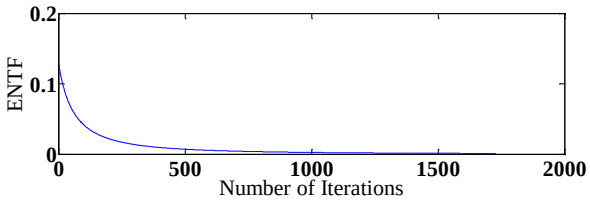


Fig. 4. Variation of ENTF with number of iterations for one set of inputs

Fig. 5 contains all the test sets. (These are shown without randomizing and without normalizing).

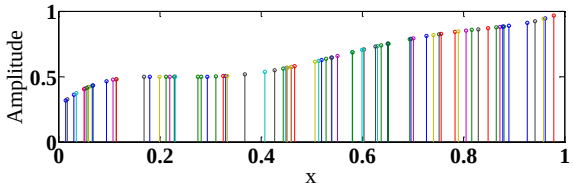


Fig. 5. A number of sets of nine test inputs

In Fig. 6 one sample test set is furnished.

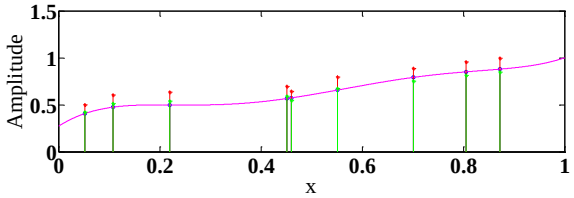


Fig. 6. A sample test set

The following are the output values for a particular set of test cases assumed.

Table 5: Output values and ENTF

Final test output vales	ENTF for nine test cases
.3524	.0263
.2396	.0967
.6517	.0197
.9724	.1560
.6532	.0201
.4214	.0070
.8668	.1000
.5789	.0056
.7459	.0486

5. Concluding Remarks

Notwithstanding the feasibility of game-theoretic pursuits indicated as regard to decision-making issues, yet another strategy can be formulated to seek decision-making optimization via artificial neural network [ANN] approach. Illustrated in this paper is an example using a conventional feedforward ANN algorithm [12]. Though, the numerical values (normalized) used are hypothetical, their relative values correspond to a typical power-utility platform.

Thus, given a set of technology options, computational efforts (algorithms) can be formulated via game-theoretic methods towards decision-making using a payoff matrix strategy. The examples indicated refer to a variety of constraints under which optimal payoff solutions are made via difference schemes of criteria-specifications. Similar decision-making can also be done by another parallel strategy built on ANN techniques. Game-theoretic decision and optimization with results correlated via ANN approach presented in this work is new and novel.

In essence the present study addresses the following:

1. Payoff matrix approach can be pursued in characterizing the game-theoretic problem.
2. Relevant decisions may indicate conflicting and competing issues to be resolved.
3. Fusion of game-theory and ANN is a new avenue for rational computation efforts.
4. The game-theoretic approach indicated can be expanded as a "fuzzy game theory" analysis and applied to green and/ or non-green ambient models.
5. Typical applications of the computational methods indicated conform to: Non-green environment modeling (such as in power utility optimization and if and when to replace coal generation methods with natural gas) which involve decision making issues

6. Appendix: Pseudo Code on ANN Computation

```
Initialize
For i = 1; Iteration start
Define F(x);
   → F(x) corresponds to regressed
     function of the payoff matrix
     (Table 4)
Plot F(x);
Call weight matrix (WM)
   → WM is the interconnect weights on
     the neuronal unit connectivity
     → Initial weight matrix is set
       with random values (-1 to +1)
Address Inputs:
   → 9x9 set
     → From normalized and randomized
       F(x) values
Compute summed outputs of hidden layer
Apply sigmoidal weighting
   → Gather output Oi
If Oi > Ti (Teacher value),
  Then get εi = (Oi - Ti), (Positive
  values)
   → Use Oi to reduce the WM elements
      → Continue iteration
If Oi < Ti, εi = (Oi - Ti) is negative
  Then use εi to augment WM elements
      → Continue iteration
If (Oi - Ti) ≈ 10^-3 (→ 0),
  Then stop iteration
      → Register the output
```

```
End
```

Note:

- Sigmoid used: y = tanh(x)
- No bias added
- Learning coefficient = 0.001
- Stop iteration when ε_i = 0.001

References

[1] M. Kutz, Editor: *Environmentally Conscious Alternative Energy Production*, John Wiley & Sons, Inc. Hoboken, NJ: 2007

[2] "The Future of Coal", Massachusetts Institute of Technology, http://web.mit.edu/coal/.

[3] "The True Cost of Coal", Greenpeace. http://www.greenpeace.org/raw/content/international/press/reports/cost-of-coal.pdf.

[4] C. Veal, "Coal Dewatering: A Review of Current Technology and New Develoments", *The Australian Coal Review*, July 1997.

[5] A. Elliott, "Shoveling Up A Future of Voltage and Steel," *Investor's Business Daily*, A15, November 29, 2010.

[6] W. Vergara, N. Hay, C. Hall, Editors: *Natural Gas Its Role and Potential in Economic Development*, Westview Press, Boulder, Colorado:2009.

[7] D. De Groff: "Green Environment: Decision-Making and Power Utility Optimization Towards Smart-Grid Options," Int. J. Smart-Grid and Renewable Energy, 2010, Published Online (http://www.scirp.org/journal/sgre/) Vol. 1, No. 1 (May 2010).

[8] P.S. Neelakanta and D.M. Baeza, *Next-generation Telecommunications and Internet Engineering*, Linus Publications, Inc. Deer Park, NY: 2009 (Chapter 8).

[9] Rau, N.S., Optimization Principles: Practical Applications to the to the Operation and Markets of the Electric Power Industry. Wiley-Interscience, Hoboken, N.J., 2003

{10] S. Harper and D. Thurston, "Electric Power Network Decision Effects", *The Engineering Economist*, 54:22-49,2009.

[11] J. L. Riggs, *Engineering Economics*, McGraw-hill Book Company, New York, NY: 1977

[12] P.S. Neelakanta and D. De Groff, *Neural Network Modeling: Statistical Mechanics and Cybernetic Perspectives* , CRC Press, Inc., Boca Raton, FL.: 1994.

Author Biography

Dolores De Groff, Ph.D., is Associate Professor of Civil, Environmental, and Geomatics Engineering at Florida Atlantic University in Boca Raton.

She received her Ph.D. in Electrical Engineering from Florida Atlantic University in 1993. She has received several teaching awards from the same institution, coauthored a book in the neural network area and authored/coauthored other journal articles and conference papers.

An Enhanced Algorithm for Offline Connected Handwritten Character Segmentation

[1]Kodituwakku S. R., [2]Rambukwella A.I

[1]Department of Statistics & Computer Science, University of Peradeniya, Sri Lanka
1salukak@pdn.ac.lk

[2]Department of Statistics & Computer Science, University of Peradeniya, Sri Lanka
2rambukwalla@yahoo.com

***Abstract*:** Handwritten character recognition has been an intensive research for last few decades. Very often even in printed text, adjacent characters tend to touch or connected. This makes the recognition of individual characters more difficult. So the successful recognition of characters heavily depends on the segmentation method used. This paper describes an enhanced segmentation algorithm for offline handwritten character segmentation, which segments two-digit handwritten character strings into meaningful segments based on junction points. The proposed algorithm splits characters in two steps: initial segmentation and total segmentation. This algorithm segments connected characters into meaningful segments so that the fuzzy characteristics can be applied on resultant segments for the recognize purpose. Therefore, this algorithm is a good candidate for developing an offline handwritten character recognition system. A prototype system based on the proposed algorithm was developed and was tested with a sample of handwritten connected English two-digit strings as well as handwritten English disconnect characters. The test results indicated 75% success rate for the sample including both upper case characters as well as lower case characters.

***Keywords*—** Character recognition; Character segmentation; connected digit strings; Fuzzy rules

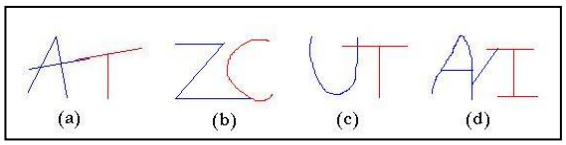
1. Introduction

The segmentation and recognition of connected characters, is a key problem in the development of OCR/ ICR systems. Many methods have been proposed in the recent years [1] – [6]. Most of the recent work examines the applicability of Fuzzy characteristics for recognition. Various segmentation algorithms have been proposed for segmentation. These include segmentation of disconnected upper case English characters and connected uppercase English characters. Usually the connection of adjacent handwritten digits can be categorized into following categories:

- Two strokes from two adjacent digits touching end to end
- The end of a stroke of the left/right digit touching the side of a stroke of the right/left digit
- Overlapping of two vertically oriented strokes from two adjacent digits
- Stroked connection between adjacent two-character strings

The existing algorithms can be used to segment and recognize characters connected as (b) and (c) shown in Figue 1. However, they do not concentrate on segmenting other types of connections. The method proposed by Jayarathna *et. al.*[6] attempts to segment connections like (d) based on junction points.

Figure 1 - Different connection styles



Early work [7] – [11] proposes different techniques based on different concepts. The work proposed by Zongkang et. al. [7] is based on the thinning of background regions. The background thinning based approach first locates several feature points on the background skeleton of a connected digit image and then possible segmentation paths are constructed by matching these feature points. The advantage of this method is the possibility of segmenting single and double touched connection in the character recognition dilemma. When considering the connected character isolation, in order to recognize a connected character input image, it is rather difficult to achieve with the background skeleton analysis. The work proposed by Akihiro et. al. [8] concentrates on touching characters in Mathematical expressions. The method needs an external comparison with the touching characters and it is difficult to extend into various different handwritten styles. A neural network based approach to deal with various types of touching in numeral strings is presented by Daekeum and Gyeonghwan [9]. Once the networks are trained, the advantages of neural networks are automatic learning and quick classification. Major drawbacks of them are the requirement of long processing time and the need of a proper dataset for learning. The method proposed by Yi Lu et. al. [10] uses statistical and structural analysis combined with peak-and-valley functions to segment machine-printed address. Performances of the statistical methods depend on the amount of data used in the definition of the statistical model parameters, and are not flexible enough to adapt new handwritten constraints. Since each and every proposed method has its own advantages as well as disadvantages, segmentation of connected hand written character is an open research area.

This paper presents an enhanced segmentation algorithm for offline handwritten character segmentation, which segments two-character handwritten character strings into meaningful segments based on junction points. The proposed algorithms are extensions of the algorithms proposed by Jayarathna et.al.[6].

.

2. Materials and Methods

The following definitions [6] are used to split connected characters into meaningful segments.

Definition 1: A major starter point is a point on the character skeleton where the traversal through the skeleton could be started.

Definition 2: A junction point is a point on the character skeleton, having three or more neighbours of value 1. It is assumed that the skeleton is in one pixel thickness (thin character).

Definition 3: The traversal direction is the direction from the current pixel to the next pixel to be visited during the traversal.

Definition 4: An end point is a pixel point in the connected character skeleton, in which there is no neighbouring pixel to visit next.

Definition 5: The written direction is the direction of a particular sequence of pixels to which they are written.

This paper presents an algorithm to segment two-character strings with unwanted connections like (a) and (d) in Figure 1. In addition, it is assumed that each connecting character image consists of only two component characters. The proposed technique is a refinement of the algorithm proposed by Jayarathna *et. al.* [6] and could be easily extended to deal with the connected character images, which consists of three or more characters. The block diagram of the proposed segmentation technique is depicted in Figure2.

2.1 Binarization

In general scanned digital images are represented as a collection of pixels having intensities within the range of 0% to 100%. Prior to identification of major starter points and junction points, the image should undergo a binarization process. Otherwise the loss of information and/or noises would result in later processing phases due to various intensity values of the pixels. In the binarization process, if the intensity of a particular pixel is greater than a particular threshold value (200) it is set to zero (0) and if the intensity value is less than or equal to the threshold, it is set to one (1). After binarization, every pixel in the image is represented as either one or zero. Binarized data is stored in a two dimensional array (data[h][w]), where h is the height and w is the width of the input connected character image.

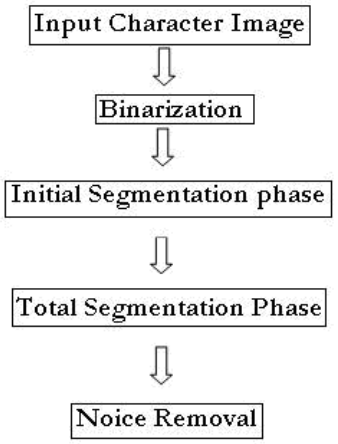


Figure 2 - Block diagram of the segmentation process

2.2 Segmentation

The most challenging task associated with this work is the segmentation of connected characters skeleton into a set of meaningful segments for the recognition purpose. A good segmentation algorithm should produce a set of segments so that each of the resulted segments can be used for recognition of individual characters.

When isolated characters are considered, it is rather easy to obtain meaningful segments. In this work, not only the individual character segments, but also the stroked connection between each individual character has to be considered for the segmentation in order to get the segments out from the connected character input.

In order to recognize individual characters accurately, each resulted segment should be a meaningful segment such as a straight line or an arc. It is almost impossible to segment connected character skeleton into a set of meaningful segments due to different connections between two-character character strings. Therefore, the segmentation stage of the proposed work consists of two phases: the initial segmentation phase and the total segmentation phase. Jayarathna *et.al.* [6] proposed two algorithms: initial split and total segmentation. In this work, we refined them as described below.

2.2.1 Initial Segmentation

Initial segmentation consists of three steps: major starter point identification, junction point identification, and major starter point and junction point based segmentation.

Identification of Major Starter Points and Junction Points - The algorithm starts with finding all the major starter points and junction points in the binarized image. The input for this initial segmentation algorithm is the 2D integer array (data[h][w]) which hold binarized values of the given handwritten connected character skeleton. In order to get the expected segmentation result, the input handwritten connected character skeleton should be in one pixel thickness and it should not contain any spurious branches.

The major starter points and junction points are identified according to the techniques given in Table 1.

Table 1 - Types of Points Identify by the Algorithm

Point Type	Identification method
Major starter point (Technique 1)	An array index in 2D array, that is having only one neighbour of value 1
Major starter point (Technique 2)	In the case of character 'O', they cannot be obtained by the technique 1, since closed 'O' like curve would have at least two neighbours of value 1. In such a case, the first array index with value 1, which is found on the binarized data (2D array), is taken as its major starter point.
Junction point	An array index in the 2D array that is having three or more neighbours of value 1.

In order to find all major starter points and junction points each index of the array is processed. When index of value 1 is found, all the eight adjacent neighbours of a current index are considered. According to the binary value of the neighbouring index, algorithm counts the total number of neighbours. Based on the number of neighbours, the system identifies major starter point and junction points and stored them in a Queue data structure.

Initial Segmentation Algorithm - The main routine of the initial segmentation algorithm is described below.

```
Function initialSegmentation (data) return
segments
Begin
       major_starts= empty
       junction_points =empty
       current_segment= empty
       current_point= empty
       current_direction=null
       segments= empty
       count = 0
       major_starts=
            find_all_major_starter_points  (data)
       junction_points=
            find_all_junction_points (data)
       for each of the unvisited major starter
       points in major_starts do
              set the point as visited
              current_direction = setDirection
                  (data, x, y)
       segments= initSplit (data, x,y)
       end for
       for each of the unvisited junction
       points in junction_points do
              set the point as visited
             current_direction = setDirection
                (data, x, y)
              segments= initSplit (data, x,y)
       end for
       return segments
End
```

Traversal through the Connected Character Skeleton - After finding all the major starter points and the junction points, the algorithm starts traversing through the connected character skeleton. First starting from the major starter points of the queue of major starter points and then starting from the junction points of the queue of junction points.

Determination of Current Traversal Direction - Let us consider the traversal through the character skeleton 'B' in Figure 3. The traversal starts with the major starter point 'a' and continues to the junction 'J1'. At the junction, the current pixel point has two unvisited neighbours. Since there is a neighbouring pixel in to the current traversal direction, the algorithm chooses that pixel (n1) between the two neighbouring pixels as the next pixel point to visit. It is clear that the other neighbouring pixel (n2) is a starter point of another path in the skeleton. In every junction in the skeleton, there may be one or more unvisited neighbours.

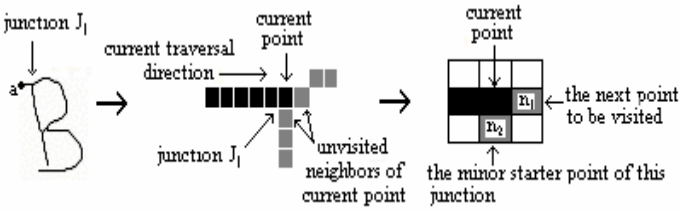


Figure 3 - Determination of current traversal direction

Determination of Closest Traversal Direction - When there is no neighbouring pixel in to the current traversal direction, the algorithm determines the closest traversal directions to the current traversal direction as described in Table 2.

Table 2 -
Determination of the Current Traversal Direction

Neighbouring pixel which is chosen as the next pixel point to be visited	Current traversal direction
x,(y-1)	U (UP)
(x-1),y	L(LEFT)
x,(y+1)	D (DOWN)
(x+1),y	R (RIGHT)
(x+1),(y-1)	RU (RIGHT_UP)
(x+1),(y+1)	RD (RIGHT_DOWN)
(x-y),(y+1)	LD (LEFT_DOWN)
(x-1),(y-1)	LU (LEFT_UP)

In Table 2, for each current traversal direction, two closest traversal directions are mentioned. Therefore, to find a neighbour in the closest traversal direction to the current traversal direction, the algorithm first checks the closest traversal direction which would be found first, when considering the directions clockwise starting from the direction. As an example (Fig. 4), if the current traversal direction is U, the algorithm first checks whether there is an unvisited neighbour in the direction RU. If it fails to find such a neighbour, then it checks whether there is an unvisited neighbour in the direction LU.

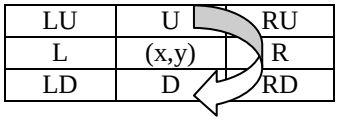


Figure 4 - Consideration of the directions in clockwise, staring from the direction up from the current pixel point (x, y)

When there is no neighbouring pixel in to the current traversal direction, as well as the closest traversal directions, the algorithm chooses an unvisited neighbour, which is in any direction from the current point, as the next point to be visited. The unvisited neighbours are processed starting with the direction U and then in clockwise (Table 3.). That is, the algorithm first checks whether there is an unvisited neighbour in the direction U to the current point (xylem). If it fails to find such a neighbour then it checks an unvisited neighbour in the RU direction to the current point and so on. Therefore the directions are considered in the order; U, RU, R, RD, D, LD, L, and finally LU.

Table 3 - Determination of the closest traversal direction to the current traversal direction

Current traversal direction	Closest traversal directions
U	RU, LU
R	RU, RD
D	RD, LD
L	LD, LU
RU	U, R
RD	R, D
LD	D, L
LU	L, U

Once the traversal reaches an end point, each point which belongs to the current traversal path are stored as a separate segment. The focus is shifted to the identified unvisited major starter points first and junction points next.

The traversal process mentioned above is continued with all the unvisited major starter points in the queue of major_starts and all the unvisited junction points in the queue of junction_points, until all the unvisited paths in the binarized character skeleton are visited. The risk of visiting the same pixel more than once during each splitting traversal is eliminated by setting the binary value of visited index to zero and only visiting the index of value one in binarized data(data[h][w]). Therefore it is guaranteed that a path in the skeleton is visited only once and hence the same segment is not identified twice. The initial splitter routing as follows.

```
Function InitSplit (data, x, y) returns
       segments
Begin
       current_segment= empty
       current_point= empty
      segments= empty
       count = 0
       if (current_direction!=null) then
          while (there is a index with value 1
          in the current direction) do
              make the point (index) as visited
             update the current_segment by
              adding visited point
         end while
       setDirection (data, x, y)
       if (current_direction!=null) then
          initSplit (data, x, y, i)
      end if
      noice_removal ()
       update the segments by adding identified
       segment
      end if
       return segments
End
```

The initial splitting process mention above is continued with all the unvisited traversal paths, and the final output of the initial segmentation phase is used as the input to the total segmentation phase, which produces complete segmentation of the connected character skeleton.

2.2.2 Total Segmentation

Although the initial segmentation algorithm works for junction points depicted in Fig. 5, it would not produce complete segmentation due to the non-junction points shown in Figure 6. For example, the handwritten character skeleton 'Z' would not be segmented into a "Negative Slanted" and two "Horizontal Line" because of the non-junction point connection in between.

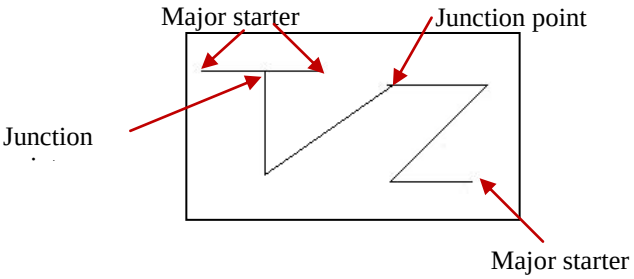


Figure 5 - Major starter points and junction points of the connected character

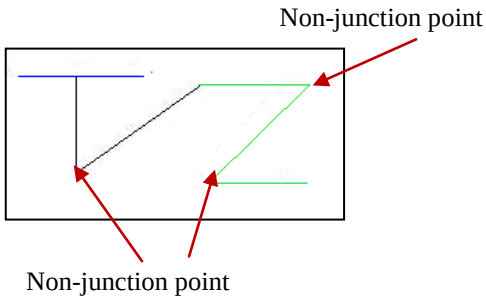


Figure 6 - Initial segmentation of the connected character

To overcome this, a separate segmentation algorithm is used with the rule based segmentation approach.

Total Segmentation Algorithm - The complete splitting process is used to analyze each initial segment and to split each of the candidate segments into complete character segments. The complete splitter routine is as follows.

```
Function total_split (segments) return
       complete_segments
Begin
    complete_segments=
        emptycurrent_segment=emptytemp_segments=
        empty
    next_segments= empty
    s_current_point=null
    end_current_point=null
    s_next_point
```

```
   end_next_point
   threshold_angle= 0.0
   set_mf_values (segments)
   set threshold angle using rule base
   s_current_point= get first point of the
       selected initial segment
   current_segment.add (current_point)
   while( no termination occur) do
    if (current_segment >1) then
      s_next_point= get the next segment point
       of the current point
      if (does neighbors in the same direction)
       then
        current_segment.add (s_next_point)
        s_current_point= s_next_point
     if (current_point is the end point)
       termination condition occurs
       complete_segments.add (current_segment)
     end if
   end if
else
 if (IsAbrubtCange (current_segment,
       temp_segments )) then
       complete_segments.add (current_segment)
       current_segment= temp_segments
       if (current_point is the end point)
              termination condition occurs
              complete_segments.add
              (current_segment)
        end if
 end if
  else
      add all the points in the temp_segments
       to current_segment
      if (current_point is the end point)
              termination condition occurs
              complete_segments.add
              (current_segment)
      end if
   end else
  end else
 end if
else
  s_next_point= get the next segment point of
       the current point
       current_segment.add (s_next_point)
       current_direction= get the current
       traversal direction
      s_current_point= s_next_point
 end else
end while
return  complete_segments
End
```

The total segmentation decision is based on the abrupt change in the written direction. According to this algorithm, as long as the current traversal direction remains unchanged (if it can find an unvisited neighbouring pixel in the current traversal direction), the algorithm considers the path, which is being visited belongs to the same segment. If the current traversal direction changes, then the algorithm goes and checks for an abrupt change in written direction. In this work, it was found that the written direction of a skeleton path could be changed after a sequence of 5 pixels. Therefore, to find out the previous written direction, the written direction of last 5 pixel points (which are stored in current_segment) in the path is considered. To find out the new written direction, the written direction of next 5 pixels in the path (which are stored in temp_segments) is considered. The determination of abrupt change in written direction can be carried out as follows.

```
Function IsAbrubtCange (current_segment,
       temp_segments) returns Boolean.
Begin
If (the size of temp_segments < 5 OR size of
       current_segment < 5)
      return false
end if
else
    prev_written_d = written direction of the
       last 5 pixel points in the
       current_segment.
    new_written_d = written direction of the
       temp_segments.
    difference = |prev_written_d -
       new_written_d|
    if (difference > 315)
      if (prev_written_d > 315 OR
       new_written_d > 315)
      difference = 360 - difference
      end if
    end if
    if (difference > threshold angle)
        return true
    end if
    else
        return false
    end else
  end else
End.
```

If there is an abrupt change in the written direction, the corresponding change in current traversal direction is considered as a starting point of a new segment. Otherwise the traversal is continued with the same segment. The suitable value for the threshold angle may vary according to the character set which is dealt with. If not, the algorithm results over-segmentation of characters or under segmentation of characters.

Calculation of Written Direction - The written direction of a given sequence of pixel points equals to the angle between the x-axis and the straight line connecting the start and end points of the pixel sequence. After calculating the angle Θ by using the Equation 1, the written direction can be calculated as described in Table 4.

$$\Theta = \tan^{-1}(|(y_s - y_e)/(x_s - x_e)|) \quad (1)$$

Table 4 - Calculation of the written direction

	written direction
$(x_s < x_e)$ and $(y_s <= y_e)$	Θ
$(x_s > x_e)$ and $(y_s <= y_e)$	180 – Θ
$(x_s > x_e)$ and $(y_s > y_e)$	180 + Θ
$(x_s < x_e)$ and $(y_s > y_e)$	360 – Θ
$(x_s = x_e)$ and $(y_s <= y_e)$	90
$(x_s = x_e)$ and $(y_s > y_e)$	270

The identification and representation of different situations - In this work, an effort was made to identify the different situations, which require different threshold angles, associated with the

meaningful segmentation of handwritten uppercase English character skeletons.

In order to differentiate the identified situations, which require different threshold angles, the following fuzzy features were calculated for each segment (segments resulted from initial splitting phase)

1. Arcness (MARC)
2. Curve Type (DLike (MDL), ULike (MUL))

Identified situation can be represented with the following rule base, where VL, LM, L and S are Fuzzy values.

```
IF (segment. MDL = "VL") THEN
            threshold_angle = 60 degrees
ELSE IF (segment. MUL = "VL") THEN
            threshold_angle = 60 degrees
ELSE IF (segment. MARC = "VL") THEN
            threshold_angle = 80 degrees
ELSE IF ((segment. MARC = "LM" OR segment. MARC
      = "L") AND segment. MDL = "S")) THEN
            threshold_angle = 60 degrees
ELSE IF (segment. MDL = "S") THEN
            threshold_angle = 20 degrees
ELSE
          threshold_angle = 40 degrees
```

Noise Removal (2);

In the process of noise removal, the segments resulted due to the noises in the skeleton area are removed. The minimum size of the segment was taken as 10 pixel points in this research. Therefore under the noise removal, the segments which contain less than 10 pixel points are discarded from the set of resulted segments.

After the segmentation process, each segment can be considered as totally spitted segments of the connected character skeleton. For the recognition, each of the isolated character should undergo proper segmentation, in order to obtain meaningful segments as the output.

3. Results and Discussion

The prototype system developed based on the proposed algorithm was tested with the connected character skeletons of thin handwritten English characters. All 52 (26 upper case and 26 lower case) alphabetic characters were tested. Figure 7 shows the complete segmentations of the connected character skeleton after segmentation process.

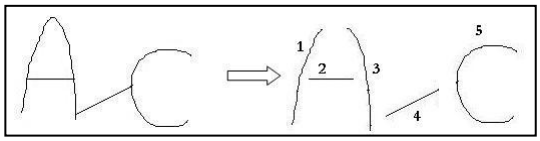


Figure 7 - Total meaningful segments of connected character after segmentation process

In the recognize process, the resultant total meaningful segments are divided in to two sets as shown in Figure 8.

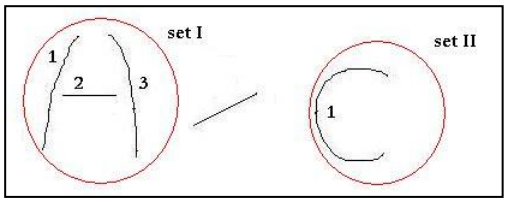


Figure 8 -Two sets of segments

The equality fuzzy values generated by the system after comparing the features of the segments of the given character with the features of the segments of the character in the database having the identical number of segments (in this case 3 segments) as the given character are shown in Tables 5 and 6.

Table 5 - The equality fuzzy values

Database Character (having 3 segments)	Segment Equalities			Minimum
	Segment1	Segment2	Segment3	
A	0.9582	0.9623	0.9901	0.9582
B	0.7321	0.5436	0.5523	0.5436
F	0.8756	0.6323	0.6423	0.6323
H	0.9563	0.9702	0.8494	0.8494
I	0.6756	0.6863	0.6746	0.6746
K	0.9563	0.7880	0.7963	0.7880
N	0.8653	0.6641	0.8675	0.6641
R	0.8634	0.4080	0.5362	0.4080
Y	0.6325	0.6531	0.4734	0.4734
Z	0.7562	0.6521	0.6600	0.6521

According to Table 5, it is clear that the equality between the given character pattern and the database character 'A' is 0.9582, which is the maximum of minimum equalities of database characters which having three segments and it is greater than 0.75 (Threshold value). Therefore, the system supposes to recognize the character pattern in segments set I in Figure 8 as character 'A'.

According to Table 6, it is clear that the equality between the given character pattern and the database character 'C' is 0.9631, which is the maximum of minimum equalities of database characters which having one segments and it is greater than 0.75 (Threshold value). Therefore, the system supposes to recognize the character pattern in segments set II in Figure 8 as character 'C'.

Table 6 - The equality fuzzy values

Database Character (having 3 segments)	Segment Equalities	Minimum
	Segment 1	
C	0.9631	0.9631
O	0.4039	0.4039
S	0.8458	0.8458
U	0.5950	0.5950

4. Conclusion

According to the test results the proposed that the recognition system based on the proposed algorithm identifies connected characters reliably and efficiently. The segmentation algorithm developed for this work was capable of separating the connections in the thin uppercase English handwritten characters into meaningful segments accurately.

The recognition rate was found to be 100%, when the same characters that were used for the training, fed to the system for identification with a connection in between. When a sample of upper case and lower case characters skeletons with unwanted connections were tested the system reported 75% success rate. So it can be concluded that an average the proposed method is 85% accurate for machine printed thin characters, which have unwanted connections.

It was also observed that it was very flexible and simple to implement, compared to other available methods for offline handwritten connected character recognition.

Although the method used in this work does not deal with the two stroked connections from two adjacent digits touching end to end, the system could still be used for the separation of any handwritten connected characters with single connections.

This algorithm was tested with different threshold angles. The major problems associated with this algorithm were the over-segmentation and under segmentation of some character skeletons. Except of these problems other character skeletons had been segmented as expected under that threshold angles.

Due to these problems, it is obvious that the use of single threshold angle for the determination of correct segmentation is not practical. The solution for this problem would be the identification of different situations, which require different threshold angles and a way to represent them.

REFERENCES

[1] K.B.M.R. Batuwita and G.E.M.D.C. Bandara (2005), "An Online Adaptable Fuzzy System for Offline Handwritten Character Recognition," in the Proceedings of Fuzzy Logic, Soft Computing and Computational Intelligence-- 11th World Congress of International Fuzzy Systems Association (IFSA 2005), China, Springer Tsinghua, (**2**), p.1185- 1190.

[2] K.B.M.R. Batuwita and G.E.M.D.C. Bandara (2005), "An improved segmentation algorithm for individual offline handwritten character segmentation," in International Conference on Computational Intelligence for Modelling, Control and Automation (CIMCA'2005), Vienna, Austria, IEEE press., (**2**), pp. 982

[3] K.B.M.R. Batuwita and G.E.M.D.C. Bandara (2005), "New Segmentation Algorithm for Individual Offline Handwritten Character Segmentation," in the Proceedings of 2nd International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2005), Changsha, China, Springer- Verlag, Part II, p. 215. (In Lecture Notes in Computer Science, http://dx.doi.org/10.1007/11540007_26.)

[4] Bandara, G.E.M.D.C., Pathirana, S.D., Ranawana, R.M. (2002), Use of Fuzzy Feature Descriptions to Recognize Handwritten Alphanumeric Characters. In: Proceedings of 1st Conference on Fuzzy Systems and Knowledge Discovery, Singapore, 269-274

[5] Bandara, G.E.M.D.C., Ranawana, R.M., Pathirana, S.D. (2004), Use of Fuzzy Feature Descriptions to Recognize Alphanumeric Characters. In: Halgamuge, S., Wang, L. (eds.): Classification and Clustering for Knowledge Discovery. Springer-Verlag 205-230.

[6] Jayarathne, U. K. S and Bandara, G. E. M. D. C , (2006), New segmentation Algorithm for Offline Handwritten Connected Character Segmentation, First International Conference on Industrial and Information Systems ICIIS 2006.

[7] Zongkang Lu, Zheru Chi, Wan_ Chi Siu, Pengfei Shi (2003), " A background thinning based approach foer separating and recognizing connected handwritten digit strings", Department of Electronic Engineering, the Hong Kong Polytechnic University, Hong Kong, Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200030, China

[8] Akihiro Nomura, Kazuyuki Michishita, Seiichi Uchida, and Masakazu Suzuki (), "Detection and Segmentation of Touching Characters in Mathematical Expressions", Graduate School of Mathematics, Faculty of Information Science and Electrical Engineering, Faculty of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, 812-8581 Japan.

[9] Daekeum Y. and Gyeonghwan K. (), "An approach for locating segmentation points of handwritten digit strings using a neural network", Dept. of Electronic Engineering Sogang University, CPO Box 1142, Seoul 100-611 Korea.

[10] Yi Lu, Beverly Haist, Laurel Harmon, John Trenkle and Robert Vogt (), "An Accurate and Efficient System for Segmenting Machine-Printed Text", Environmental Research Institute of Michigan P. O. Box 134001 Ann Arbor, MI 48113-4001

[11] LI Guo-hong SHI Peng-fei (2003), "An approach to offline handwritten Chinese character recognition based on segment evaluation of adaptive duration", Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200030, China

A controlling Strategy for Industrial Machines using Bluetooth Technology

Arafat Zaidan[1], Mutamed Khatib[2] and Basim Alsayid [3]

[1]Department of Electrical Engineering, Palestine Technical University-Kadoorie (PTU), Palestine
arafatzaidan@yahoo.co.uk
[2]Department of Telecommunication Engineering and Technology, Palestine Technical University-Kadoorie (PTU), Palestine
mutamed.khatib@gmail.com
Dean, College of Engineering and Technology, Palestine Technical University-Kadoorie (PTU), Palestine
b.alsayid@ptuk.edu.ps

***Abstract*:** Wireless technology is used to control industrial applications. The control system uses wireless communication to transmit commands between the control station and the controlled object or process. By using a laptop, mobile unit or PC, one can control the equipments and observe the industrial processes using Bluetooth communication which enhance flexibility. This system will control a DC motor status (on or off), speed and direction of rotation. The motor controller takes its commands from the laptop by a wireless link. To make the control operation easier for the user, a Windows based interface is programmed.

***Keywords*:** Control, Wireless, Bluetooth, PIC.

1. Introduction

Nowadays, wireless communications are becoming more and more popular; it gives opportunity to wireless data transfer technology, such as Infra-red, radio frequency (RF) and Bluetooth [1]. With the advancement of RF, researchers are concentrating efforts in developing electronic devices that can communicate within a few meters, the result is Bluetooth. It is a practical idea to control a motor through wireless connection, as wireless technology becomes increasingly more available. One of these technologies is Bluetooth technology [2].

Bluetooth technology provides an unlicensed band that is ISM (Industrial Scientific Medical) band which ranges from 2.4 GHz to 2.4835 GHz enable the goals of global applicability, low power and high aggregate capacity to be met [3]. This is suitable for a wireless home or office environment and industrial applications. Bluetooth wireless technology is a short-range communications system intended to replace the cables connecting portable and/or fixed electronic devices. The key features of Bluetooth wireless technology are robustness, low power, standardized protocol, Upgradeable and low cost [4].

The proposed system can be divided into the control circuit that controls the DC motor, the Bluetooth module, the adaptor, and the user interface. The control parameter controls the DC motor speed; direction of rotation and status (on-off) is the value of the duty cycle. By changing the duty cycle value, the user can change the magnitude of the input voltage and so the power delivered to the motor, which changes the speed. So to do this, a control unit and a power electronic circuit are needed. The control unit is a Peripheral Interface Controller (PIC) that generates the necessary output signal that controls the duty cycle value which is a Pulse Width Modulation (PWM) signal. The power electronic circuit is H-bridge which is attached to the terminals of the DC motor and the H-bridge output is controlled by the PWM signal coming from the PIC.

The user will change the duty cycle value and the Bluetooth wireless system is responsible of transmitting the new value of the duty cycle to the PIC. So the design of the PIC microcontroller, the wireless Bluetooth system and the user interface is as shown in Figure 1.

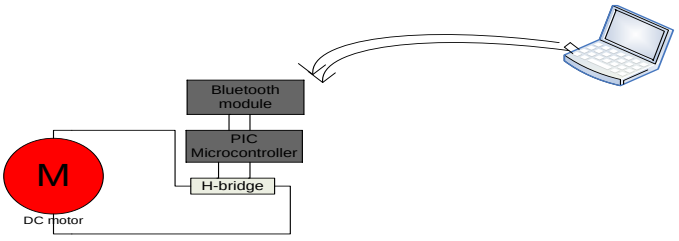


Figure 1: system description

2. System model

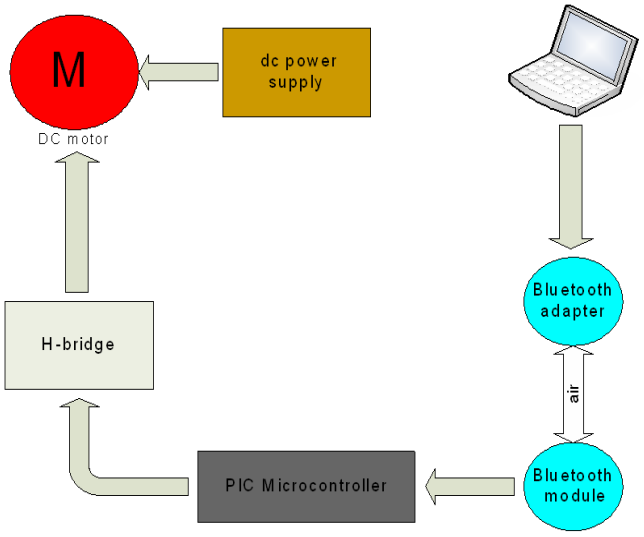


Figure 2: Block diagram

The general view of the system can be described in the Figure 2

The design depends on the duty cycle value, how to transmit it from the user interface to the PIC microcontroller and how to generate the corresponding PWM signal. The control circuit consists of two main devises: the PIC microcontroller and the H-bridge, the H-bridge –as shown in Figure 3- is a power electronic circuit consists of switches that controls the voltage applied to the motor with freewheeling diodes not shown for simplicity, this circuit takes its control signal (PWM signal) from the PIC [5].

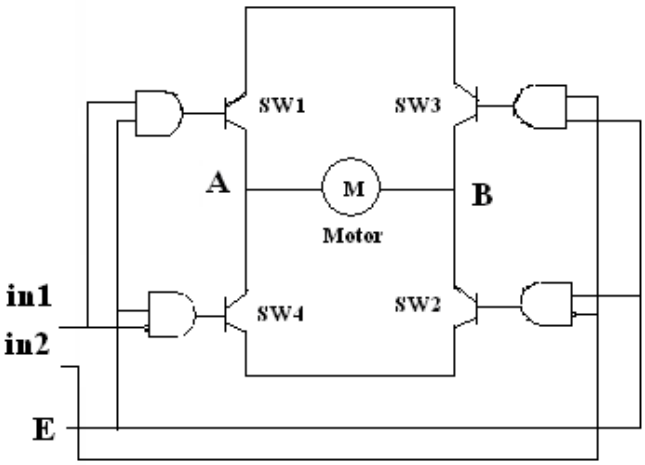


Figure 3: H-bridge connected to motor

Pulse width modulation is the most efficient method of generating control signal with constant volts and frequency control requirement, PWM involves taking a square wave and, without changing the frequency, changing how long in each cycle the wave spends high and low [6]. Thus by changing the duty cycle the average voltage can be changed as will be shown in the results.

If a PWM signal is applied to sw_1 and sw_2 with a specific duty cycle, the motor may rotate clockwise with speed corresponding to the duty cycle value, if sw1 and sw2 are turned off, and the same PWM signal is applied to sw_3 and sw_4, the motor will rotate counterclockwise at the same speed. To achieve this, a square wave is applied on in_1 with duty cycle ρ and the opposite one on in_2; The H-bridge that will be used in the design is L298N which is a good choice because it is small so it is suitable to a 12v DC motor. This H-bridge can take an input voltage range from 2.5-46 volt [7].

The microcontroller acts like the brain of the DC motor speed control system. The microcontroller chip that has been selected for the purpose of controlling the speed of DC motor is PIC16F877A manufactured by Microchip. The function of the PIC is to generate a PWM output signal corresponding to the received value from the Bluetooth module. PIC16F877A can be programmed to receive an input signal from the Bluetooth transmitter.

RS232 is a popular communications protocol for connecting modems and data acquisition devices to computers. It can be plugged straight into the computer's serial port (also known as the COM port). The PIC provides a serial RS232 transceiver built-in function; which can communicate any device that provides this communication protocol which PARANI ESD200 provides. This function identifies the receiver pin, the transmitter pin and the baud rate. Another built in function that corresponds to the RS232 function is getc( ). When an 8-bit character is sent to the PIC, this function waits for a character to come over the RS232 receiver pin and returns the character value in hexadecimal. The PIC is programmed by using C-programming language, after writing the program compiling and checking the program of errors can be done using PIC C Compiler software. This software compiles the program and if there was no errors it generates many files one of them the HEX file; this file is used by the programmer kit to download the program on the PIC.

In order to have a wireless data transmission, the system must include a transmitter unit and a receiver unit, in this system, the transmission unit may be a laptop built in device or a Bluetooth device connected serially or via a USB port to the computer, the receiver is a Bluetooth module (chip); this module has its own system that should be programmed to maintain the communication process.

The chosen Bluetooth module is Parani-ESD200. This Series is designed for integration into user devices by on-board installation. They are connected to the device via built-in UART interface and communicate with other Bluetooth device. Parani-ESD is a module device for wireless serial communication using Bluetooth technology that is international a standard for short range wireless communications. Parani-ESD can communicate with other Bluetooth devices that support the Serial Port Profile.

A Bluetooth device can play a role as a master or slave. Master tries to connect itself to other Bluetooth devices, and slave is waiting to be connected from other Bluetooth devices. A Bluetooth connection is always made by a pair of master and slave devices. A slave can be in two modes, Inquiry Scan or Page Scan mode. Inquiry Scan mode is waiting for a packet of inquiry from other Bluetooth device and Page Scan mode is waiting for a packet of connection from other Bluetooth device. Every Bluetooth device has its unique address, called BD (Bluetooth Device) address, which is composed of 12 hexadecimal numbers.

Parani-ESD supports two security options, Authentication and Encryption. If the Authentication option is enabled, a Pin Code value must be entered. If the authentication is enabled, the connection, between the Master and Slave device must share the same Pin Code. In case that Parani-ESD connects to another Bluetooth device, that requires authentication, you must know the other device's Pin Code. In general, most Bluetooth devices have a pin code of 1234 or 0000. If you check Encryption option, the Parani-ESD will encrypt packets and send it to the device. The Encryption options works well in case that only one of the devices between Master and Slave use the Encryption option.

The user interface is a program that allows the user to control the motor. The operation of this program is to send the new value of the motor speed and the direction of rotation which means to send the duty cycle value. It is a program that allows the user to change the status and the speed of the

motor by clicking an Icon making it user friendly. The interface is a window with four icons these include an icon to increase the speed, another to decrease it. A third icon is used to start the motor and a fourth to stop the motor. The user can increase and decrease the speed of the motor by pressing the relevant icon and each press changes the motor speed by 0.1 of the rated speed; in other words 0.1 increase in the duty cycle value.

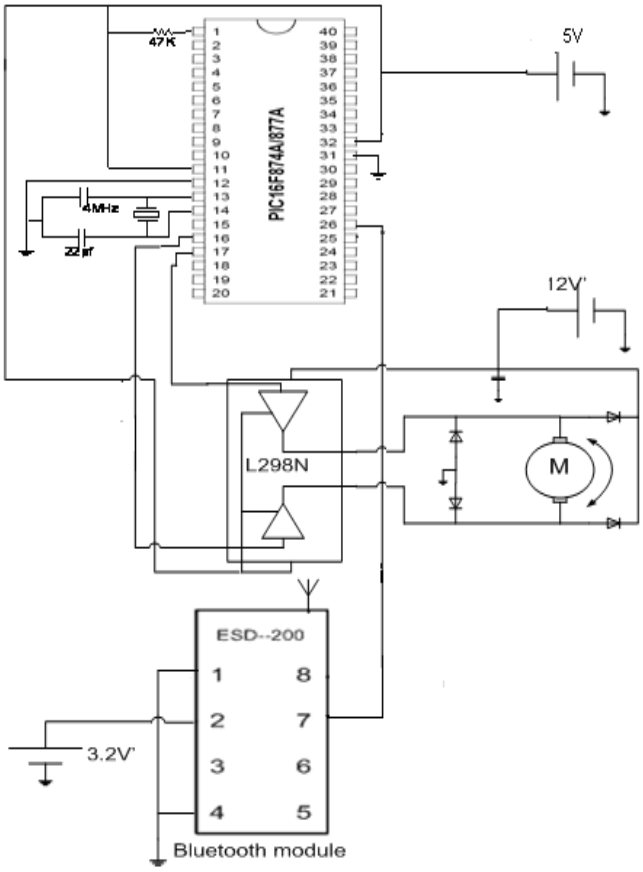


Figure 4: Full Schematic

In order to change the direction of rotation, the motor must be stopped first and then the direction of rotation is changed, if the user sends a command to change the direction while the motor is rotating at a high speed, the system will not respond as it may damage the control circuit. So to prevent this, the program doesn't respond to a sudden change, the program allows a change in direction just when the motor speed is zero then the user can increase the speed in the other direction. This interface is programmed in java programming language, java language provides an easy to use graphical interface functions which is very important in our program also it provides a class that can communicate and identify any Bluetooth device attached to the computer and identifying this class in the program you can send and receive data to and from the Bluetooth device.

The full schematic diagram is shown in Figure 4, while the flow chart of the program that has been downloaded to the PIC is shown in Figure 5.

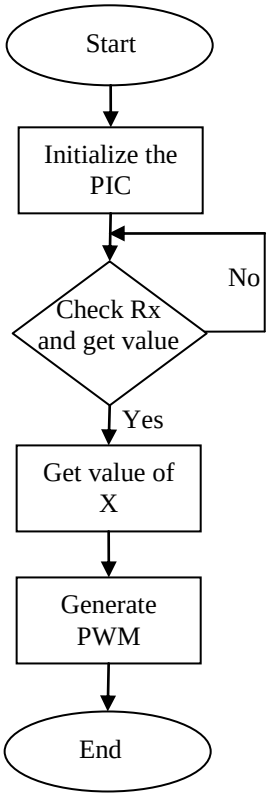


Figure 5: Program flow chart

3. Results and discussion

A control unit that is able to run a DC motor wirelessly using Bluetooth technology is designed, to transmit control signals from the user interface (PC side) to the receiver on the motor side. This controller functions through a software application on a laptop or desktop computer from within 10 meters of the motor. Controlling a dc motor means controlling the status of the motor (on or off); changing the direction of rotation and the speed.

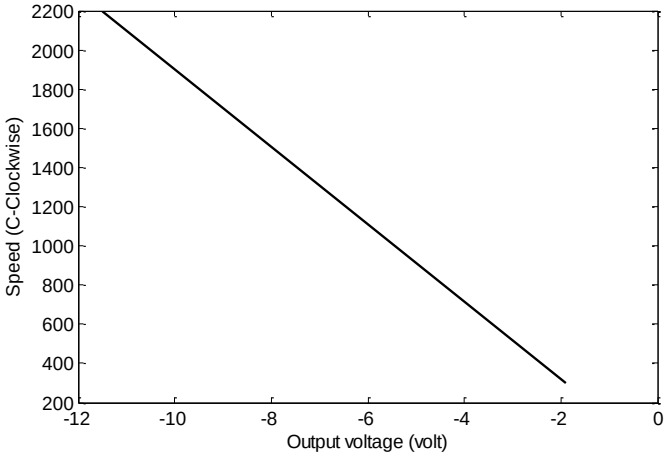


Figure 6: Relationship between speed and voltage in counterclockwise

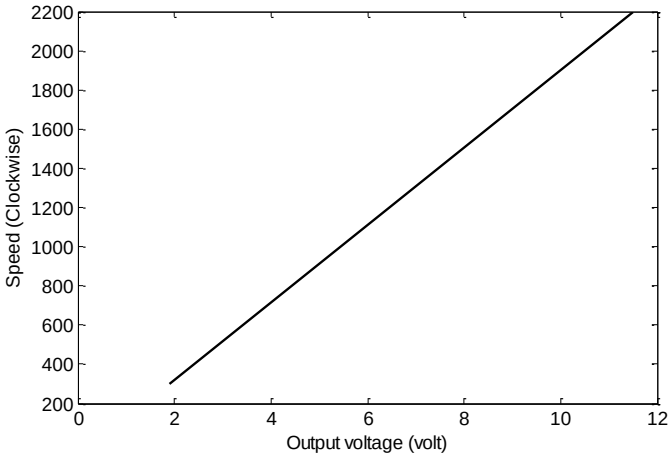


Figure 7: Relationship between speed and voltage in clockwise

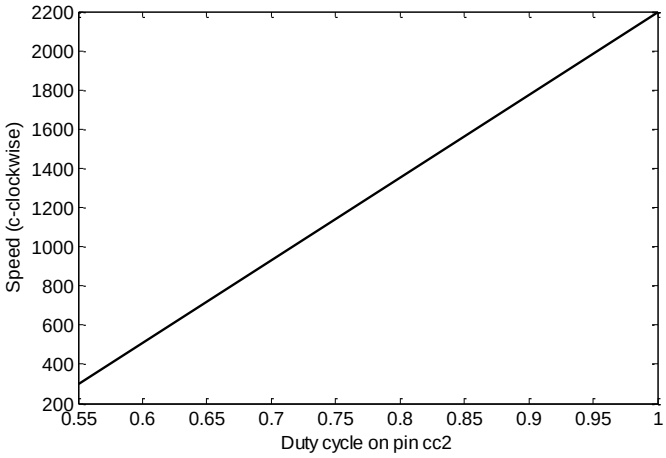


Figure 9: Relationship between the duty cycle and the speed in c-clock wise

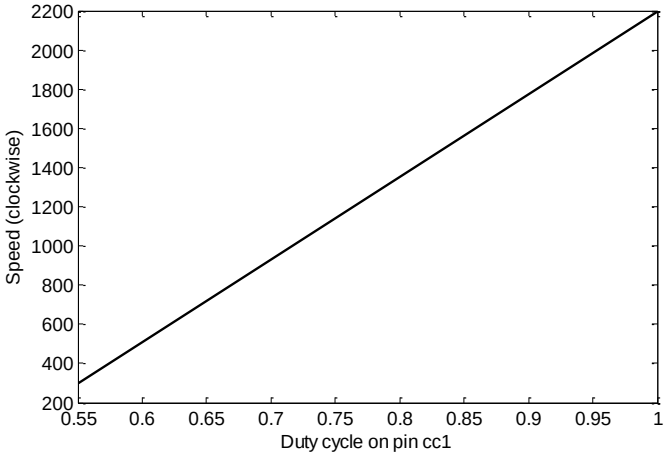


Figure 8: Relationship between the duty cycle and the speed in clock wise

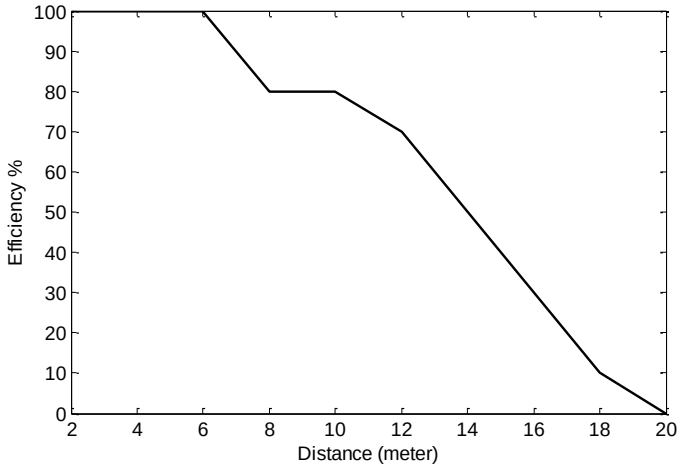


Figure 10: Relationship between distance (in meters) and efficiency of the system

Figures 6 and 7 show the relationship between the speed and the average voltage at the output of the H-bridge in both counterclockwise and clockwise respectively. It is clear that reducing the voltage will increase the speed in counterclockwise, while it will give the opposite effect for clockwise

In Figures 8 and 9, the relationship between the duty cycle and the speed are plotted. Applying the duty cycle on Pin cc1 in the PIC, the direction of the rotation is in clock wise while it is counterclockwise if the signal is applied on cc2. In both cases, the speed will increase by increasing the duty cycle.

The system efficiency is checked in Figure 10. It shows relationship between the transmitting distance and the response of the system. This is done by transmitting 10 orders from each distance and count how many times the system responds correctly to the orders. The efficiency was introduces as the number of correct responds divided by the number of total tries.

4. Conclusion

In this design, a prototype of a wireless Bluetooth control of DC motor is developed and implemented. It is a system that controls a DC motor speed and direction of rotation, A small 12v dc motor is used as a sample of the industrial applications where the project goal to be used in. This model provides an easy to use graphical interface that the user can control from a PC or laptop**.**

5. Acknowledgments

Authors would like to thank Palestine Technical University-Kadoorie (PTU) for supporting this research and allowing us to conduct this work in the university labs.

Sincere thanks to the undergraduate students: Ali Badwan, Bashshar Asaad, Leena Younis, Sahar Ali and Mujahed

Sawalha for their help in building the hardware model and performing the results readings.

References

[1] Mutamed Khatib, Mohd Fadzil Ain, Farid Ghani and Syed Idris Syed Hassan, "Study of Parameters Effect on the Performance of Precoding and Equalizer Sharing Systems," NAUN International Journal of Communications, Issue 1, Volume 2, pp. 54-64, December 2008.

[2] J. Y. Khan, J. Wall, M. A. Rashid, "Bluetooth-Based Wireless Personal Area Network for Multimedia Communication," Electronic Design, Test and Applications, IEEE International Workshop on, pp. 47, The First IEEE International Workshop on Electronic Design, Test and Applications (DELTA '02), 2002.

[3] Sailesh Rathi, "Blue Tooth Protocol Architecture," Dedicated Systems Magazine, Embedded Internet - 00q4 - p. 28 , 2000

[4] Qi Yuan, Zhao Xiaokang, Zhang Qiong, "Key Technology and System Design in Mobile Supply Chain Management," International Symposium on Electronic Commerce and Security(isecs), pp.258-262, 2008

[5] Stephen Chapman, "Electric Machinery Fundamentals," McGraw-Hill, 4th edition, 2003.

[6] Simon Haykin, Michael Moher, "An Introduction to Analog and Digital Communications," Wiley, 2nd edition, 2006

[7] Muhammad Almazidi, Rolin D., Danny Causy "PIC microcontroller and embedded systems using assembly and C for PIC18", Pearson international edition, 2008

Author Biographies

Arafat Zaidan received his B.Eng. in Electrical & Electronic Engineering from University of Leicester in 1993, and the MPhil/PhD in Digital Control Engineering, University of Salford in 2000. The research was primarily concerned with deriving mathematical models and implementing a pole placement controller for a powered orthosis He was a control System Engineer with research experience in modern control strategies and plant supervision, and worked for five years at the Hashemite University in Jordan as an assistant professor in the department of Mechatronics. Currently working as an assistant professor in the Department of Electrical Engineering, Palestine Technical University (Kaddorie), Tul Karm, Palestine. Dr. Arafat Zaidan has a number of publications to his credit in various international journals and conference proceedings. He is a member of IEEE, Palestinian Engineers Association.

Mutamed Khatib received B.Sc. in Telecommunication Engineering from Yarmouk University, Irbid, Jordan in 1996 and M.Sc. in Electrical & Electronic Engineering from Jordan University for Science & Technology, Irbid, Jordan in 2003. He received his PhD Degree in wireless and mobile systems from University Sains Malaysia (USM), Malaysia in 2009. From 1996 to 2005, he worked as Transmission, Outside Broadcasting & Studio Engineer in Palestinian Broadcasting Corporation (PBC). From 2005 to 2009 he worked as an Instructor in the Department of Electrical Engineering, Palestine Technical University (Kadoorie), Tul Karm, Palestine. Since September 2009, Dr Mutamed Khatib is working as Assistant professor in the same university. Dr. Khatib has a number of publications to his credit in various international journals and conference proceedings. He is a member of IEEE, Palestinian Engineers Association and Arab Engineers Association.

Basim Alsayid received B.Sc. in Electrical Engineering from Studies University of Bologna, Bologna, Italy in 1991. He received his PhD Degree in Electrical Drives Engineering from University of Bologna, Bologna in 2002. From 2002 to 2007 he worked as Assistant professor in the Department of Electrical Engineering, Palestine Technical University (Kadoorie), Tul Karm – Palestine. From 2007 to 2009 he worked as the head of the electrical engineering department and from 2009 till now he is the dean of the college of engineering and technology at the same university. He is a member of IEEE, Palestinian Engineers Association. He is now involved in a 2 years research program about design and control of photovoltaic systems with a French research group.

Block Transmission Systems in Synchronous Multiuser Communications

Mutamed Khatib[1] and Farid Ghani[2]

[1]Palestine Technical University – Kadoorie (PTU), Department of Telecommunication Engineering and Technology, Palestine.
mutamed.khatib@gmail.com
[2] Universiti Malaysia Perlis (UniMAP), School of Computer and Communication Engineering, Perlis, Malaysia.
faridghani@unimap.edu.my

***Abstract*:** This paper presents equalization and coding techniques of signals in a CDMA system to transmit digital data over time varying channels such as the HF mobile channels. The receiver equalization is used to improve the performance of the system by transmitting the data in blocks, and the coding is designed such that no signal processing is required at the receiver except testing the received signal against appropriate threshold. The signals at the transmitter are arranged in groups and adjacent groups are separated by zero-level elements to avoid interference between consecutive transmitted groups. The impulse response is required to be known at the transmitter which is the requirement for all systems that employ coding at the transmitter.

***Keywords*:** Coding, equalization, CDMA, mobile, ISI.

1. Introduction

Block Linear Equalizer (BLE) has been proposed for transmitting digital data over time varying and time dispersive channels [1-3]. It is a synchronous serial data transmission system that employs transmission of alternating blocks of data and training symbols, where each data block is detected as a unit [3]. This system requires that the channel impulse response is known with the assumption that it remains constant during the transmission of the block.

In BLE, the channel is always perfectly equalized with no error extension effects. This may be introduced as the main advantage in comparison with the conventional linear and nonlinear equalizers. Although the transmission efficiency is reduced due to the addition of training symbol blocks between consecutive data blocks, this disadvantage is more than offset in comparison to the advantages offered by the system.

In mobile systems, the goal of maintaining low cost and complexity, especially at the mobile unit, is very important for the designer [4, 5]. It is obvious that the new designed systems should have high capacity, flexibility, can provide the required data rates and services the users need both for speech and high end applications such as video. Low power transmission results in high capacity, but the signal will be very sensitive to disturbances, which may be either noise or interference from other users. High speed can be achieved by reducing the symbol period, but then reflection problems from buildings, mountains, cars etc. will arise. Finally, high flexibility can be achieved by designing a system that supports different user requirements, but then it is important not to lose efficiency in the transmission.

As a result, researchers have recently begun investigating signal processing techniques that eliminate the effect of the channel, and move computational complexity from the mobile unit to the base station, where it can be managed more efficiently [6, 7]. In such techniques, a transmitter-based interference cancellation is done at the base station and just simple linear processing, *e.g*., threshold decision at the mobile unit. This technique is called precoding or pre-equalizing.

Many researches have tried to simplify the receiver unit, for example, Reynolds, et al. [7] have proposed a precoding technique that simplifies the receiver. They use a sophisticated channel estimation method [8] to have knowledge of the channel elements, *i.e*, the delayed version of the spreading waveform, and the complex channel fading gain for each user in each path. The original information can be retrieved at the mobile unit using a matched filter. Vojčić, et al. [9] and Esmailzadeh, et al. [10] suggested precoding techniques for synchronous CDMA over AWGN channel. In their design, they used a RAKE receiver. The disadvantage for RAKE reception is that it is sensitive to channel mismatch and its performance is generally inferior to MMSE or decorrelator based multiuser interference rejection [7].

In a K-user multipath CDMA system with time disruptive channels, intersymbol interference (ISI) is introduced when the delay spread is large resulting in increased bit error rate (BER). ISI can be removed by inserting guard intervals between symbols to insure that the delayed version of the pulse will not affect the other pulses from other paths. When the multipath delay spread is less than the symbol interval, ISI can be neglected because the delayed pulse will not affect the pervious or the next pulse from the other paths [11]. The precoding technique can achieve both portable unit simplicity and ISI reduction.

An important assumption for precoding in multipath channels is that the transmitter has information about all channels between it and active receivers. This information can be obtained from receivers via feedback channels [12]. Another important requirement is that the multipath channel is slow, *i.e*., that it remains constant over the block of precoded bits. Though, the length of the precoding block can be adjusted to match the channel dynamics.

The practical applications of transmitter precoding can be found in wireless local loop, wireless LAN's and indoor communications in general, as well as any other wireless scenario where the precoding block size can be made sufficiently small so that the channel appears slow [9].

In this paper, we introduced an equalization technique where an equalizer is used in the receiver to reduce the effect of the channel as an alternative approach of the block linear equalizers and block decision feedback equalizers introduces earlier [1, 3, 13]. Also, we proposed a precoding technique for CDMA downlink in synchronous multipath fading channel that reduces the complexity of the receiver in which the detection process needs only a threshold decision to retrieve the transmitted data. In this technique, there is no need for match filtering or any other processing is needed.

The mobile unit simplification depends on using a precoding technique at the base-station that reduces the receiver at the mobile unit to a decision process due to a certain threshold testing. It depends on the channel's prior knowledge at the base station, so, the channel characteristics are assumed to be known both at the transmitter and the receiver.

This paper is organized as follows. In Section II, we present the system model. The design and analysis of the precoder are presented in Section III. In Section IV, numerical results are presented and the system performance is compared with those without transmission in blocks. Finally, Section **V** presents the conclusions of our study.

Notation: All bold faces variables in this paper denote vectors and matrices.

2. Block Linear Equalizer

The system model of the BLE is shown in Figure 1. The input to the channel is the corresponding antipodal and statistically independent signal elements, after being grouped in vectors of size *m*, and it may be either binary or multilevel. The baseband channel has an impulse response $y(t)$, which includes the transmitter and receiver filters used for pulse shaping and linear modulation and demodulation. The impulse response $h(t)$ of the transmitter and receiver filters in cascade is assumed to be such that:

$$h(t)=\begin{cases}1 & t=0\\ 0 & t\neq 0\end{cases} \tag{1}$$

In the frequency domain, $h(t)$ can be written as [11]:

$$H(f)=\begin{cases}\frac{1}{2}T(1+\cos(fT)) & -\frac{1}{T}<f<\frac{1}{T}\\ 0 & elsewhere\end{cases} \tag{2}$$

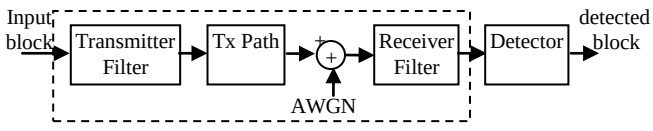


Figure 1: Model of the Block transmission system.

During transmission, Additive White Gaussian Noise (AWGN) with zero mean and a two sided power spectral density of σ^2 is added, giving the zero mean Gaussian waveform $w(t)$ at the output of the receiver filter, hence the received signal is:

$$r(t)=\sum_i s_i y(t-iT)+w(t) \tag{3}$$

The received signal is sampled at time instant $t=iT$, where T is the symbol interval. Here, consecutive blocks of m information symbols at the input to the transmitter filter are separated by blocks of g zero level symbols as shown in Figure 2, where g is the largest memory length of the channel $y(t)$, and $y=\begin{bmatrix}y_o & y_1 & \dots & y_g\end{bmatrix}$ is the sampled impulse response.

$$\underbrace{0\ \cdots\ 0}_{g\text{ zero-level elements}}\ \Big|\ \underbrace{s_1\ \cdots\ s_m}_{m\text{ signal elements}}\ \Big|\ \underbrace{0\ \cdots\ 0}_{g\text{ zero-level elements}}\ \Big|$$

Figure 2: Structure of transmitted signal elements in a block transmission system.

For each received group of m signal-elements, there are $n=m+g$ sample values at the detector input that are dependent only on the m elements, and independent of all other elements. The detector uses these n values in the detection of the symbol block, then, the detected values are used for the estimation of the channel sampled impulse response using the same equipment.

If only the i^{th} signal-element in a group is transmitted, in the absence of noise and with s_i set to unity, the corresponding received n sample values used for the detection of m elements of a group are given by the n-component row vector

$$\mathbf{Y}_i=\overbrace{0\ \dots\ 0}^{i-1}\ \overbrace{y_0\ \ y_1\ \dots\ y_g}^{g+1}\ \overbrace{0\ \dots\ 0}^{m-i} \tag{4}$$

Where y_h must be non-zero for at least one h in the range $0:g$. The sum of the m received signal elements in a group in the absence of noise is given by:

$$\mathbf{R}=\sum_{i=1}^{m}s_i\mathbf{Y}_i=\mathbf{SY} \tag{5}$$

where **S** is the m-component row vector whose i^{th} component is s_i and represents the transmitted signal block. **Y** is an $m\times n$ matrix whose i^{th} row $\mathbf{Y}_i$ is given by Equation 4. Since at least one of the y_h is non–zero, the rank of the matrix **Y** is always m, and hence, the m rows of the matrix **Y** are linearly independent. Note that the sampled impulse response of the channel completely determines the matrix **Y**.

When AWGN is present, the sample values corresponding to a received signal block at the detector input is given by the n components of the row vector **R** where:

$$\mathbf{R}=\mathbf{SY}+\mathbf{W} \tag{6}$$

Where **W** is the n-component noise vector whose components are sample values of statistically independent Gaussian random variable with zero mean and variance σ^2.

The vectors **R**, **SY** and **W** can be represented as points in the n-dimensional Euclidean signal space. Assume that the detector has prior knowledge of $\mathbf{Y}_i$, but has no prior knowledge of the s_i or σ^2. A knowledge of the $\mathbf{Y}_i$ of course implies a knowledge of the channel impulse response. Since the detector knows **Y**, it knows the m-dimensional

subspace spanned by $\mathbf{Y}_i$ and hence the subspace containing the vector **SY**, for all s_i. Since the detector has no prior knowledge of s_i, it must assume that any value of **S** is as likely to be received as any other, and in particular, as far as the detector is concerned, s_i need not be ± 1. For a given vector **R** the most likely value of **SY** is now at the minimum distance from **R**. Clearly, if **R** lies in the subspace spanned by the $\mathbf{Y}_i$, then the most likely value of SY is **R**. In general, **R** will not lie in this sub-space, and in this case, the best estimate the detector can make of **S** is the *m*-component vector **X**, whose components may have any real values and are such that **XY** is at the minimum distance from **R**. By the projection theorem [14], **XY** is the orthogonal projection of **R** onto the *m*-dimensional subspace spanned by the $\mathbf{Y}_i$. It follows that $\mathbf{R}-\mathbf{XY}$ is orthogonal to each of the $\mathbf{Y}_i$, so that:

$$(\mathbf{R}-\mathbf{XY})\mathbf{Y}^T = 0 \tag{7}$$

In other words,

$$\mathbf{X} = \mathbf{R}\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1} \tag{8}$$

$\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}$ is a real $n \times m$ matrix of rank *m*. Since the $m \times n$ matrix **Y** has rank *m*, the $m \times m$ matrix $\mathbf{Y}\mathbf{Y}^T$ is symmetric positive definite and its inverse will always exist.

Thus if the received signal vector **R** is fed to the *n* input terminals of the linear network $\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}$, the signals at the *m* output terminals are the components x_i of the vector **X**, where **X** is the best linear estimate the detector can make of **S**, under the assumed conditions. Thus:

$$\mathbf{X} = \mathbf{R}\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1} \tag{9}$$

$$= (\mathbf{SY}+\mathbf{W})\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1} \tag{10}$$

$$= \mathbf{S}+\mathbf{U} \tag{11}$$

The *m*-component row-vector **U** is the noise vector at the output of the network $\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}$. Each component u_i of the noise vector **U** is a sample value of a Gaussian random variable and a variance which is not normally equal to σ^2, and which normally differ from one component to another.

In the final stage of the detection process, the receiver examines the signs of the x_i and allocates the appropriate binary values to the corresponding signal elements, to give the detected value of **S**. The detector requires no prior knowledge of the received signal level and is linear up to the decision process just mentioned. It can be seen that in the linear $n \times m$ network $\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}$, $\mathbf{Y}^T$ represents a set of *m* matched filters or correlation detectors tuned to the *m* $\mathbf{Y}_i$ whose *m* outputs feed the inverse network represented by the matrix $\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}$ as shown in Figure 3.

The probability of error for the block linear equalizer is given by:

$$P_e = \frac{1}{2}erfc\left(\frac{1}{\eta}\sqrt{\frac{E_b}{N_o}}\right) \tag{12}$$

where η^2 is the effect of the linear network matrix $\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}$ on the AWGN vector.

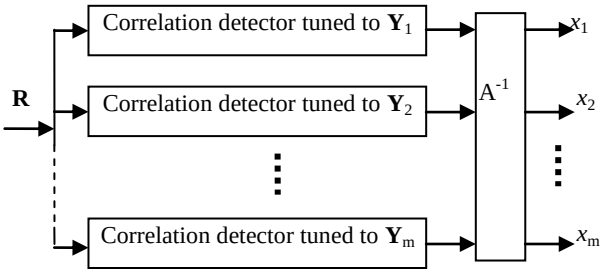


Figure 3: Optimum Linear Detector

3. BLE with Precoding

In this section, we developed a technique for CDMA downlink in synchronous multipath fading channel that reduces the complexity of the receiver in which the detection process needs only a threshold decision to retrieve the transmitted data, no match filtering or any other processing is needed. In the base station, a precoder is used to generate a code from the transmitted signal that makes it immune to the channel, so, there is no need for any further equalization process in the receiver. This reduces the mobile unit receiver to a decision process due to a certain threshold testing. When comparing the cost of adding a coder at the base station with the savings at the receiver units, it will be acceptable because few base stations serve many receiver units in the downlink.

The system considered is shown in Figure 4. The signal at the input to the transmitter is a sequence of *k*-level element values $\{s_i\}$, where k = 2, 4, 8,… and the $\{s_i\}$ being statistically independent and equally likely to have any of the possible values. The buffer-store at the input to the transmitter holds *m* successive element values $\{s_i\}$. In the coder, the *m* $\{s_i\}$ are converted into the corresponding *m* coded signal-elements. The coder performs a linear transformation on the *m* $\{s_i\}$ to generate the corresponding sequence of impulses that is fed to the baseband channel $y(t)$ which is assumed that it is either time invariant or varies slowly with time.

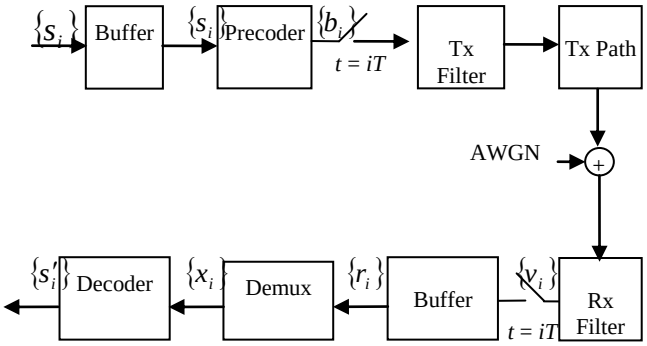


Figure 4: The downlink of a CDMA system

White Gaussian noise, with zero mean and variance σ^2, is assumed to be added to the data signal at the output of the transmission path, giving the Gaussian waveform $w(t)$ added to the data signal.

The sampled impulse-response of the baseband channel in Figure 4 is given by the $(g+1)$ component row vector: [3, 11, 15]

$$y_i = y(iT) = y_o \quad y_1 \quad \dots \quad y_g \tag{13}$$

where $y_o \neq 0$, and $y_i = 0$ for $i < 0$ and $i > g$.

The received waveform $r(t)$ at the output of the baseband channel is sampled at the time instants $\{iT\}$, for all integers $\{i\}$. The $\{r_i\}$ are fed to the buffer store which contains two separate stores. While one of these stores holds a set of the received $\{r_i\}$ for a detection process, the other store is receiving the next set of $\{r_i\}$ in preparation for the next detection process. A group of *m* multiplexed signal-elements are detected simultaneously in a single detection process, from the set of $\{r_i\}$ that depends only on these elements. The receiver uses the knowledge of the $\{y_i\}$ and the possible values of $\{s_i\}$ in the detection of the *m* element values $\{s_i\}$ from the received samples $\{r_i\}$. A period of *nT* seconds is available for the detection process, *n* is given by:

$$n = m + g \tag{14}$$

where *m* is the block length, and *g* is the channel length –1.

Except where otherwise stated, the decoder in Figure 4 determines from the appropriate set of received $\{r_i\}$ the *m* estimated $\{x_i\}$ of the *m* element-values $\{s_i\}$ in a received group of elements. Each x_i is an unbiased estimate of the corresponding s_i such that:

$$x_i = s_i + u_i \tag{15}$$

where u_i is a zero mean Gaussian random variable. The detector detects each s_i by testing the corresponding x_i against appropriate thresholds. The detected value of s_i is designated as s_i'.

In the transmitter, using buffer store, an $1 \times m$ vector $\mathbf{S} = [s_1 \quad s_2 \quad \cdots \quad s_m]$ is formed from the symbols to be transmitted. This vector is coded at the transmitter. The coder accepts the input vector **S** and codes it to form the $1 \times n$ signal vector **B**, which is the convolution between the input vector **S** and the $m \times n$ coder matrix **F**, *i.e.*:

$$\mathbf{B} = \sum_{i=1}^{m} \mathbf{s_i F_i} = \mathbf{SF} \tag{16}$$

This convolution process will add a time gap of gT seconds between each pair of adjacent groups of *m* signal-elements. Then, the output values from the coder are fed to the baseband channel. The sampled impulse response of the baseband channel is given by the $g+1$ component row vector as given in Equation 13.

At the receiver, the sample values of the received signal, corresponding to a single group of *m* signal elements, will normally be a sequence of $n+g$ non-zero sample values. The sequence of these $n+g$ sample values in the absence of noise is:

$$v_i = \sum_{j=1}^{n} b_j y_{i-j} \qquad i = 1, 2, \dots, n+g \tag{17}$$

Taking a practical example to clarify the convolution here, if $m = 2$, and $g = 1$, so $n = 3$ and $n + g = 4$. The output of the channel will be the 1×4 vector **V** whose elements are:

$$\mathbf{V} = [b_1 y_o + b_2 y_{-1} + b_3 y_{-2} \quad b_1 y_1 + b_2 y_o + b_3 y_{-1} \quad \cdots \quad \cdots \quad b_1 y_2 + b_2 y_1 + b_3 y_o \quad b_1 y_3 + b_2 y_2 + b_3 y_1]$$

Applying the limitations on the channel impulse response given in Equation 13, we may write **V** as:

$$\mathbf{V} = [b_1 y_o + b_2 0 + b_3 0 \quad b_1 y_1 + b_2 y_o + b_3 0 \quad \cdots \quad \cdots \quad b_1 0 + b_2 y_1 + b_3 y_o \quad b_1 0 + b_2 0 + b_3 y_1]$$

So, this result looks like the multiplication of the vector **B** by a 3×4 matrix **C** that depends on the channel information:

$$\mathbf{C} = \begin{bmatrix} y_o & y_1 & 0 & 0 \\ 0 & y_o & y_1 & 0 \\ 0 & 0 & y_o & y_1 \end{bmatrix}$$

In vector form, it may be written as:

$$\mathbf{V} = \mathbf{BC} \tag{18}$$

where $\mathbf{V} = [v_1 \quad v_2 \quad \cdots \quad v_{n+g}]$ is the $1 \times (n+g)$ received signal, and **C** is the $n \times (n+g)$ channel matrix and its i^{th} row is:

$$\mathbf{C_i} = \overbrace{0 \quad \dots \quad 0}^{i-1} \quad \overbrace{y_o \quad y_1 \quad \dots \quad y_g}^{g+1} \quad \overbrace{0 \quad \dots \quad 0}^{n-i} \tag{19}$$

Assume now that successive groups of signal-elements are transmitted and one of these groups is that just considered, where the first transmitted impulse of the group occurs at time *T* seconds. Figure 5 shows the $n+g$ received samples which are the components of **V**.

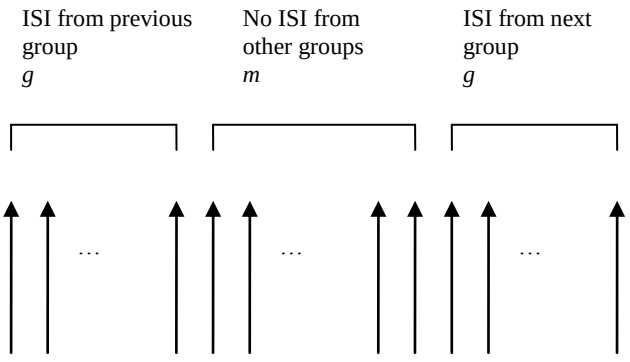


Figure 5: Sequence of $n+g$ samples for one received block.

Due to the Inter Block Interference (IBI), the first *g* components of **V** are dependent in part on the preceding received group of *m* signal-elements, and the last *g* components of **V** are dependent in part on the following received group of *m* elements. Thus there is Intersymbol Interference (ISI) from adjacent received groups of elements in both the first and the last *g* components of **V**. However, the central *m* components of **V** depend only on the corresponding transmitted group of *m* elements, and can therefore be used

for the detection of these elements without ISI from adjacent groups.

Returning back to the same practical example of $m=2$ and $g=1$, the central *m* components of **V** are:

$$\mathbf{V}_{central}=\begin{bmatrix}b_1y_1+b_2y_o+b_3 0 & b_1 0+b_2y_1+b_3y_o\end{bmatrix}$$

which also looks like the multiplication of the vector **B** by a 3×2 matrix that depends on the channel information too, and equal to:

$$\frac{\mathbf{V}_{central}}{\mathbf{B}}=\begin{bmatrix}y_1 & 0\\ y_o & y_1\\ 0 & y_o\end{bmatrix}$$

Mathematically, if we want to "receive" only the central *m* components of **V**, this matrix now represents the channel (mathematically only). To make this matrix somehow looks like the matrix **C**, this matrix is the transpose of a new 2×3 matrix **D** that is equal to:

$$\mathbf{D}=\begin{bmatrix}y_1 & y_o & 0\\ 0 & y_1 & y_o\end{bmatrix}$$

In general, the central *m* components of the vector **V**, $v_{g+1}\quad v_{g+2}\quad \dots\quad v_{g+m}$, can be obtained by introducing a new matrix $\mathbf{BD^T}$ where **D** is the $m\times n$ matrix of rank *m* whose i^{th} row is:

$$\mathbf{D_i}=\overbrace{0\ \dots\ 0}^{i-1}\ \overbrace{y_g\quad y_{g-1}\quad \dots\quad y_o}^{g+1}\ \overbrace{0\ \dots\ 0}^{m-i} \tag{20}$$

Thus,

$\mathbf{BD^T}$ is a $1\times m$ vector where each row of it gives information about the received symbols at that row:

$$\mathbf{BD^T}=\begin{bmatrix}v_{g+1} & v_{g+2} & \dots & v_{g+m}\end{bmatrix} \tag{21}$$

When noise is present, the received vector is:

$$\mathbf{R}=\mathbf{BD^T}+\mathbf{W} \tag{22}$$

where **W** is the zero mean AWGN

Thus the detector can now detect the values of the signal elements by comparing the corresponding $\{r_i\}$ with the appropriate thresholds.

To maximize the tolerance to noise at the detector input, the elements of **B** should be selected such that the total transmitted energy of all the symbols is minimized. *i.e.* $\mathbf{BB}^T=|\mathbf{B}|^2$ must be minimized for the given vector **S**. Thus the problem is to find an $m\times n$ linear network **F** representing the coder, which minimizes the transmitted element energy and, at the same time, satisfies $\mathbf{BD}^T=\mathbf{S}$.

As shown in Appendix A, the coder matrix **F** has to be:

$$\mathbf{F}=\left(\mathbf{DD}^T\right)^{-1}\mathbf{D} \tag{23}$$

Thus, under the assumed conditions, the linear network **F** representing the transformation performed by the coder is such that it makes the *m* signal elements of a group orthogonal at the input of the detector and also maximizes the tolerance to additive white Gaussian noise in the detection of these signal elements.

Now we may redraw the block diagram of the precoding system using the new assumptions about the precoder and the channel matrix as in Figure 6.

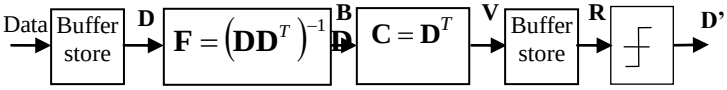


Figure 6: Block diagram of the precoding system in vectors form.

Assume that the possible values of s_i are equally likely and that the mean square value of **S** is equal to the number of bits per element. Suppose that the *m* vectors $\{\mathbf{D}_i\}$ have unit length. Since there are *m* *k*-level signal elements in a group, the vector **S** has k^m possible values each corresponding to a different combination of the *m* *k*-level signal-elements. So, the vector **B** whose components are the values of the corresponding impulses fed to the baseband channel, has k^m possible values. If *e* is the total energy of all the k^m values of the vector **B**, then in order to make the transmitted signal energy per bit equal to unity, the transmitted signal must be divided by:

$$\ell=\sqrt{\frac{e}{nk^m}} \tag{24}$$

The *m* sample values of the received signal from which the corresponding $\{s_i\}$ are detected, are the components of the vector:

$$\mathbf{R}'=\frac{1}{\ell}\mathbf{BD}^T+\mathbf{W} \tag{25}$$

Then, the *m* sample values which are the components of the vector **V** (after taking only the central *m* components), must first be multiplied by the factor ℓ to give the *m*-component vector:

$$\begin{aligned}\mathbf{R}&=\ell\mathbf{V}\\ &=\mathbf{BD}^T+\ell\mathbf{W}\\ &=\mathbf{S}+\tilde{\mathbf{W}}\end{aligned} \tag{26}$$

where $\tilde{\mathbf{W}}$ is an *m* component row vector that represents the AWGN vector after being multiplied by ℓ.

The mean of the new noise vector $\tilde{\mathbf{W}}$ is zero and its variance is:

$$\eta_T^2=\ell^2\sigma^2 \tag{27}$$

Now, the block diagram can be finally drawn as:

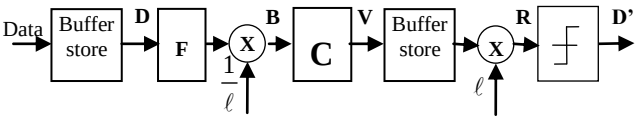


Figure 7: Final block diagram of the precoding system.

Thus, the tolerance to noise of the system is determined by the value of η_T^2. When there is no signal distortion from the channel, $\left(\mathbf{DD}^T\right)^{-1}$ is an identity matrix. Under these conditions, $\ell=1$, so that $\eta_T^2=\sigma^2$.

Note that the $m\times n$ linear network transforms the transmitted signal such that under the assumed conditions,

the corresponding sample values at the receiver are the best linear estimates of the $\{s_i\}$.

Taking into consideration now that the variance now is η_T instead of σ, the bit error rate may be calculated as:

$$P_e = \frac{1}{2} erfc\left[\frac{\sqrt{\xi_b}}{\sqrt{2}\eta_T}\right] = \frac{1}{2} erfc\left[\frac{\sqrt{\xi_b}}{\sqrt{2}\ell\sigma}\right] = \frac{1}{2} erfc\left[\frac{1}{\ell}\sqrt{\frac{\xi_b}{N_o}}\right] \quad (28)$$

Although the result of Equation 28 will show some enhancement on the performance of the system in comparison with the block linear equalizer, the main goal obtained in this system is that no processing to eliminate the effect of the channel is done in the receiver. That leaves the receiver quite simple, and will save a lot through the designing process. Although there will be a little complication in the transmitter (*i.e.* base station), but comparing the savings in the manufacturing of the receivers (*i.e.* handsets) will show that the precoding in the base station is nothing

.

4. Numerical Results

In the Block Linear Equalizer, it has been shown that the best linear estimate of the transmitted group of *m* signal-elements is given when the received signal due to a group of *m* signal-elements is processed by the linear network $\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}$, where the matrix $\mathbf{Y}$ is the $m \times n$ matrix that represent the channel characteristics and its i^{th} row is given by:

$$\mathbf{Y_i} = \overbrace{0 \quad \ldots \quad 0}^{i-1} \quad \overbrace{y_o \quad y_1 \quad \ldots \quad y_g}^{g+1} \quad \overbrace{0 \quad \ldots \quad 0}^{m-i} \quad (29)$$

where $y_i = y_o \quad y_1 \quad \ldots \quad y_g$ is the channel impulse response.

The main goal of BLE is to implement an equalization technique that removes the effect of the channel. All the process is done in the receiver, but it uses the same block techniques as in the precoding system. So, it will be significant to make a comparison between those two systems.

The bit error rate curves for the two systems are shown in Figure 8. The signal elements are binary antipodal having possible values as +1 or −1. There are four elements in a group (block length $m = 4$) and these are equally likely to have any of the two values. The sampled impulse response of the channel is $\{y_i\} = [0.408 \quad 0.817 \quad 0.408]$. This channel has a second order null in the frequency domain and introduces severe signal (amplitude) distortion [3]. For the sake of comparison the bit error rate for the Linear Transversal Filter is also given.

The precoding system has better performance than the block linear equalizer, each one of them provides the best linear estimate of a received group of *m* signal elements and in the block linear equalizer, all the signal processing is carried out at the receiver while in the proposed precoding system all the processing is done at the transmitter and leaves the receiver quite simple.

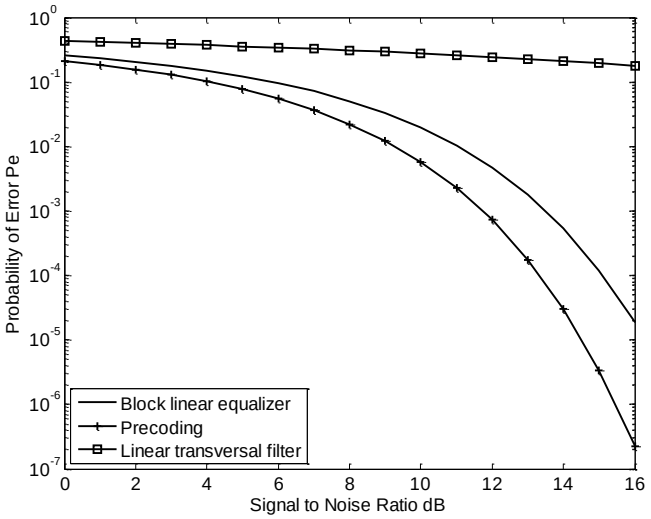


Figure 8: Probability of bit error versus SNR for the precoding system.

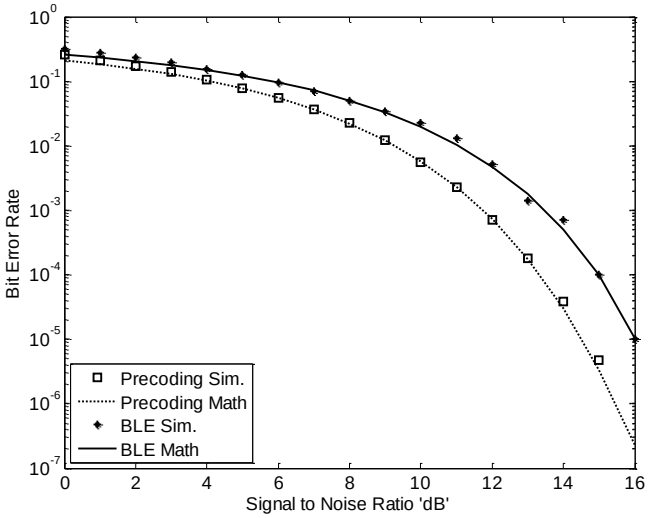


Figure 9: Mathematical and simulation results for the precoding system.

In simulation, we used Matlab as a simulation program. We built two command programs that follow the steps of the systems block diagrams. We assumed that the channel characteristics are known, and fixed for all the transmission procedure. Of course, channel impulse response may vary through the transmission, but it must be fixed within the block, and it should be known all the time. We didn't suggest a certain estimation method, but literature is rich with many methods, and any adaptive one may be used. Here, we'll give a brief description for the program of the precoding system as an example. Before starting the transmission loop, the program will construct the channel matrix **C**, the matrix **D** (that is another form of **C**, and is the main component of the precoder matrix **F**), the precoder matrix **F**, the factor ℓ that makes the energy per bit is unity at the input of the channel, and at last, the input data stream. Then, the program starts the first loop by selecting the first *m* components from the data

stream and multiply it by the coder, then it will divide it by the value of ℓ given by Equation 24 to normalize the energy per bit. After that, the coded and normalized block **B** is multiplied by the matrix channel **C**, and AWGN with the pre-identified SNR is added to the vector to form the received data vector. Another stage of the program will act as a buffer to select only the central *m* components of the received vector, and this will be multiplied again by the factor ℓ to reverse the division process done earlier at the transmitter. The last stage of the program will act as a comparator that will compare the received data with a decision level (0) to give the output data again in the form of $+1$ or -1. Then, a new loop will start again.

After that, another part will compare the input data with the output data to determine the number of bits in error and the bit error rate.

In order to make a comparison between the mathematical results for the systems presented in Figure 8, and the simulation program results, we introduced Figure 9, which clarify that the behavior is the same.

5. Acknowledgment

Authors would like to thank Palestine Technical University (PTU) for supporting this research.

References

[1] S. Crozier, D. Falconer, and S. Mahmoud, "Reduced Complexity Short-Block Data Detection Techniques for Fading Time-Dispersive Channels," IEEE Transactions on Vehicular Technology, vol. 41, no. 3, pp. 255-265, 1992.

[2] K. Hayashi and H. Sakai, "Single Carrier Block Transmission without Guard Interval," presented at the 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Finland, 2006.

[3] G. Kaleh, "Channel Equalization for Block Transmission Systems," IEEE Journal on Selected Areas in Communications, vol. 13, no. 1, pp. 110-121, 1995.

[4] T. Rappaport, Wireless Communications: Principles and Practice, 2nd ed. Singapore: Prentice Hall, 2002.

[5] C. Teuscher, "Low Power Receiver Design for Portable RF Applications: Design and Implementation of an Adaptive Multiuser Detector for an Indoor, Wideband CDMA Application." PhD Thesis, University of California, USA, 1998.

[6] J. Lee and W. Zhuang, "Channel Precoding with Small Envelope Variations for π/4-QPSK and MSK over Frequency-Selective Slow Fading Channels," IEEE Transactions on Vehicular Technology, vol. 52, no. 1, pp. 24-36, 2003.

[7] D. Reynolds, A. Host-Madsen, and X. Wang, "Adaptive Transmitter Precoding for Time Division Duplex CDMA in Fading Multipath Channels: Strategy and Analysis," EURASIP Journal of Applied Signal Processing, vol. 2002, no. 12, pp. 1365-1376, 2002.

[8] X. Wang and H. V. Poor, "Blind Equalization and Multiuser Detection in Dispersive CDMA Channels," IEEE Transaction on Communications, vol. 46, no. No. 1, pp. 91-103, 1998.

[9] B. R. Vojcic and W. M. Jang, "Transmitter Precoding in Synchronous Multiuser Communications," IEEE Transaction on Communications, vol. 46, no. 10, 1998.

[10] R. Esmailzadeh, E. Sourour, and M. Nakagawa, "Prerake Diversity Combining in Time-Division Duplex CDMA Mobile Communications," IEEE Transaction on Vehicular Technology, vol. 48, no. 3, 1999.

[11] J. Proakis, Digital Communications, 3rd ed. New York: McGraw Hill, 1995.

[12] D. Borah, "Estimation of Frequency-Selective CDMA Channels with Large Possible Delay and Doppler Spreads," IEEE Transactions on Vehicular Technology, vol. 55, no. 4, pp. 1126-1136, 2006.

[13] S. Crosier, D. Falconer, and S. Mahmoud, "Least Sum of Squared Error (LSSE) Channel Estimation," Institute of Electrical Engineering Proceedings, vol. 138, pp. 371-378, 1991.

[14] R. Varga, Matrix Iterative Analysis. Englewood Cliffs , New Jersey: Prentice Hall, 1962.

[15] F. Ghani, "Block Data Communication System for Fading Time Dispersive Channels," presented at the 4th National Conference on Telecommunication Technology, Malaysia, 2003.

Appendix A

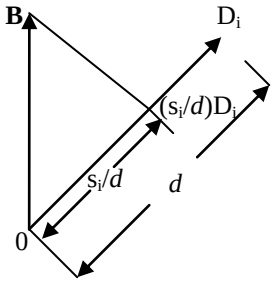


Figure 10: vectors **B** and $\mathbf{D}_i$ for $d > 1$ and $s_i = 1$

Assuming that d is the length of the vector $\mathbf{D}_i$, where $i = 1, 2, \ldots, m$, (*i.e.* the distance of the point $\mathbf{D}_i$ from the origin in an nT dimensional vector space). It can be easily shown that $\mathbf{D}_i$ is independent of *i*. $\mathbf{B}\mathbf{D}_i^T$ is the inner product of the vectors **B** and $\mathbf{D}_i$ so that it is d times the value of the orthogonal projection of **B** onto the vector $\mathbf{D}_i$ [14]. Thus, **B** lies on the hyper plane ($n-1$ dimensional subspace) which contains the point $(s_i/d)\mathbf{D}_i$ and which is orthogonal to the vector given by this point, so that the hyper plane is orthogonal to the line joining the origin to $(s_i/d)\mathbf{D}_i$. The vectors **B** and $\mathbf{D}_i$ are shown in Figure 10, for the case where $d > 1$ and $s_i = 1$. The vector **B** must, therefore, lay on each of the *m* hyper planes and as illustrated in Figure 10. Thus, the required vector **B** is the point on these *m* hyper planes at the minimum distance from the origin. By the Projection Theorem [14], **B** is the orthogonal projection of the origin on

to the $nT-m$ dimensional subspace formed by the intersection of the m hyper planes. Thus **B** is the intersection of the m dimensional subspace spanned by the m $\{\mathbf{D_i}\}$ (each of which is orthogonal to the corresponding hyper plane) with the $nT-m$ dimensional subspace formed by the intersection of the m hyper planes. Clearly, **B** can be represented as a linear combination of the m $\{\mathbf{D_i}\}$, so that

$$\mathbf{B}=\sum_{i=1}^{m} e_i \mathbf{D_i} \tag{30}$$

where $\mathbf{E}=[e_1 \quad e_2 \quad \cdots \quad e_m]$

Then, it can be easily shown that

$$\mathbf{S}=\mathbf{B}\mathbf{D}^T=\mathbf{E}(\mathbf{D}\mathbf{D}^T) \tag{31}$$

Thus,

$$\mathbf{E}=\mathbf{S}(\mathbf{D}\mathbf{D}^T)^{-1} \tag{32}$$

and,

$$\mathbf{B}=\mathbf{S}(\mathbf{D}\mathbf{D}^T)^{-1}\mathbf{D} \tag{33}$$

From the previous equation, and knowing that $\mathbf{B}=\mathbf{S}\mathbf{F}$, it is clear that **F** can be given by

$$\mathbf{F}=(\mathbf{D}\mathbf{D}^T)^{-1}\mathbf{D} \tag{34}$$

So,

$$\mathbf{R}=\mathbf{B}\mathbf{D}^T+\mathbf{W} \tag{35}$$

$$=\mathbf{S}(\mathbf{D}\mathbf{D}^T)^{-1}\mathbf{D}\mathbf{D}^T+\mathbf{W} \tag{36}$$

$$=\mathbf{S}+\mathbf{W} \tag{37}$$

Author Biographies

Mutamed Khatib received B.Sc. in Telecommunication Engineering from Yarmouk University, Irbid, Jordan in 1996 and M.Sc. in Electrical & Electronic Engineering from Jordan University for Science & Technology, Irbid, Jordan in 2003. He received his Ph.D.Degree in wireless and mobile systems from University Sains Malaysia (USM), Malaysia in 2009.
From 1996 to 2005, he worked as Transmission, Outside Broadcasting & Studio Engineer in Palestinian Broadcasting Corporation (PBC).
From 2005 to 2009 he worked as an Instructor in the Department of Electrical Engineering ,Palestine Technical University (Khadoury) ,Tul Karm – Palestine. Since September 2009, Dr Mutamed Khateeb is working as Assistant professor in the same university. Dr. Khatib has a number of publications to his credit in various international journals and conference proceedings. He is a member of IEEE, Palestinian Engineers Association and Arab Engineers Association.

Farid Ghani received B.Sc. (Engg) and M.Sc. (Engg) degrees from Aligarh Muslim University, India and M.Sc. and Ph.D. degrees from Loughborough University Of Technology (U.K).
From 1982 to 2007 and for varying durations, he worked as Professor of Communication Engineering at Aligarh Muslim University, India, Professor and Head of the Department of Electronics Engineering, Al-Fateh University, Libya, and Professor at Universiti Sains Malaysia. He is currently working as Professor in the School of Computer and Communication Engineering ,Universiti Malaysia Perlis, Malaysia. He is actively engaged in research in the general areas of digital and wireless communication, digital signal processing including image coding and adaptive systems. He has a large number of publications to his credit and is a reviewer for several international research journals.
Professor Ghani is Fellow of Institution of Engineering and Technology (IET) UK, Fellow of Institution of Electronics and Telecommunication Engineers(IETE) India, and Fellow of National Telematic Forum (NTF) India. He is also registered with the Council of Engineers(C.Eng) UK as Chartered Engineer.

A Perspective on the Cloud

Sandeep Keshav*, Mradul Pandey* and Shailendra Raj **
*VIET,Greater Noida Phase-II
**RGGI, Meerut- 250004
{sandeep.keshav77@gmail.com,mradulpande@gmail.com,srtyagi@rggi.edu.in}

Abstract- The natural rhythm in an IT organization is to view cloud as a technology disruption and assess it from a delivery orientation: how this new technology enables them to deliver solutions to the business, better, faster, and/or cheaper. Most IT practitioners, suppliers and advisors are approaching cloud exclusively from this perspective, generating more hype and confusion than actual results. I assert that an equally important but often ignored dimension to pay attention to is what cloud services are and how they are consumed: these services are ready to run, self-sourced, available anywhere, subscription-based or pay as you go.

"Cloud computing and in-house virtualization offers a wide range of complementary capabilities that can help your organization not only scale smoothly, but also simplify your IT infrastructure, reduce data center complexity, and cut costs."

This paper will discuss about the past, present and future with virtualization of the cloud and The Standards of Cloud Computing Service and more.

1. Introduction

a) What Cloud Means?

Clouds are a new way of building IT infrastructures from dynamic pools of virtualized resources that are operated as low-touch IT services and are consumed in a modern, web-savvy way.

Conceptually, cloud computing can be thought of as building resource abstraction and control on top of the hardware abstraction provided by virtualization.

In most cases, cloud computing infrastructures will evolve in scope and complexity over time. Functions such as elastic provisioning, metering, and self-service will often be added as the environment matures and goes into full production.

Cloud computing includes delivering services from multiple levels of the software stack. The most widely-used taxonomy specifes Infrastructure-as-a-Service (e.g. compute and storage), Platform-as-a-Service (e.g. middleware and infrastructure automation), and Software-as-a-Service (applications). These levels may layer on top of each other but can also exist independently.

Cloud computing can take place either on-premises (private cloud), as a shared, multi-tenant, off-premises resource (public cloud), or some combination of the two (hybrid cloud).

Cloud computing is not just another name for virtualization. It builds on virtualization, and constructing a virtualized infrastructure will be the frst step to a private cloud for many customers.

A private cloud improves effciency, helps organizations save money, and improves service levels relative to less fexible and dynamic IT infrastructures.

Public clouds offer a pay-as-you-go pricing model for computing resources that customers do not need to own or operate themselves.

Clouds take many forms because different customers or even different business units and applications within a single customer have vastly different requirements. One size doesn't ft all.

Cloud infrastructure should support interoperability, open standards, and the ability to run existing applications in many different environments and on many different clouds.

Clouds should provide fexibility for your organization and not lock you into a single vendor's solution.

b) Yesterday, Today and Tomorrow

The natural rhythm in an IT organization is to view cloud as a technology disruption and assess it from a delivery orientation: how this new technology enables them to deliver solutions to the business, better, faster, and/or cheaper. Most IT practitioners, suppliers and advisors are approaching cloud exclusively from this perspective, generating more hype and confusion than actual results. I assert that an equally important but often ignored dimension to pay attention to is what cloud services are and how they are consumed: these services are ready to run, self-sourced, available anywhere, subscription-based or pay as you go.

This might seem like a subtle distinction, a delivery orientation versus a consumer orientation. After all, they are two sides of the same coin. But in fact, if IT shops continue to focus on the delivery aspect only, they'll miss the boat altogether, because the consumption model has changed so profoundly: businesses now expect to consume services, not technologies.

As businesses re-orient themselves around a cloud model and services mature and grow, new capabilities will emerge. Businesses will be able to solve new problems that weren't addressable before. In addition, they will be able to solve old problems quicker, cheaper, and with higher- quality results.

Because of this, the cloud will ultimately usher in new business models which in turn will force IT to reinvent itself in order to remain relevant to the business in the emerging service oriented economy. This means that IT must move away from its exclusive focus on delivery and management of assets and toward creating a world-class supply chain for managing supply and demand of business services.

Contextually, the cloud actually becomes a forcing function – forcing IT to transform. It may not be next year or the year after but we will arrive relatively soon at a place where the business can source services on its own from any provider it chooses (fulfilling Nicholas Carr's 2003 assertion that "IT doesn't matter"). CIOs that ignore this fact and fail to take transformative measures are likely to see their role devolve from innovation leader to CMO – Chief Maintenance Officer – responsible for maintaining all of the legacy systems and services sourced by the business.

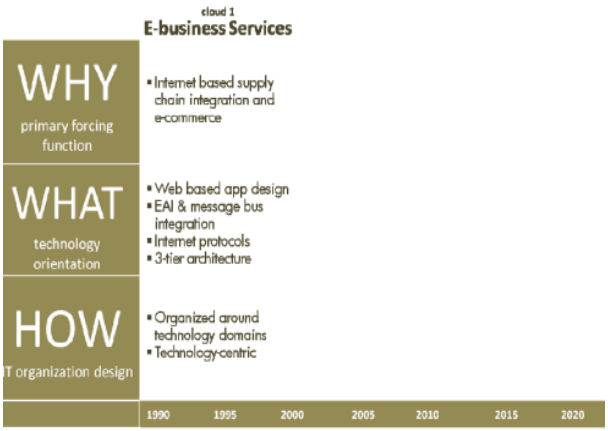
2. How The Cloud Occur, And Offer An Alternative Strategy

a) Cloud 1

News flash: the cloud isn't new. It's been around for years now, starting with what many now refer to as the "internet era." Remember when we as IT people used to talk about what was being hosted "in the cloud?"

This was the first generation/version of cloud. Let's call it cloud 1. It was an enabler that originated in the enterprise. Commercial use of the internet as a trusted platform revolutionized supply chain management processes and brought about new shopping experiences for consumers. Over time, these technologies became mainstream and fundamentally changed the IT architectural landscape.

We moved from two-tier to three-tier architectures, embraced externally hosted applications and transitioned to internet protocols and development techniques internally as a means of delivering solutions. And this internet-enabled, cloud 1 ecosystem made its way through the entire portfolio of IT capabilities, solving new problems and driving efficiency gains into old ones.



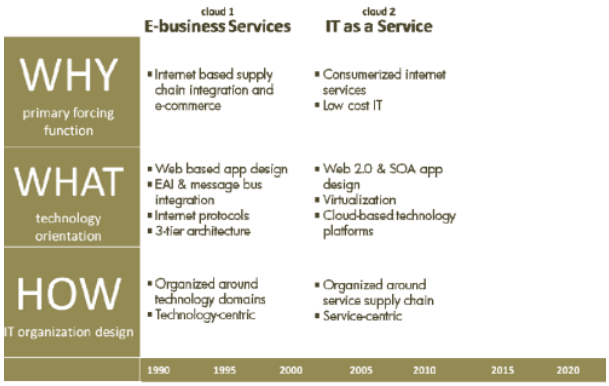
As an offshoot of this, the enterprise oriented technologies moved out of the computer room and into the living room of the employees as they became internet shoppers and searchers. How long did this take? Maybe five years – give or take a year.

As this model evolved and matured it gave birth to thousands of consumer class services. These services use next generation internet technologies on the front-end and massive scale architectures on the back end deliver, and are low cost, good enough services as far as the service consumer is concerned. And while these consumer-class services don't necessarily pass the smell test for compliance and security concerns of large enterprises, they are good enough for small companies and individuals and are maturing rapidly. The services and delivery techniques have ushered in a more advanced generation or version of cloud.

b) Cloud 2

The current generation of cloud services is driven by the consumer experiences that blossomed from cloud 1 living rooms. Internet-based shopping, search, and countless other services have brought about a new economic model and introduced new technologies. Services can be self-sourced, from anywhere and virtually any device and delivered immediately. Infrastructure and applications can be sourced as services in an on-demand manner. This consumer-oriented model is forcing its way back into the enterprise on the backs of employees.

Most of the attention around cloud services in the enterprise today is focused on the new techniques and sourcing alternatives for IT capabilities – IT as a service. Using standardized, highly virtualized infrastructure and applications, IT can drive higher degrees of automation and consolidation, thus reducing the cost of maintaining existing solutions and delivering new ones. In addition, externally supplied infrastructure, software, and platform services are delivering capacity augmentation and a means of using operating expense funding instead of capital.



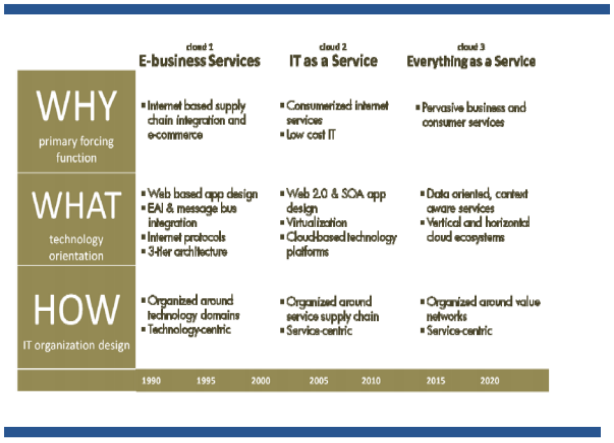
This cloud service orientation is also spawning new ideas for business innovation in the enterprise. For example, think of the impact that publicly available GPS services and mapping technologies have had on logistics processes. With these technologies and services, integrated into logistics management systems, tracking of assets, shipments and personnel has reached a higher

degree of accuracy and improved cost efficiency than ever before. Meanwhile, business people are attempting to further exploit this service orientation to eliminate complexity and barriers in their monolithic processes. They read about examples; they visualize new opportunities and approaches but are typically "held hostage" by their IT department and the inherent latency of project driven processes required to make it happen. And missing the IT planning window can add years to the cycle time of getting the desired results. This is adding stress to the already tenuous relationship between IT and the business it services.

How long is this going to last? Here's a prediction – don't blink. While many companies are wrestling with the technology transitions required to move from cloud 1and 2, the volume of services in the commercial cloud marketplace is increasing, propagation of data into the cloud is occurring and web 3.0/semantic web technology is maturing. And because of this we are just starting to see the next generation of cloud materialize.

c) Cloud 3

That data and information is growing at a break-neck pace is just the tip of the iceberg. Within three years there will be more information produced and consumed than in the history of man. The next version of cloud will enable access to information through services that are set in the context of the consumer experience. This is significantly different: it means that data will be separated from the applications. Because of this processes can be broken into smaller pieces and automated through a collection of services, woven together with access to massive amounts of data. It eliminates the need for large scale, complex applications that are built around monolithic processes. Changes can be accomplished by refactoring service models and integration achieved by subscribing to new data feeds. This will create new connections, new capabilities, and new innovations surpassing those of today.



For example, think about the problem of trying to understand the salmonella outbreak in the U.S. a couple of years ago. We were all originally led to believe by the experts that it was caused by tomatoes. You couldn't buy a tomato for six weeks and the entire industry almost went under. Then they revised their position and said "well, no, it's the jalapenos." And the actual root cause was never really figured out. Solving this kind of problem with traditional approaches would mean that you have to integrate with supply-chain management systems of every participant in an extended ecosystem which is very expensive, complex, and won't happen. This means that when these things happen, people are looking through manual reports, trying to do the correlation by hand, which delivers answers three months to late – if ever.

In 2009, a cloud service was developed to enable the GS1 Canada Product Recall program. GS1 Canada is the Canadian arm of GS1 – a leading, global, not-for-profit organization dedicated to improving supply chain efficiencies. This program runs on HP's cloud computing platform for manufacturing, allowing companies to expediently exchange product recall information in a secure and direct manner using GS1 supply chain identification keys. Participants in the ecosystem simply propagate data to the program and then all of the analytics happen in the cloud. This approach helps to eliminate the challenges and inaccuracies that result from manual product recall processes.

3. Which Cloud Should IT Focus On Today?

IT organizations are still struggling to understand the how to utilize technologies associated with cloud 2. Discussions and debates about how/what/why virtualization, private cloud, Infrastructure as a Service, Software as a Service and Platform as a Service are occurring each and every day. And as seasoned IT folks, we are comfortable dealing with technology disruptions and have industrial strength processes to respond to them.

The reality is that by focusing attention here, we'll miss the forest for the trees. Taking a technology-oriented approach to cloud will only exacerbate existing problems over time. One reason is because technology can only be calibrated to cost – thus the anchor IT metric: Total Cost of Ownership. Cloud 3 is not so much about delivering cost reduction through technology as it is about value. It's about services that people are willing to pay for because of the direct and visible benefit they provide.

Cloud 3 services offer businesses opportunities to solve existing problems in new ways or to solve problems, not solvable up until now. As more cloud 3 services materialize, businesses will not wait in line for IT projects to deliver them. Instead, equipped with two very powerful tools – an internet connection and a credit card – they will be able to circumvent IT to get the desired results - diverting precious innovation funding away from IT budgets.

Therefore, it's vital that IT leaders start focusing on reinventing themselves. Transforming to a model where 90 percent of the deliverables still flow through a project-oriented pipeline will not lend itself to competing in the new economy. In the not too distant future, the business will not want IT projects that deliver applications, it won't want to invest in IT assets – what will it want?

4. Clouds Today And Tomorrow

The National Institute of Standards and Technology (NIST) defnes cloud computing as "a model for enabling convenient, on-demand network access to a shared pool of confgurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." This defnition, together with associated service and deployment models, has emerged as a tough industry consensus of where cloud computing is headed.

Characteristics such as the following will become the norm over time in mature, production cloud implementations. However, not all will necessarily be present with early implementations — and, indeed, not all will necessarily be valued by a given customer. For example, while pay-as-you-go utility pricing is often associated with cloud computing, organizations that implement internal clouds may not care about highly granular chargeback to individual departments or users. Ultimately, "cloud computing" is the set of capabilities that solves a customer's problems or provides them with new business value rather than something described by an infexible, abstract defnition.

a) Resource Abstraction And Pooling

Pooled computing resources serve multiple consumers using a multi-tenant model (whether different internal groups within one company or different organizations within a shared, public resource) with physical and virtual resources dynamically assigned and reassigned depending on demand. The consumer of the service generally has limited control over or knowledge of the exact location of the provided resources but may be able to specify location at a higher level of abstraction — such as to create high-availability domains or to meet regulatory requirements around data location. Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

b) Network-Centric

Whether implemented within a single organization or at a public cloud provider, cloud computing is network centric. Services are made available over the network and accessed through standard mechanisms, typically lightweight web protocols.

c) Simple, Fast Provisioning of Resources

One of the ways that cloud computing makes an IT infrastructure more agile is by enabling new resources to be brought online quickly. This may include over time a degree of self-service — meaning that a user can provision computing capabilities, such as server capacity and storage, as needed without having to interact with a human. (Typically, policies that limit total resource limits or credit without an additional level of authorization would limit the scope of self-service requests.)

d) Rapidly and Elastically Provisioned Resources

In a cloud environment, resources can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out based on pre-set policies and the demands of an application. Just as importantly, resources can also be rapidly decreased when they are no longer needed, avoiding the familiar situation of unused servers sitting idle after the task they were initially purchased for ends.

e) Utility Pricing

Cloud computing is often associated with "utility pricing" or pay-per-use, although today, this is something that is largely specifc to public cloud providers. Metering at a level of abstraction appropriate to the type of service (such as storage, CPU usage, bandwidth, or active user accounts) will become more widespread over time as organizations learn the types of data that are most useful. Even if this information is not used to directly charge for use, it can be applied to capacity planning and other purposes.

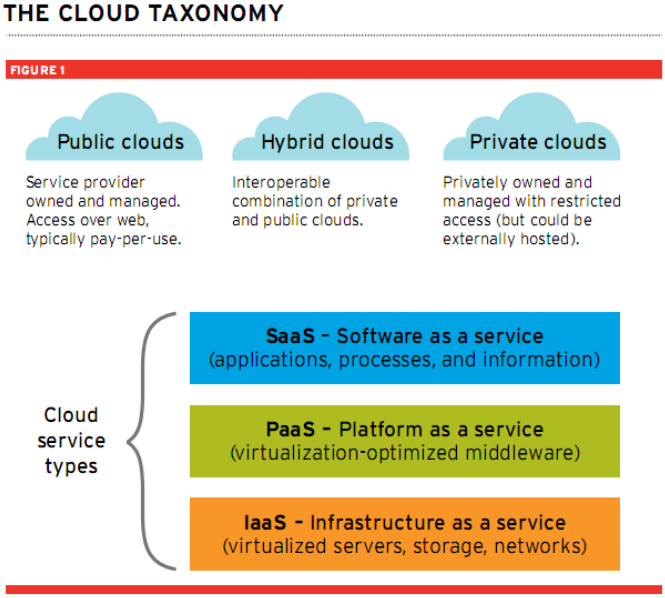


Figure 1 The fgure above shows what's come to be a widely accepted framework for viewing and discussing the cloud computing space. It shows the different types of cloud services and the different types of cloud deployments.

f) Infrastructure-as-a-Service (Public Cloud examples: Amazon Web Services, IBM Cloud)

IaaS (Infrastructure as a Service) lets the user or creator of the service provision processing, storage, networking, and other computing resources on which they can then run operating systems and applications. The deployer of the service does not manage or control the underlying compute infrastructure except select networking confgurations or perhaps physical location of the resources at a gross geographical level. On the other hand, the deployer is responsible for confguring and maintaining their own operating systems and software and, to a large degree, scaling their application and provisioning all the services required to run it.

g) Platform-as-a-Service (Public Cloud examples: Google App Engine, force.com)

PaaS (Platform as a Service) usually describes an additional level of services layered on top of an IaaS foundation. These services effectively provide an additional level of abstraction, taking care of tasks so that a developer or operator doesn't have to. Examples include widely used middleware, such as application servers and databases that would otherwise have to be added to a base-level infrastructure on a case-by-case basis. Alternatively, grid software middleware serves to automate common tasks such as scaling applications. The lines between what is a platform and what is just infrastructure and what is end-user hosted software can blur. But platforms are generally a higher level of abstraction than infrastructure while still not being an application —that is something directly interacted with by a line-of-business user.

SOFTWARE-AS-A-SERVICE

h) (Public Cloud examples: salesforce.com, Intuit Quickbooks Online)

SaaS (Software as a Service) is perhaps the most familiar face of cloud computing, referring as it does to the direct consumption of a cloud resource — an application — by end users rather than by developers or operators. There is no inherent relationship between the hosted application and the nature of the infrastructure on which it runs. However, SaaS applications can and often do run on cloud infrastructure and middleware; after all, cloud computing virtues such as fexibility and effciency are ultimately in service of applications. Thus, although SaaS does not necessarily imply IaaS or PaaS, the underlying infrastructure very much affects how effciently the application runs and how it is experienced by users.

5. Different Ways Of Delivering Cloud.

With public clouds, the cloud infrastructure or platform is made available to the general public or a large industry group and is owned and operated by the organization selling cloud services. (To be absolutely true to the described framework, any hosted SaaS application is technically a public cloud, but in practice the term is usually reserved for infrastructure and platform offerings.)

A private cloud, by contrast, is operated solely for an organization, either behind an internal frewall or operated by a third party for the exclusive use of that organization. It may be managed by the organization or a third party and may exist on premises or off-premises. (In other words, this defnition really refers to the boundaries of control and trust rather than who employs the infrastructure's operators or who holds title to the equipment.)

A hybrid cloud blends the public and private cloud models. It assumes some level of interoperability between a private cloud and public cloud implementation. This approach might be of interest to a customer who wants to own and operate a cloud infrastructure sized for typical loads but reserve the option of renting additional capacity on a pay-as-you-go basis to handle load spikes.

Possible variants on these basic approaches include community clouds, which are shared by several organizations and support a specifc community that has shared concerns such as security, compliance, or other needs that could be specifc to a particular industry.

6. Forecast : Cloudy With A Chance Of Virtualization

a) Is cloud computing or in-house virtualization the answer to your scalability challenges? Maybe both.

Just about everything your organization does these days generates a data trail. Multiply millions of interactions times daily or weekly backups times years of history and—well, you get the picture. Your data is never going to shrink. In fact, chances are that you'll be managing hundreds of terabytes within just a few years. And as your volume of information grows, the computing power required by the systems leveraging that information will likely expand as well.

Scaling your data center to handle increasingly intensive computing requirements is no mean feat. If you wait until your servers and storage systems are bursting at the seams, the chances of ad hoc expansion and unnecessary infrastructure complexity are high. But by thinking ahead, you can lay the foundation to scale with ease.

b) Understanding the Environment

A variety of strategies can help manage the data growth headed your way. Two strategies in particular—cloud computing and in-house virtualization—can provide a one-two punch that helps you simultaneously lighten your computing workload and pack more processing power into your existing data center footprint.

Cloud computing allows enterprises to extend the power of their data centers by providing access to computing infrastructure and application platforms as a service over the Internet. But while the term cloud computing is relatively new, the technologies behind it are not.

Way back when, in the middle of the twentieth century, organizations that could not fit all their work onto a single mainframe began networking servers together to form clusters; clustering techniques that have been refined over the years now form the basis for cloud computing. Large clusters like the ones offered by cloud computing providers can offer virtually unlimited scalability—so whether your workload sees steady growth or the occasional spike in demand, the cloud can keep up. Alternatively, large enterprises may establish an internal computing cloud

when that approach is suitable to specific organizational requirements.

Virtualization is also a major enabler of the cloud. Because it allows providers to partition large servers into many smaller virtual servers that have been rightsized for their assigned tasks, virtualization is the key to the flexibility and cost advantages of cloud computing.

7. What Cloud Computing Is Not

Before getting into the benefts of cloud computing, it's worth briefy discussing what cloud computing isn't but to which it's often equated.

a) Cloud CompuTing Is Not Just Virtualization Or Grid

Cloud computing also abstracts resources and provides self-service and automated facilities to control the allocation of resources. Does this mean virtualization and grid are irrelevant to cloud computing? Hardly. Hardware abstraction is a key enabler of cloud computing and, in fact, virtualizing/consolidating is a logical frst step in a cloud infrastructure project. Virtualization also means that there's an evolutionary path to the cloud.

b) Nor Is Cloud Computing Just About Using External Resources

Books have popularized the public cloud, presenting it as a revolutionary way to consume computing, akin to the development of the electric grid. Public and hybrid clouds do offer a great way to deal with unpredictable workload demands and a way to get applications up and running and scaling quickly without CAPEX or time-consuming hardware procurement cycles. However, for regulatory, compliance, and a host of other reasons, many organizations will not want to move their applications — or at least all of their applications — to a public service provider. And, for them, private cloud infrastructures can provide many of the same benefts as a public cloud without the loss of control.

c) Cloud CompuTing Isn't JusT A Buzz Word Hyping Computing As Usual

Yes, like any hot technology, there is hype around cloud computing. And various aspects of cloud computing do have historical antecedents. Accessing applications over the network recollects application service providers and remote storage storage service providers. Utility pricing for rented compute resources recalls timesharing. However, collectively, the technologies that make up cloud computing represent the infusion of characteristics — like abstraction, elasticity, automated, and dynamic — into traditional computing, which is better described by words like hardwired, fxed, manual, and static.

d) Who Creates Clouds And Who Uses Them

Earlier, we covered the different types of clouds and delivery models. Now, we'll take a look at what that means in terms of who sells what to whom. IT infrastructure vendors like Red Hat sell the products needed to build clouds to end-user organizations and to public cloud providers like Amazon. They may also sell services to assist those organizations with their implementations and certify specifc cloud providers as destinations for their products. For example,

Red Hat has worked with its Red Hat Certifed Cloud Providers to develop an industry-leading certifcation and testing program, making it simple, safe, and cost-effective to move and manage applications between on-premise and public clouds.

However, IT infrastructure vendors do not generally operate their own public clouds. (IBM is an exception; it both sells hardware and software to build clouds, and it operates a cloud for consumption by end-user customers.) Public cloud vendors may internally develop value-add cloud services that they offer to their customers and that they use to monitor, manage, and run their implementation. However, especially as you move beyond the very largest public cloud providers, which have large in-house development staffs, service providers are primarily consumers of IT infrastructure software and hardware.

Application vendors/developers may themselves offer their software in the form of SaaS — in which case they will directly use infrastructure software and hardware developed by others — or they may use a third-party cloud or other service provider, in which case they consume PaaS, SaaS, or some form of hosting. End users — whether consumers or business users — typically will only consume SaaS, whether it's an inter-nally-hosted business application or one that's on the public web (which is why cloud means "software delivered over the Internet" to so many people).

Operators of enterprise IT infrastructure will purchase the infrastructure needed to build private clouds along with infrastructure services from external cloud providers.

8. The Seven Standards of Cloud Computing Service Delivery

The Seven Standards of Cloud Computing Service Delivery technology industry has changed dramatically over the last 10 years. In economic times like these, leading companies are looking to cloud-computing platforms to deliver business functions ranging from packaged business applications to custom application development at a fraction of the time and cost of traditional on-premises platforms. With this growth in enterprise use of cloud-computing comes a corresponding increase in responsibility on the part of vendors to provide cloud-computing platforms that ofer outstanding service delivery.

The CIOs of the largest financial and network security companies require that cloud-computing platforms meet the highest standards of service. After all, they're entrusting them with critical corporate data.

To meet these requirements and ensure we can make customers of any size successful, the Force.com platform adheres to the seven standards outlined below. Tey're the building blocks of the best practices every successful cloud-computing platform should follow:

a) Platform Standard

1) World-Class Security :Provision world-class security at every level.

Security is more than just user privileges and password policies. It's a multidimensional business imperative, especially for platforms that are responsible for customer data. Cloud-computing platforms must have detailed, robust policies and procedures in place to guarantee the highest possible levels of:

- Physical security
- Network security
- Application security
- Internal systems security
- Secure data-backup strategy
- Secure internal policies and procedures
- Third-party certification

2) Trust and Transparency:Provide transparent, real-time, accurate service performance and availability information.

Cloud-computing platforms should provide customers with detailed information about service delivery and performance in real time, including:

- Accurate, timely, and detailed information about service performance data and planned maintenance activities
- Daily data on service availability and transactionperformance
- Proactive communications regarding maintenance activities.

3) True Multitenancy :Deliver maximum scalability and performance to customers with a true multitenant architecture.

Leading Web applications—including Google, eBay,and Salesforce CRM—run on a single code base and infrastructure shared by all users. A multitenant architecture allows for high scalability and faster innovation at a lower cost. Single-tenant systems, on the other hand, are not designed for large-scale cloud-computing success. The internal inefciencies of maintaining a separate physical infrastructure and/or separate code lines for each customer make it impossible to deliver quality service or innovate quickly. Multitenancy provides customers with the following benefts:

- Efcient service delivery, with a low maintenance and upgrade burden
- Consistent performance and reliability based on an efcient, large-scale architecture
- Rapid product release cycles.

4) Proven Scale: Support millions of users with proven scalability.

With any cloud-computing service, customers benefit from the scale of the platform. A larger scale means a larger customer community, which can deliver more and higher-quality feedback to drive future platform innovation. A larger customer community also provides rich opportunities for collaboration between customers, creating communities that can share interests and foster best practices. Cloud-computing platforms must have:

- Proof of the ability to scale to hundreds of thousands of subscribers
- Resources to guarantee the highest standards of service quality, performance, and security to every customer
- The ability to grow systems and infrastructure to meet changing demands
- Support that responds quickly and accurately to every customer
- Proven performance and reliability as customer numbers grow.

5) High Performance:Deliver consistent, high-speed performance

globally.Cloud-computing platforms must deliver consistent, high-speed systems performance worldwide and provide detailed historical statistics to back up performance claims, including:

- Average page response times
- Average number of transactions per day.

6) *6. Complete Disaster Recovery:*

Protect customer data by running the service on multiple, geographically dispersed data centers with extensive backup, data archive, and failover capabilities.Platforms providing cloud-computing services must be fexible enough to account for every potential disaster. A complete disaster recovery plan includes:

- Data backup procedures that create multiple backup copies of customers' data, in near real time, at the disk level
- A multilevel backup strategy that includes disk-to-disk-to-tape data backup in which tape backups serve as a secondary level of backup, not as the primary disaster-recovery data source. This disk-oriented model ensures maximum recovery speed with a minimum potential for data loss in the event of a disaster.

7) High Availability:Equip world-class facilities with proven highavailability infrastructure and application software.

Any platform ofering cloud-computing applications needs to be able to deliver very high availability. Requirements for proving high availability include:

- Facilities with reliable power, cooling, and network infrastructure
- High-availability infrastructure: networking, server infrastructure, and software

- N+1 redundancy
- Detailed historical availability data on the entire service, not just on individual servers.

9. Future Threat

Cloud computing and virtualization, while offering significant benefits and cost-savings, move servers outside the traditional security perimeter and expand the playing field for cybercriminals. The industry already witnessed Danger/Sidekick's cloud-based server failure that caused major data outages in November 2009, highlighting cloud-computing risks that cybercriminals will likely abuse. We believes cybercriminals will either be manipulating the connection to the cloud, or attacking the data center and cloud itself.

10. Extract / Conclude:

Some may look at this whole area of cloud computing and say, "What is old is new again," because we are simply going back to the service bureaus of days gone by. While on some level that is true, the benefits are far greater this time around. Whether you call it cloud computing or SaaS (software as a service) or utility-based computing or any number of near-synonyms, the concept has been around for quite some time but has never quite been perfected, and the reasons for that vary.

References

1. **Cloud** Computing - An Agile Approach to IT www.hds.com/assets/pdf/cloud-computing
2. GTSI - ideas Newsletter - Volume IV www.gtsi.com/eblast/corporate/cn
3. Grid Computing: Conferences, **Seminars** and Training www.gridsummit.com/training.htm
4. **Cloud** Computing and **Virtualization** 2010 – Conference in Singapore www.itevent.net/cloud-computing-and-virtualization
5. Event Details | dctechsource.com - Walk in the **Cloud Seminar ...** www.dctechsource.com
6. Program - Specialist **Seminars** www.itcspecialistseminar.com/program.php
7. ISACA Virtual **Seminar** to Help Clear the **Cloud** Confusion www.sourcews.com/isaca-virtual-seminars
8. Web **Seminars** www.sqlmag.com/resources/web-seminars
9. Parallels **Virtualization** for **Cloud** Services Providers www.parallels.com/spp/parallelsvirtualaziation
10. **Cloud** computing « **Seminars** For You www.seminars4you.wordpress.com/2008
11. IBM developerWorks : **Cloud** computing : Events www.ibm.com/developerworks/views/clou...
12. Scientific **Seminar** www.salsahpc.indiana.edu/CloudCom2010

A New Low-Voltage RMS Converter based on Current Conveyors in MOS Technology

Ebrahim Farshidi

Department of ElectricalEngineering, Faculty of Engineering, University of Shoushtar Azad, Shoushrar, Iran,
e_farshidi@hotmail.com

***Abstract*:** In this paper, a new current mode true RMS-to-DC converter circuit using second generation current conveyors (CCII) is presented. The second generation current conveyor employs MOSFET transistors that are operating in strong inverted saturation region. The proposed methodology provides the following advantages: a complete current-mode circuit makes it suitable for current-mode signal processing applications; modular design of the converter is achieved by using of CCIIs as the basic building block; and using of CCII that is active building blocks of the converter features low voltage supply and low circuit complexity, which lead to low voltage applications. Hspice simulation results shows high performance of the converter and validate the workability of the proposed circuit.

***Keywords*:** current-mode, converter, current conveyor, MOS.

1. Introduction

The RMS measurement device is widely used in electrical engineering for different purposes such as instrumentation devices [1-3]. The RMS value of a voltage or a current represents the effective energy that is transferred to a load by a periodic source. Different methods have been reported for the precision measurement of the RMS value of an ac current. The on-going trend towards lower supply voltages has brought the area of analog integrated circuits into limelight. In conventional circuits implementation techniques using op-amp and transconductance, the supply voltage restricts the attainable maximum dynamic range. Also, it is well known that the use of current-mode active devices has some advantages such as greater linearity, wider bandwidth, larger dynamic range, and less power consumption when compared to their voltage counterparts: operational amplifiers. The most popular current-mode active device is second generation current conveyor (CCII).

In recent years, some integrated design forms of true RMS-to-DC converter have been proposed. Some of them are designed based on bipolar transistors [4, 5]. However, in many situations, particularly in mixed A/D systems, it is desirable to implement the circuits in the MOS technology. One attempt is to employ sigma-delta converter [6], but the extra needed circuit leads to a large number of transistors and high power consumption.

In this work, a new design to overcome the above problems is presented. Using a novel synthesis, a squarer/divider circuit based on CMOS second generation current conveyor (CCII) is presented. The CCII is a very interesting, attractive building block, which can easily compete with standard operational amplifiers. (in any way ,it can be considered as an excellent alternative to standard operational amplifier). The proposed squarer/divider is modular and with low supply voltage. The time average operation is performed by a current-mode low pass filter which consists of one capacitor and one MOS transistor. The complexity of this filter is much less than those proposed before [7, 8]. The proposed converter requires very low voltage with minimum supply voltage of one V_{gs} plus two V_{ds} and extra current and voltage biasing are not needed.

The paper organized as follows. In section 2, the basic principle of current-mode true RMS-DC converter operation is discussed. In section 3 the proposed squarer/divider and low-pass filter as the basic building blocks of the converter are presented. In section 4 the circuit and simulation results of the RMS-DC converter that demonstrate the validity of the proposed method are presented; and finally, in section 5 concluding remarks are provided.

2. Basic Principle

The definition of the root means square value of an input signal with a period of T is given by:

$$I_{out} = \sqrt{\frac{1}{T}\int_{t}^{t+T} I_{in}^{2}(t)dt} \qquad (1)$$

where $I_{in}(t)$ and $I_{out}(t)$ are input and output currents of the RMS-to-DC converter, respectively.

A mathematically equivalent expression, but more precise considering the offset of the system [4], is given by:

$$I_{out} = \left\langle \frac{I_{in}^{2}(t)}{I_{out}} \right\rangle \qquad (2)$$

where $\langle .. \rangle$ represents the averaging operation.

The block diagram of proposed converter is shown in Fig.1, in which it consists of a squarer/divider and a low pass filter [4].

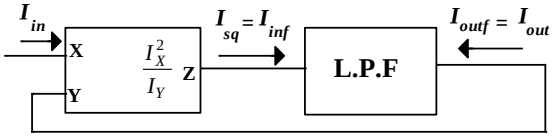


Figure 1. Block diagram of the RMS-to-DC converter[4]

3. Circuit Design

3.1 Current Conveyor

The CCII is a tree terminal (X, Y, and Z) device whose responses I_Y, V_X and I_Z to the applied excitations; V_Y, I_X and V_Z are given by the following matrix representation [9]:

$$\begin{bmatrix} I_x \\ V_x \\ I_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \pm 1 & 0 \end{bmatrix} \begin{bmatrix} V_y \\ I_x \\ V_x \end{bmatrix} \quad (3)$$

The node Y is an infinite impedance node with the voltage applied to it is replicated at the node X. Also when a current is injected in the node X, the same current is injected in to node Z. The notation CCII+ denotes a positive Z output current conveyor (Z+) whereas CCII- denotes a negative Z output current conveyor (Z-). This performance can be achieved by combining a voltage follower placed between terminals Y and X, with non-inverting and inverting current buffers (mirrors) between terminals X and Z (+ and -).

Fig. 2 shows block diagram of the proposed current conveyor. The unity gain feedback of op-amp at the front end forces Vx to be equal to Vy and the current mirror at the back end lead to Iz to be equal to Ix in CCII+ (and –Ix in CCII-).

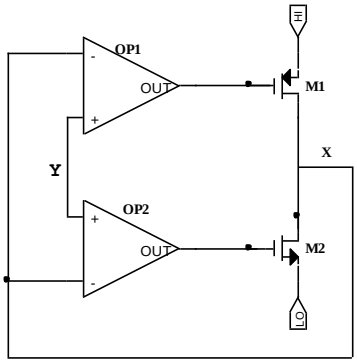


Figure 2. Block diagram of the input of CCII

Fig. 3 shows circuit diagram of the proposed current conveyor. The conveyor consists of a voltage follower and a current mirror. The voltage follower itself is realized by two stages: a pair of complimentary differential amplifiers and a complimentary common source amplifier. Transistors M1-M5 and M6-M10 are employed for NMOS-input and PMOS-input differential amplifiers, respectively. Transistors M9 and M10 are used for common source amplifier, in which the output of the common source amplifier is fed back to the inputs of differential amplifiers to insure of voltage follower operation. The CCII is designed in MOS technology and features low voltage supply and low circuit complexity, which lead to low voltage applications.

3.2 Squarer/divider

In this subsection a modular scheme of the current-mode squarer/divider by using of CCIIs as the basic active building block is presented. Figure 4 shows the proposed circuit for the squarer/divider employing CCII. The circuit consist of two CCII block and 4 diodes.

The I-V relationship for diode is given by:

$$I_D = I_s \exp\left(\frac{V_D}{\eta U_T}\right) \quad (4)$$

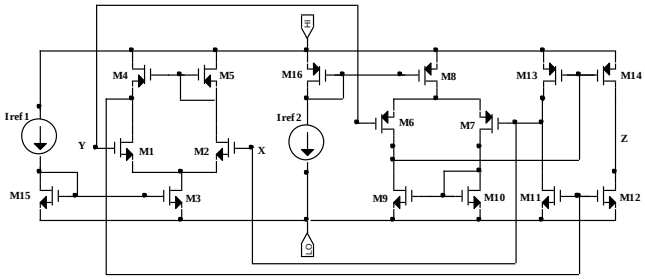


Figure 3. Circuit diagram of the proposed CCII

Using Kerchief voltage law (KVL) for the voltage of the diodes in the loop and the I-V relationship of diodes it is obtained:

$$I_{D_4} = \frac{I_{D1} \cdot I_{D_2}}{I_{D_3}} \quad (5)$$

By replacing $I_{D_1} = I_{D_2} = I_{in}$, $I_{D_3} = I_{rms}$, $I_{D_4} = \pm I_Z$ it results:

$$I_Z = \frac{I_{in}^2}{I_{rms}} \quad (6)$$

Equation (6) shows the squarer/divider behaviour of the the proposed circuit in the Fig. 4.

3.3 Filter

For a current-mode first-order low-pass filter, output current I_{outf} and input current I_{inf} in Laplace domain are related as [10, 11]:

$$\frac{I_{outf}(s)}{I_{\inf}(s)} = \frac{1}{1 + \frac{s}{\omega_c}} \quad (7)$$

in which, $\omega_c = 1/\tau$ is the cut-off frequency of the filter.

Fig. 5 shows the basic circuit of the proposed low-pass filter. For this circuit the I-V relationships of transistor Mf1 in saturation region and its derivation can be expressed by [10]:

$$I_{outf} = \frac{\beta}{2}(V_{cap} - V_{th})^2 \longrightarrow$$

$$\frac{dI_{outf}}{dt} = \sqrt{2\beta I_{outf}} \frac{dV_{cap}}{dt} \quad (8)$$

in which, $V_{cap} = V_{gs}$ is the voltage of capacitor C.

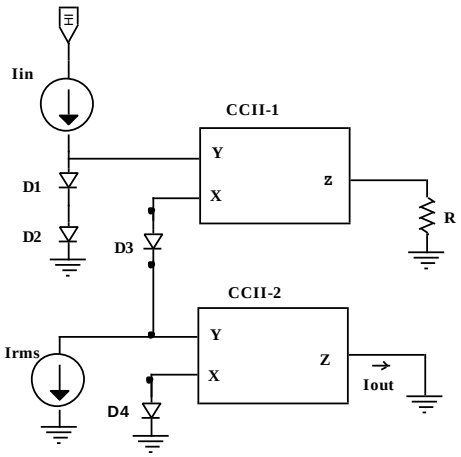


Figure 4. Block diagram of the proposed squarer/divider

Using (8) and also equation $I_{cap} = \frac{dV_{cap}}{dt}$, it results:

$$I_{cap} = \sqrt{\frac{C\omega_c}{2\beta I_{out}}}(I_{\inf} - I_{outf}) \tag{9}$$

Considering the cut-off frequency of the filter as:

$$\omega_c = \frac{\sqrt{2\beta I_{outf}}}{C} \tag{10}$$

and then, substituting (10) into (9) gives:

$$I_{cap} = I_{\inf} - I_{outf} \tag{11}$$

Fig. 6 shows the proposed low-pass filter circuit based on (11), which consists of a capacitor and only one transistor.

Comparing this filter with other proposed filters which are realized based on the switched capacitor technique [6] or is implemented by translinear loops [7], or by two class-AB transconductors [8], shows that this analysis gives a much simpler circuit. To obtain this simple circuit, as equation (10) indicates, the cut-off frequency of the filter is allowed to vary with output current. This dependency to I_{out}, and also the maximum acceptable output ripple determine the suitable values for capacitor C. This subject is discussed next. The output current of the RMS-to-DC converter is a DC current $I_{true-RMS}$ with an ac ripple I_{ripple} on it. In order that converter gives a good performance the amplitude of this ripple should be small compared to $I_{true-RMS}$. This requires suitable values for the cut-off frequency of the filter.

As it is mentioned above, ω_c is proportional to the output current of the converter; however the value of the capacitor can be calculated in such a way that it eliminates the effect of this dependency and also achieves enough accuracy in the output of the converter [12, 13]. To this end and for the first step, (7) in time domain can be written as:

$$\frac{dI_{outf}}{dt} = \omega_c\left(I_{\inf} - I_{outf}\right) \tag{12}$$

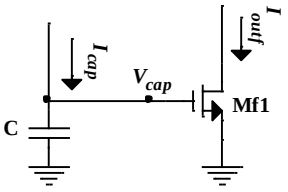


Figure 5. Current-mode low-pass filter principle

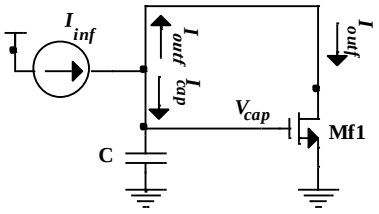


Figure 6. Circuit diagram of the proposed current-mode low-pass filter

As the output current of the filter converter I_{outf} is equal to the output current of the converter I_{out} ($I_{outf} = I_{out}$) and from (8) the input current of the filter I_{inf} is equal to $\frac{I_{in}^2}{I_{out}}$, then substituting these equations into (7) that results:

$$(13)\ \frac{dI_{out}}{dt} = \omega_c\left(\frac{I_{in}^2}{I_{out}} - I_{out}\right)$$

Multiplying both sides of (13) by I_{out} gives:

$$I_{out}\frac{dI_{out}}{dt} = \omega_c\left(I_{in}^2 - I_{out}^2\right) \longrightarrow$$

$$(14)\ \frac{d\left(I_{out}^2\right)}{dt} = 2\omega_c\left(I_{in}^2 - I_{out}^2\right)$$

Equation (14) is time domain representation of a current-mode filter with input of I_{in}^2, output of I_{out}^2 and cut-off frequency of $2\times\omega_c$. In Laplace domain, the equation can be rewritten as:

$$\frac{I_{out}^2(s)}{I_{in}^2(s)} = \frac{1}{1+\frac{s}{2\omega_c}}. \tag{15}$$

Assuming that the input current is a sinusoidal function we have:

$$I_{in} = \sqrt{2}I_{true-RMS}\cos(\omega t) \tag{16}$$

where ω is the frequency of the input current of the converter. Squaring both sides of (16) gives the input signal as follows:

$$I_{in}^2 = 2I_{true-RMS}^2 \cos^2(\omega t) \longrightarrow$$

$$I_{in}^2 = I_{true-RMS}^2 (1+\cos(2\omega t)). \quad (17)$$

Using input signal of (17) in low pass filter of (15) gives:

$$I_{out}^2 = I_{true-RMS}^2 (1+B\cos(2\omega t+\phi)) \quad (18)$$

where

$$B = \frac{1}{\sqrt{1+(2\omega/2\omega_c)^2}} \quad , \quad \phi = -tg^{-1}\left(\frac{2\omega}{2\omega_c}\right). \quad (19)$$

Assuming that $\omega_c << \omega$ and using Taylor series expansion gives:

$$I_{out} = I_{true-RMS}\sqrt{(1+B\cos(2\omega t+\phi))} \cong$$

$$I_{true-RMS}\left(1+\frac{B}{2}\cos(2\omega t+\phi)-\frac{B^2}{8}\cos^2(2\omega t+\phi)+\ldots\right)$$

$$(20) \cong I_{true-RMS}\left(1-\frac{B^2}{16}+\frac{B}{2}\cos(4\omega t+2\phi)+\ldots\right)$$

From (20), the values of RMS and peak-to-peak ripple at the output current are expressed as:

$$I_{RMS,out} = \left(1-\frac{B^2}{16}\right)I_{true-RMS} \longrightarrow$$

$$I_{RMS,out} = \left(1-\frac{1/16}{1+(2\omega/2\omega_c)^2}\right)I_{true-RMS} \quad (21a)$$

$$I_{ripple,out} = \frac{B}{2}I_{true-RMS} \longrightarrow$$

$$I_{ripple,out} = \frac{1/2}{\sqrt{1+(2\omega/2\omega_c)^2}}I_{true-RMS} \quad (21b)$$

From (21) it can be concluded that if the frequency of the input signal is at least five times bigger than the cut-off frequency of the filter $(\omega \geq 5\omega_c)$, then the accuracy of more than 1% can be obtained at the output.

In such case, the output current of the RMS-to-DC converter is a DC current I_{rms} with an ac ripple on it. The amplitude of this ripple is small compared to I_{rms}, i.e., $I_{out} \cong I_{rms}$. Considering these conditions and also using (10) it results that the values of the capacitor to achieve the accuracy of more than 1% is:

$$C \geq \frac{5\sqrt{2\beta I_{rms,\max}}}{\omega_{\min}} \quad (22)$$

in which, $\omega_{\min}$ is the lower end of the frequency range and $I_{rms,\max}$ is the output current higher end of the converter.

From (22), it can be seen more accuracy at the output of the converter can be achieved by decreasing the cut-off frequency of the filter by employing larger capacitances or reducing linear transconductance of the transistor of the filter.

4. Simulation Results

Fig. 7 shows the complete circuit diagram of the proposed RMS-to-DC converter based on block diagram of Figs. 1 and 4 and circuit diagrams of Figs. 3 and 6. The circuit was simulated using HSPICE with 0.18um AMS CMOS process parameters. V_{dd}=0.93V, . V_{dd}=-0.93V and C=900nF were employed. Fig. 8 show the steady-state time response of the RMS-to-DC converter for a 60uA peak-to-peak and 10kHz frequency of sinusoidal input currents. The relative error of the output current of converter for different amplitudes of the sinusoidal input currents is calculated, which shows the error less than 3% is achieved for amplitudes between 10uA and 50uA.

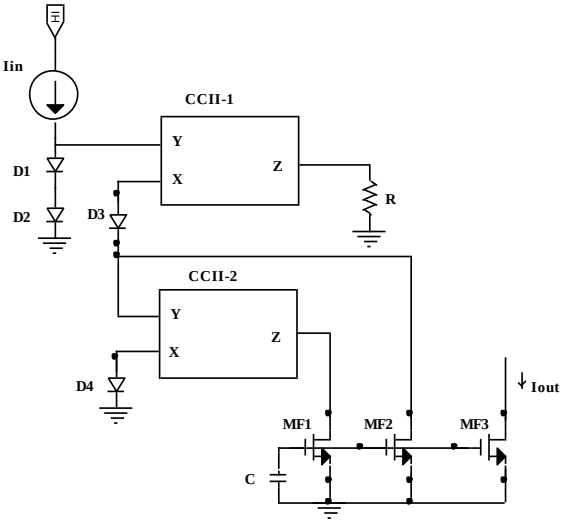


Figure 7. Complete circuit diagram of the proposed RMS-to-DC converter

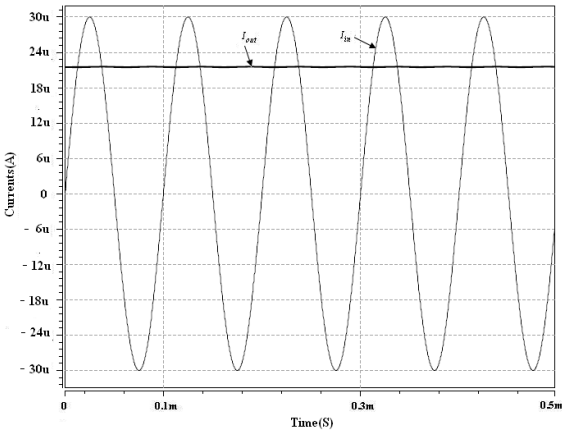


Figure 8. Time response of a sinusoidal input current and output current of the converter

5. Conclusion

A RMS-to-DC converter by employing a new squarer/divider based on current conveyor is designed. The proposed squarer/divider circuit uses MOS transistors that

are operating in strong inversion region and works in low supply voltage. The results of implemented converter designed based on the proposed squarer/divider and simplified filter show that it can be used in RMS measuring integrated circuits.

References

[1] Analog Devices, AD636 Data Sheet, "Low level True RMS-to-DC converter," Available: http://www. selectronic.fr/includes_selectronic/pdf/ Analog Device/ AD636JH.pdf.

[2] National-Semiconductor, LH0091, "True RMS to DC converter,"Available:http://www.chipdocs.com/datashe ets/datasheet-pdf/National Semi conductor/ LH0091. html.

[3] Maxim, MX536 Data Sheet, "True RMS-to-DC converter," Available: http://www.maxim-ic.com/ datasheet/index.mvp/id/1197.

[4] J. Mulder, W. A. Serdijn, A. C. V. D. Woerd, and A. H. M. V. Roermund, "Dynamic translinear RMS-DC converter," Electron. Lett., vol. 32, no. 22, pp. 2067-2068, Oct. 1996.

[5] J. Mulder, A. C. V. D. Woerd, W. A. Serdijn,and A. H. M. V. Roermund, "An RMS-DC converter based on the dynamic translinear principle," *IEEE J. Solid-State Circuits*, vol. 32, no.7, pp. 1146-1150, Jul. 1997.

[6] W. S. Wey, and Y. C. Huang, " CMOS delta-sigma true RMS converter,"IEEE J. Solid-state Circuits, vol. 35, no. 2, pp. 248-257, Feb. 2000.

[7] C. Psychalinos and S. Vlassis, "A systematic design procedure for square-root -domain circuits based on the signal flow graph approach," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl*., Vol. 49, No. 12, pp. 1702-1712, Dec. 2002.

[8] A. J. Lopez-Martin, and A. Carlosena, "A 1.5V current-mode CMOS RMS-to-DC converter," *Analog Integrated Circuits Signal Process*., vol. 36, pp. 137-143, 2003.

[9] A. Fabre, O. Saaid, F. Wiest and C. Boucheron, "High frequency applications based on a new current controlled conveyor" I*EEE Transactions on Circuits and Systems-I*, 31, pp. 82-91, 1996.

[10] C. A. DE La Cruz-Blas, A. J. Lopez-Martin, A. Carlosena and J. Ramirez-angulo, "1.5v current-mode CMOS true RMS-DC converter based on class-AB transconductors," *IEEE Trans. Circuits Syst. II, Express Briefs*, vol. 52, no. 7, pp. 376-379, Jul. 2005.

[11] E. Farshidi, S. M. Sayedi, "A 1.2V current-mode true RMS–DC converter based on the floating gate MOS translinear principle," *Microelectronics Journal, Elsevier*, vol. 39, no. 2, pp. 293–298, Feb. 2008.

[12] E. Farshidi and S. M. Sayedi, "A micropower multi decade dynamic range current-mode true rms-to-dc," *MWSCAS'07, Proceedings of the 50th IEEE Circuits and Systems International Midwest Symposium*, pp. 1493-1496, Aug. 2007.

[13] R. Frey, "Current mode class-AB second order filter," *Electron. Lett.* , vol. 30, pp. 205–206, Feb. 1994.

Author Biography

Ebrahim Farshidi was born in Shoushtar, Iran, in 1973. He received the B.Sc. degree in 1995 from Amir Kabir University, Iran, the M.Sc. degree in 1997 from Sharif University, Iran and the Ph. D. degree in 2008 from electrical engineering at IUT, Iran, all in electronic engineering. He worked for Karun Pulp and Paper Company during 1997–2002. From 2002 he has been with shahid chamran university, Ahvaz, where he is currently assistant professor and head of electrical engineering department. He is author of more than 26 technical papers in electronics. His areas of interest include current-mode circuits design, low power VLSI circuits, and data converters.

A Temporally Hybrid Method for Integrating Overlapped Mobile Ad Hoc Networks with the Internet

Young-Chul Shim

Hongik University, Department of Computer Engineering
72-1 Sansudong, Mapogu, Seoul, Korea
shim@cs.hongik.ac.kr

***Abstract*:** It is an important task to connect various access networks to the core internet and build an integrated network in which they can interact with each other. Among many different types of access networks several difficult problems should be solved to connect overlapped mobile ad hoc networks to the core Internet. In this paper we propose a protocol which connects overlapped mobile ad hoc networks to the core Internet using the mobile IP protocol and enables efficient message exchanges. The protocol includes mechanisms to find and maintain the optimal gateway for internet connection, hand over with the ping-pong effect minimized, and efficiently recover from link failures. Mobile nodes can act as either clients or servers and efficiently exchange messages with correspondent nodes which reside within or outside a mobile ad hoc network. We also evaluate the performance and overhead of the proposed protocol through simulation.

***Keywords*:** mobile ad hoc network, internet connection, gateway, mobile IP, protocol, performance evaluation**.**

1. Introduction

As wireless communication technology advances, many different types of wireless networks coexist and these wireless networks are interconnected with each other through the connection to the core Internet [1]. Wireless networks can be classified into networks with infrastructure and networks without infrastructure. Wireless networks with infrastructure include cellular networks and wireless LANs configured as an infrastructure mode. Their connection to the Internet is well understood and they are widely used commercially.

Nodes in ad hoc networks are configured in a network without infrastructure, they act as both hosts and routers, and messages among them are delivered using multi-hop methods. If nodes in an ad hoc network can move, such a network is called a MANET(Mobile Ad hoc NETwork). Research on connecting MANETs to the Internet is currently being performed but the results are not sufficient compared with wireless networks with infrastructure [2], [3]. Gateways are used to connect MANETs to the Internet. Packets, which come from the Internet to a node within a MANET, first arrive at a gateway and then are delivered to the destination node within the MANET using an ad hoc routing protocol. If a packet is delivered from a node within a MANET to the outside Internet, it is delivered though the reverse process. Therefore, if a node within a MANET wants Internet connection, maintaining an efficient connection path to a gateway becomes an important issue. If a MANET is connected to the Internet through one gateway, then the problem becomes simple. But if many MANETS are overlapped and they are connected to the Internet through many gateways, then many complicated problems occur including the problem of how a mobile node within a MANET finds the best gateway for itself.

In this paper we assume an environment in which overlapped MANETS are connected to the Internet through many gateways and describe a protocol which enables for a mobile node within a MANET to efficiently exchange messages with a correspondent node in either the MANET or outside Internet. We use the mobile IP protocol for connecting MANETs to the Internet. Not only a mobile node within a MANET can find a node in the outside Internet and send messages to it but also a node in the outside Internet can find a mobile node within a MANET and send messages to it. To make the above possible, a mobile node uses its home IP address issued from its home network wherever it moves. Unlike most researches which assume that a mobile node can act as only a client, using the proposed protocol a mobile node can not only act as a server and but also be used in a peer-to-peer network.

When studying the problem of connecting overlapped MANETs to the Internet, the most fundamental research issue is how a mobile node, which has entered a MANET and wants Internet connection, finds the best access gateway. If we do not consider the quality of service, the optimal access gateway for a mobile node is the gateway to which a mobile node can access with the minimal number of hops. Methods for finding the optimal access gateway are classified into proactive methods and reactive methods. In proactive methods, a gateway advertises its existence by periodically broadcasting a message with a short interval and upon receiving this broadcast message, a mobile node can find the nearest gateway easily [4], [5]. Moreover when a mobile node moves into a new MANET with a nearer access gateway than its current access gateway, it can find this new gateway promptly and, therefore, perform the handover efficiently. But because the gateway should periodically broadcast a message with a short interval, the overhead due to broadcasting is excessively high. In reactive methods, a gateway does not broadcast any advertisement messages and, therefore, a mobile node which wants Internet connection should find the nearest gateway using an ad hoc routing protocol [6], [7]. Compared with the proactive methods, reactive methods do not incur any overhead due to broadcasting advertisement messages. But in

reactive methods the process of finding a nearest access gateway takes more time in general. Moreover, it is difficult to find the proper handover time, and, therefore, the average distance between a mobile and an access gateway can become very large when a mobile node moves among many MANETs. Some researchers combined proactive and reactive methods and proposed hybrid methods [8], [9]. In hybrid methods the advertisement message of a gateway is broadcast within a certain range from the gateway and a mobile node within this range can find the proper access gateway and perform handovers using advertisement messages. But if a mobile node is located outside this range, it should rely on the reactive method to find the access gateway. In this paper we call hybrid methods of this type to be spatially hybrid methods.

We propose a new type of hybrid methods called temporally hybrid methods in this paper. Upon entering a MANET, a mobile node finds an access gateway using a reactive method. And when it wants to send a message to the outside Internet but finds the route to the access gateway to be invalid, it finds the nearest access gateway using a reactive method. But because gateways advertise themselves by broadcasting messages with a long interval, mobile nodes can perform handovers at proper time. By combining the advantages of both proactive and reactive methods, the proposed method reduces the overhead due to advertisement messages broadcast by gateways and minimizes the distance between a mobile node and a gateway. Moreover, when a link failure occurs during message delivery, the proposed method either performs a local recovery process or initiates handover if necessary. If a mobile node performs handover as soon as it crosses a MANET boundary, inefficiency can arise. If, after moving to a new MANET, a mobile node returns to the old MANET immediately, it should perform handover twice. Inefficiency due to this is called a ping-pong effect. In order to minimize the ping-pong effect, in the proposed protocol a mobile node performs handover only after it enters a new MANET deeply enough.

The rest of the paper is organized as follows. Section 2 explains the overall network structure for connecting MANETs to the Internet and message flows within this structure and Section 3 describes the proposed protocol. Section 4 presents simulation results and is followed by the conclusion in Section 5.

2. Overall Structure for Connecting MANETs to the Internet

Figure 1 shows the overall structure for connecting MANETs to the Internet. A mobile node MN gets its home address I_H from its home network. When it visits a foreign network MANET1, it finds the nearest gateway GW1. MN gets a care-of-address I_{CoA1} from GW1 and registers this address to its home agent. Most of other researches assume that a mobile node does not have a home address and gets a temporary IP address as client when it enters a MANET and has messages to send. But in this case when an outside node wants to send a message to a mobile node within a MANET, it cannot do so because the outside node cannot know the permanent address of the mobile node. Therefore, in our approach a mobile node has a permanent home address and gets a new care-of-address

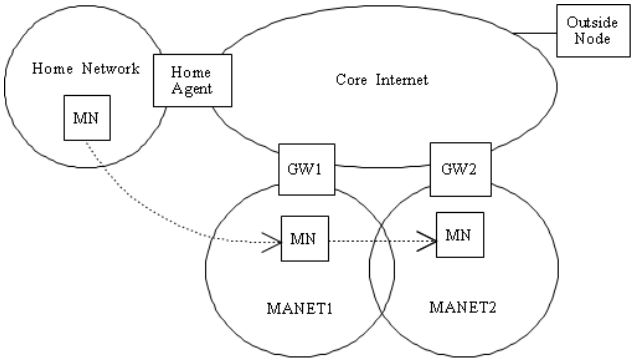


Figure 1. Structure for Integrating MANETs with Internet

whenever it enters a new MANET. We also assume that a gateway manages one subnetwork.

When MN moves from MANET1 to another overlapped MANET called MANET2, it selects GW2 as an access gateway, gets a new care-of-address I_{CoA2}, and registers this address to its home agent. There are two cases when MN hands over from MANET1 to MANET2. First, MN loses connection with GW1 and finds a new route with GW2 and second, while moving from the area maintained by GW1 to the area maintained by GW2, MN finds GW2 closer than GW1. The distance between a mobile node and a gateway is defined to be the number of hops between these two nodes. If a mobile node moves from MANET1 to MANET2 and hands over to MANET2 as soon as the distance to GW1 becomes one hop longer than to GW2, inefficiency can arise. If MN moves back to MANET1 right after it moved to MANET2 and got and registered a care-of-address in MANET2, it has to get a new care-of-address in MANET1 and register it to its home agent again. This ping-pong effect causes much overhead. To avoid this overhead, MN hands over to GW2 when it judges that it entered deep enough into MANET2. In this paper MN hands over to GW2 when the distance to GW2 is at least 2 hops shorter than the distance to GW1.

Next we explain 3 cases of message flows assuming that MN is in MANET1.

- If an outside node wants to send a message to MN, the message is first delivered to the home agent using the mobile IP protocol, forwarded to GW1 using IP tunneling with I_{COA1} as the destination address, and finally sent to from GW1 to MN using tunneling and an ad hoc routing protocol.
- When a mobile node has a message to be sent to a node located in an outside network, the message is first sent to GW1 using tunneling with the address of GW1 as the destination address and then sent to the outside node directly with the outside header removed.
- Finally we consider the case where the correspondent node is located in the same network, MANET1. Because MN does not know that the correspondent node is in the same MANET, it first forwards messages to GW1 using tunneling as in the previous case. Because GW1 knows that the destination mobile node is in MANET1, it sends the message to the destination mobile node using tunneling with the destination node's care-of-address as the destination address and an ad hoc routing protocol. At the same time, it notifies MN of the destination node's care-of-address. Upon receiving the care-of-address, MN can send messages directly to the correspondent node

using an ad hoc routing protocol without relying on GW1.

3. Description of the Proposed Protocol

In this section, we describe the proposed protocol for connecting overlapped MANETs to the core Internet

3.1 Power-on in a foreign MANET

When the power is turned on in a foreign MANET, the mobile node multicasts a RREQ(Route REQuest) message with All Gateways as the destination address. Upon receiving this message, a gateway unicasts a RREP(Route REPly) message to the mobile node. The RREP message includes the care-of-address for the mobile node and the distance to the gateway which is initialized to 0. An intermediate node which receives the RREP message increases the distance to the gateway by 1 and forwards the message to the mobile node. After receiving one or more RREP messages, the mobile node selects the nearest gateway as its access gateway and stores the care-of-address provided by that gateway and the distance to that gateway. The information on the distance to the gateway has validity period, will be invalidated after this validity period, and can be used as only hint after expiration. The mobile node sends to its home agent the registration message which includes the home address and the care-of-address of the mobile node and the address of the home agent. This registration message is delivered through the access gateway. Upon receiving this registration message, the gateway stores the home address and the care-of-address of the mobile node in its list of managed mobile nodes. Other gateways which do not receive a registration message within a certain time period know that they are not selected and return the care-of-addresses that they have proposed to their shared care-of-address pools.

3.2 Receiving a router advertisement message

A gateway manages one MANET and advertises its identity by periodically broadcasting a router advertisement message within its MANET. This message has the address of the gateway, the network address managed by the gateway, and the distance to the gateway. Upon receiving the message, a mobile node stores the information carried by the message, increases the distance to the gateway by 1, and forwards the message to its neighboring nodes. If a mobile node has other gateway as its access gateway and its gateway is nearer than the advertising gateway, then it drops the advertisement message. But the advertising gateway is much closer than the current access gateway, the mobile node hands over to the advertising gateway.

Using router advertisement messages, every mobile node has one gateway as its access gateway, receives a care-of-address from this gateway, and can access to the Internet. All the mobile nodes having the same access gateway forms a tree with the access gateway as the root. To reduce the overhead due to the router advertisement message, the broadcasting interval is much longer than in proactive methods.

After receiving a router advertisement message, a mobile node behaves as follow:

```
if (not connected to Internet yet) {
 store the address of and the distance to the gateway
      temporarily;
 if (has not sent a RREQ message yet)
   send RREQ(All Gateways) with the TTL value set to
     the distance to the advertising gateway;
 increase the distance to the gateway by 1 and broadcast
    a router advertisement message to neighbors}
else if (the advertising gateway is the same as mobile node's
        access gateway) {
 update the distance to the gateway if necessary;
 increase the distance to the gateway by 1 and broadcast
    a router advertisement message to neighbors;
 send an "I'm Alive" message to the gateway}
else /* the advertising gateway is different from */
      /* the access gateway of the mobile node */ {
 if (distance to access gateway – 2 >=
     distance to the advertising gateway)
   hand over to the new gateway
 else if (distance to the access gateway – 2 <
          distance to the advertising gateway <=
          distance to the access gateway) {
   store the address of and the distance to the advertising
      gateway;
   increase the distance to the gateway by 1 and broadcast
     the router advertisement message to neighbors}}
```

If a mobile node decides to hand over to the advertising gateway, it behaves as follow:

```
get a new care-of-address from the new gateway;
change the access gateway to the new gateway;
register to its home agent through the new gateway;
if (it has a valid path to the old access gateway)
 notify the old access gateway of the handover
```

3.3 Sending a message

To keep the description simple, in this paper we explain about the case where a mobile node sends a message to a node in the outside Internet.

If a valid path to the access node exists, the mobile node forwards the message to the access node using tunneling with the access gateway's address as the destination address. Then the access gateway removes the outside header and forwards the message to the outside destination node. If a valid path to the access node does not exist, messages are delivered as follows:

```
send RREQ(All Gateways) with TTL set to distance to the
   access gateway + 2;
/* the reason why an access gateway is found in a wider */
/* region by adding 2 to the distance to the access */
/* gateway is to consider the case in which a mobile */
/* node moves further away after it has received the */
/* distance information to the access gateway */
if (a mobile node receives RREP only from
      the access gateway)
 update distance to the gateway if necessary
else (RREP comes only from a new gateway)
 hand over to the new gateway
else /* RREP comes from both the access gateway and */
      /* a new gateway */
 If (distance to the access gateway – 2 >=
     distance to the new gateway)
   hand over to the new gateway;
send message to the access gateway;
```

If a RREP message does not arrive, TTL is increased properly and the above process is repeated. If a RREP message does not arrive even after increasing TTL several times, the mobile node is notified that access attempt to Internet has failed.

A RREQ message with All Gateways as the destination address was used in not only this subsection but also subsection 3.1. To differentiate these two cases, when a mobile node sends a RREQ message to send a data packets, it also includes the address of the access node in the RREQ message. Upon receiving this RREQ message, the gateway behaves as follows. If the address of the gateway is the same as that in the RREQ message, it returns a RREP message including only its address. Otherwise, the gateway returns a RREP message including not only its address but also a care-of-address which the mobile node can use in case of handover. With this feature a mobile node can hand over to a new gateway quickly because it does not have to send an additional message requesting a care-of-address.

3.4 Link failure

If a link failure occurs while sending messages to an access gateway, the child node of the failed link tries to recover a path to the access gateway. The child node starts the recovery process by sending RREQ(Access Gateway) with TTL set to the distance to the access gateway + 2. If a RREP message is received from the access gateway, the child node updates its distance to the access gateway, distributes this new distance information to its descendant node, and continues to send messages through the access gateway.

If the recovery process initiated by the child node fails, it notifies the link failure to its descendants by sending RERR(Access Gateway) to them. Among these descendant nodes all the nodes that have been sending messages through the access gateway invalidate their path to the access gateway and find a valid access gateway by using the protocol explained the subsection 3.3.

3.5 Analysis of the proposed protocol

This subsection presents qualitative analysis results of the proposed protocol.

In the proposed protocol as soon as a mobile node enters a MANET, it finds an access gateway and gets a care-of-address. Therefore, unlike other research works on wireless network integration, outside node can find and send messages to a mobile node in MANETs. Moreover, two processes of finding an access gateway and getting a care-of-addresses are overlapped and the number of control messages is reduced.

When a mobile node sends messages to outside node, it can send messages right away if it has a valid path to an access gateway. Even if it does not have a valid path, it can continue to use the same care-of-address and send messages without much delay if it is confirmed that the access gateway is a valid access gateway. If a mobile node using a proactive method gets a care-of-address when it sends a message, a mobile node using the proposed protocol can begin to send messages quicker than proactive methods. If in a proactive method a mobile node gets a care-of-address as soon as it enters a foreign MANET, a proactive method can begin to send messages quicker than the proposed protocol on average. But the proposed protocol can always begin to send messages quicker than a reactive method.

The proposed protocol periodically broadcasts router advertisement messages like a proactive method. But because the proposed protocol has much longer broadcasting interval, it has much less overhead. Using this advertisement message, a mobile node can find immediately the moment of entering a new MANET and perform handover promptly. Therefore, the average distance between a mobile node and an access gateway is not much bigger than in a proactive method. In a reactive method handover can occur only when a mobile node begins to send messages or a link failure occurs during the message delivery. But in the proposed protocol, a mobile node can perform handover whenever it finds a better access gateway. Moreover the proposed protocol includes a mechanism to avoid ping-pong effects and can avoid unnecessary mobile IP registrations

4. Simulation Results

This section presents performance evaluation results for the proposed protocol obtained from experimentations using the NS2 simulator [10]. Experimentations compare performance of a temporally hybrid method which is the proposed protocol with a proactive method, a reactive method, and a spatially hybrid method. Metrics for performance evaluation are the distance between a mobile node and an access gateway and the number of control packets required for the network maintenance.

4.1 Simulation environments

Figure 2 shows the experimentation environment. The network has a home agent for the mobile node, one correspondent node located in the outside Internet, and two overlapped MANETs, MANET1 and MANET2. MANETs use AODV as the ad hoc routing protocol and mobile nodes use 802.11 CSMA/CA as the MAC protocol. The topology of a MANET is assumed to be a 1000m*1000m square and the number of mobile nodes in a MANET is set to be 25, 50 or 100 depending on the experimentation. A mobile node moves randomly with the speed of 0~30m/s. Nodes exchange CBR packets and the size of each packet is 1024Kbyte. The broadcasting interval for router advertisement messages is chosen randomly between 1 sec and 1.5 sec in case of proactive and spatially hybrid methods and is chosen randomly between 9.5 sec and 10 sec in case of the proposed method. Router advertisement messages of spatially hybrid methods are distributed within 3 hops from a gateway.

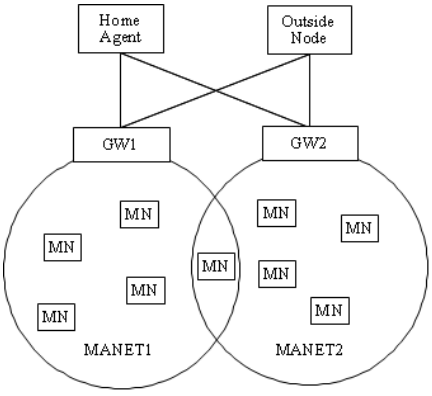


Figure 2. Network Configuration for Experimentation

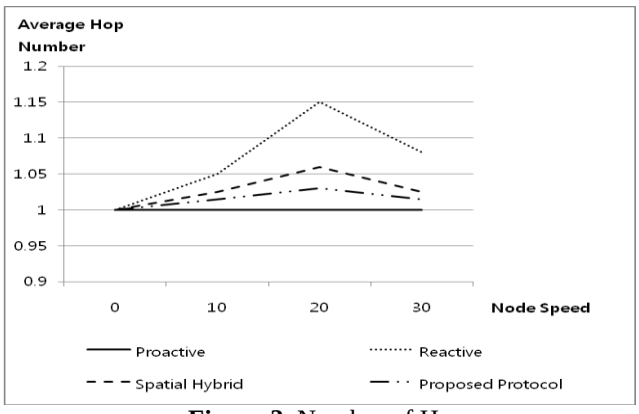


Figure 3. Number of Hops

4.2 Analysis of hop numbers to gateways

The number of hops is the number of nodes that a message from a mobile node passes before arriving at a gateway and, therefore, is the distance to the gateway from a mobile node. A gateway distance affects packet delay and the initial connection time to the gateway. Less distance means less packet delay and smaller initial connection time. In the first experimentation, the gateway distance of the proposed method is compared with proactive methods, reactive methods, and spatially hybrid methods. In the experiments, a mobile node sends packets with 20% probability and its speed is increased from 0m/s to 30m/s. Figure 3 shows the relative distance values for reactive methods, spatially hybrid methods, and the proposed method with the value for proactive methods set to 1.

When a mobile node does not move, there is no difference among different methods. As the speed increases, the difference begins to appear. A proactive method has the least distance value at the cost of overhead due to frequent router advertisement messages. In case of a reactive method, the distance increases as a mobile node moves faster because it is difficult for a mobile node to find an optimal gateway and find proper timing for handover. When a mobile node is close from a gateway, a spatially hybrid method behaves similarly to a proactive method but when a mobile node is far away from a gateway, it behaves similarly to a reactive method. Therefore, its distance value lies between a proactive method and a reactive method. In the case of the proposed protocol, the local recovery protocol finds the proper access gateway promptly and router advertisement messages broadcast with a long interval help the path optimization process. Therefore, the gateway distance of a proposed method is slightly bigger than that of a proactive method but the difference is not big at all.

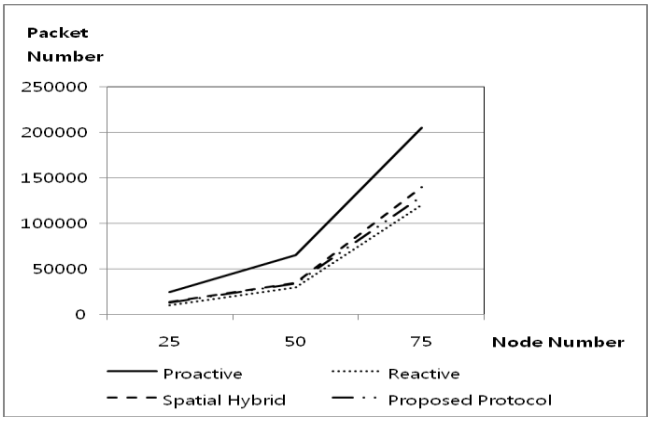


Figure 4. Number of Control Packets

4.3 Analysis of the number of control packets

Figure 4 shows the number of control packets when mobile nodes move at 20m/s and send packets with 20% probability. As the number of mobile nodes increases, the curve for a proactive method rises sharply while curves for a reactive method, a spatially hybrid method, and the proposed method rise much slowly. The graph shows that the proposed protocol generates much less control packets than proactive methods. It also shows that the proposed method generates similar number of control packets as spatially hybrid methods while generating slightly more packets than reactive methods.

5. Conclusion

It is an important task to connect various access networks to the core internet and build an integrated network in which they can interact with each other. Among many different types of access networks several difficult problems should be solved to connect overlapped mobile ad hoc networks to the core Internet.

In this paper we proposed a temporally hybrid protocol which connects overlapped MANETs to the core Internet using the mobile IP protocol and enables efficient message exchanges. The protocol enables a mobile node to find and maintains the optimal gateway for internet connection. The proposed method mixes proactive and reactive methods temporally and, therefore, reduces overhead due to router advertisement messages of gateways and minimizes the distance between a mobile node and an access gateway. It includes mechanisms to handover with the ping-pong effect minimized and efficiently recover from link failures. A mobile node can act as not only a client but also a server. A mobile node efficiently exchanges messages with a corresponding node locating in the same MANET as well as the outside Internet.

Using the NS-s simulator, the performance of the proposed protocol is evaluated with proactive methods, reactive methods, and spatially hybrid methods. The simulation results showed that the proposed protocol has the similar gateway distance to proactive methods and the similar overhead due to control packets to reactive methods. It was shown that the proposed protocol has very good performance with very low overhead.

References

[1] D. Cavalcanti, D. Agrawal, C. Cordeiro, B. Xie, & A. Kumar, "Issues in Integrating Cellular Networks, WLANs, and MANETs: A Futuristic Heterogeneous Wireless Network," IEEE Wireless Communications, pp. 30-41, June 2005.

[2] P.M. Ruiz, F.J. Ros, & A. Gomez-Skarameta, "Internet Connectivity for Mobile Ad Hoc Networks: Solutions and Challenges," IEEE Communications Magazine, pp. 118-125, October 2005.

[3] F.M. Abduljalil & S.K. Bodhe, "A Survey of Integrating IP Mobility Protocols and Mobile Ad Hoc Networks," IEEE Communications Surverys, Vol. 9, No. 1, pp. 14-30, 1st Quarter 2007.

[4] M. Ergen & A. Puri, "MELWANA-MOBILE IP: Enriched Wireless Local Area Network Architecture," Proc. IEEE Vehicular Technology Conference, pp. 2449-2453, Sep. 2002.

[5] B. Xie & A. Kumar, "Integrated Connectivity Framework for Internet and Ad Hoc Networks," 1st IEEE International Conference on Mobile Ad-Hoc and Sensor Systems, Oct. 2004.
[6] Y. Sun, E.M. Belding-Royer, & C.E. Perkins, "Internet Connectivity for Ad Hoc Mobile Networks," Int. Journal of Wireless Information Networks, Vol. 9, No. 2, pp. 75-88, April 2002.
[7] M. Michalak & T. Braun, "Common Gateway Architecture for Mobile Ad-Hoc Networks," 2nd Annual Conference on Wireless On-Demand Network Systems and Services, pp. 70-75, 2005.
[8] B.A. Kock & J.R. Schmidt, "Dynamic Mobile IP Routers in Ad Hoc Networks," Int. Workshop on Wireless Ad-Hoc Networks, June 2004.
[9] M. Benzaid et al, "Integration of Mobile IP and OLSR for a Universal Mobility," Wireless Networks, Vol. 10, No. 4, pp. 377-388, July 2004.
[10] The Network Simulator (ns-2), http://www.isi.edu/nsman/ns.

Author Biographies

Young-Chul Shim was born in Seoul, Korea. He received Ph.D in computer science from the University of California, Berkeley, U.S.A. He is a professor at the department of computer engineering in Hongik University in Seoul, Korea. His major fields of studies include mobile network protocols and computer and network security.

Fuzzy Arithmetic with and without using α-cut method: A Comparative Study

Palash Dutta[1], Hrishikesh Boruah[2], Tazid Ali[3]

Dept. of Mathematics, Dibrugarh University, Dibrugarh,-786004, India

palash.dtt@gmail.com, h.boruah07@gmail.com, tazidali@yahoo.com

***Abstract*:** In [2] and [6] the authors proposed a method for construction of membership function without using α-cut. They claim that the standard method of α-cut [1] fails in certain situations viz in determining square root of a fuzzy number. We would like to counter their argument by proving that α-cut method is general enough to deal with different type of fuzzy arithmetic including exponentiation, extracting nth root, taking logarithm. Infact we illustrate with examples to show that α-cut method is simpler than their proposed method.

***Keywords*:** Fuzzy membership function, fuzzy number, alpha-cut

1. Introduction:

α-cut method is a standard method for performing different arithmetic operations like addition, multiplication, division, subtraction. In [2] and [6] the authors argue that finding membership function for square root of X where X is a fuzzy number, is not possible by the standard alpha-cut method. They have proposed a method of finding membership function from the simple assumption that the Dubois-Prade left reference function is a distribution function and similarly the Dubois-Prade right reference function is a complementary distribution function. In this paper we are going to show that alpha-cut method can be used for finding n^{th} root of fuzzy number and infact this method is simpler than that proposed by them. However we do acknowledge that the proposed method has more mathematical beauty than the existing alpha-cut method.

2. Basic Concept of Fuzzy Set Theory:

In this section, some necessary backgrounds and notions of fuzzy set theory are reviewed.

Definition 2.1: Let X be a universal set. Then the fuzzy subset A of X is defined by its membership function

$$\mu_A : X \to [0,1]$$

which assign a real number $\mu_A(x)$ in the interval [0, 1], to each element $x \in X$, where the value of $\mu_A(x)$ at x shows the grade of membership of x in A.

Definition 2.2: Given a fuzzy set A in X and any real number $\alpha \in [0, 1]$, then the α-cut or α-level or cut worthy set of A, denoted by ${}^{\alpha}A$ is the crisp set

$${}^{\alpha}A = \{x \in X: \mu_A(x) \geq \alpha\}$$

The strong a cut, denoted by ${}^{\alpha+}A$ is the crisp set

${}^{\alpha+}A = \{x \in X: \mu_A(x) > \alpha\}$

For example, let A be a fuzzy set whose membership function is given as

$$\mu_A(x) = \begin{cases} \dfrac{x-a}{b-a}, a \leq x \leq b \\ \dfrac{c-x}{c-b}, b \leq x \leq c \end{cases}$$

To find the α-cut of A, we first set α ∈ [0,1] to both left and right reference functions of A. That is, $\alpha = \dfrac{x-a}{b-a}$ and $\alpha = \dfrac{c-x}{c-b}$.

Expressing x in terms of α we have $x = (b-a)\alpha + a$ and $x = c-(c-b)\alpha$.

which gives the α-cut of A is

$$^{\alpha}A = [(b-a)\alpha + a, c-(c-b)\alpha]$$

Definition 2.3: The support of a fuzzy set *A* defined on *X* is a crisp set defined as

Supp (*A*) = {$x \in X$: $\mu_A(x) > 0$}

Definition 2.4: The height of a fuzzy set *A*, denoted by *h*(*A*) is the largest membership grade obtain by any element in the set.

$h(A) = \sup_{x \in X} \mu_A(x)$

Definition 2.5: A fuzzy number is a convex normalized fuzzy set of the real line *R* whose membership function is piecewise continuous.

Definition 2.6: A triangular fuzzy number A can be defined as a triplet [*a*, *b*, *c*]. Its membership function is defined as:

$$\mu_A(x) = \begin{cases} \dfrac{x-a}{b-a}, a \le x \le b \\ \dfrac{c-x}{c-b}, b \le x \le c \end{cases}$$

Definition 2.7: A trapezoidal fuzzy number *A* can be expressed as [*a*, *b*, *c*, *d*] and its membership function is defined as:

$$\mu_A(x) = \begin{cases} \dfrac{x-a}{b-a}, a \le x \le b \\ 1, b \le x \le c \\ \dfrac{d-x}{d-c}, c \le x \le d \end{cases}$$

3. Arithmetic operation of fuzzy numbers using α-cut method:

In this section we consider arithmetic operation on fuzzy numbers using α-cut cut method considering the same problem as done in [2] and [3] and hence make a comparative study.

3.1. Addition of fuzzy Numbers:

Let X= [a, b, c] and Y= [p, q, r] be two fuzzy numbers whose membership functions are

$$\mu_X(x) = \begin{cases} \dfrac{x-a}{b-a}, a \le x \le b \\ \dfrac{c-x}{c-b}, b \le x \le c \end{cases}$$

$$\mu_Y(x) = \begin{cases} \dfrac{x-p}{q-p}, p \le x \le q \\ \dfrac{r-x}{r-q}, q \le x \le r \end{cases}$$

Then $^{\alpha}X = [(b-a)\alpha + a, c-(c-b)\alpha]$ and $^{\alpha}Y = [(q-p)\alpha + p, r-(r-q)\alpha]$ are the α–cuts of fuzzy numbers X and Y respectively. To calculate addition of fuzzy numbers X and Y we first add the α–cuts of X and Y using interval arithmetic.

$$^{\alpha}X + {}^{\alpha}Y$$

$$= [(b-a)\alpha + a, c-(c-b)\alpha] + [(q-p)\alpha + p, r-(r-q)\alpha]$$

$$= [a + p + (b-a+q-p)\alpha, c + r - (c-b+r-q)\alpha] \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

To find the membership function $\mu_{X+Y}(x)$ we equate to x both the first and second component in (3.1) which gives

$x = a + p + (b-a+q-p)\alpha$ and

$x = c + r - (c+r-b-q)\alpha$

Now, expressing α in terms of x and setting $\alpha = 0$ and $\alpha = 1$ in (3.1) we get α together with the domain of x,

$$\alpha = \frac{x-(a+p)}{(b+q)-(a+p)}, (a+p) \le x \le (b+q)$$

and

$$\alpha = \frac{(c+r)-x}{(c+r)-(b+q)}, (b+q) \le x \le (c+r)$$

which gives

$$\mu_{X+Y}(x)=\begin{cases}\dfrac{x-(a+p)}{(b+q)-(a+p)},(a+p)\le x\le(b+q)\\ \dfrac{(c+r)-x}{(c+r)-(b+q)},(b+q)\le x\le(c+r)\end{cases}$$

3.2. Subtraction of Fuzzy Numbers:

Let X= [a, b, c] and Y= [p, q, r] be two fuzzy numbers. Then ${}^{\alpha}X=[(b-a)\alpha+a,c-(c-b)\alpha]$ and ${}^{\alpha}Y=[(q-p)\alpha+p,r-(r-q)\alpha]$ are the α–cuts of fuzzy numbers X and Y respectively. To calculate subtraction of fuzzy numbers X and Y we first subtract the α–cuts of X and Y using interval arithmetic.

$${}^{\alpha}X-{}^{\alpha}Y$$

$$=[(b-a)\alpha+a,c-(c-b)\alpha]$$
$$-[(q-p)\alpha+p,r-(r-q)\alpha]$$

$$=[(b-a)\alpha+a-(r-(r-q)\alpha),$$
$$c-(c-b)\alpha-((q-p)\alpha+p)]$$

$$=[(a-r)+(b-a+r-q)\alpha,$$
$$(c-p)-(c-b+q-p)\alpha].....(3.2)$$

To find the membership function $\mu_{X-Y}(x)$ we equate to x both the first and second component in (3.2) which gives
$x=(a-r)+(b-a+r-q)\alpha,$

and $x=(c-p)-(c-b+q-p)\alpha$

Now, expressing α in terms of x and setting $\alpha=0$ and $\alpha=1$ in (3.2) we get α together with the domain of x,

$$\alpha=\frac{x-(a-r)}{(b-q)-(a-r)},(a-r)\le x\le(b-q)$$

and

$$\alpha=\frac{(c-p)-x}{(c-p)-(b-q)},(b-q)\le x\le(c-p)$$

which gives

$$\mu_{X-Y}(x)=\begin{cases}\dfrac{x-(a-r)}{(b-q)-(a-r)},(a-r)\le x\le(b-q)\\ \dfrac{(c-p)-x}{(c-p)-(b-q)},(b-q)\le x\le(c-p)\end{cases}$$

3.3. Multiplication of Fuzzy Numbers:

Let X= [a, b, c] and Y= [p, q, r] be two positive fuzzy numbers. Then ${}^{\alpha}X=[(b-a)\alpha+a,c-(c-b)\alpha]$ and ${}^{\alpha}Y=[(q-p)\alpha+p,r-(r-q)\alpha]$ are the α–cuts of fuzzy numbers X and Y respectively. To calculate multiplication of fuzzy numbers X and Y we first multiply the α–cuts of X and Y using interval arithmetic.

$${}^{\alpha}X*{}^{\alpha}Y$$

$$=[(b-a)\alpha+a,c-(c-b)\alpha]*$$
$$[(q-p)\alpha+p,r-(r-q)\alpha]$$

$$=[((b-a)\alpha+a)*((q-p)\alpha+p),$$
$$(c-(c-b)\alpha)*(r-(r-q)\alpha)].....(3.3)$$

To find the membership function $\mu_{XY}(x)$ we equate to x both the first and second component in (3.3) which gives

$$x=(b-a)(q-p)\alpha^2+((b-a)p+(q-p)a)\alpha+ap$$
and
$$x=(c-b)(r-q)\alpha^2-((r-q)c+(c-b)r)\alpha+cr$$

Now, expressing α in terms of x and setting $\alpha=0$ and $\alpha=1$ in (3.3) we get α together with the domain of x,

$$\alpha x = \frac{-((b-a)p+q-p)a)+\sqrt{((b-a)p+q-p)a)^2-4(b-a)(q-p)(ap-x)}}{2(b-a)(q-p)}, ap \le x \le bq$$

and

$$\alpha x = \frac{((r-q)c+(c-b)r)-\sqrt{((r-q)c+(c-b)r)^2-4(c-b)(r-q)(cr-x)}}{2(c-b)(r-q)}, bq \le x \le cr$$

which gives,

$$\mu_{XY}(x) = \begin{cases} \dfrac{-((b-a)p+q-p)a)+\sqrt{((b-a)p+q-p)a)^2-4(b-a)(q-p)(ap-x)}}{2(b-a)(q-p)}, ap \le x \le bq \\ \dfrac{((r-q)c+(c-b)r)-\sqrt{((r-q)c+(c-b)r)^2-4(c-b)(r-q)(cr-x)}}{2(c-b)(r-q)}, bq \le x \le cr \end{cases}$$

3.4. Division of Fuzzy Numbers:

Let X= [a, b, c] and Y= [p, q, r] be two positive fuzzy numbers. Then ${}^{\alpha}X = [(b-a)\alpha + a, c-(c-b)\alpha]$ and ${}^{\alpha}Y = [(q-p)\alpha + p, r-(r-q)\alpha]$ are the α–cuts of fuzzy numbers X and Y respectively. To calculate division of fuzzy numbers X and Y we first divide the α–cuts of X and Y using interval arithmetic.

$$\frac{{}^{\alpha}X}{{}^{\alpha}Y} = \frac{[(b-a)\alpha + a, c-(c-b)\alpha]}{[(q-p)\alpha + p, r-(r-q)\alpha]}$$

$$= \left[\frac{(b-a)\alpha + a}{r-(r-q)\alpha}, \frac{c-(c-b)\alpha}{(q-p)\alpha + p}\right] .....(3.4)$$

To find the membership function $\mu_{X/Y}(x)$ we equate to x both the first and second component in (3.4) which gives

$$x = \frac{(b-a)\alpha + a}{r-(r-q)\alpha} \text{ and } x = \frac{c-(c-b)\alpha}{(q-p)\alpha + p}.$$

Now, expressing α in terms of x and setting $\alpha = 0$ and $\alpha = 1$ in (3.4) we get α together with the domain of x,

$$\alpha = \frac{xr-a}{(b-a)+(q-r)x}, a/r \le x \le b/q$$

and

$$\alpha = \frac{c-px}{(c-b)+q-p)x}, b/q \le x \le c/p$$

which gives

$$\mu_{X/Y}(x) = \begin{cases} \dfrac{xr-a}{(b-a)+(q-r)x}, a/r \le x \le b/q \\ \dfrac{c-px}{(c-b)+q-p)x}, b/q \le x \le c/p \end{cases}$$

3.5. Inverse of fuzzy number:

Let *X*= [a, b, c] be a positive fuzzy number. Then ${}^{\alpha}X = [(b-a)\alpha + a, c-(c-b)\alpha]$ is the α-cut of the fuzzy numbers *X*. To calculate inverse of the fuzzy number *X* we

first take the inverse of the α -cut of *X* using interval arithmetic.

$$\frac{1}{{}^{\alpha}X} = \frac{1}{[(b-a)\alpha + a, c-(c-b)\alpha]}$$

$$= \left[\frac{1}{c-(c-b)\alpha}, \frac{1}{(b-a)\alpha + a}\right] \text{.....(3.5)}$$

To find the membership function $\mu_{1/X}(x)$ we equate to x both the first and second component in (3.5), which gives

$$x = \frac{1}{c-(c-b)\alpha} \text{ and } x = \frac{1}{(b-a)\alpha + a}$$

Now, expressing α in terms of x and setting $\alpha = 0$ and $\alpha = 1$ in (3.5) we get α together with the domain of x,

$$\alpha = \frac{cx-1}{x(c-b)}, \frac{1}{c} \le x \le \frac{1}{b} \text{ and}$$

$$\alpha = \frac{1-ax}{x(b-a)}, \frac{1}{b} \le x \le \frac{1}{a} \text{ which gives,}$$

$$\mu_{1/X}(x) = \begin{cases} \frac{cx-1}{x(c-b)}, \frac{1}{c} \le x \le \frac{1}{b} \\ \frac{1-ax}{x(b-a)}, \frac{1}{b} \le x \le \frac{1}{a} \end{cases}$$

3.6. Exponential of a Fuzzy number:

Let X= [a, b, c] > 0 be a fuzzy number. Then ${}^{\alpha}A = [(b-a)\alpha + a, c-(c-b)\alpha]$ is the α -cut of the fuzzy numbers A. To calculate exponential of the fuzzy number *A* we first take the exponential of the α -cut of A using interval arithmetic.

$$\exp({}^{\alpha}A) = \exp([(b-a)\alpha + a, c-(c-b)\alpha])$$

$$= [\exp((b-a)\alpha + a), \exp(c-(c-b)\alpha)] \text{.....(3.6)}$$

To find the membership function $\mu_{\exp(X)}(x)$ we equate to x both the first and second component in (3.6) which gives

$$x = \exp((b-a)\alpha + a) \text{ and}$$

$$x = \exp(c-(c-b)\alpha)$$

Now, expressing α in terms of x and setting $\alpha = 0$ and $\alpha = 1$ in (3.6) we get α together with the domain of x,

$$\alpha = \frac{\ln(x)-a}{b-a}, \exp(a) \le x \le \exp(x) \text{ and}$$

$$\alpha = \frac{c-\ln(x)}{c-b}, \exp(b) \le x \le \exp(c)$$

which gives

$$\mu_{\exp(X)}(x) = \begin{cases} \frac{\ln(x)-a}{b-a}, \exp(a) \le x \le \exp(x) \\ \frac{c-\ln(x)}{c-b}, \exp(b) \le x \le \exp(c) \end{cases}$$

3.7. Logarithm of a fuzzy number:

Let X= [a, b, c] > 0 be a fuzzy number. Then ${}^{\alpha}X = [(b-a)\alpha + a, c-(c-b)\alpha]$ is the α -cut of the fuzzy number X. To calculate logarithm of the fuzzy number X we first take the logarithm of the α -cut of X using interval arithmetic.

$$\ln({}^{\alpha}X) = \ln([(b-a)\alpha + a, c-(c-b)\alpha])$$

$$= [\ln((b-a)\alpha + a), \ln(c-(c-b)\alpha)] \text{.....(3.7)}$$

To find the membership function $\mu_{\ln(X)}(x)$ we equate to x both the first and second component in (3.7) which gives

$$x = \ln((b-a)\alpha + a) \text{ and}$$

$$x = \ln(c-(c-b)\alpha) \text{ .}$$

Now, expressing α in terms of x and setting $\alpha = 0$ and $\alpha = 1$ in (3.7) we get α together with the domain of x,

$$\alpha = \frac{\exp(x)-a}{b-a}, \ln(a) \le x \le \ln(b)$$ and

$$\alpha = \frac{c-\exp(x)}{c-b}, \ln(b) \le x \le \ln(c)$$

which produces

$$\mu_{\ln(X)}(x) = \begin{cases} \frac{\exp(x)-a}{b-a}, \ln(a) \le x \le \ln(b) \\ \frac{c-\exp(x)}{c-b}, \ln(b) \le x \le \ln(c) \end{cases}$$

4. Square root of fuzzy number by α-cut method

In this section we determine square root of a fuzzy number by α-cut cut method and compare with that done in [1] and [2].

Let X= [a, b, c] > 0 be a fuzzy number. Then ${}^{\alpha}X = [(b-a)\alpha + a, c-(c-b)\alpha]$ is the α-cut of the fuzzy numbers X. To calculate square root of the fuzzy number X we first take the square root of the α-cut of X using interval arithmetic.

$$\sqrt{{}^{\alpha}A} = \sqrt{[(b-a)\alpha + a, c-(c-b)\alpha]}$$
$$= \left[\sqrt{(b-a)\alpha + a}, \sqrt{c-(c-b)\alpha}\right] ...(4)$$

To find the membership function $\mu_{\sqrt{X}}(x)$ we equate to x both the first and second component in (4), which gives

$$x = \sqrt{(b-a)\alpha + a}$$ and
$$x = \sqrt{c-(c-b)\alpha}$$

Now, expressing α in terms of x and setting $\alpha = 0$ and $\alpha = 1$ in (4) we get α together with the domain of x,

$$\alpha = \frac{x^2-a}{b-a}, \sqrt{a} \le x \le \sqrt{b}$$ and

$$\alpha = \frac{c-x^2}{c-b}, \sqrt{b} \le x \le \sqrt{c}$$

Which gives

$$\mu_{\sqrt{X}}(x) = \begin{cases} \frac{x^2-a}{b-a}, \sqrt{a} \le x \le \sqrt{b} \\ \frac{c-x^2}{c-b}, \sqrt{b} \le x \le \sqrt{c} \end{cases}$$

5. n[th] root of a Fuzzy number:

Let X= [a, b, c] > 0 be a fuzzy number. Then ${}^{\alpha}X = [(b-a)\alpha + a, c-(c-b)\alpha]$ is the α-cut of the fuzzy numbers X. To calculate nth root of the fuzzy number *X* we first take the nth root of the α-cut of X using interval arithmetic.

$$({}^{\alpha}X)^{1/n} = ([(b-a)\alpha + a, c-(c-b)\alpha])^{1/n}$$

$$= \left[((b-a)\alpha + a)^{1/n}, (c-(c-b)\alpha)^{1/n}\right] ...(5)$$

To find the membership function $\mu_{\sqrt[n]{X}}(x)$ we equate to x both the first and second component in (5) which gives

$$x = ((b-a)\alpha + a)^{1/n}$$ and

$$x = (c-(c-b)\alpha)^{1/n}$$

Now, expressing α in terms of x and setting $\alpha = 0$ and $\alpha = 1$ in (5) we get α together with the domain of x,

$$\alpha = \frac{x^n-a}{b-a}, \sqrt[n]{a} \le x \le \sqrt[n]{b}$$ and

$$\alpha = \frac{c-x^n}{c-b}, \sqrt[n]{b} \le x \le \sqrt[n]{c}$$

which gives

$$\mu_{\sqrt[n]{X}}(x) = \begin{cases} \frac{x^n-a}{b-a}, \sqrt[n]{a} \le x \le \sqrt[n]{b} \\ \frac{c-x^n}{c-b}, \sqrt[n]{b} \le x \le \sqrt[n]{c} \end{cases}$$

6. A comparison with the proposed method:

In this section we solve the same problems as done in [2] and [3] by α-cut method and hence make a comparative study.

6.1. Addition of fuzzy Numbers:

Let X= [1,2,4] and Y= [3,5,6] be two fuzzy numbers whose membership functions are

$$\mu_X(x)=\begin{cases}x-1, 1\le x\le 2\\ \dfrac{4-x}{2}, 2\le x\le 4\end{cases}$$

$$\mu_Y(x)=\begin{cases}\dfrac{x-3}{2}, 3\le x\le 5\\ 6-x, 5\le x\le 6\end{cases}$$

Then ${}^{\alpha}X=[1+\alpha, 4-2\alpha]$ and ${}^{\alpha}Y=[2\alpha+3, 6-\alpha]$ are the α –cuts of fuzzy numbers X and Y respectively. Therefore

$${}^{\alpha}X+{}^{\alpha}Y$$

$$=[1+\alpha, 4-2\alpha]+[2\alpha+3, 6-\alpha]$$

$$=[3\alpha+4, 10-3\alpha].....(6.1)$$

We take $x=3+4\alpha$ and $x=10-3\alpha$,
Now, expressing α in terms of x by setting $\alpha=0$ and $\alpha=1$ in (6.1) we get α together with the domain of x,

$$\alpha=\frac{x-4}{3}, 4\le x\le 7 \text{ and}$$

$$\alpha=\frac{10-x}{3}, 7\le x\le 10$$

which gives

$$\mu_{X+Y}(x)=\begin{cases}\dfrac{x-4}{3}, 4\le x\le 7\\ \dfrac{10-x}{3}, 7\le x\le 10\end{cases}$$

6.2 Subtraction of Fuzzy Numbers:

Let X= [1,2,4] and Y= [3,5,6] be two fuzzy numbers. Then ${}^{\alpha}X=[1+\alpha, 4-2\alpha]$ and ${}^{\alpha}Y=[2\alpha+3, 6-\alpha]$ are the α –cuts of fuzzy numbers X and Y respectively. Therefore

$${}^{\alpha}X-{}^{\alpha}Y$$

$$=[1+\alpha, 4-2\alpha]-[2\alpha+3, 6-\alpha]$$

$$=[2\alpha-5, 1-4\alpha].....(6.2)$$

we take
$x=2\alpha-5$ and $x=1-4\alpha$
Now, expressing α in terms of x by setting $\alpha=0$ and $\alpha=1$ in (6.2) we get α together with the domain of x,

$$\alpha=\frac{x+5}{2}, -5\le x\le -3 \text{ and}$$

$$\alpha=\frac{1-x}{4}, -3\le x\le 1$$

which gives

$$\mu_{X-Y}(x)=\begin{cases}\dfrac{x+5}{2}, -5\le x\le -3\\ \dfrac{1-x}{4}, -3\le x\le 1\end{cases}$$

6.3 Multiplication of Fuzzy Numbers:

Let X= [1,2,4] and Y= [3,5,6] be two fuzzy numbers. Then ${}^{\alpha}X=[1+\alpha, 4-2\alpha]$ and ${}^{\alpha}Y=[2\alpha+3, 6-\alpha]$ are the α –cuts of fuzzy numbers X and Y respectively. Therefore,

$${}^{\alpha}X*{}^{\alpha}Y$$

$$=[1+\alpha, 4-2\alpha]*[2\alpha+3, 6-\alpha]$$

$$=[2\alpha^2+5\alpha+3, 2\alpha^2-16\alpha+24].....(6.3)$$

We take
$x=2\alpha^2+5\alpha+3$ and $x=2\alpha^2-16\alpha+24$
Now, expressing α in terms of x by setting $\alpha=0$ and $\alpha=1$ in (6.3) we get α together with the domain of x,

$$\alpha=\frac{\sqrt{1+8x}-5}{4}, 3\le x\le 1 \text{ and}$$

$$\alpha=\frac{8-\sqrt{16+2x}}{2}, 10\le x\le 24$$

which gives,

$$\mu_{XY}(x)=\begin{cases}\dfrac{\sqrt{1+8x}-5}{4}, 3\le x\le 10\\ \dfrac{8-\sqrt{16+2x}}{2}, 10\le x\le 24\end{cases}$$

6.4 Division of Fuzzy Numbers:

Let X= [1,2,4] and Y= [3,5,6] be two fuzzy numbers. Then ${}^{\alpha}X = [1+\alpha, 4-2\alpha]$ and ${}^{\alpha}Y = [2\alpha+3, 6-\alpha]$ are the α –cuts of fuzzy numbers X and Y respectively. Therefore,

$$\frac{{}^{\alpha}X}{{}^{\alpha}Y} = \frac{[1+\alpha, 4-2\alpha]}{[2\alpha+3, 6-\alpha]}$$
$$= \left[\frac{1+\alpha}{6-\alpha}, \frac{4-2\alpha}{2\alpha+3}\right] \text{.....(6.4)}$$

We take

$$x = \frac{1+\alpha,}{6-\alpha} \text{ and } x = \frac{4-2\alpha}{2\alpha+3}.$$

Now, expressing α in terms of x by setting $\alpha = 0$ and $\alpha = 1$ in (6.4) we get α together with the domain of x,

$$\alpha = \frac{6x-1}{1+x}, \frac{1}{6} \le x \le \frac{2}{5}, \text{ and}$$

$$\alpha = \frac{4-3x}{2(1+x)}, \frac{2}{5} \le x \le \frac{4}{3}$$

which gives

$$\mu_{X/Y}(x) = \begin{cases} \frac{6x-1}{1+x}, \frac{1}{6} \le x \le \frac{2}{5} \\ \frac{4-3x}{2(1+x)}, \frac{2}{5} \le x \le \frac{4}{3} \end{cases}$$

6.5 Addition of a triangular and a non-triangular fuzzy number:

Let X=[2,3,4] and Y=[4,16,25] be triangular and non triangular fuzzy numbers respectively whose membership functions are given as,

$$\mu_Y(x) = \begin{cases} \frac{x-2}{2}, 2 \le x \le 4 \\ 5-x, 4 \le x \le 25 \\ 0, otherwise \end{cases} \text{ and}$$

$$\mu_Y(x) = \begin{cases} \frac{\sqrt{x}-2}{2}, 4 \le x \le 16 \\ 5-\sqrt{x}, 16 \le x \le 2 \\ 0, otherwise \end{cases}$$

Then ${}^{\alpha}X = [2+2\alpha, 5-\alpha]$ and ${}^{\alpha}Y = [(2+2\alpha)^2, (5-\alpha)^2]$ are the α –cuts of fuzzy numbers X and Y respectively. Therefore,

$${}^{\alpha}X + {}^{\alpha}Y$$
$$= [2+2\alpha, 5-\alpha] + [(2+2\alpha)^2, (5-\alpha)^2]$$
$$= [4\alpha^2 + 10\alpha + 6, \alpha^2 - 11\alpha + 30] \text{.....(6.5)}$$

To find the membership function $\mu_{X+Y}(x)$ we equate to x both the first and second component in (6.5), which gives

$x = 4\alpha^2 + 10\alpha + 6$ and

$$x = \alpha^2 - 11\alpha + 30$$

Now, expressing α in terms of x and setting $\alpha = 0$ and $\alpha = 1$ in (6.5) we get α together with the domain of x,

$$\alpha = \frac{\sqrt{1+4x}-5}{4}, 6 \le x \le 20 \text{ and}$$
$$\alpha = \frac{11-\sqrt{1+4x}}{2}, 20 \le x \le 30$$

which gives

$$\mu_{X+Y}(x) = \begin{cases} \frac{\sqrt{1+4x}-5}{4}, 6 \le x \le 20 \\ \frac{11-\sqrt{1+4x}}{2}, 20 \le x \le 30 \\ 0, otherwise \end{cases}$$

By comparing the solutions of the above examples as done in [6] we have seen that the α –cuts method is simpler than the method proposed by them. Infact for performing division (and subtraction), their method is lengthy compared to α –cuts method, because they have to first perform inverse (negative) of the divisor (fuzzy number to be subtracted). In example 5 of [6] the result obtained should be

$$\mu_{X+Y}(x) = \begin{cases} \frac{\sqrt{1+4x}-5}{4}, 6 \le x \le 20 \\ \frac{11-\sqrt{1+4x}}{2}, 20 \le x \le 30 \\ 0, otherwise \end{cases}$$

instead of

$$\mu_{X+Y}(x) = \begin{cases} \frac{\sqrt{1+4x}-2}{4}, 6 \le x \le 20 \\ \frac{11-\sqrt{1+4x}}{2}, 20 \le x \le 30 \\ 0, otherwise \end{cases}$$

5. Conclusion:

In this paper we have shown that α-cut method is a method general enough to deal with all kinds of fuzzy arithmetic including nth root, exponentiation and taking log. We have solved problems using this method and compared them with the method in proposed [2] and [6]. We have seen that the alpha-cut method is simpler than the proposed method. As a passing remark we would like to mention that example 3 in [2] is confusing. Because there they started to extract nth root of X =[a, b, c], (a,b,c)>0 but ended up performing exponentiation of the fuzzy number X.

Acknowledgement

The work done in this paper is under a research project funded by Board of Research in Nuclear Sciences, Department of Atomic Energy, Govt. of India.

References:

[1] Bojadziev, G., Bojadziev, M (1995), *Fuzzy sets, Fuzzy logic, application*, World Scientific.

[2] Chutia, R., Mahanta, S., and Baruah, H.K. (Dec.2010) "An alternative method of finding the membership of a fuzzy number", International journal of latest trends in computing, Vol. 1, Issue 2, 69-72.

[3] Dubois, D., and Prade, H., (2000) "Fundamentals of Fuzzy sets", Kluwer Academic Publishers, Boston.

[4] Kaufman A., and Gupta, M. M., (1984) "Introduction to Fuzzy Arithmetic, Theory and applications", Van Nostrand Reinhold Co. Inc., Workingham, Berkshire.

[5] Klir, G. J., and Yuan B., (2005) "Fuzzy Sets and Fuzzy Logic", Prentice Hall of India Private limited, new Delhi.

[6] Mahanta, S., Chutia, R., and Baruah, H.K. (Dec. 2010) "Fuzzy Arithmetic without using the method of α-cut" International journal of latest trends in computing, Vol. 1, Issue 2, 73-80.

Analysis and Comparison of Texture Features for Content Based Image Retrieval

S.Selvarajah [1] and S.R. Kodituwakku [2]

[1] Department of Physical Science, Vavuniya Campus of the University of Jaffna, Vavuniya, Sri Lanka

[2] Department of Statistics & Computer Science, University of Peradeniya, Sri Lanka.

{shakeelas@mail.vau.jfn.ac.lk, salukak@pdn.ac.lk}

***Abstract*:** Texture is one of the important features used in CBIR systems. The methods of characterizing texture fall into two major categories: Statistical and Structural. An experimental comparison of a number of different texture features for content-based image retrieval is presented in this paper. The primary goal is to determine which texture feature or combination of texture features is most efficient in representing the spatial distribution of images. In this paper, we analyze and evaluate both Statistical and Structural texture features. For the experiments, publicly available image databases are used. Analysis and comparison of individual texture features and combined texture features are presented.

***Keywords*:** Image Retrieval, Feature Representation, First Order Statistics, Second Order Statistics, Gabor Transform, Wavelet Transforms.

1. Introduction

In CBIR systems [1]-[9], a feature is a characteristic that can capture a certain visual property of an image either globally for the entire image or locally for regions or objects [5]. Texture is the main feature utilized in image processing and computer vision to characterize the surface and structure of a given object or region. Since an image is made up of pixels, texture can be defined as an entity consisting of mutually related pixels and group of pixels. This group of pixels is called as texture primitives or texture elements (texels). As the texture is a quantitative measure of the arrangement of intensities in a region, the methods to characterize texture fall into two major categories: Statistical and Structural. Statistical methods characterize texture by the statistical distribution of the image intensity. Spatial distribution of gray values is one of the defining qualities of texture. Statistical methods analyze the spatial distribution of gray values, by computing local features at each point in the image, and deriving a set of statistics from the distributions of the local features [10]. Structural methods describe texture by identifying structural primitives and their placement rules. They are suitable for textures where their spatial sizes can be described using a large variety of properties. Gabor Filters and Wavelet Transforms are widely used for describing structural primitives [11], [12].

Many approaches have been proposed for texture based image retrieval. These range from first order statistics, second order statistics, higher order statistics and multiresolution techniques such as wavelet transform. In this paper, we try to analysis and compare the performances of both statistical and structural approaches. Among the statistical features the first-order statistics and the second-order statistics based features are considered. Two-dimensional wavelet transform and Gabor transform are considered as structural features. This paper presents the evaluation of the effectiveness and efficiency of such texture features in CBIR.

2. Methods and Materials

In order to evaluate the effectiveness and efficiency of texture features the following materials and methods are used.

2.1 Materials

To analyze and evaluate the performance of texture descriptors, following structural and statistical features are considered.

2.1.1 Structural Methods

Two different structural methods are considered: two-dimensional wavelet transform and Gabor transform.

(a) Gabor Transform

Gabor filters are a group of wavelets. For a given image $I(x, y)$ with size M×N, its discrete Gabor wavelet transform is given by the following formula.

$$g(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right] \quad (1)$$

where W is called the modulation frequency.

Frequency response of Gabor function $\psi(x, y)$, i.e., its 2-D Fourier transform is given by:

$$\Psi(u,v)=\exp\{-\frac{1}{2}[\frac{(u-W)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}]\} \quad (2)$$

where $\sigma_u=(2\pi\sigma_x)^{-1}$ and $\sigma_v=(2\pi\sigma_y)^{-1}$

After applying Gabor transformation on the image with different orientation at different scale, the following array of magnitudes can be obtained:

$$E(m,n)=\sum_x\sum_y |G_{mn}(x,y)| \quad (3)$$

where m = 0, 1, …, M-1; n = 0, 1, …, N-1

These magnitudes represent the energy content at different scale and orientation of the image.

In order to identify the homogenous texture, the following mean and standard deviation of the magnitude of the transformed coefficients are used to represent the homogenous texture feature of the region:

$$\mu_{mn}=\frac{E(m,n)}{P\times Q} \quad (4)$$

$$\sigma_{mn}=\frac{\sqrt{\sum_x\sum_y (|G_{mn}(x,y)|-\mu_{mn})^2}}{P\times Q} \quad (5)$$

where P and Q are the image sizes.

A feature vector f (texture representation) is created using μ_{mn} and σ_{mn} as the feature components [12]. Five scales and six orientations are used in common implementation and the feature vector is given by:

$$f=(\mu_{00},\sigma_{00},\mu_{01},\sigma_{01},.....,\mu_{45},\sigma_{45}). \quad (6)$$

(b) 2-D wavelet transforms

Region-based image retrieval was proposed to extend content-based image retrieval so that it can cope with the cases in which users want to retrieve images based on information about their regions. The region-based systems which use wavelet transform are classified into three categories: a hierarchical block, a moving window and a pixel. Since these methods are subject to some limitations, several improvements have been proposed. The method proposed in [11] segments an image into some regions by clustering pixels. Then the texture features are calculated from wavelet coefficients of all regions (subbands). The segmented regions are indexed by the averaged features in the regions.

After decomposing the image into non-overlapping sub subbands as shown in Figure 1, the mean and standard deviation of the decomposed image portions are calculated. These mean and standard deviation are treated as the texture features for image comparison.

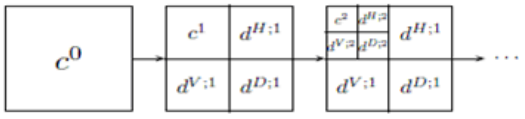


Figure 1: Wavelet decomposition of an image.

2.1.2 Statistical Methods

Among the statistical features the following first-order statistics and second-order statistics are used as texture features in representing images.

(a) First-order statistics

First-order texture measures are calculated from the original image values. They do not consider the relationships with neighborhood pixel. Histogram-based approach to texture analysis is based on the intensity value concentrations on all or part of an image represented as a histogram. Features derived from this approach include moments such as mean, standard deviation, average energy, entropy, skewness and kurtosis [10]. The histogram of intensity levels is a simple summary of the statistical information of the image and individual pixels are used to calculate the gray-level histogram. Therefore, the histogram contains the first-order statistical information about the image (or sub image). These statistics are defined by the following equations.

$$mean(\mu_i)=\frac{\sum_{x=1}^{M}\sum_{y=1}^{N} I_i(x,y)}{M\times N} \quad (7)$$

$$Sandard\ Deviation(\sigma_i)=\sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I_i(x,y)-\mu)^2}{M\times N}} \quad (8)$$

$$Energy(e_i)=\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} I_i{}^2(x,y) \quad (9)$$

$$Entropy_i=\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} I_i\ (x,y)(-\ln I_i(x,y)) \quad (10)$$

$$Skewness=\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I_i(x,y)-\mu)^3}{M\times N\times\sigma^2} \quad (11)$$

$$Kurtosis(k)=\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I_i(x,y)-\mu)^4}{M\times N\times\sigma^4}-3 \quad (12)$$

(b) Autocorrelation function

An autocorrelation function measures the linear spatial relationships between spatial sizes of texture primitives. Autocorrelation-based approach to texture analysis is based on the intensity value concentrations on all or part of an image represented as a feature vector. Calculation of the autocorrelation matrix involves individual pixels.
The set of autocorrelation coefficients are defined by the following function.

$$C_{ff}(p,q)=\frac{MN\sum_{i=1}^{M-p}\sum_{j=1}^{N-q} f(i,j)f(i+p,j+q)}{(M-p)\times(N-q)\sum_{i=1}^{M}\sum_{j=1}^{N} f^2(i,j)} \quad (13)$$

where p and q are the positional difference in the $i^{th}j^{th}$ direction, and M and N are image dimensions.

(c) Run length matrices

Another method characterizes texture images based on run-lengths of image gray levels. Galloway [13], Chu *et. al.* [14] and Dasarathy and Holder [15] introduced different run-length matrices as feature representatives. For a given image, a run-length matrix $p(i, j)$ is defined as the number of runs with pixels of gray level i and run-length j. Let M be the number of gray levels and N be the maximum run length. Then following three matrices define the traditional run-length features.

1) Gray Level Run-Length Pixel Number Matrix:

$$P_p(i, j) = p(i, j) \cdot j$$

Each element of the matrix represents the number of pixels of run-length j and gray level i.

2) Gray-Level Run-Number Vector:

$$P_g(i) = \sum_{j=1}^{N} p(i, j)$$

This vector represents the sum distribution of the number of runs with gray level i.

3) Run-Length Run-Number Vector:

$$P_r(j) = \sum_{i=1}^{M} p(i, j)$$

This vector represents the sum distribution of the number of runs with run length j.

Galloway [13] introduced five original features of run-length statistics as defined as follows.

$$\textit{Short Run Emphasis (SRE)} = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{p(i,j)}{j^2}$$

$$= \frac{1}{n_r} \sum_{j=1}^{N} \frac{p_r(j)}{j^2} \qquad (14)$$

$$\textit{Long Run Emphasis (LRE)} = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} p(i, j) \cdot j^2$$

$$= \frac{1}{n_r} \sum_{j=1}^{N} p_r(j) \cdot j^2 \qquad (15)$$

$$\textit{Gray-Level Nonuniformity (GLN)} = \frac{1}{n_r} \sum_{i=1}^{M} \left(\sum_{j=1}^{N} p(i,j) \right)^2 = \frac{1}{n_r} \sum_{i=1}^{M} p_g(i)^2 \qquad (16)$$

$$\textit{Run Percentage (RP)} = \frac{n_r}{n_p} \qquad (17)$$

where n_r is the total number of runs and n_p is the number of pixels in the image.

Chu et al [14] proposed the following two features.

$$\textit{Low Gray-Level Run Emphasis (LGRE)} = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{p(i,j)}{i^2} = \frac{1}{n_r} \sum_{i=1}^{M} \frac{p_g(i)}{i^2} \qquad (18)$$

$$\textit{High Gray-Level Run Emphasis (HGRE)}$$

$$= \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} p(i, j) \cdot i^2$$

$$= \frac{1}{n_r} \sum_{i=1}^{M} p_r(j) \cdot i^2 \qquad (19)$$

Dasarathy and Holder [15] introduced the following features using the joint statistical measure of gray level and run-length.

$$\textit{Short Run Low Gray-Level Emphasis (SRLGE)}$$

$$= \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{p(i,j)}{i^2 . j^2} \qquad (20)$$

$$\textit{Short Run High Gray-Level Emphasis (SRHGE)}$$

$$= \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{p(i,j) . i^2}{j^2} \qquad (21)$$

$$\textit{Long Run Low Gray-Level Emphasis (LRLGE)}$$

$$= \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{p(i,j) . j^2}{i^2} \qquad (22)$$

$$\textit{Long Run Hgh Gray-Level Emphasis (LRHGE)} = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} p(i,j) . i^2 . j^2 \qquad (23)$$

(d) Second-order statistics

The gray-level co-occurrence matrix (GLCM) or gray-level spatial dependence matrix based calculations fall under the category of second-order statistics. Haralick *et. al.* [16] suggested a set of 14 textual features which can be extracted from the co-occurrence matrix, and which contain information about image textural characteristics such as homogeneity, contrast and entropy.

A gray level co-occurrence matrix (GLCM) contains information about the positions of pixels having similar gray level values. It is a two-dimensional array, P, in which both the rows and the columns represent a set of possible image values. A GLCM Pd [i, j] is defined by first specifying a displacement vector d = (dx, dy) and counting all pairs of pixels separated by d having gray levels i and j.

$P_d[i, j] = n_{ij}$, where n_{ij} is the number of occurrences of the pixel values (i, j) lying at distance d in the image.

The co-occurrence matrix P_d has dimension n x n, where n is the number of gray levels in the image. From this co-occurrence matrix P_d we can derive the following statistics as texture features.

$$\text{Contrast} = \sum_{i\,j=1}^{n} P_d (i - j)^2 \quad (24)$$

Contrast returns a measure of the intensity contrast between a pixel and its neighbor over the whole image.

$$\text{Dissimilarity} = \sum_{i\,j=1}^{n} P_d |i - j| \quad (25)$$

$$\text{Homogeneity} = \sum_{i\,j=1}^{n} \frac{P_d}{1 + |i - j|} \quad (26)$$

Homogeneity returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$\text{ASM (Energy)} = \sum_{i\,j=1}^{n} P_d{}^2 \quad (27)$$

ASM returns the sum of squared elements in the GLCM.

$$\text{Entropy} = \sum_{i\,j=1}^{n} P_d (-\ln P_d) \quad (28)$$

Entropy is a measure of information content. It measures the randomness of intensity distribution.

$$GLCM \text{ mean} = \begin{cases} \mu_i = \sum_{i\,j=1}^{n} i(P_d) \\ \mu_j = \sum_{i\,j=1}^{n} j(P_d) \end{cases} \quad (29)$$

$$\text{GLCM Standard Deviation} = \begin{cases} \sigma_i = (\sum_{i\,j=1}^{n} P_d (i - \mu_i)^2)^{1/2} \\ \sigma_j = (\sum_{i\,j=1}^{n} P_d (j - \mu_j)^2)^{1/2} \end{cases} \quad (30)$$

$$\text{Inverse Difference Moment} = \sum_{i\,j=1}^{n} \frac{P_d}{|i - j^2|} \qquad i \neq j \quad (31)$$

GLCM correlation returns a measure of how correlated a pixel is to its neighbor over the whole image. Correlation is 1 or -1 for a perfectly positively or negatively correlated image.

2.2.3 Similarity measurements

Eight similarity measurements have been proposed [9]. In this work, we use fixed threshold, Sum-of-Squared-Differences (SSD) and sum-of-Absolute Differences (SAD) methods.

$$SSD(f_q, f_t) = \sum_{i=o}^{n-1} (f_q[i] - f_t[i])^2 \quad (32)$$

$$SAD(f_q, f_t) = \sum_{i=o}^{n-1} (1 f_q[i] - f_t[i]) \quad (33)$$

Where $f_q f_t$ represent query feature vector, and database feature vectors and n is the number of features in each vector.

***2.2* Methodology**

Since the spatial distribution of gray values defines the qualities of texture, we use the gray-scale model for our experiments. In order to evaluate the efficiency of the texture features, we use 29 features defined in the previous section. Fixed threshold, sum-of-squared-difference (SSD) and sum-of-absolute-difference (SAD) are used to measure the similarity between query image and the database images. The Brodatz image database that contains 800 purely periodic and homogeneous images is used for the retrieval.

The images are stored in GIF format and for each image texture features have been computed and stored in a database. The database images are subjected to histogram equalization to prevent bias towards images processing similar grey levels. The histogram based features such as mean, standard deviation, energy etc are computed for images in the database and the computed values are stored in a database. The absolute difference of the feature vector values of the query image and database images are also calculated. After that, in order to identify the relevant images, fixed threshold, SAD and SSD are used. In case of fixed threshold, the threshold values are computed for different query images and from the observation of the relevant retrieved images. The best threshold value is chosen as the threshold of that particular texture feature.

The same procedure that is used for first-order statistics is used for autocorrelation based approach as well as the run-length matrix approach. In the autocorrelation based approach, 100 features are undertaken for analysis. Total of eleven features are used for run-length matrix approach.

In calculating the GLCM, we defined an array of offsets that specify four directions (horizontal, vertical, and two diagonals) and one distance. In this case, the input image is represented by four GLCMs. First the statistics from these four GLCMs are calculated and then the average of those four GLCMs is taken. The features such as contrast, homogeneity, correlation etc are computed for database images and the computed values are stored in a database. The SAD and SSD methods are used to measure the similarity. After retrieving images based on individual texture features, combinations of such features are also considered. In comparing the combination of image features, sum-of-Absolute Differences (SAD) similarity measurement is used.

The procedure used for statistical feature extraction was applied for structural feature extraction as well. After retrieving images based on individual texture features, combinations of such features are also considered. In comparing the combination of image features, sum-of-Absolute Differences (SAD) similarity measurement is used.

3. Results and Discussion

The texture features mentioned earlier were used to retrieve image and to compare the performance of individual and combined texture features. In this experiment, the top 50 images were chosen as the retrieval results. Some of the test results for individual texture features mentioned under statistical features are shown in Figures 1, 2 and 3. Test results for Wavelet and Gabor transformations are shown in Figures 4 and 5 respectively. The first image of these resultant images is the query images. The results for combinations of features are given in Table 1.

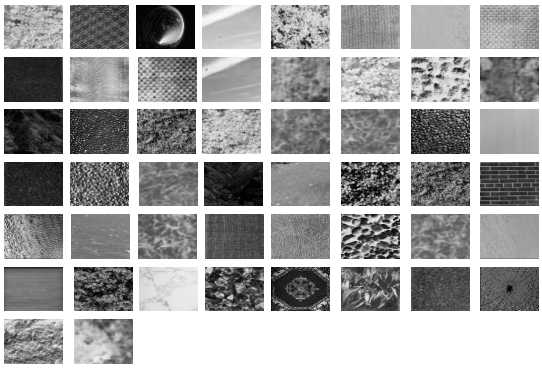


Figure 1: First 50 retrieved images for the feature mean.

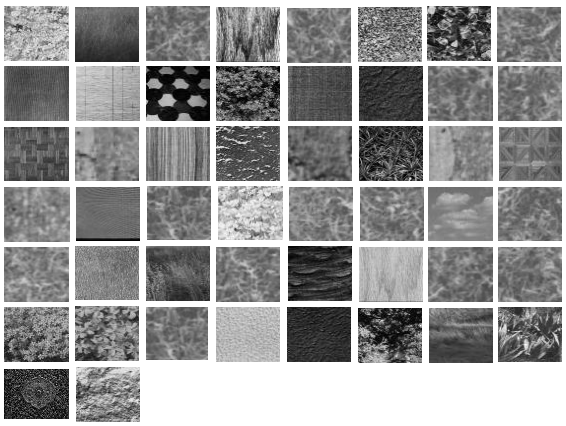


Figure 2: First 50 retrieved images for the feature

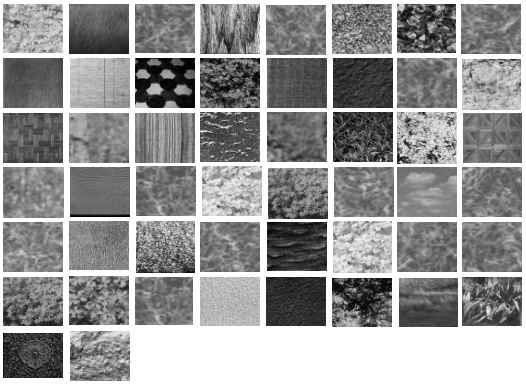


Figure 3: First 50 retrieved images for the feature GLCM.

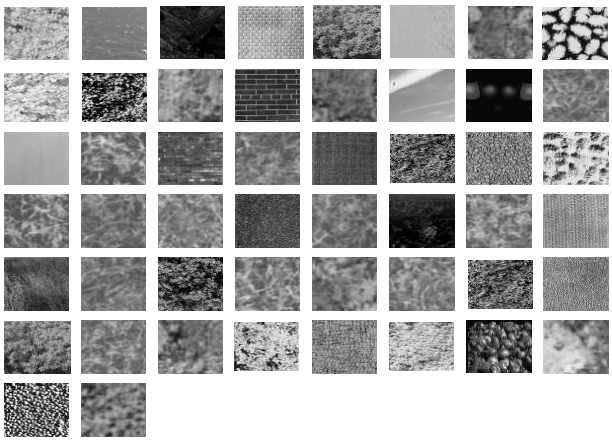


Figure 4: First 50 retrieved images for the feature Gabor Features.

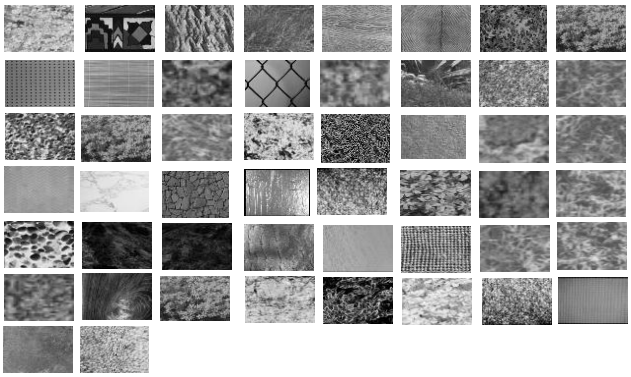


Figure 5: First 50 images retrieved for the feature 2-DWT.

Table 1: An average retrieval accuracy of combined texture features

Combined Feature	Average Precision
First Order Statistics(FS)	0.3400
Autocorrelation (AC)	0.4200
Gray Level Run Length Matrices(GLRLM)	0.4000
Gray Level Co occurrence Matrix(GLCM)	0.4400
FS+AC	0.5200
FS+AC+GLRLM	0.6000
FS+AC+GLCM	0.7500
FS+GLRLM	0.5200
FS+GLRLM+GLCM	0.6500
FS+GLCM	0.6600
AC+GLRLM	0.5600
AC+GLCM	0.6800
AC+GLRLM+GLCM	0.8000
GLRLM+GLCM	0.7400
FS+AC+GLCM+GLRLM	0.8100
Gabor Transform	0.7600
2 D wavelet Transform	0.7400
Combination of all features	0.7800

4. Conclusions

An experimental comparison of a number of different texture features for content-based image retrieval was carried out. The First-order statistics, second-order statistics, Gabor transform and 2D Wavelet transforms were considered for retrieval. The retrieval efficiency of the texture features was investigated by means of relevance. According to the results obtained it is difficult to claim that any individual feature is superior to others. The performance depends on the spatial distribution of images. The test results indicated that Gray Level Co occurrence Matrix performs well compared to other features when images are homogeneous. In most of the image categories, Autocorrelation feature also shows similar performance. It is also noted that the structural texture features are more effective than the statistical texture features. In case of combination features, combinations recorded better retrieval rate compared to the performances of those individual texture features.

5. Reference

[1]. Chang S.K., and Hsu A. (1992), "Image information systems: where do we go from here?" IEEE Trans. On Knowledge and Data Engineering, Vol. 5, No. 5, pp. 431-442.

[2]. Tamura, H. and Yokoya N. (1984), "Image database systems: A survey," Pattern Recognition, Vol. 17, No. 1, pp. 29-43.

[3]. Zachary J. M., Jr. and Sitharama S. I. Content Based Image Retrieval Systems, Journal of the American Society for Information Science and Technology, 2001.

[4]. Goodrum A. A. (2000), Image Information Retrieval: An overview of Current Research.

[5]. Smulders A.M.W., Worring M., Santini, A. S., Gupta, and R. Jain (2000), "Content-based image retrieval at the end of the early years", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.22, No. 12, pp. 1349-1380.

[6]. Faloutsos C., Barber R., Flickner M., Hafner J., Niblack W., Petkovic D.and Equitz W. (1994), "Efficient and effective querying by image content, "Journal of intelligent information systems, Vol.3, pp.231-262.

[7]. Fuhui L., Hongjiang Z. and David D. F. (2003), Multimedia Information retrieval and Management.

[8]. Rui Y. and Huang T. S (1997), Image retrieval: Past, Present, and Future, Journal of Visual Communication and Image Representation.

[9]. Huang J., Kumar S. R., Mitra M., Wei-Jing Z., and Zabih R. (1997), Image indexing using color correlograms. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 762--768.

[10]. Brenard J., (2005), Digital Image Processing, Springer-Verlag Berlin, germany. Nobuo Suematsu, Yoshihiro Ishida, Akira Hayashi and Toshihiko Kanbara, Region-Based Image Retrieval using Wavelet Transform.

[12]. Dengsheng Zhang, Aylwin Wong, Maria Indrawan, Guojun Lu, Content-based Image Retrieval Using Gabor Texture Features.

[13]. Galloway M. M. (1975), Texture analysis using gray level run lengths, Computer Graphics and Image Processing, Volume 4(2), Pages 172-179

[14]. Chu A., Sehgal C.M.A. and Greenleaf J.F., (1990), Use of gray value distribution of run lengths for texture analysis, Pattern Recognition Letters, Volume 11(6), Pages 415-419

[15]. Belur V. Dasarathy and Edwin B. Holder, Image characterizations based on joint gray level—run length distributions. (1981), Pattern Recognition Letters, Volume 12 (, Issue 8, August 1991, Pages 497-502

[16]. Haralick R. M., Shanmugum K. and Dinstein I. (1973), Texture features for image classicification, IEEE Transactions on System, Man and. Cybernetics, SMC-3(6), Brodatz Textures: A photographic Album for Artists & Designers, New York: Dover, 1966.

Author Biographies

Saluka Ranasinghe Kodituwakku is an associate professor at the Statistics and Computer Science Department, University of Peradeniya, Sri Lanka. His research interests are database systems, distributed computing and software engineering.

S. Selvarajah is a lecturer at the Department of Physical Science, Vavuniya Campus of the University of Jaffna, Vavuniya, Sri Lanka and a Mphil student at the Postgraduate Instituteof Science, university of Peradeniya, Sri Lanka.

Real Time Outlier Detection in Wireless Sensor Networks

M.Syed Mohamed[1] T. Kavitha[2]
[1]M.E. (CSE) Final Year, Anna University of Technology, Tirunelveli, Tamilnadu , India.
E-mail: seyadumohamed@gmail.com
[2]Assistant Professor, Department of Computer Science and Engineering,
Anna University of Technology, Tirunelveli, Tamilnadu, India.
E-mail: kavi_gce@rediffmail.com

Abstract: In the field of wireless sensor networks, the sensed patterns that significantly deviate from the normal patterns are considered as outliers. The possible sources of outliers include noise and errors, events, and malicious attacks on the network. Wireless sensor networks (WSNs) are more likely to generate outliers due to their special characteristics, like constrained available resources, frequent physical failure, and often harsh deployment area. In general, the outlier data are ignored, but in some critical applications like fire detection, hoax and intruder detection, these outlier data are much important to take decision in real time. The usual outlier detection techniques are not directly applicable to wireless sensor networks due to the nature of sensor data and specific requirements and limitations of the wireless sensor networks. In this paper we are going to present the outlier detection method for real time events in the wireless sensor networks. Our proposed method will train and test the data in real time. So, this proposed method is much useful in finding the events in real time more efficiently with the help of outliers.

Keywords: Outliers, Wireless Sensor Network(WSN), Support Vector Machine(SVM)

1. Introduction

1.1. Wireless Sensor Network

Wireless sensor nodes are tiny, low-cost sensor devices integrated with sensing, processing and short-range wireless communication capabilities. The popular WSN applications are industrial, business and military domains, environmental and habitat monitoring, object and Inventory tracking, medical monitoring, battlefield observation, industrial safety and controlling etc.

1.2 Architecture of Wireless Sensor Network

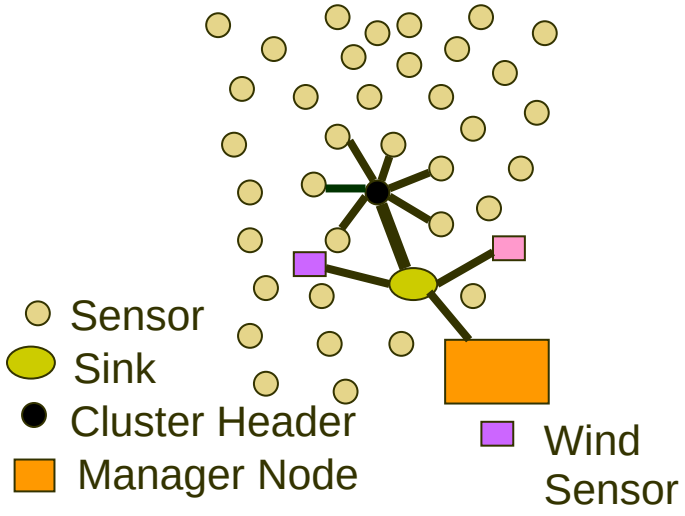


Figure 1: Architecture of Wireless Sensor Network

- In WSN, the tiny sensors are deployed all over the Field.
- The Networks are usually composed of few sinks and large quantity of sensor nodes.
- Sensor nodes are organized into clusters.
- Each node has a corresponding cluster header.
- Each sensor node can measure temperature, smoke and relative humidity.
- Nodes location information can be obtained by equipments such as Global Positioning System (GPS).

The remainder of this paper is organized as follows. Section II provides an overview of outliers and advantages of support vector machines for outlier detection. Section III introduces the problem description and the proposed outlier detection technique for real time sensor data. Section IV presents experiment data sets, data analysis, and lessons learned. Conclusion and future enhancements will be addressed in Section V.

2. Outliers

In the field of wireless sensor networks, the sensed patterns that significantly deviate from the normal patterns are considered as outliers.

2.1 Type of Outliers

Local Outliers:

i. Each node identifies the anomalous values only depending on its historical values
ii. Each sensor node collects readings of its neighboring nodes to collaboratively identify the anomalous values.

Global Outliers:

i. The identification of global outliers can be performed at different levels in the network
ii. In a centralized architecture, all data is transmitted to the sink node for identifying outliers.

2.2 Requirements for outlier detection in WSNs

- It must distributively process the data to prevent unnecessary communication overhead and energy consumption and to prolong network lifetime.
- It must be an online technique to be able to handle streaming or dynamically updated sensor data.
- It must have a high detection rate while keeping a false alarm rate low.
- It should be unsupervised as in WSN the pre- classified normal or abnormal data is difficult to obtain. Also, it should be non-parametric as no a priori knowledge about the input sensor data distribution may be available.
- It should take multivariate data into account.
- It must be simple, have low computational complexity, and be easy to implement in presence of limited resources.
- It must enable auto-configurability with respect to dynamic network topology or communication failure.
- It must scale well.
- It must consider dependencies among the attributes of the sensor data as well as spatio-temporal correlations that exist among the observations of neighboring sensor nodes.
- It must effectively distinguish between erroneous measurements and events

2.3 Advantages of SVM-Based Classification Techniques

It is the popular classification-based approaches in the data mining and machine learning communities. It is widely used to detect outliers due to the following main advantages:

i. Do not require an explicit statistical model
ii. Provide an optimum solution for classification by maximizing the margin of the decision boundary
iii. Avoid the dimensionality problem

3. Problem Description

We assume that sensor data are correlated in both time and space and the sensor nodes are densely deployed with time synchronized. The network topology is modeled as an undirected graph.

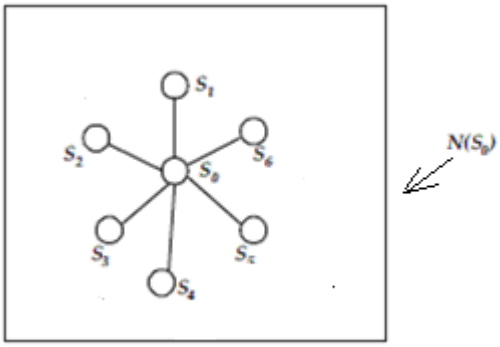


Figure 2: A closed neighborhood of the sensor node S_0

The Figure 2 representing a closed neighborhood of the sensor node S_0 as centered with the radio transmission range of S_0. At each time interval Δi, each sensor node in the set $N(S_0)$ measures a data vector. At every time interval Δi,. Let $x^i_0, x^i_1, x^i_2, \ldots, x^i_k$ denote the data vector measured at S_0, S_1, $S_2, \ldots, S_k$, respectively. Our aim is to identify every real time measurements collected by S_0 as outlier or normal. This process can be implemented on each node in the network and scales to large wireless sensor networks.

3.1 The Proposed Method

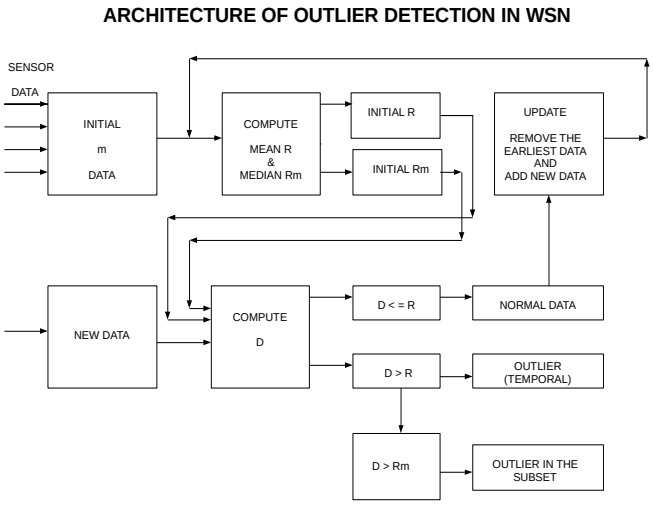


Figure 3: Architecture of Outlier Detection in WSN

First, every node using its *m* sequential data measurements find the local radius R. The learned radius information of each node is then transmits to its neighboring nodes. After receiving the radii from all of its neighbors, each node computes a median radius *Rm* of its neighboring nodes in the sub-network. Subsequently, when a new sensor measurement x^t is collected at time *t*, the node first compares the distance of x^t from the origin with the radius *R* learned with respect to its *m* previous measurements. The measurement x^t will be classified as normal data if $d(x^t)$ <=R. If the measurement $x^t > R$ then x^t is classified as outlier in the node. So, x^t is compared with median radius Rm of neighboring nodes, ie. if $d(x^t)$ >Rm then x^t is classified as outlier in the subset. Otherwise it is identified as the local outlier of the node.

4. Results and Discussion

4.1 Experimental Dataset

The real data are collected from a closed neighborhood from a WSN deployed in *Grand-St-Bernard* as shown in Figure 4. The results are obtained using MATLAB.

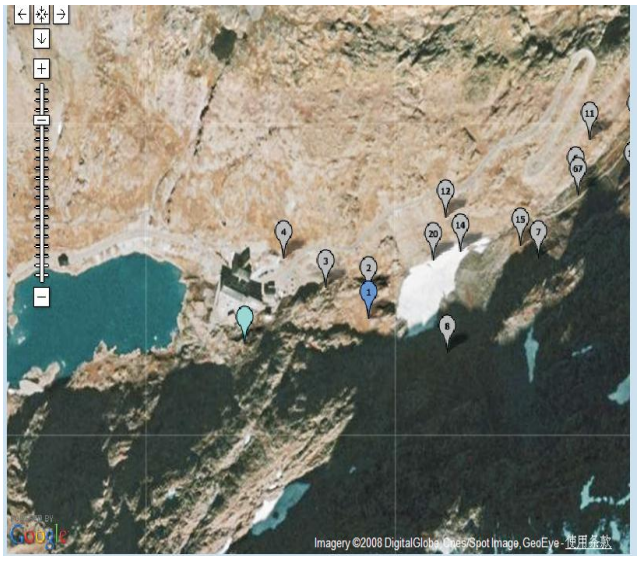


Figure 4: Grand-St-Bernard Deployment

The closed neighborhood contains the node 14 and its 6 spatially neighboring nodes, namely nodes 8, 12, 20, 7, 15, 67. The network recorded ambient temperature, relative humidity, soil moisture, solar radiation and watermark measurements at 2 minutes intervals. In our experiments, we use a 6am-6pm period of data recorded on 20th September 2007 with the attribute ambient temperature for each sensor measurement.

4.2 Results and Analysis

The Figure-5 represents the training set of node-1 data. It finds some outliers and normal data. The Figure-6 represents the classified data of node-1 which identifies local outliers of node-1.

The Figure-7 represents the subset outliers of node-1. It clearly represents that some of the local outliers of node-1 is also the outliers of the subset.

The Figure-8 represents the classified data of node-6 which identifies local outliers of node-6.

The Figure-9 clearly represents that the outliers of node-6 are only the local outliers. There is no data of node-6 is classified as subset outliers. This result concludes that our outlier detection method classifies the sensor data in real time as normal or outlier in an efficient manner.

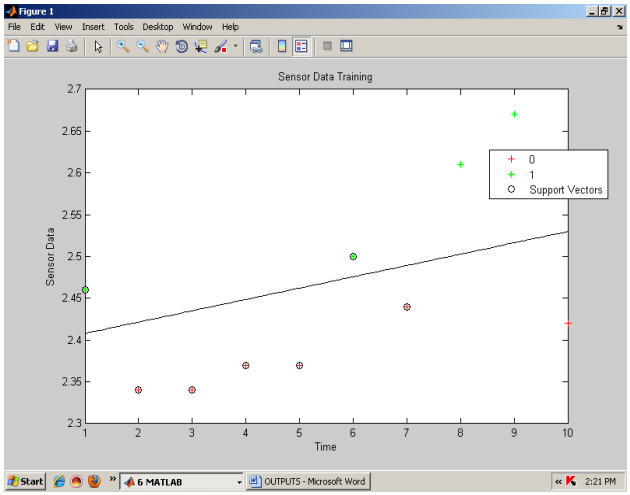


Figure 5: Training Set of Node-1

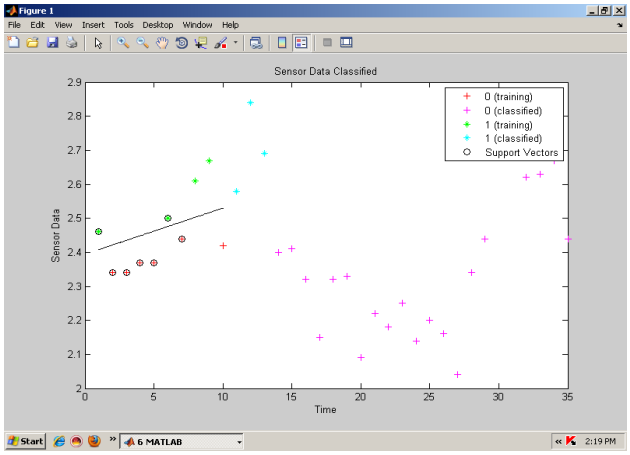


Figure 6: Local Outliers in the Node-1

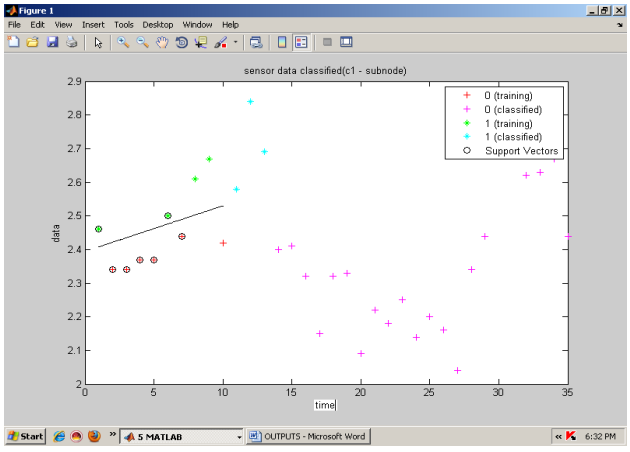


Figure 7: Subset Outliers in the Node-1

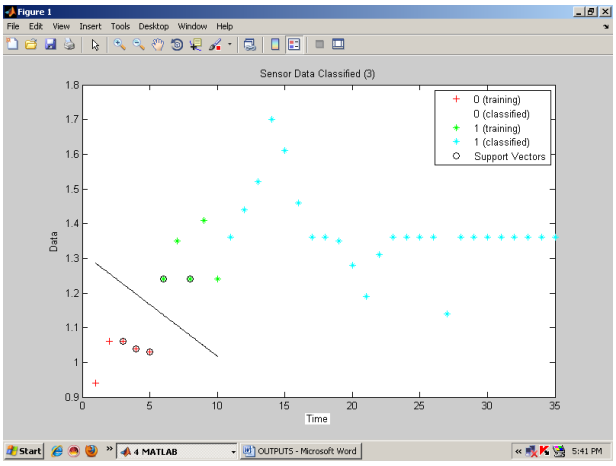


Figure 8: Local Outliers in the Node-3

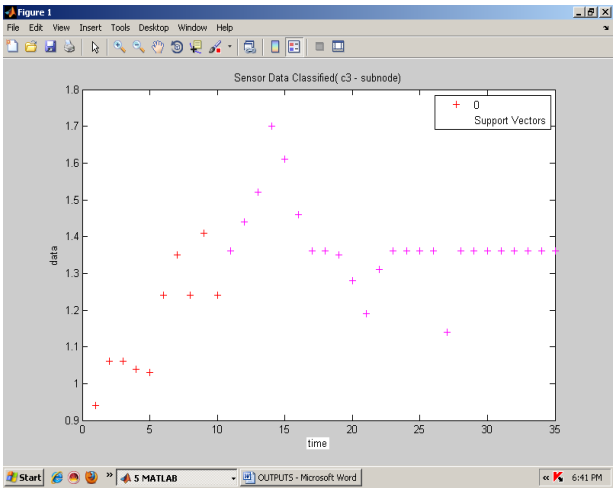


Figure 9: Local Outliers in the Node-3 (No Subset Outliers Exists in Node-3)

5. Conclusion

In this paper, we have examined the real time wireless sensor data as outlier or normal based on SVM based outlier detection technique. We compared the performance of the technique with synthetic data and real data of the SensorScope System. We found that our method finds the outlier with better accuracy and minimum false alarm rate.

5.1 Future Work

In our future proposals, we wish to find the solution to minimize the computational complexity and the memory requirements. Further we wish to classify the sensed data as an event or an error in the subset. The results may also be tested with other traditional methods. If needed, we wish to

apply multi-classification method to keep the accuracy and to reduce the false alarm rate. Further, time complexity to train and test the real time data is also to be verified.

References

[1] Chandola V, A. Banerjee, and V. Kumar. Outlier detection: A survey. *Technical Report, University of Minnesota*, 2007

[2] Ding M, D. Chen, K. Xing, and X. Cheng, "Localized fault-tolerant event boundary detection in sensor networks," in *Proceedings of IEEE Conference of Computer and Communications Societies*, 2005, pp. 902-913.

[3] *http: //sensorscope.epfl.ch/index.php/MainP age*.

[4] Luo X, M. Dong, and Y. Huang, "On distributed fault-tolerant detection in wireless sensor networks", *IEEE Transactions on Computers*, vol. 55, pp. 58-70, 2006.

[5] Marin-Perianu M. and P. J. M. Havinga, " D-FLER - A Distributed Fuzzy Logic Engine for Rule-Based Wireless Sensor Networks," Springer Verlag, Germany, 2007, pp. 86-101.

[6] Rajasegarar S., C. Leckie, M. Palaniswami, and J. C. Bezdek. Quarter sphere based distributed anomaly detection in wireless sensor networks. *IEEE International Conference on Communications*, June 2007.

[7] Segal M.L., F. P. Antonio, S. Elam, J. Erlenbach, and K. R. de Paolo, "Method and apparatus for automatic event detection in a wireless communication system," U. Patent, Ed. USA, 2000.

[8] Tax D.M.Jand R. P. W. Duin. Support vector data description. *Machine Learning*, 54(1):45-66, 2004.

[9] Vu C. T,, R. A. Beyah, and Y. Li, "Composite Event Detection in Wireless Sensor Networks," in *Performance, Computing, and Communications Conference, 2007. IPCCC 2007 . IEEE International*, 2007.

[10] Werner-Allen G, K. Lorincz, M. Ruiz, O. Marcillo, J. Johnson, J. Lees, and M.Welsh, "Deploying a wireless sensor network on an active volcano" *IEEE Internet Computing*, vol. 10, pp. 18-25, 2006.

[11] Xue .W, Q. Luo, L. Chen, and Y. Liu, "Contour map matching for event detection in sensor networks," in *International Conference on Management of Data*, Chicago, IL, USA 2006

[12] Zhang Y, N. Meratnia, and P. Havinga. An online outlier detection technique for wireless sensor networks using unsupervised quarter-sphere support vector machine. *ISSNIP 2008*

[13] Zhang Y., N. Meratnia, and P. Havinga. Outlier detection techniques for wireless sensor networks: A survey. *Technical Report, University of Twente*, 2007.

Sophomores Use of Productivity Tools across Departments

Paul A. Walcott[1], Jamillah M. A. Grant[2], Troy Lorde[3], Colin Depradine[1], Elisabeth Bladh[4]

[1]Department of Computer Science, Mathematics and Physics, The University of the West Indies, Cave Hill Campus,
[2]College of Education, Northcentral University,
[1]Department of Economics, The University of the West Indies, Cave Hill Campus,
[4]Department of Languages and Literatures, University of Gothenburg.

Corresponding Address
paul.walcott@cavehill.uwi.edu

***Abstract*:** It is expected that university students are competent, to some degree, in the use of productivity tools such as word processors, spreadsheets, presentation software and databases. The level of competency that most students possess in these productivity tools however is seldom reported in the literature. To determine these levels, a survey was administered to sophomore students who self-reported the activities they could perform. Descriptive statistics were calculated for the percentage of students that could perform each activity per productivity tool. Competency levels were ranked by five departments, namely Computer Science; Economics; Government; History and Philosophy; and Language, Linguistics, and Literature. Statistical tests were then used to determine if there were significant differences across these departments. The findings indicate that Computer Science and Economic students have comparable levels of competency, while Language, Linguistics, and Literature students have significantly lower competencies in the productivity tools examined.

***Keywords*:** Technology in education, Computer competencies, computer skills, productivity tools, Microsoft Office.

1. Introduction

Colleges and Universities have had "to stretch the range of their teaching programs due to the broadening of the quality of students and the curriculum designed to attract them (Keller, 2005). As the student population at the college and university level increases, so does the heterogeneity of the student body with respect to computer knowledge and skills. One school of thought is that most young university students are tech-savvy (Oblinger and Hawkens 2006; Prensky 2001a, 2001b); while another school of thought is that "incoming freshman do not have the information technology skills needed for college-level work (Kaminski, Seel and Cullen 2003)."

Keller (2005) noted that in light of universities growing heterogeneous student populations, institutions needed to become more aware of student qualities so that programs of learning are comparable to the students' nature. Researchers such as Gay, Mahon, Devonah, Alleyne, and Alleyne (2005) have studied students' attitudes toward the use of technology, however, the students' levels of competency in those skills assessed were not addressed. Our earlier papers (Walcott, Grant and Depradine 2008; Grant and Walcott 2009) specifically address the competencies possessed by students, but do not report on the levels of these competencies. Other studies have examined the perceived competencies of freshman and senior students in the use of productivity tools (Kvavik and Caruso 2005; Haywood et al 2004). None of these studies specifically examine the computer competencies required by sophomore students in order for them to complete their course of study. This study fills that important gap in the current literature by addressing sophomore students' level of skill using basic technological tools, namely: word processing, database, presentation, and spreadsheet tools.

Determining the tech-savvy of university students with respect to productivity tools makes it plausible to set standards or expectations for students' use of technology in academia. For this reason, this research will inform students' technological abilities in relation to productivity tools and will assist in establishing a benchmark of proficiency for students' use of such technological skills. This study particularly addressed those skills that students should need in academia and might need in the workplace.

To assess whether students were equipped with requisite computer skills, it was decided that a survey would be conducted. After administering the survey, further discussion lead to the determining of the level of competency of sophomore students in the use of the four productivity tools.

This paper provides a theoretical background on what is meant by computer literacy and/or competencies. Next a contextual framework of the survey study is presented. The methodology of the survey study and instruments used are detailed. The results are presented statistically and analyzed dialogically, then recommendations are made and conclusions are drawn. For unanswered questions future research studies are suggested.

1.1 Importance of the study

The Barbados National Strategic Plan 2006-2025 (http://www.barbados.gov.bb/Docs/NSP_Final%202006-2025.pdf) has over 103 references to technology. The plan emphasises the use of technology in private and public sectors for businesses as well as government. In particular, the plan calls for establishing "state of the art" technology facilities; developing an Information Economy; improving the use of "ICT" in youth education and training programmes; and,

facilitating and encouraging the acquisition, adaptation and use of Science and Technology. According to Brandford's article, "Strategies such as expanding the community development programme, the removal of [tax] duties on computers, the liberalisation of the telecommunications sector, the EduTech programme, and the setting up of kiosks will assist greatly in bridging the digital divide, …" (Brandford 2006). Yet the implementation of the plan has come at a considerable cost.

The aim of enhancing the use of technology and bridging the digital gap are some of the important objectives of the Barbados National Strategic Plan. To this end, millions of Barbadians dollars have been spent to insure that the country has state of the art technology. Government officials have lamented about the large sums of money spent compared to the outcomes. One of the teachers at the Alexander Secondary School stated that, "for three years the schoolchildren had a useless piece of hardware to tote around. I never got to teach the students with that technology and for the last three years the laptop has been at my home used as a doorstop," (Barbados Nation Newspaper 2005a). This clearly indicates a need for educators and students to learn to use technology. For instance, one of the Sports Ministers commented that, "We (Barbados) have spent millions of dollars in the Ministry of Education on the EduTech programme and I have never gone to any of the schools and seen any of that interactive material being used (Barbados Nation Newspaper 2005b)."

Recent headlines magnified the problem by stating that thus far, "The government has spent over 170 million dollars in the Edutech programme (Nation Newspaper 2007)." This is only part of the cost since the total cost allocated is about 213 million US dollars (Barbados Nation Newspaper 2005a).

Since there has been a large investment by the government in providing information technology education, the underlying assumption is that students should enter university with at least some proficiency in the use of productivity tools; namely, word processors, spreadsheets, presentation software, and databases; especially since information technology courses are offered at secondary and tertiary level institutions in Barbados. To determine whether or not this was the case, a study of the competencies of sophomore students in the use of these productivity tools was conducted.

1.2 Statement of purpose

The purpose of this survey study is to determine the students' levels of competency in four productivity tools and to compare the competencies of students in five departments.

2. Theoretical Framework

The term computer literacy seems synonymous with the terms computer competency and computer skills. Unfortunately, for decades the term computer literacy has defied a precise definition. Presently there is an ongoing debate about what is computer literacy, competency, or proficiency. One issue of these debates is whether computer literacy should encompass the social impact of computers, as well as the ethical and moral impact (Childers, 2003).

Childers' (2003) article defines computer literacy in terms of achieving a desired level of computer proficiency. In his Computer Proficiencies chart the tasks that a user must be able to do is recorded in three distinct levels, namely level 1 (baseline), level 2 (desired) and level 3 (target). The competencies required for level 1 include "know[ing] the location of the power buttons" and "know[ing] how to open and close a [web] browser." For level 2, the user's competencies include the ability "to differentiate between legitimate security threats and those that are not, such as hoaxes." In Level 3, the required competencies include "how to clear a paper jam", and "understand how the security software protects [the] computer." These actions address troubleshooting and technical adeptness.

Tsai (2002), cited in (Poynton 2005) defines computer literacy differently and considers it to be "the basic knowledge, skills, and attitudes needed by all citizens to be able to deal with computer technology in their daily life" (p. 69) This definition removes the need for the user to be familiar with the internals of the computer, its architecture and the ability to program it. Productivity tools are designed with exactly this in mind; that is to make tedious task, which in the past had to be done by software programmers, easy**.** One can only imagine creating a presentation, one screen at a time using Word Star which required the use of individual macro commands to format the text.

For the purpose of this work the definition of computer literacy is based on the ability to perform activities described in a productivity tools competencies chart. The competency levels section of this article provides a detailed description of this chart. Computer competency and proficiency refers to the students' performance at certain levels of ability. Therefore, proficiency and competency will be used interchangeably in this work.

3. Contextual Framework

The University of the West Indies is made up of three campuses, located in Jamaica (Mona Campus), Trinidad and Tobago (St Augustine Campus) and Barbados (Cave Hill Campus), and the UWI 12 which is comprised of twelve centres, one in each of the other contributing countries.

This study was set at the UWI Cave Hill campus when it had four faculties; Humanities and Education, Pure and Applied Sciences, Social Sciences, and Law. In addition the campus had a School of Clinical Medicine and Research. The student population stood at nearly 7000 with projections for 15000 or more students by 2015.

The majority of its students attended on a part-time basis, however, the Faculty of Law and the School of Clinical Medicine and Research generally had full-time students. The female population outnumbered the male population on campus in three out of the four faculties. Only Pure and Applied Sciences had more males than females taking its courses.

In this study forty-five percent of the students had an age range from 16-20; thirty-four percent ranged from 21-25; ten percent, 26-30; eight percent, 31-40 and the remainder were in the 41-50 age range. No students over 50 were included in the sample.

Barbados is the native country for most of the students with a small representation from the nearby Caribbean islands. The United States of America represented one and a third percent of the international student population with Canada at one-half percent, and both Britain and Kenya at one-third percent.

Several computer laboratories were available to students at the university; these computers typically ran Microsoft Windows 2000 or XP and the Microsoft Office Suite 2003.

4. Methodology

This survey study utilized a questionnaire. Thirteen (13) faculty members from five (5) departments agreed to have the survey administered in their sophomore class.

In the survey, students were asked to select from a list, which activities they were capable of performing.

4.1 Participants

Two hundred and fifty-seven (257) sophomore students agreed to participate in this study; however four of these surveys were unusable because pertinent information related to this study was missing. Five departments participated in administering the surveys. Twenty students were drawn from the Department of History and Philosophy (Hist. & Phil.); fifty-one from Language, Linguistics and Literature (LLL); Sixty-two from Computer Science (Comp. Sci.); Eight-seven from Economics; and thirty-three from Government.

4.2 Data collection procedures

Prior to the main data collection phase, a pilot study was conducted on twenty five (25) students to test the suitability, reliability and validity of the survey items and measures. Participants in the pilot study were conveniently selected since the emphasis was on instrument testing and design.

In the main data collection phase, students were informed about the nature and purpose of the study, and were assured of the confidentiality and anonymity of the research. The questionnaire was a closed response survey self-reported by students. The questionnaires were distributed and collected before or after lecture sessions during the second week of the first semester of the 2006-2007 school year.

4.3 Survey instrument

The survey instrument created by the Scholarship of Teaching through Action Research on Teaching (START) group investigated the demographic questions of gender, country of origin, language and age group. In addition, information about student's academic experience, current course of study, total years of study and whether part-time or full-time, was also collected. To gain an insight into the computer literacy skills of students, questions focusing on computer experience, Internet access and technical skills were included. This closed response survey allowed students to indicate the activities that they were able to perform.

A basic set of skills (for example, the use of the keyboard and mouse) was required to use each of the productivity tools. The tools were arranged in order of increasing difficulty; word processing tools were considered the easiest, with spreadsheets, PowerPoint, and database tools being increasingly more difficult. This ordering was based on the requisite knowledge required by the student to use the tools properly. For each tool, the associated activities were also listed in increasing order of difficulty; this ordering was determined by the university's computer science faculty members. Again, each activity's prerequisite knowledge was considered in the general ordering.

The prerequisite knowledge for word processing is based on the reading and writing skills gained throughout primary and secondary education. For example, spelling and grammar checking are fundamental concepts taught throughout pre-tertiary level education. Therefore, students only need to understand how to perform these tasks using the word processor.

Similarly, spreadsheets require basic knowledge of mathematics. At the elementary level, spreadsheets are data entry tools that utilize arithmetic functions, such as addition and subtraction. Spreadsheets would be considered harder to use than word processors since some concepts would be new to most students; for example, the concept of a cell, the relationships between cells and the dragging and dropping of cell contents. Also, creating formulae requires a basic understanding of the concepts of fields, rows, and columns since such formulae make direct references to them.

Unlike overhead projectors, which are used to display static information, PowerPoint provides dynamic overhead presentations. A certain skill level is required to create templates, headers and footers and slide transitions. A thorough understanding of the concept of the master slide and its global effect on all slides is also required. The inclusion of animation employs higher order application skills, such as, order, behaviour, timelines, auto-timing and, the manipulation of sound and video files.

Database skills are considered the most difficult because they require an in-depth understanding of computer science concepts such as tables, fields and records. These concepts are from the field of relational database theory developed by E. F. Codd (1990). As discussed in Date (2000), database systems can be described using a three level model: internal (how the data is stored); external (how the user views the data); and, conceptual (how to represent the contents of the database). The difficulty of constructing a database is increased by the need to create relationships, rows, columns, tables, forms and reports. For example, in order to formulate queries one must understand the relationships between fields in a table and the relationships between tables.

4.4 Competency levels

After the survey instrument was devised, a productivity tools competencies chart was formulated which categorized the productivity tools activities into four levels. These levels are basic, elementary, intermediate, and advanced (Figure 1):

- *Basic Skills*: The user has acquired the minimum set of skills required to use the tool but is still uncomfortable using it.
- *Elementary*: The user is somewhat comfortable with the tool.
- *Intermediate*: The user is comfortable with the tool and can use some of its advanced features.
- *Advanced*: The user is considered a power user; that is, the user is fully comfortable with the tool and is able to use its advanced features.

Level	Word Processor	Spreadsheet	PowerPoint	Database
Basic	Cutting and pasting	Creating fields Inserting / deleting rows or columns	Toggling between views Saving file as PowerPoint show	Creating tables Creating fields Adding records
Elementary	Spell checking Using thesaurus Setting language	Sorting	Changing templates	
Intermediate	Creating tables Creating bulleted and numbered lists Changing bullet symbols	Creating charts / graphs	Transferring or downloading templates from external sources	Creating forms Sorting
Advanced	Setting alignments/tabs Creating templates	Creating formulae Creating macros	Adjusting the timing of slides for presentations Adding animation Adding sound and/or video	Performing queries

Figure 1. Productivity tools competencies chart

For each technical skill, the steps required to perform the relevant activity determined the level in the productivity tools competencies chart; the ordering of these activities are justified in the following sections.

1) Word Processing

- *Cutting and pasting*: This is a basic activity.
- *Spelling and Thesaurus*: This requires the ability to select the appropriate language from a list.
- *Creating tables and bulleted lists*: In order to create a table, concepts such as rows and columns, table headings and formats need to be understood. This is more difficult than selecting a language. In addition, changing bullet symbols requires image identification and possibly manipulation.
- *Setting Alignment/Tabs*: This skill has been given a high level of difficulty because setting the different levels of alignment does not always produce the expected result and often requires troubleshooting.

2) Spreadsheets

- *Creating fields and inserting/deleting rows or columns*: The first task a user must perform is the entering of data. This requires the creation of fields and often the insertion and deletion of rows or columns.
- *Sorting*: The actions required for sorting a range of data requires the user to select the range, determine the sort order (ascending/descending) and determine the primary (and if necessary, the secondary) sort columns.
- *Creating Graphs*: This requires more steps than sorting and hence appears at a higher level.
- *Creating Formulae and Macros*: These two skills have both been grouped at level 4 since they both can contain programming constructs.

3) PowerPoint

- *Saving shows*: This is a basic operation.
- *Changing templates*: The user has to select the template from a list.
- *Downloading templates:* The user must first find the template (whether online or on an internal or external media), then must copy the file into the

appropriate folder for use.

- *Adjusting timing, Adding animation, Sound and video*: All of these skills require similar levels of competency and are generally used together in the formulation of a presentation.

4) Databases

- *Creating tables, fields, and record*: Before a database can be used, a user must create at least one table then add fields and records.
- *Creating Forms*: The creation of forms requires a similar set of skills as creating graphs and charts using spreadsheet software and so both appear at the same level.
- *Sorting*: Sorting within databases requires a higher skill level than spreadsheet sorting due to various advanced concepts, such as the use of primary keys and indexes.
- *Performing queries*: This can be a very complex task especially when tables must be joined in order to get the desired results.

4.5 Data Analysis

In this study, the independent variables were sophomore students and academic departments. The dependent variables were productivity tool activities and levels of competency. The main statistical tools used included percentages, Pearson Chi-square tests and a one-way ANOVA; Alpha was set at less than .05 *($p < .05$)*.

The percentages and Pearson Chi-square tests were computed to determine:

1. The percentage of students in each department who could perform a particular activity.
2. Whether any significant differences existed between students, who performed the given activities, in the given departments.

A one-way ANOVA was computed to determine:

1. The average number of basic, elementary, intermediate and advanced activities which students from each department could perform.
2. Whether any significant differences existed between students in the given departments, who performed these activities.

5. Results

In the following sections answers to the following questions are provided:

1. Which activities, per productivity tools, were most students able to perform and least able to perform across and within the given departments (Tables 1 - 4)?
2. What is the mean number of activities students could perform for each productivity tools for each department (Table 5)?
3. What is the mean number of activities students could perform across the four competency levels for each department (Table 6)?
4. For each productivity tool, were there significant differences in the perceived abilities of students to perform the activities across the given departments (Table 7)?
5. Across the competency levels, were there significant differences in the perceived abilities of students to perform the activities across the given departments (Table 8)?

5.1 Activities that most/least students were able to perform

The activities that students could perform when using word processors, spreadsheets, PowerPoint and databases are presented below (Tables 1, 2, 3 and 4, respectively).

1) Word Processing

When examining the departments collectively, the activity that most students could perform was spell checking in a word processor; notably, 100% of the students from the History and philosophy, Computer Science and Government departments indicated that they could perform this activity. Conversely, the activity that least students could perform was creating templates; only 69% of Computer Science students and 46% of Government students were able to perform this activity (Table 1).

There was a general downwards trend of students being able to perform the specified activities within the five departments; from cutting and pasting to creating templates (Table 1). Some variations did occur however, at the spell checking and the creating bulleted and numbered lists activities. A trend was not clearly seen in the Government, or History, and Philosophy departments.

The Pearson Chi-square values (Table 1) show that there were significant differences in the percentage of students being able to perform the creating tables, changing bullet symbols, setting alignments/tabs and creating templates activities across the departments.

Table 1. The word processing activities students indicated that they could perform

Activity	Departments					Chi-Square χ^2
	Hist & Phil	LLL	Comp. Sci.	Economics	Government	
Cutting and pasting	90%	94%	98%	97%	94%	3.34
Spell checking	100%	98%	100%	95%	100%	5.41
Using thesaurus	95%	88%	95%	92%	85%	3.89
Setting language	85%	86%	90%	90%	73%	6.95
Creating tables	90%	80%	97%	91%	82%	9.68*
Creating bulleted and numbered lists	90%	88%	98%	98%	91%	8.92
Changing bullet symbols	90%	75%	98%	90%	82%	16.31*
Setting alignments/tabs	85%	73%	98%	87%	76%	17.85*
Creating templates	65%	47%	69%	67%	46%	10.57*
N=253						

Note: * $p<0.05$

2) Spreadsheets

For all of the departments, the spreadsheet activity that most students could perform was the inserting/deleting rows or columns activity; the percentages ranged from 73% for Government students to 95% for Computer Science students. Conversely, the activity that the least number of students could perform, by a significant margin, was creating macros; the percentage of students ranged from 5% for History and Philosophy to 27% for Computer Science (Table 2).

Science who had a 2% increase. This drop was as large as 35% for the History and Philosophy department.

There was also a large percentage drop in the percentage of students being able to perform the creating formulae and creating macros activities. This drop was highest for Computer Science at 59%. Only five percent of the students in the History and Philosophy department; eight percent in the Language, Linguistics, and Literature department; and 12 percent in the Government department were able to create macros.

Table 2. The spreadsheet activities students indicated that they could perform

Activity	Departments					Chi-Square χ^2
	Hist. & Phil	LLL	Comp.Sci.	Economics	Government	
Creating fields	60%	51%	90%	68%	55%	24.14*
Inserting/deleting rows or columns	70%	82%	95%	90%	73%	14.72*
Sorting	55%	37%	89%	79%	55%	43.78*
Creating charts/graphs	65%	69%	84%	89%	67%	14.13*
Creating formulae	30%	43%	86%	76%	42%	43.86*
Creating macros	5%	8%	27%	24%	12%	12.22*
N=253						

Note: * $p<0.05$

A clearly identifiable downward trend of students being able to perform specific activities within the Computer Science Department is present (Table 2); a spike did exist, however at the inserting/deleting rows or columns activity. Trends were not as easily identifiable within the other departments.

There was a large drop in the percentage of students being able to perform the creating charts/graphs and the creating formulae activities in all of the departments except Computer

Pearson Chi-square tests indicated that there were significant differences across departments for all of the spreadsheet activities.

3) PowerPoint

When examining the departments collectively, the PowerPoint activity that most students could perform was saving file as PowerPoint show; the percentages of students being able to perform this activity ranged from 49% for

Language, Linguistics and Literature to 78% for Economics (Table 3).

The activity that least students could perform was transferring or downloading templates from an external source; the percentages of students being able to perform this activity ranged from 12% for Language, Linguistics and Literature to 55% for Computer Science.

There were no recognizable trends in students' abilities to complete PowerPoint activities.

The Pearson Chi-Square values (Table 3) indicate that there are significant differences in the abilities of students across the departments for toggling between views; saving file as PowerPoint show; changing templates; and transferring or downloading templates from an external source.

In general there was a downward trend of students being able to perform Database activities across departments. Some variations did occur, however for the activities adding records and sorting.

Chi-square values show significant differences for each activity related to database use across the departments (Table 4).

5.2 Mean number of activities students could perform across the productivity tools

Mean values were used to compare each department across the four productivity tools. Computer Science and Economics students ranked first and second across all productivity tools. Government students were consistently ranked third with the exception of Word Processing activities and Language, Linguistics, and Literature consistently ranked fifth (Table 5).

Table 3. The PowerPoint activities students indicated that they could perform

Activity	Departments					Chi-Square χ^2
	Hist. & Phil	LLL	Comp.Sci.	Economics	Government	
Toggling between views	50%	28%	76%	61%	55%	27.93*
Saving file as PowerPoint show	75%	49%	76%	78%	73%	19.95*
Changing templates	40%	35%	61%	61%	55%	11.71*
Transferring or downloading templates from external sources	45%	12%	55%	51%	46%	26.03*
Adjusting timing of slides for presentations	45%	45%	61%	62%	61%	5.67
Adding animation	50%	45%	63%	67%	64%	7.53
Adding sound or video	45%	37%	61%	55%	55%	7.42
N=253						

Note: * $p<0.05$

4) Database

The database activity that most students could perform, across all the departments, was creating tables; 77% of Computer Science students could perform this activity while only 43% of Language, Linguistics and Literature students could perform this activity. Conversely, the activity that the least number of students could perform was performing queries (Table 4).

5.3 Mean number of activities students could perform across the competency levels

In terms of competency levels, (Table 6) Computer Science and Economics students ranked first and second, respectively. History and Philosophy students ranked third except for the Advanced level; and Language, Linguistics, and Literature students ranked fifth.

Table 4. The database activities students indicated that they could perform

Activity	Departments					χ^2
	Hist. & Phil	LLL	Comp. Sci.	Economics	Government	
Creating tables	55%	43%	77%	68%	55%	16.05*
Creating fields	35%	31%	79%	64%	46%	32.88*
Adding records	50%	37%	79%	62%	36%	26.82*
Creating forms	35%	26%	68%	56%	33%	26.63*
Sorting	25%	26%	69%	55%	36%	29.51*
Performing queries	20%	26%	68%	53%	33%	29.68*
N=253						

Note: * $p<0.05$

Table 5. Mean number of activities across the four productivity tools by department

Department	Word Processing/9		Spreadsheet/6		PowerPoint/7		Database/6	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Hist. & Phil.	7.90 [3]	1.83	2.85 [5]	1.95	3.50 [4]	2.67	2.20 [4]	2.26
LLL	7.26 [5]	2.06	2.90 [4]	1.85	2.51 [5]	2.70	1.88 [5]	2.32
Comp. Sci.	8.45 [1]	1.18	4.71 [1]	1.43	4.53 [1]	2.80	4.43 [1]	2.42
Economics	8.06 [2]	1.82	4.25 [2]	1.75	4.34 [2]	2.74	3.59 [2]	2.58
Government	7.27 [4]	2.10	3.03 [3]	2.28	4.06 [3]	2.86	2.39 [3]	2.60
N=253								

Note: The numbers after each forward slash (/) indicate the total number of activities for a given productivity tool. Numbers in square brackets indicate the rankings of departments for a given productivity tool.

5.4 Significance across departments

One-way ANOVA rendered significant differences between and within groups for each productivity tool. Therefore, follow-up tests were made to compare the means of computer science students to students in each of the other departments. The results showed the alpha values were significantly different between departments for each productivity tool. Scrutinizing the comparisons between departments pinpointed areas which were significantly different.

Word processing activities were significantly different across the departments (*F=4.150, df 4, p=.003).* Post hoc comparisons show that Language, Linguistics, and Literature and Government students were significantly lower than Computer Science students; Language, Linguistics, and Literature *p=.001* and Government *p=.002.*

Spreadsheet activities were significantly different *(F=11.462, df 4, p=<.001)* for History and Philosophy, Language, Linguistics, and Literature, and Government. Post hoc comparisons with Computer Science students show History and Philosophy *p=<.001,* Language, Linguistics, and Literature *p=<.001,* and Government *p=<.001.*

PowerPoint activities showed a significant difference *(F=4.745, df 4, p=.001).* Language, Linguistics, and Literature was significantly lower than Computer Science students; *p=<.001.*

Database activities were significantly different *(F=9.615, df 4, p=<.001).* History and Philosophy, Language, Linguistics, and Literature, Economics, Government were significantly lower than Computer Science students; History and Philosophy *p=.001,* Language, Linguistics, and Literature *p=<.001,* Economics *p=.043,* and Government *p=<.001.*

There were significant differences across the four productivity tools (Table 7) and competency levels (Table 8).

Table 6. Mean number of activities across the four competency levels by department

Department	Basic/8		Elementary/5		Intermediate/7		Advanced/8	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Hist. & Phil.	4.85 [3]	2.41	4.10 [3]	0.97	4.40 [3]	1.67	3.45 [4]	2.24
LLL	4.16 [5]	2.33	3.59 [5]	1.17	3.75 [5]	1.86	3.24 [5]	2.26
Comp. Sci.	6.66 [1]	2.06	4.50 [1]	0.82	5.69 [1]	1.53	5.34 [1]	2.30
Economics	5.87 [2]	2.31	4.34 [2]	1.16	5.29 [2]	1.70	4.91 [2]	2.59
Government	4.85 [3]	2.75	3.85 [4]	1.25	4.36 [4]	2.23	3.88 [3]	2.65
N=253								

Note: The numbers after each forward slash (/) indicate the total number of activities for each competency level. Numbers in square brackets indicate the rankings of departments for each competency level

6. Discussion

It is not surprising that Computer Science students had the highest competencies in all of the productivity tools since: students are required to submit typewritten assignments; some of the students complete courses in database technologies; and in some courses students are required to give presentations, either in groups or individually. It is unclear; however, why students' spreadsheet competencies are high since the use of spreadsheets is not an integral part of the current computer science curriculum.

Computer Science and Economics students' competencies were comparable; but, higher than those of the student in other departments. Economics students unlike students from the various other departments were required to take Introduction to Computers.

Table 7. Significant differences across the four productivity tools

Productivity Tools	Word Processing	Spreadsheets	PowerPoint	Database
F-Value	4.15; p=0.001	11.46; p<0.001	4.75; p<0.001	9.62; p<0.001
N=253				

This course introduced basic concepts of computer evolution and classification as well as computer hardware and software. Additionally, students had the opportunity to learn productivity tools' competencies in the course Microcomputer Applications which teaches modelling with spreadsheets; human communication with word processor; and database development and management.

Although a large percentage of Language, Linguistics, and Literature students were able to perform word processing activities, less than half of them could perform PowerPoint, spreadsheets, and database activities. Generally, Language, Linguistics, and Literature students ranked lower than the other departments. We interviewed several lecturers in Language, Linguistics, and Literature and found that most lecturers do not require students to use productivity tools other than word processors. In some cases, hand-written assigned are accepted.

Table 8. Significant differences across the four competency levels

Competency levels	Basic	Elementary	Intermediate	Advanced
F-Value	9.72; p<0.001	6.34; p<0.001	10.71; p<0.001	7.24; p<0.001
N=253				

As far as specific tools, spreadsheets and PowerPoint tools revealed a curious trend. A low percentage of students were able to create macros in spreadsheets. Even computer science students were below expectations in their self-reported ability to create macros. This is probably a result of the curriculum not including the creation of macros.

The curious trend in using the PowerPoint tool is that only 12 percent of Language, Linguistics, and Literature students were able to perform the transferring or downloading templates from external sources activity. Integrating technology may be suggested by the lecturer for group presentations but is not a requirement in students' coursework.

While competencies were divided into four different skill levels, students' ability may vary across activities within the same tool. Interestingly enough, it was found that a large number of students could perform most of the elementary level skills without reporting they had competencies in basic level skills. For the most part, many of the students were unable to perform basic database skills. The use of tools such as Microsoft Access provides user-friendly interfaces which allow students to query predefine databases which may be one reason for students' lack of basic skills.

Other than Computer Science and Economics students, the competency of students was below average at the advanced level. Students from all the departments were average or above average at the intermediate level.

7. Conclusion

In this study, we examined sophomore students' abilities to perform various activities of four productivity tools. The results show that Computer Science students had the highest competency levels in all of the tools and were closely followed by Economics students. Language, Linguistics, and Literature students were proficient in the use of word processors, but were unable to perform many of the activities of the other tools.

The deficit of opportunities to learn to proficiently use productivity tools is most prominent with Language, Linguistics, and Literature students. This may be due to the nature of the discipline; the type of assignments; handwritten coursework, for example workbooks, or the lack of rubrics for assessing the use of technology.

The findings suggest that faculty members are not requiring students to use technology in order to complete their course work. Experience of working with technology is not included in the core curriculum but hidden within specific departmental curriculum. The findings indicate that not all students are equally tech-savvy due in part to the lack of integration of technology into the curriculum. This study clearly does not support the supposition that most young people are tech-savvy.

Some students are at-risk due to their inability to compete in a technological environment either at university or in the workplace. This lack of skills may contribute negatively to employers' perception of students' abilities.

8. Recommendations

As a result of these findings, the authors recommend that:

1. The creation of spreadsheet macros be included in an information systems course for computer science students.
2. Students in all departments should be required to create and deliver multimedia presentations (including animation, sound and video).

3. Faculty members need to expose their full expectations of students' use of technology in their courses and give students credit for fulfilling those expectations.
4. Faculty members need to open a dialog within and across departments to discuss and establish levels of technology competencies desired of students.
5. A core curriculum should be created that includes courses that teach the use of productivity tools.

References

Barbados Nation Newspaper. 2005a. Perry: Waste in EduTech, September 11, http://bararchive.bits.baseview.com/archive_detail.php?archiveFile=./pubfiles/bar/archive/2005/September/11/Politics/8204.xml&start=0&numPer=20&keyword=EduTech&sectionSearch=&begindate=9%2F1%2F2005&enddate=9%2F30%2F2005&authorSearch=&IncludeStories=1&pubsection=&page=&IncludePages=1&IncludeImages=1&mode=allwords&archive_pubname=Daily+Nation%0A%09%09%09 (accessed November 21, 2007).

Barbados Nation Newspaper. 2005b. EduTech tools 'going to waste', December 11, http://bararchive.bits.baseview.com/archive_detail.php?archiveFile=./pubfiles/bar/archive/2005/December/11/LocalNews/12563.xml&start=0&numPer=20&keyword=EduTech&sectionSearch=&begindate=12%2F1%2F2005&enddate=12%2F30%2F2005&authorSearch=&IncludeStories=1&pubsection=&page=&IncludePages=1&IncludeImages=1&mode=allwords&archive_pubname=Daily+Nation%0A%09%09%09 (accessed November 21, 2007).

Barbados Nation Newspaper. 2007. $170m spent on EduTech, May 7, http://bararchive.bits.baseview.com/archive_detail.php?archiveFile=./pubfiles/bar/archive/2007/May/07/LocalNews/37859.xml&start=0&numPer=20&keyword=EduTech&sectionSearch=&begindate=5%2F1%2F2007&enddate=5%2F30%2F2007&authorSearch=&IncludeStories=1&pubsection=&page=&IncludePages=1&IncludeImages=1&mode=allwords&archive_pubname=Daily+Nation%0A%09%09%09 (accessed November 21, 2007).

Brandford, Albert. 2006. Plan to bridge digital divide. *Barbados Nation Newspaper*, July 25, http://bararchive.bits.baseview.com/archive_detail.php?archiveFile=./pubfiles/bar/archive/2006/July/25/Business/23450.xml&start=0&numPer=20&keyword=EduTech&sectionSearch=&begindate=7%2F1%2F2006&enddate=7%2F31%2F2006&authorSearch=&IncludeStories=1&pubsection=&page=&IncludePages=1&IncludeImages=1&mode=allwords&archive_pubname=Daily+Nation%0A%09%09%09 (accessed November 21, 2007).

Childers, S. 2003. Computer literacy: necessity or buzzword? *Information Technology Libraries* 22, no. 3: 100-4.

Codd, E.F. 1990. *The relational model for database management: version 2*. Reading Mass.: Addison-Wesley.

Date, C.J. 2000. *An introduction to Database Systems*. United States: Addison Wesley.

Gay, Glenda, Sonia Mahon, Dwayne Devonish, Philmore Alleyne and Peter G. Alleyne. 2005. Perceptions of information and communication technology among undergraduate management students in Barbados. *International Journal of Education and Development using information and Communication Technology* 2, no. 4: 6-17.

Grant, J. M. A., Walcott, P. A.. 2009. Sizing-up Students' Competencies for Successful Blended Learning. *Higher Education in the Anglophone Caribbean, Educación superior y sociedad*, nueva época, año 14, Numero 2, 159-170.

Haywood, J., D. Haywood, H. Macleod, R. Baggetun, , A.P. Baldry, E. Harskamp, J. Teira, P. Tenhonen, Abo Akademi. 2004. A comparison of ICT skills and students across Europe. *Journal of eLiteracy* 1: 69-81.

Kaminski, Karen, Pete Seel and Kevin Cullen. 2003. Technology literate students? Results from a survey. *Educause Quarterly*, no. 3: 34-40.

Keller, G. 2005. The new demand for heterogeneity in college teaching. In *To Improve the Academy: Resources for Faculty, Instructional and Organizational Development*, Ed. Sandra Chadwick-Blossey and Douglas Reimondo Robertson, 62-9. Bolton Massachusetts: Anker Publishing Company.

Kvavik, Robert B., Judith B. Caruso. 2005. ECAR study of students and information technology, 2005: Convenience, connection, control and learning. *ECAR Research Study* 6 (September): 1-140.

Oblinger, D.G. and B.L. Hawkins. 2006. The myth about student competency: our students are technologically competent. *Educause review* (March/April): 12-3.

Poynton, T.A. 2005. Computer literacy across the lifespan: a review with implications for educators. *Computers in Human Behavior* 21: 861-72

Prensky, Marc. 2001a. Digital Natives, Digital Immigrants. *On the Horizon* 9, no. 5: 1-6.

Prensky, Marc. 2001b. Digital Natives, Digital Immigrants, Part II: Do they really think differently? *On the Horizon* 9, no. 6: 1-6.

Walcott, P. A., Grant, J. M. A., Depradine, C., "Are Students Ready for a Virtual University?", in Proceedings of ED-MEDIA World Conference on Educational Multimedia, Hypermedia & Telecommunications, Vienna, Austria, June 30-July 4, 2008, pp. 946-951.

Statistical Implementation of Segmentation of Dermoscopy Images using Multistep Region Growing

S.Zulaikha Beevi [1], M.Mohammed Sathik [2], K.Senthamaraikannan [3]

[1] Assistant Professor, Department of IT, National College of Engineering, Tamilnadu, India.

[2] Associate Professor, Department of Computer Science, Sathakathullah Appa College,Tamilndu, India.

[3] Professor & Head, Department of Statistics, Manonmaniam Sundaranar University, Tamilnadu, India.

***Abstract*-**A method for segmentation of skin cancer images, firstly the algorithm automatically determines the compounding colors of the lesion, and builds a number of distance images equal to the number of main colors of the lesion (reference colors). These images represent the similarity between reference colors and the other colors present in the image and they are built computing the CIEDE2000 distance in the L*a*b* color space. Texture information is also taken into account extracting the energy of some statistical moments of the L* component of the image. The method has an adaptative, N-dimensional structure where N is the number of reference colors. The segmentation is performed by a texture-controlled multi-step region growing process. The growth tolerance parameter changes with step size and depends on the variance on each distance image for the actual grown region. Contrast is also introduced to decide the optimum value of the tolerance parameter, choosing the one which provides the region with the highest mean contrast in relation to the background.

***Keywords*-** region growing, segmentation, texture-controlled

1. Introduction

Skin cancers are one of the most common forms of cancers in humans. Skin cancers can be classified into melanoma and non-melanoma. Although melanomas are much less common than non-melanomas, they account for most of the mortality from skin cancers [1].Automatic image segmentation applied to the detection of this kind of lesions could result in the detection of the disease at an early stage and a subsequently increment in the likelihood that the patient will survive. Although color is the most important information in this kind of images, it is not the only source of knowledge available. Because these types of pigmented lesions are rich in color and texture, those segmentation processes that take into account only color information will probably fail in giving us a proper result. That is why texture information must be included.

2. Existing Algorithm Description

a. Selection Box

In the first step the user selects an area (selection box) from the image with the mouse. The algorithm will segment all pixels in the image with colors and textures similar to the ones present in this selected region.

b. Color Information

Wavelet denoising

The color image is converted to grey level image and stationary wavelet is applied to decrease noise We use Stationary wavelet Transform (SWT) for decreasing noise in image. For this, we used r.r.coifman et al [7] work. Their algorithm is as follow: Image transformed to wavelet coefficients. Soft or hard thresholding is applied to detail coefficients. Therefore, coefficients smaller than threshold are eliminated. At last, inverse Stationary wavelet transform is applied to approximation and detail coefficients.

$L^*a^*b^*$ color space

In a further step the color of each pixel will be substituted by its distance to the reference colors.Therefore, a perceptually uniform color space is needed so that distances between colors measured in this space are correlated with color differences according to human perception. We have chosen the L*a*b* color space [2].

c. Reference Colors

We consider all the colors present in the selection box, and we will call them reference colors. In order to find these reference colors, we perform a clustering operation with the well known FCM algorithm in the

L*a*b* color space. To obtain the value k (number of clusters) automatically, we use Dunn's coefficient [3]

$$D = \min_{1<i<j} \left\{ \min_{1<j<k} \left\{ \frac{d(ci,cj)}{\max_{1<n<k}(d'(cn))} \right\} \right\} \quad (1)$$

where d(ci,cj) is the distance between cluster I and cluster j, that is, the inter-cluster distance. d'(cn) is the intra-cluster distance. We assume that, in skin lesions, the number of different colors is less or equal to 16. So, we perform 16 clusterings beginning from k=1 to k=16. Each time, the Dunn's coefficient (1) is computed and stored. When the last clustering is done, the obtained D coefficients are compared, selecting the k value that provides the highest value of D, which leads to a maximum inter-cluster distance and a minimum intra-cluster distance. Then, the k reference colors are defined as the centroids of the k clusters in L*a*b* color space. Figure. 1 shows an example.

d. Distance Images

Once the reference colors are obtained, the distances between every single pixel of the image and each of the reference colors are calculated. We have chosen the distance metric. This measure has been extensively tested and outperformed other existing color difference formulae [4]. As a consequence, the CIE 2000 is color difference formula. Then, a new set of images is built, where each pixel value will be the color difference to each of the reference colors. In order to obtain a better visualization, we invert this image, that is, those pixels whose values are similar to the reference ones, will appear light in a dark background. These inverted images are called the distance images. We can see an example in Figure 1.

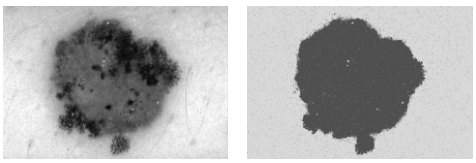


Figure 1: Two distance images obtained with CIEDE2000color distance formulae, (a). The original image ,(b). the two reference colors using FCM.

3. Proposed Approach

a. Conventional FCM

Clustering is the process of finding groups in unlabeled dataset based on a similarity measure between the data patterns (elements) [12]. A cluster contains similar patterns placed together. The fuzzy clustering technique generates fuzzy partitions of the data instead of hard partitions. Therefore, data patterns may belong to several clusters, having different membership values with different clusters. The membership value of a data pattern to a cluster denotes similarity between the given data pattern to the cluster. Given a set of n data patterns, X = $x_{1,\ldots},x_k,\ldots,x_n$, the fuzzy clustering technique minimizes the objective function, O(U,C):

$$O_{fcm}(U,C) = \sum_{k=1}^{n}\sum_{i=1}^{v}(u_{ik})^m d^2(x_k, c_i) \quad (2)$$

where x_k is the k-th D-dimensional data vector, c_i is the center of cluster i, u_{ik} is the degree of membership of x_k in the i-th cluster, m is the weighting exponent, d (x_k, c_i) is the distance between data x_k and cluster center c_i, n is the number of data patterns, v is the number of clusters. The minimization of objective function J (U, C) can be brought by an iterative process in which updating of degree of membership u_{ik} and the cluster centers are done for each iteration.

$$u_{ik} = \frac{1}{\sum_{J=1}^{V}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{1}{m-1}}} \quad (2)$$

$$c_i = \frac{\sum_{k=1}^{n}(u_{ik})^m x_k}{\sum_{k=1}^{n}(u_{ik})^m} \quad (3)$$

where $\forall i\ u_{ik}$ satisfies: $u_{ik} \in [0,1]$, $\forall k \sum_{i=1}^{v} u_{ik} = 1$ and $0 < \sum_{k=1}^{n} u_{ik} < n$

Thus the conventional clustering technique clusters an image data only with the intensity values but it does not use the spatial information of the given image.

b. Initialization

The theory of Markov random field says that pixels in the image mostly belong to the same cluster as their neighbors. The incorporation of spatial information[8][9] in the clustering process makes the

algorithm robust to noise and blurred edges. But when using spatial information in the clustering optimization function may converge in local minima, so to avoid this problem the fuzzy spatial c means algorithm is initialized with the Histogram based fuzzy c-means algorithm. The optimization function for histogram based fuzzy clustering is given in the equation 5

$$O_{hfcm}(U, C) = \sum_{l=1}^{L}\sum_{i=1}^{v}(u_{il})^{m} H(l)\, d^{2}(l, c_i) \quad (5)$$

where H is the histogram of the image of L-gray levels. Gray level of all the pixels in the image lies in the new discrete set G= {0,1,…,L-1}. The computation of membership degrees of H(*l*) pixels is reduced to that of only one pixel with *l* as grey level value. The member ship function u_{il} and center for histogram based fuzzy c-means clustering can be calculated as.

$$u_{il} = \frac{1}{\sum_{J=1}^{V}\left(\frac{d_{li}}{d_{lj}}\right)^{\frac{1}{m-1}}} \quad (6)$$

$$c_i = \frac{\sum_{l=1}^{L}(u_{il})^{m} H(l) l}{\sum_{l=1}^{L}(u_{il})^{m}} \quad (7)$$

Where d_{li} is the distance between the center i and the gray level l.

c. *Proposed ISFCM*

The histogram based FCM algorithm converges quickly since it clusters the histogram instead of the whole image. The center and member ship values of all the pixels are given as input to the fuzzy spatial c-means algorithm. The main goal of the FSCM is to use the spatial information to decide the class of a pixel in the image.

The objective function of the proposed ISFCM is given by

$$O_{fcm}(U, C) = \sum_{k=1}^{n}\sum_{i=1}^{v}(u_{ik}^{s})^{m} d^{2}(x_k, c_i) \quad (8)$$

$$u_{ik}^{s} = \frac{P_{ik}}{\left(\sum_{J=1}^{V}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{1}{m-1}}\right)\left(N_k\sum_{z=1}^{N_k}\left(\sum_{J=1}^{V}\left(\frac{d_{iz}}{d_{jz}}\right)^{\frac{1}{m-1}}\right)\right)} \quad (9)$$

Two kinds of spatial information are incorporated in the member ship function of conventional FCM. Apriori probability and Fuzzy spatial information

Apriori probability: This parameter assigns a noise pixel to one of the clusters to which its neighborhood pixels belong. The noise pixel is included in the cluster whose members are majority in the pixels neighborhood.

Fuzzy spatial information: In the equation (9) the second term in the denominator is the average of fuzzy membership of the neighborhood pixel to a cluster. Thus a pixel gets higher membership value when their neighborhood pixels have high membership value with the corresponding cluster. The figure 2 shows the distance images using ISFCM, our proposed algorithm.

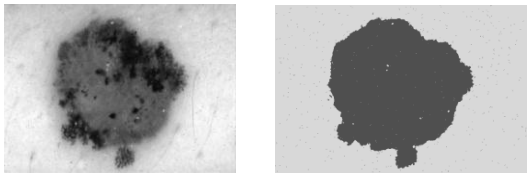


Figure 2: Two distance images obtained with CIEDE2000color distance formulae, (a). The original image, (b). The two reference colors using ISFCM.

d. *Texture Information*

The proposed method extracts texture features only from the luminance component (L*) of the original image. These features are based on some local low statistical moments [5]. The algorithm calculates for every pixel, four statistical moments mpq with p,q={0,1} by processing the L* component with local masks expressed in a normalized coordinate system. A formal expression of these moments is shown in equation (2).

$$m_{pq} = 1 - \frac{1}{w^2} \sum_{m=i-w/2}^{i+w/2} \sum_{n=j-w/2}^{j+w/2} f(m,n)\ x_m^p\ y_n^q$$

$$x_m = \frac{m-i}{w/2},\quad y_n = \frac{n-j}{w/2} \qquad (10)$$

i,j ∈ image,p,q=0,1

where W is the window width, (i,j) are the pixel coordinates for which the moments are computed, (m,n) the coordinates of another pixel which falls within the window, (xm yn) are the normalized coordinates for (m,n), and f(m,n) is the value of the L* component at the pixel with coordinates (m,n). This normalized expression leads us to compare among pixel moments and it is equivalent to the finite convolution of the image with a mask. The sizes of these masks have been fixed to the size of the selection box. Usually, for each segmentation this size will be different, so CTREG will be automatically adapted to the texture we want to isolate. With all these parameters, we can build four new images Mpq with p,q={0,1} corresponding to each statistical parameter. To this purpose we assign to each pixel a value equal to the previously calculated moment mpq. For example, in the case of pixel (30, 20) if we want to build the image M11 we define the value at position (30,20) as the moment m11, calculated with a window centered in that pixel. Afterwards, we defined new images calculated from the energy of the moments. We called them energy images,E00, E01, E10 and E11, and they represent the strength of each moment around the pixel location. The computation of the energies follows equation (3).

$$E_{pq}(i,j) = 1 - \frac{1}{w^2} \sum_{m=i-w/2}^{i+w/2} \sum_{n=j-w/2}^{j+w/2} m_{pq}^2\,(m,n) \qquad (11)$$

where $E_{pq}(i,j)$ is the energy corresponding to the pixel with coordinates (i,j) in the image M_{pq}, W is the window width, $M_{pq}(m,n)$ is the value of the pixel with coordinates (m,n) in the moment image M_{pq} and p,q={0,1} Each pixel is now characterized with four values, one from each energy image. They are considered as coordinates in a four-dimensional space. Subsequently, in order to assign each pixel to one texture in the image, we apply the same clustering procedure previously described in section 2.2.3 but in this four-dimensional texture space. We again assume that, the number of different textures is less or equal to 16. Once each pixel in the image has been classified, we select only those pixels whose texture is equal to the desired one, obtaining a black and white image in which white pixels are those with the desired texture, as shown in Figure 3. This image will be used afterwards in the region-growing process.

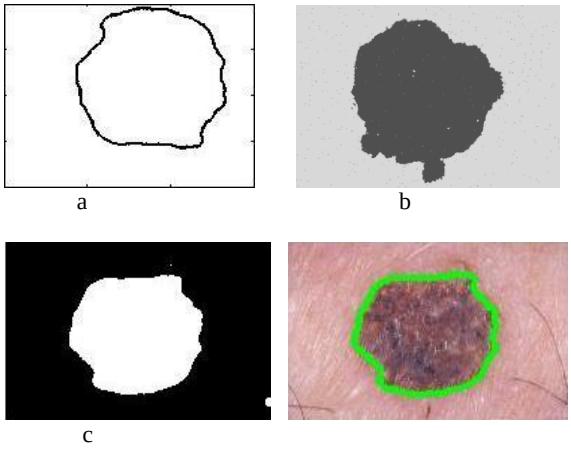


Figure 3 The original image is processed inorder to isolate the lesion. Image (a) is the result of the ISFCM algorithm for k=2. The value of D is 0.94. Image (b) is the result for k=3 ,(c).image with texture information, (d). Image with boundary .

e. Multistep Region Growing Algorithm

Once the color distance images and the texture information are obtained, the region-growing process starts. As explained before, region-growing techniques have two critical tasks: the seed selection and the choice of the belonging condition.

f. Seed selection

The seed selection is the very critical task in region growing segmentation. This seed determines the region. To identify possible seeds we take advantage of the knowledge about how the distance images have been built. Those pixels more similar to the reference color have been assigned a high value (note that we have inverted the distance image). In order to select the seeds, the next three steps are followed for each distance image: 1) election of the local maxima of the image, which represent the candidates to seeds. Not all these candidates will be seeds for the region growing, because these local maxima do not belong necessarily to the region of interest. 2) Application of a threshold to these candidate seeds. The threshold is determined from the histogram of the distance image, more specifically, the threshold will be the position of the peak closest to the right part of the histogram, as seeds should have high values as explained before. In images where different objects are present, the histogram typically presents

different modes, each representing an object or the background, and each mode contains at least a local maximum. In the distance image, the rightmost mode corresponds to the object to be segmented. Therefore, any pixel with values belonging to this mode is a valid seed for the region growing. One way of assuring that seeds belong to the rightmost mode is to choose pixels with values on the right of the rightmost peak in the histogram, because this peak will correspond for sure to this mode. Obviously, the only thing what matters for the good performance of the algorithm is that seeds belong to the region to be segmented and, provided this condition is met, the choice of a particular threshold for the seeds is not decisive in the success of the segmentation. The procedure to find significant peaks and valleys and the threshold follows an automatic algorithm [6]. 3) Finally, texture information is applied to reject some of the seed candidates: the final seeds must have, not only the desired color, but also the desired texture. That is, among the group of color seeds, only those pixels that appear white in the texture image are selected. The group of final seeds will be formed by the seeds obtained with the described method and with each of the color distance images.

g. Multistep Region Growing Refinement

In an ordinary region growing, the belonging condition is always the same. For each seed, the algorithm grows a region with a determined condition. With this multi-step technique, the belonging condition automatically changes in order to find its optimum value, which will correspond to the highest value of the contrast parameter explained later on in this subsection.Let us take a particular seed. The process begins with a region growing with three conditions: 1) Not belonging to another region grown before. 2) The texture of the pixel must be the desired one. That means that a pixel only will be added to the region if it has a value equal to one (for normalized values) in the texture image.3) The new pixel must be similar to the pixels that already are in the region for all the distance images. This similarity is measured according to (4):

$$\frac{F_{max,n}+F_{min,n}}{2}-\tau \le F_{ij,n} \le \frac{F_{max,n}+F_{min,n}}{2}+\tau \quad (12)$$

and n=1,…N, where n refers to the distance image corresponding to the reference color n, N is the number of reference colors, Fmax,n and Fmin,n are the maximum and minimum values ofthe pixels in the distance image n inside the region, i and j are the coordinates of the pixel, F is the value of the pixel in the distance image n, and 2 is the tolerance step, which will be iteratively increased. It must be emphasized that the region growing does not depend on the position of the seed within the region. Although a boundary is encountered on one side, the algorithm will continue growing with the same parameter of tolerance in the other directions until no other pixel can be included in the region; and only then, the contrast is calculated to determine if we should increase the tolerance and continue growing.

$$\text{Contrast}=\frac{inside\ edge-outside\ edge}{inside\ edge+outside\ edge} \quad (13)$$

where Inside edge and Outside edge represent the mean values of the pixels belonging to the inner border and outer border of a region respectively. We then use the mean of the contrast values to determine whether the region is the best or not.At the beginning, the region growing has a very restrictive belonging condition. This will lead us to obtain a small region. While repeating the process, the contrast parameter of equation (5) is calculated. While the grown region is inside the object, the contrast parameter increases its value in a smooth way, because pixels belonging to the inner border and to the outer border of the region are similar. When the region matches the object , the contrast parameter has a high value because pixels surrounding the region will differ from those inside the region. If we continue growing, the contrast parameter will be low again because both the inner border and the outer border are similar. Therefore when the contrast parameter reaches its maximum we have obtained the best region. A steep slope in the contrast parameter evolution corresponds to those values of tolerance for which boundaries are reached. This increase may be either because the whole boundary of the lesion has been reached or because the boundary of the region grown matches in part the boundary of the lesion. In the second situation, the tolerance will continue being increased until the whole boundary is reached. During this increase of the parameter D, the contrast parameter never decreases and, as a consequence, the stop condition is not reached, for the increase of the tolerance is not high enough to overcome a boundary. Once the whole boundary is reached, if the tolerance is being enlarged again the region will exceed the limits of the lesion and, therefore, the contrast will decrease. In such a situation the region growing will stop because the stop condition has been attained.

4. Experimental Results

The algorithm has been validated with many skin cancer lesion images achieving high quality results in all cases. We can see some examples in Figure 4. It is important to note the similarity in color between the normal skin and the lesion in some of the images. In this sense, the

inclusion of texture descriptors is crucial for the good performance of the algorithm.

5. Conclusions

An automatic method for segmentation of skin cancer images is presented. We take into account all the colors in the change in colors, performing a statistical based multistep region growing procedure which has an adaptative structure. For each of the distance images built, contrast is introduced to decide whether a region is the best or not, and the step which provides a region with the highest contrast in relation to the background is chosen. Color and texture information are used in order to fulfill the requirements of pigmented skin color changes due to disease. Figure shows the segmentation result.

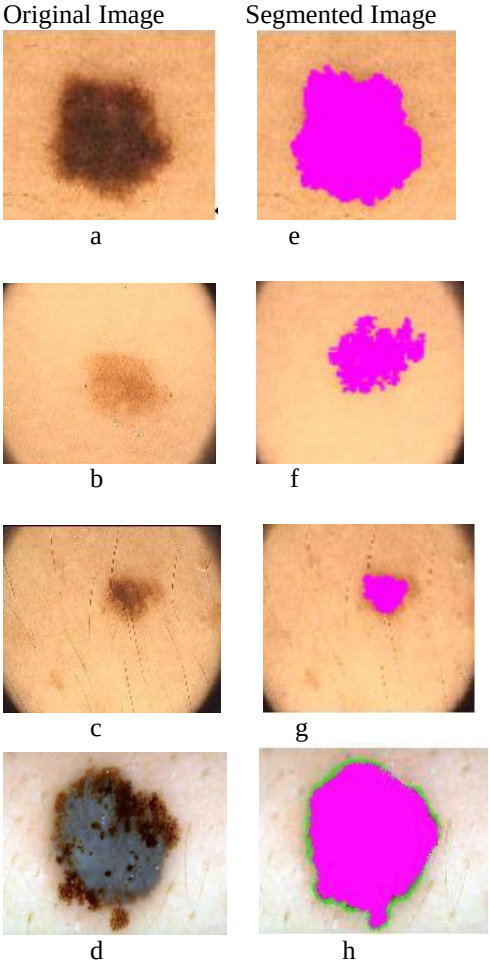


Figure 4. Example of the segmentation. Images (a)-(d) are the original images and images (e)-(h) are the results, in pink, obtained with our algorithm.

REFERENCES

[1] L. Xua, M. Jackowskia, A. Goshtasbya,* , D. Rosemanb, S. Binesb, C. Yuc, A. Dhawand, A. Huntleye, "Segmentation of skin cancer images," Image and Vision Computing 17 65–74 (1999).

[2] K. N. Plataniotis and A. N. Venetsanopoulos, Color Image Processing and Applications, Springer, Berlin 35-37 (2000).

[3] U. Maulik and S. Bandyopadhyay, "Performance evaluation of someclustering algorithms and validity indices," IEEE Trans. PatternAnal. Mach. Intell. 24 (12), 1650–1654 (2002

[4] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," Color Res. Appl. 26 (5) 340–350 (2001).

[5] M. Tuceryan, "Moment based texture segmentation," Pattern Recogn.Lett. 15 (7) 659– 668 (1994).

[6] I. Fondón, C. Serrano, B. Acha. "Color-Texture Image Segmentation based on Multi-Step Region Growing," Optical Engineering, 45 (5) 057002_1- 057002_9 (2006).

[7] R. R. Coifman, D. L. Donoho, "Translation invariant denoising," Lecture Notes in Statistics, volume. 103, pp. 125-150, (1995).

[8] S.Zulaikha Beevi, Dr.M.Mohamed Sathik, Dr.D.Senthamarai Kannan,"A Robust Fuzzy Clustering Technique with Spatial Neighborhood Information for Effective Medical Image Segmentation", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 3, March 2010

[9] S.Zulaikha Beevi, Dr.M.Mohamed Sathik," A Robust Segmentation Approach for Noisy Medical Images Using Fuzzy Clustering With Spatial Probability" European Journal of Scientific Research, Vol.41 No.3 (2010), pp.437-451

A New Method for Finding the Cost of Assignment Problem Using Genetic Algorithm of Artificial Intelligence

Shaikh Tajuddin Nizami[1], Jawaid Ahmed Khan[2] ,
Fozia Hanif Khan[3] , Nasiruddin Khan[4] and Syed Inayatullah[5]

[1&2] Department of computer Science & Information Technology , NED University of Engineering & Technology, Karachi, Pakistan ,
[3]Department of Mathematics, Sir Syed University of Engineering & Technology, Karachi, Pakistan
[4&5] Department of Mathematics, University of Karachi, Pakistan
{ [1]nizamitaj@yahoo.com, [2] jawaid_bilal@yahoo.com, [3]mf_khans@hotmail.com
[4]drkhan.prof@yahoo.com [5]inayat_ku@yahoo.com }

***Abstract*:** In this paper we present a new method for minimizing the cost of Assignment Problem of Operation Research. The proposed method uses the well-known Genetic Algorithm of Artificial Intelligence as a basic ingredient.

***Keywords*:** Artificial Intelligence, Assignment Problem, Genetic Algorithm, Hungarian Method, Operation Research.

1. Introduction

The Assignment Problem of Operations Research is usually handled by the Hungarian Method (see [1], [6]). In this paper, we present the Genetic Algorithm as a worthy alternative (see [2], [3], [4], [5]) for generalities on Genetic Algorithm). This Algorithm, despite its alien and outlandish trappings, is a much faster and more efficient tool to handle the Assignment problem than the Hungarian Method. In section 2, we recall some definitions from Genetics so that the Algorithm in question may make sense. Section 3 contains a brief description of the Assignment problem. Our section 4 is the central one the describes how to apply Genetic Algorithm to the Assignment Problem. In Section 5, we illustrate the Algorithm by working out the test example.

2. Definitions

Genetic Algorithm: The terminology of genetic algorithm is an odd mixture of Computer Science and Genetics. Genetic Algorithms are search algorithms for finding optimal or near optimal solutions .For an excellent introduction to genetic Algorithms, we refer the reader to [2], [3] and [5].

Chromosomes : A structure in the nucleus containing a linear thread of DNA which transmits genetic information and associated with RNA and his tones. An individual's genetic structure is described by bit strings as a list of 1's and 0's. These strings are called chromosomes.

Alleles: One of two or more alternative forms of a gene at corresponding site (loci) of homogeneous chromosomes which determine alternative characters in inheritance .Chromosomes strings containing bits are called alleles.

Genotype: The entire genetic constitution of an individual, also, the alleles present at one or more specific loci. The bit string associated with a given individual is called the individual's genotype.

Generation:
(i) The process of reproduction
(ii) A class composed of all individuals removed by the same number of successive ancestors from common predecessors.

Reproduction: The production of offspring by organized bodies, individuals from one generation are selected for next generation.

Crossover: The exchanging of material between homologous chromosomes during the first meiotic division resulting in new combination of genes. Genetic material from one individual is exchanged with genetic material of another individual.

Mutation: A permanent transmissible change in genetic material.

3. Assignment Problem.

A company has four machines and four jobs to be completed. Each machine must be assigned to complete one job. The required to set up each machine for completing each job is shown in table below. M stand for Machine and J stands for jobs.

	$J-1$	$J-2$	$J-3$	$J-4$
$M-1$	14	5	8	7
$M-2$	2	12	6	5
$M-3$	7	8	3	9
$M-4$	2	4	6	10

X_{ij} if machine i is assigned to meet the demand of job j. i and j are row and column indices. This problem is solved by Hungarian method in [6]. The optimal

solution is 15. That is X_{12} , X_{24} , X_{33} , X_{41}.

4. Solution of Assignment Problem Using Genetic Algorithm

1. The string is defined as $_{ij}|C$ where i and j are row and column and C is cost.
2. Write $_{ij}|C$ in bit string, a list of 1's and 0's.
3. The selection of individual's genotype depends on the largest value of cost and genotype is allowed two off springs and the smallest value of cost is not fit, put this right adjacent to highest value.
4. Put the remaining genotypes to right and left according to value of cost in decreasing order.
5. Three genetic operators are applied to produce the next generation solution (chromosome).These operators are selection / reproduction, crossover and mutation. After reproduction, crossover and mutate the new individual by a small number of bits, arrows with dotted line indicate mutation and arrows with smooth lines indicate copying. We have generation $n+1$. Choose highest value of bits to make the circle $_{ij}|C$ in the cost matrix and draw the vertical line of j^{th} column of matrix. If there are two same highest values, consider the lowest cost.
6. Again take the genotype for next generation of second row except $2j$ of above return to step 5.
7. put the remaining genotypes of third and fourth row for reproduction, after completion of all steps, choose the highest value of bits to make the circle $_{ij}|C$ Cost and draw vertical line of j^{th} column. Now take value of cost of remaining row.

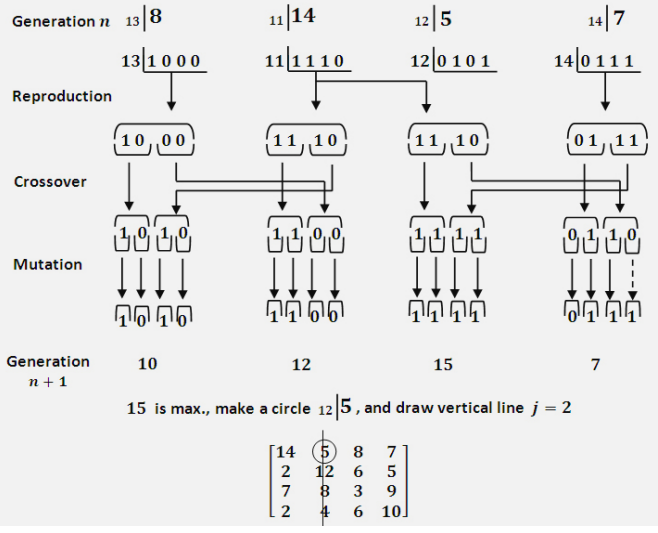
5. Test Examples

Examples-1

Consider the following cost matrix of Assignment problem, see [6]

$$\begin{bmatrix} 14 & 5 & 8 & 7 \\ 2 & 12 & 6 & 5 \\ 7 & 8 & 3 & 9 \\ 2 & 4 & 6 & 10 \end{bmatrix}$$

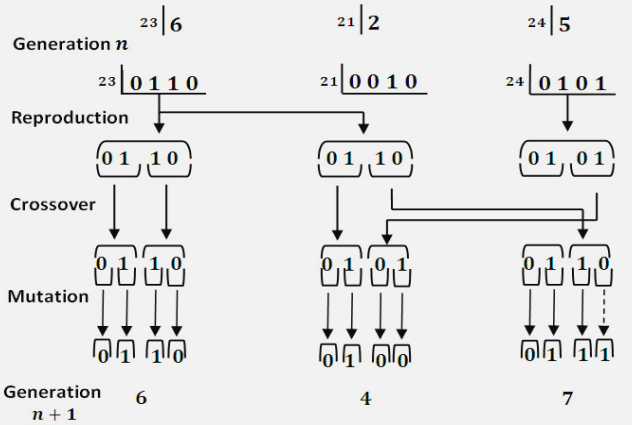
Row-1



15 is max., make a circle $_{12}|5$, and draw vertical line $j=2$

$$\begin{bmatrix} 14 & \textcircled{5} & 8 & 7 \\ 2 & 12 & 6 & 5 \\ 7 & 8 & 3 & 9 \\ 2 & 4 & 6 & 10 \end{bmatrix}$$

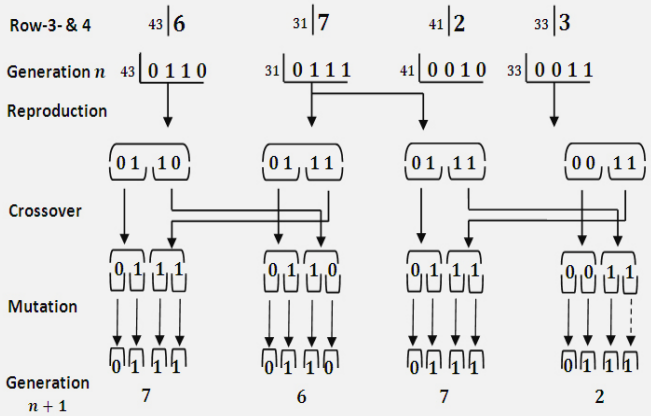
Row-2 Take second row for generation except $_{22}|12$



7 is max., make a circle $_{24}|5$, and draw vertical line $j=4$

$$\begin{bmatrix} 14 & \textcircled{5} & 8 & 7 \\ 2 & 12 & 6 & \textcircled{5} \\ 7 & 8 & 3 & 9 \\ 2 & 4 & 6 & 10 \end{bmatrix}$$

Take the third & fourth row for generation except $_{32}|8$, $_{34}|9$, $_{42}|4$, $_{44}|10$

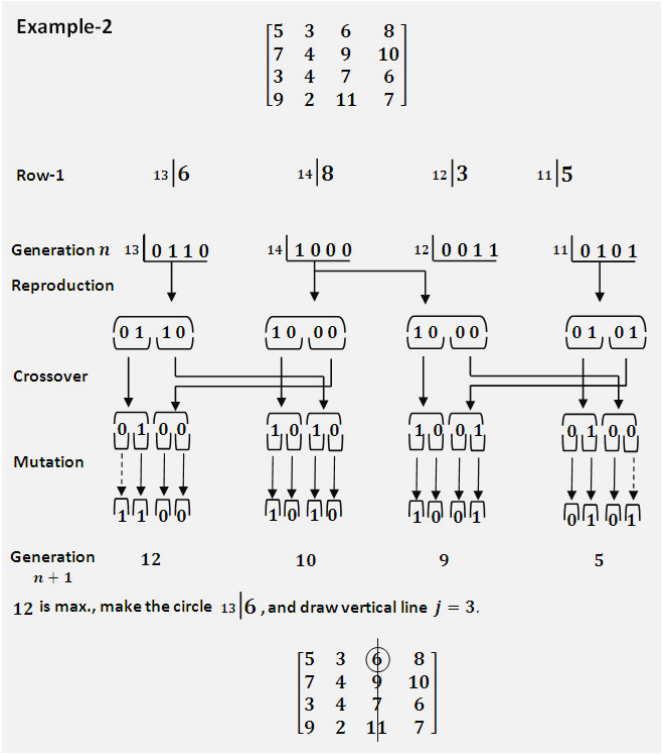


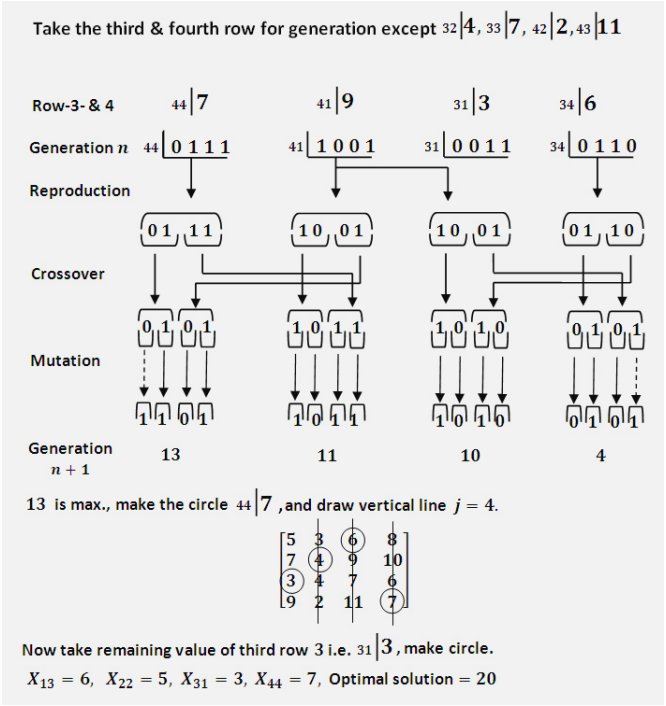
there are two max. values i.e. 7, we choose lowest cost i.e. make a circle $_{41}|2$, draw vertical line $j=1$.

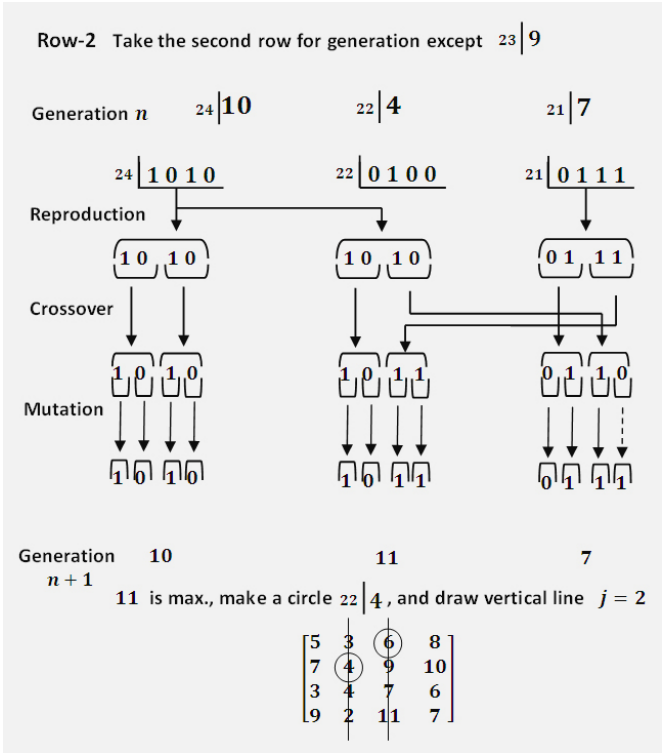
$$\begin{bmatrix} 14 & \textcircled{5} & 8 & 7 \\ 2 & 12 & 6 & \textcircled{5} \\ 7 & 8 & \textcircled{3} & 9 \\ \textcircled{2} & 4 & 6 & 10 \end{bmatrix}$$

Now take remaining value of third row i.e. $_{33}|3$, make circle.

$X_{12}=5,\ X_{24}=5,\ X_{33}=3,\ X_{41}=2,$ Optimal solution $=15$







6. Summary and Conclusion

From the general description in section 4 of the use of Genetic Algorithm in the Assignment Problem and from the test problem solved in section 5, a little efforts shows that these two problems yield the same results with the use of Hungarian Method . It is evident that this method is shorter and more efficient than Hungarian method. The Hungarian Method requires in general, a succession of steps to take care of zeros, whereas Genetic Algorithm uses the Genetic Operators only thrice in 4 order cost matrix. Compare to Genetic Algorithm method, the Hungarian method, when applied to the Assignment Problem is something of a long haul.

References

[1] Fredric S. Hillier "Introduction to Mathematical Programming", Stanford University, Mac Graw-Hill Inc. (1995).

[2] Goldberg .D.E. "Optimization and Machine Learning", Genetic Algorithm in Search, (1989).

[3] Addison-Wesley, Mitchell ,"An introduction to Genetic Algorithm" MIT Press, (1996).

[4] Patricia D. Novak, "Dorland's Pocket Medical Dictionary" W.B. Saunders Company.

[5] Thomas Dean, James Allen, "Artificial Intelligence Theory & Practice", The Benjamin/Cumming Publishing Company Inc. Redwood City, California 94665. (1995).

[6] Wayne L. Winston, "Operation Research, application & algorithm" Indiana University, California 94065, Duxbury Press, Boston, (1987).

Statistical Analysis of Web Quality of Virtual Interfaces of University Grants Commission – Academic Staff Colleges

Dr. Vikas Sharma
International Centre for Distance Education and Open Learning (ICDEOL)
Himachal Pradesh University, Shimla, H.P. India
dr.vikas.hpu@gmail.com

Abstract: In this competitive world, the Information Technology (IT) revolution has been facilitating the higher education service providing organisations to make deeper reach among masses and present themselves in a more competitive and integrated way. Our IT enabled education system has become a global village where flow of information from one corner to another is fast and seamless. This IT enabled academic world is using web as a strategic tool for fast flow of information, creation and dissemination of knowledge in convenient, flexibility and cost effective manner. Almost all University Grant Commission- Academic Staff Colleges in India have their own web presence. A major concern of these educational organisations is whether these virtual interfaces (websites) are projecting themselves effectively or not? In general, an effective website is that which meets the expectations of its target audiences. This paper discusses the quality of websites of University Grant Commission- Academic Staff Colleges (ASCs) functioning in Himachal Pradesh, Jammu and Kashmir, Haryana, Uttrakhand and Punjab states to highlight their potential strengths and weaknesses in terms of physical appearance, design, information availability, presentation, etc. The results of the study show that all ASCs are providing basic web service to advance level service but still there is a gap in the expectation of the audiences and actually available web features.

Keywords: ASC-UGC, NPE, ICT, Website, World Wide Web, Download, HTTP, HTML.

1. Introduction

A big revolution came in India, after implementation of globalisation and liberalisation policies during 90s. These changes were further escorted by rapid revolution in communication technologies. A major development in ICT (Information and Communication Technology) was the convergence of computing and communication technologies, which gave birth to Internet and its allied technologies. The modern state-of-the-art technology based services of the private sector universities has been forcing the traditional Indian Universities to adopt modern technologies and come out from the clutches of old educational system where pen and paper were the strategic tools. In the present scenario, almost all universities have their web presence and making use of Internet and its allied technologies for creation, extraction and exchange of information in the form of knowledge. The Internet is a platform where different categories of users are able to interact and share information without any geographical and time constraint. A website is a collection of web pages, images, videos or other digital assets that is hosted on one or several web server(s), usually accessible via the Internet, cell phone or a Local Area Network (LAN). A web page is a document, typically written in HTML (Hyper Text Markup Language) that is almost accessible via HTTP (Hyper Text Transfer Protocol), a protocol that transfers information from the web server to display it in the user's web browser [10]. Quality of a website is an important factor that differentiates similar services offered by different service providers. Such a measure would allow web service clients to choose and bind to a suitable web service at run time [8]. Any particular website will have its own unique scale and meaning for "quality", because quality is a measurement of a complex relationship of goals, purposes, audience needs, execution of design and other issues specific to that site. Without a clear direction and goals, a web site cannot deliver a consistent, steady message to customers; hardly a platform for quality [7]. So, there is no use of deployment of web services to the targeted users without clear-cut objectives or aims. Every website should have some minimum standards or benchmarks to meet the requirements of different users. It has emerged as a prime communication channel among various categories of users in this modern IT enabled world.

2. Role of University Grants Commission-

Academic Staff Colleges

The University Grant Commission- Academic Staff Colleges (UGC-ASCs) came into existence after implantation of

National Policy on Education (NPE) 1986. These colleges were established to develop a sound linkage between teacher and quality of education. The NPE stressed on providing professional and career development opportunities to the teachers so that they can play effective role and responsibility within higher education system using service of UGC-ASCs. These colleges were setup to enhance the motivation skills and knowledge through systematic orientation in specific subjects, techniques and methodologies [9]. The objectives of UGC-ASCs are: 1) to organise orientation programmes in pedagogy, educational psychology and philosophy, and socio-economic and political concerns for all new entrants at the level of Assistant Professors, 2) to organise orientation/refresher courses for serving teachers, covering every teacher at least once in three to five years, 3) to organise specially designed orientation programmes/refresher courses in IT for new entrants as well as for in-service teachers; and 4) to encourage teachers to participate in seminars, symposia, workshops, etc. [9]. As on May 2010, there were sixty six UGC-Academic Staff Colleges functioning at national level in India in which Andhra Pradesh had the highest number of Academic Staff Colleges (6) followed by Uttar Pradesh (5) and Maharastra (5), Tamilnadu (4), Madhya Pradesh (4), West Bengal (4), Delhi (3), Gujrat (3), Karnataka (3), Kerla (3), Rajasthan (3), Punjab(3), Haryana(3), Jammu & Kashmir (2), Orisa (2), Bihar(2), Chattisgarh (2), Pondicherry (1), Assam (1), Goa(1), Himachal Pradesh(1), Manipur(1), Jharkhand (1), Uttrakhand (1), Mizoram (1), and Meghalaya (1). These UGC-ASCs have been functioning in all the nook and corner of the country to provide modern academic development trainings to the young teachers so that they can inculcate quality education among students at higher education level.

3. Review of Literature

Lemieuxm [4] observed that websites should look attractive using: 1) graphics & layouts, and 2) programming pages but both of these require memory of computer when downloading. If one or the other is too large then this may result in negative impact on visitors. The good web designing is based on good combination of graphics not more than 30-50K in size and attractive programming page layouts.

Reis[5] observed that the main objective of website designing was to attract users and provide them proper guidelines. One thing is necessary to keep in mind is that users are alone there, without any supportive staff to guide them and explain how to navigate information. The website should stand on its own merit. It should have authentic information with appropriate images, format, size and colours in addition to user-friendly navigation facilities.

Sisson[7] stressed on detail study on audience's interaction with website's information, space and design of presenting information to meet users' mode related to interaction. It is better to get feedback and involve audience in design process. A critical element of a good website is to develop around well-understood audience because the success of any website depends on meeting the needs and requirements of the people who use the site.

Crystalcoastech [2] concluded that a visitor to web site took about 8 seconds to decide to stay on that website or click the “Back” button. Today, people want instant results. If website does not offer something of benefit, the users won't bookmark it and they won't come back. The size of file should not be too large that it prevents the web page from loading quickly. There is a need to make good balance between image qualities and file size.

EUPORA[3] observed five broad areas of Quality of Service for website as: 1) ability to connect, that is how quick and easy it is to connect to the ISP -Internet Service Provider, 2) downstream connectivity, that is how quick and reliable the connection between the user and their ISP is, 3) upstream connectivity, that is how quick and reliable the connection between their ISPs and the rest of the Internet is, 4) cost, that is the cost of Internet connectivity and presence; and 5) others, that is a selection of other criteria on which the individual or SMEs experience of the Internet can be judged.

The above review of literature is mainly centric on basic principles for designing of website. There is no thorough study on websites dealing with teachers’ training at national level such as UGC-ASCs. Since, the emerging competition in our education system has been putting huge pressure on the universities to replace their traditional communication channels with computers, Internet, Intranet, mobile phones and various other modes. So, it becomes essential to study the quality parameters which play vital role in designing of effective website which deal with teachers’ training. This research has tried to analyse the quality of virtual interfaces (websites) of UGC-ASCs in terms of features opined important by the new entrant teachers serving at colleges and universities level. The main objective of this study is to help UGC-ASCs to know about the effectiveness of their virtual faces in this IT-Savvy world.

4. Objective of Study

1. To evaluate the web quality parameters considered important for designing of virtual interfaces of University Grants Commission- Academic Staff Colleges.

2. To compare the web quality features of different University Grants Commission- Academic Staff Colleges.

5. Research Design and Methodology

The research methodology of this study has been divided into four main parts, namely: 1) Scope of Study, 2) Population, 3) Sample, and 4) Research Tools.

5.1 Scope of Study

The study is based on ten University Grants Commission-Academic Staff Colleges (UGC-ASCs) functioning in Jammu & Kashmir, Himachal Pradesh, Uttrakhand, Haryana and Punjab states of India.

5.2 Population

Population of this study includes ten Academic Staff Colleges of different universities who have their virtual interfaces (websites). These are the Himachal Pradesh University Shimla(www.ascshimla.org), Kumaon University Nanital (www.ascnainital.org), Guru Jameshwar University Hisar (www.gju.ernet.in/asc/asc.htm), BPS Mahila Vishavvidyala Sonipat(www.bpswomenuniversity.ac.in/division/default.aspx?stream=College&type=15), Kurukshetra University Kurukshetra (www.kukinfo.com/), Punjabi University Patiala (www.punjabiuniversity.ac.in/pages/asc.htm), Punjab University Chandigarh (www.acs-pu.puchd.ac.in/about.html), Guru Nanak Dev University, Amritsar, Punjab (www.department.gndu.ac.in/asc.asp), University of Kashmir, (www.kashmiruniversity.net/researchcentre.aspx?collect=3) and University of Jammu, Jammu and Kashmir state of India (www.jammuuniversity.in/academics_colleges.asp).

5.3 Sample

The study is based on convenient sample survey of 35 participants of Refresher Course (RC-218) on 'IT Awareness" conducted at UGC-ACS, Himachal Pradesh University, Shimla during the period 01.04.2010 to 21.04.2010. These participants rated the qualities of above ten Academic Staff colleges' website. Total 350 cases (35 participants X 10 ASCs) were considered for analysis of effectiveness of websites.

5.4 Research Tool

The data collection tool used here was self designed questionnaire having two parts: 1) first part was used to rate the importance of each quality parameter as general web feature which should be present in any teachers training information dealing website, 2) second part was used to observe the actual available quality features of UGC-ASCs website. A 5-point Likert Scale (5 = highly important and 1 = not important at all) was used to observe the opinion of the participants. To build an initial list of web qualities for UGC-ASCs, five technical sessions (lectures) on attractive web interface related information such as "Human Computer Interaction & Ergonomics", "Internet Usage", "IT and Society", "IT and Management" and "Latest Trends in Web Technologies" were delivered to the participants by the resource persons on different days followed by brainstorming sessions to conclude a raw list of web qualities expected from UGC-ASCs website. A raw list of items on web qualities emerged during this Refresher Course. These items were arranged then in logical order to give a questionnaire format. Using the literature on information quality and by looking carefully for overlap of web qualities, the items in the questionnaire were reduced to a more manageable 20 questions. In conjunction with defining the qualities, a description was also prepared to provide readily available remarks of each quality question to the observers while completing their questionnaires. The data collection process was carried out firstly by using the first one of the questionnaire meant for rating the importance of each quality element as general web quality feature for any UGC-ASC website on five point Likert scale. This was followed by using of second part of the questionnaire in computer laboratory having high speed Broadband Internet facility to observe the actually available quality features for all the ten websites.

6. Results and Discussion

6.1 Analysis of Quality Parameters

In a sample of 35 participants/observers, 40.00 percent participants were from Physical Sciences background, followed by 28.57 percent participants from Social Sciences background, 11.43 percent participants each from Languages and Computer Science & Applications background and rest 8.57 percent participants from Management & Commerce background. Table 1 shows list of quality parameters opined important by the participants after instrumenting the first part of the questionnaire.

Table 1. List of Quality Parameters Opined Important by the Participants

Q1.	The site has an attractive appearance	Q11.	Facility to contact Academic Staff College using email
Q2.	Colours are used appropriately	Q12.	Latest information about Orientation Programmes, Refresher Courses, Workshop and Seminars
Q3.	Layout is sharp and not crowded with words	Q13.	The website provides updated list of selected candidates for different courses
Q4.	The site is easy to use	Q14.	Availability of Information about lodging arrangements during stay as participants
Q5	The site is easy to navigate	Q15.	Availability of news letters on the web site
Q6.	Interaction with the site is clear and understandable	Q16.	Availability of web links to useful websites
Q7.	Figures and tables are aligned correctly	Q17.	Contents within the site are easy to follow
Q8.	Images are relevant, interesting and not crowding	Q18.	Presents information in an appropriate format
Q9.	The design is appropriate to the type of site	Q19.	Facility to download application forms
Q10.	The site conveys a sense of competency	Q20.	Facility to download articles or research papers

These total 20 data items cover some broad areas related to website quality such as appearance, usability, information availability, information downloading, etc. The table 2 displays the summary averages for weighted and unweighted data sets.

Firstly, the importance scores (IP) shows the average importance ranking for each question as rated by 35 participants. Secondly, the average scores per data item for each UGC-ASC website are given. This is displayed in two modes: 1) raw score (RS) as unweighted ratings (with a theoretical Likert Scale range of 1 to 5), and 2) weighted score (WS). The weighted score (theoretically ranging from 1 to 25) is obtained by multiplying the unweighted score by the importance score for each respondent. The web qualities considered most important by the participants, e.g. above the upper quartile (4.89) are all about latest information about refresher courses and orientation programmes, list of selected candidates who had applied for above courses and attractiveness of website. It was found that question numbers 12, 13, 19 and 1 are of higher importance – above upper quartile. The web qualities considered least important, e.g. below the lower quartile (4.41) are based around web designing, arrangement of figures, tables, images, availability of news letters, competency of website. Specifically, the questions numbers 9, 10, 15, 7 and 18 are in ascending order of their importance. Other questions are in between and the median for above importance score is 4.74. The above results show that there are specific priorities in the qualities demanded by the new entrant teachers. Web features such as availability of updated information regarding courses or

Table 2. Summary Averages for Weighted and Unweighted Data Sets

Q. No.	IP*	H P University, Shimla, H.P.		Kumaon University, Nanital, Uttrakhand		Kurukshetra University Kurukshetra, Haryana		BPS Mahila Vishavvidyala, Sonipat, Haryana		Guru Jameshwar Univ. Hisar, Haryana		Jammu University Jammu, J & K		University of Kashmir, Srinagar, J & K		Punjab University Chandigarh, Punjab		Punjabi University, Patiala, Punjab		G N D U, Amritsar, Punjab	
		RS*	WS*	RS*	WS*	RS*	WS*	RS*	WS*	RS*	WS*	RS*	WS*	RS*	WS*	RS*	WS*	RS*	WS*	RS*	WS*
Q.1	4.91	4.7	23.1	4.9	24.0	2.6	12.8	4.9	23.9	4.9	24.1	4.8	23.7	1.4	6.7	1.5	7.3	4.2	20.6	3.7	18.4
Q.2	4.80	4.7	22.6	4.9	23.5	2.5	12.0	5.0	24.0	4.9	23.6	4.8	23.2	3.5	16.8	1.5	7.2	4.2	20.2	3.7	18.0
Q.3	4.49	4.9	22.0	5.0	22.5	2.5	11.3	5.0	22.5	4.9	22.0	4.8	21.7	1.5	6.9	2.6	11.7	4.2	18.9	3.7	16.8
Q.4	4.83	5.0	24.2	5.0	24.2	3.0	14.3	3.5	17.1	4.9	23.7	4.8	23.3	1.7	8.2	4.3	20.7	4.2	20.3	3.7	18.1
Q.5	4.89	4.7	23.0	5.0	24.5	2.8	13.7	5.0	24.5	4.9	24.0	4.8	23.6	1.7	8.3	4.3	21.0	4.2	20.5	3.7	18.3
Q.6	4.60	4.8	22.1	5.0	23.0	2.5	11.5	4.5	20.7	4.9	22.5	4.8	22.2	1.0	4.6	4.3	19.6	4.2	19.3	3.7	17.2
Q.7	4.34	4.7	20.4	5.0	21.7	3.4	14.8	5.0	21.7	4.9	21.3	4.8	21.0	1.7	7.3	4.3	18.5	4.2	18.2	3.3	14.5
Q.8	4.49	4.7	21.1	5.0	22.5	1.5	6.8	4.5	20.2	4.9	22.0	4.8	21.7	1.7	7.6	4.3	19.1	4.2	18.9	3.3	15.0
Q.9	3.94	4.7	18.6	5.0	19.7	1.5	5.9	4.5	17.7	4.9	19.3	4.8	19.0	3.1	12.1	4.3	16.9	4.2	16.5	3.7	14.7
Q.10	4.14	4.7	19.5	5.0	20.7	1.1	4.6	4.5	18.6	4.9	20.3	4.8	20.0	1.0	4.1	1.2	4.8	4.2	17.4	1.0	4.1
Q.11	4.80	5.0	24.0	5.0	24.0	5.0	24.0	5.0	24.0	5.0	24.0	5.0	24.0	1.0	4.8	5.0	24.0	5.0	24.0	1.0	4.8
Q.12	5.00	5.0	25.0	5.0	25.0	5.0	25.0	5.0	25.0	5.0	25.0	5.0	25.0	1.0	5.0	5.0	25.0	5.0	25.0	5.0	25.0
Q.13	5.00	5.0	25.0	1.0	5.0	1.0	5.0	1.0	5.0	1.0	5.0	1.0	5.0	1.0	5.0	1.0	5.0	1.0	5.0	1.0	5.0
Q.14	4.89	5.0	24.5	5.0	24.5	1.0	4.9	1.0	4.9	1.0	4.9	1.0	4.9	1.0	4.9	1.0	4.9	1.0	4.9	1.0	4.9
Q.15	4.26	5.0	21.3	5.0	21.3	1.0	4.3	1.0	4.4	1.0	4.3	1.0	4.3	1.0	4.3	1.0	4.3	1.0	4.3	1.0	4.3
Q.16	4.83	5.0	24.2	5.0	24.2	1.0	4.8	4.4	21.1	5.0	24.2	5.0	24.2	1.0	4.8	4.3	20.6	1.0	4.8	1.0	4.8
Q.17	4.69	4.7	22.1	5.0	23.5	3.0	13.9	5.0	23.5	4.9	23.0	4.8	22.7	4.8	22.4	4.3	20.1	4.2	19.8	4.5	21.1
Q.18	4.37	4.7	20.6	5.0	21.9	2.0	8.9	4.3	18.6	4.9	21.5	4.8	21.1	3.6	15.6	4.3	18.6	4.2	18.5	1.0	4.4
Q.19	5.00	5.0	25.0	5.0	25.0	5.0	25.0	5.0	25.0	5.0	25.0	5.0	25.0	1.0	5.0	5.0	25.0	5.0	25.0	4.5	22.5
Q.20	4.46	5.0	22.3	3.1	13.6	1.0	4.5	1.0	4.5	1.0	4.5	1.0	4.5	1.0	4.5	1.0	4.50	1.0	4.5	1.0	4.5
TS*	**92.7**	**97.1**	**450.6**	**93.8**	**433.9**	**48.4**	**227.8**	**79.0**	**366.6**	**82.9**	**384.1**	**82.0**	**379.9**	**34.6**	**158.8**	**64.2**	**298.7**	**70.5**	**326.6**	**54.8**	**256.2**

*** IP- Importance Score, * RS- Raw Score, * WS-Weighted Score and * TS-Total Score**

programmes currently being run or available in near future, confirmation about selection, facility to download application forms, contact details seem of higher importance whereas readable blocks, complete sentences, design, paragraph setting, etc. are of least importance. So, primary focus of the young teachers seems to be information-oriented rather than physical design and format of website. The total scores for weighted and unweighted data sets were calculated to give a broader picture of users' rating on effectiveness of UGC-ASCs website on different quality parameters. The above total scores make it difficult to analyse the quality of individual website, so Web Quality Index (WQI) was calculated for each website using weighted score against the total possible score (Importance Score X 5). A summary of these calculations and totals are given in Table 3.

Table 3. Comparative Web Quality Index (WQI) for UGC-ASCs Website (in descending order)

Sr. No.	Academic Staff Colleges	Weighted Score	Max. Score	WQI
1.	Himachal Pradesh University Shimla, H.P.	450.56	463.65	0.97
2.	Kumaon University Nanital, Uttrakhand	433.93	463.65	0.94
3.	Guru Jameshwar University Hisar, Haryana	384.12	463.65	0.83
4.	University of Jammu, Jammu, J &K	379.95	463.65	0.82
5.	B P S Mahila Vishavvidyala Sonipat, Haryana	366.57	463.65	0.79
6.	Punjabi University, Patiala, Punjab	326.57	463.65	0.70
7.	Punjab University, Chandigarh	298.72	463.65	0.64
8.	Guru Nanak Dev University. Amritsar	256.15	463.65	0.55
9.	Kurukshetra University Kurukshetra, Haryana	227.80	463.65	0.49
10.	University of Kashmir, Srinagar, J &K	158.75	463.65	0.34

Overall, it appears that Academic Staff College of Himachal Pradesh University scored (0.97 points) followed by Academic Staff College of Kumaon University (0.94 points), Academic Staff College of Guru Jameshwar University (0.83 points), Academic Staff College of University of Jammu (0.82 points), Academic Staff College of BPS Mahila Vishavvidyala (0.79 points), Academic Staff College of Punjabi University (0.70 points), Academic Staff College of Punjab University (0.64 points), Academic Staff College of Guru Nanak University (0.55 points), Academic Staff College of Kurukshetra University (0.49 points) and Academic Staff College of University of Kashmir (0.34 points). This indicates that website of Academic Staff College of Himachal Pradesh University has higher quality features as compared to others UGC-ASCs website. Three states namely Punjab, Haryana and Jammu & Kashmir have more than one Academic Staff Colleges. The Academic Staff College of Guru Jameshwar University, Hisar, Academic Staff College of Jammu University, Jammu and Academic Staff College of Punjabi University, Patiala scored highest WQI points as compared to others Academic Staff Colleges functioning in similar states. It is essential to know here about how the websites differ in quality. A detailed discussion of scores for each and every question would be more cumbersome. But a website can be assessed in a number of meaningful, reliable question sub-groupings. The next section drives a number of sub-groupings and applies them for better analysis in terms of reliability and validity of data items used in the questionnaire.

6.2 Scale Reliability and Question Groupings

In order to verify the reliability of the instrument used for data collection, a statistical reliability analysis was conducted using Cronbach's Alpha (α). The overall average α value was 0.87 which indicates a quite reliable scale. For better analysis of difference in qualities, reliability analysis was extended to different question sub-groupings. The reliability analysis for questionnaire items is summarised in Table 4.

Table 4. Reliability Analysis for Questionnaire Items

Subcategories of Question Items	Alpha (α) value
Physical Appearance (Q 1. to Q 3.)	0.93
Website Usability (Q 4. to Q 6.)	0.94
Website Design (Q 7. to Q 10.)	0.91
Information Availability (Q 11. to Q 15.)	0.78
Information Presentation (Q 16. to Q 18.)	0.70
Core Web Services (Q 19. to Q 20.)	0.75
Overall Alpha (α) value =0.87	

These grouping are explained as: 1) the physical appearance of any website is the first impression which indicates the strength to attract the viewers at first attempt. This includes the colour combination, layout, composition of words, etc., 2) usability refers to being able to get around a site and explore information. This includes the user-friendly interaction and easy navigation within the site, 3) the design part of a website includes the stylish composition of text with images, usage of graphics, alignment, heading within the pages and a appeal of attraction, 4) information availability is fulfillment of users' requirements for specific information 5) presentation of information in the stylish way that attracts the users and force him/her to stay a while on the page. The format of information presentation is important. It should be easy to read, understandable, relevant, current, reliable with requisite level of detail and format, 6) Core Web Service is the way by which the user can avail some online service such as downloading of information and communication with ASCs. Factor analysis was also conducted on the set of 350 cases. The Varimax Factor Rotation converged in 8 iterations and a relatively simple factor structure emerged. Total six factors emerged in the Principal Components Analysis as given in Table 5. Factor loadings in excess of 0.7 can be considered excellent [1]. As shown in the above table that all the values are higher or equal to 0.70, this indicates that there is something common in grouping of above variables. The physical appearance quality consists of questions 1 to 3, usability consists of questions 4 to 6, design consists of questions 7 to 10, information availability consists of questions 11 to 15, presentation of information consists of questions 16 to 18 and core web service consists of question 19 to 20. These categories provide some

useful criteria to analyse the quality features of individual Academic Staff College website. This is discussed in the next section.

Table 5. Exploratory Factor Analysis

Physical Appearance	1	2	3	4	5	6
1. The site has an attractive appearance					0.82	
2. Colours are used appropriately					0.90	
3. Layout is sharp and not crowded with words					0.72	
Usability						
4. The site is easy to use	0.70					
5. The site is easy to navigate	0.71					
6. Interaction with the site is clear and understandable	0.72					
Design						
7. Figures and tables are aligned correctly		0.76				
8. Images are relevant, interesting and not crowding		0.73				
9. The design is appropriate to the type of site		0.71				
10. The site conveys a sense of competency		0.70				
Information Availability						
11. Facility to contact Academic Staff College using email			0.81			
12. Latest information about RCs and OPs.			0.96			
13. Provides information about selected candidates			0.85			
14. Availability of Information about lodging arrangements during stay as participants			0.92			
15. Availability of news letters on the web site			0.92			
Presentation of Information						
16. Availability of web links to useful websites				0.71		
17. Contents within the forms are easy to follow				0.71		
18. Presents information in an appropriate format				0.87		
Core Web Service						
19. Facility to download application forms						0.97
20. Facility to download Articles or Research Papers						0.97
(Principal Components Method with Varimax Rotation) KPO Value =0.87						

6.3 Website Analysis Using Question Subcategories

Using the above question groupings, a profile of an individual website was built and used for comparison with other websites. It has become essential to know why some websites fared better than others on the Web Quality Index (WQI). The WQI data is summarised around the above six question subcategories. Since, there are total 10 UGC-ASCs, so it becomes very cumbersome to explain the WQI on subcategories basis using a single chart for all the websites. So, intra-state comparison of websites was done firstly for Academic Staff Colleges functioning in the Jammu and Kashmir, Haryana and Punjab states. The highest ranking WQI websites were chosen from this intra-state comparison for further inter-state comparison. Figure 1 demonstrates very clearly that the website of Academic Staff College, Guru Jameshwar University, Hisar has better-quality features than other two Academic Staff Colleges located at Kurukshetra University Kurukshetra and B P S Mahila Vishavvidyala, Sonipat. The website of Academic Staff College, Kurukshetra University Kurukshetra found less attractive with poor usability, design and presentation of information.

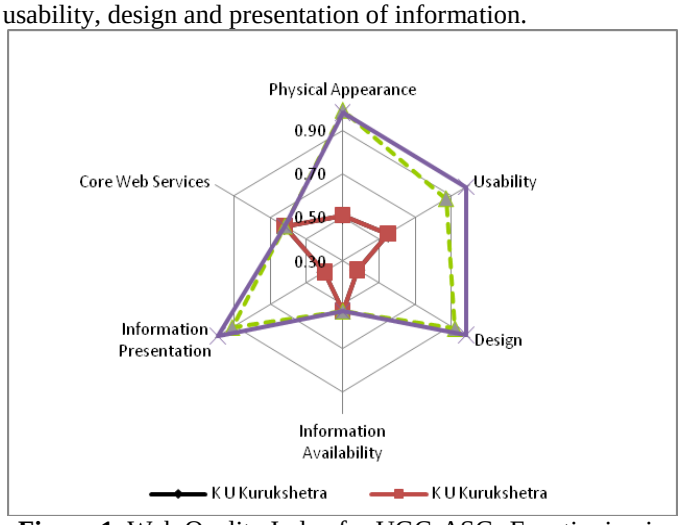


Figure 1. Web Quality Index for UGC-ASCs Functioning in Haryana State

Similarly, figure 2 demonstrates the comparison of WQI for the websites of academics staff colleges functioning in Punjab state. The website of Academic Staff College, Punjabi University Patiala is covering a larger area as shown in the radar chart as compared to the two others websites. Further, the designing features of Academic Staff College, Guru Nanak Dev University Amritsar and presentation of information features of Punjab University Chandigarh have edges over their counterpart Academic Staff Colleges functioning in similar state.

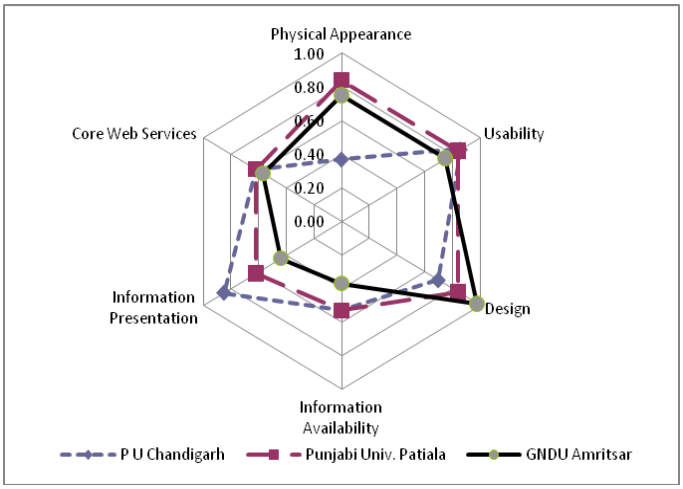


Figure 2. Web Quality Index for UGC-ASCs Functioning in Punjab State

Figure 3 indicates the Web Quality Index for UGC-ASCs functioning in the Jammu & Kashmir state. The website of Academic Staff College, University of Jammu is better in all spheres than its counterpart website of Academic Staff College, University of Kashmir, Srinagar.

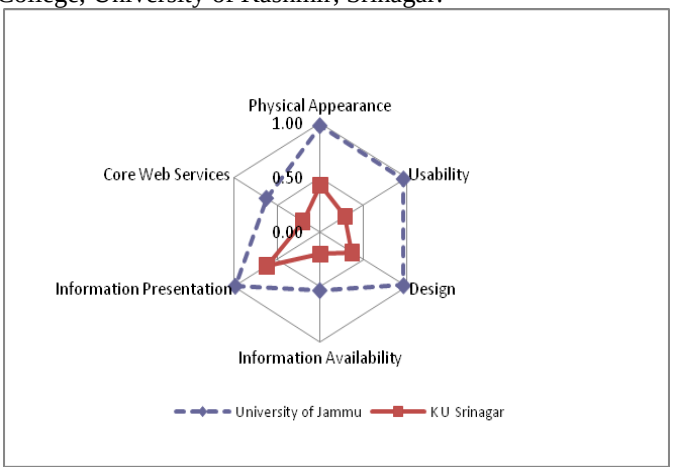


Figure 3. Web Quality Index for UGC-ASCs Functioning in Jammu & Kashmir State

From the above comparison, three websites of ASC Guru Jameshwar University, Hisar, ASC of Jammu University, Jammu and ASC of Punjabi University Patiala were selected for inter-sate comparison with other ASCs functioning in Himachal Pradesh and Uttrakhand states. The figure 4 demonstrates the Web Quality Index for top scoring Academic Staff Colleges of different universities. Though there is no much more difference in physical appearance, usability and designing part of websites of different ASCs but the WQI for the Academic Staff College, Himachal Pradesh University makes a clear circle around the other four websites on availability of useful information and core web services. The website of Academic Staff College, Himachal Pradesh University stands head and shoulders above its four rivals. The presentation of information web features of Academic Staff College, Kumoun University, Nanital is better than the website of Academic Staff College, Himachal Pradesh University. The website of Punjabi University Patiala is lacking behind than its contemporaries Academic Staff Colleges on entire quality feature subcategories.

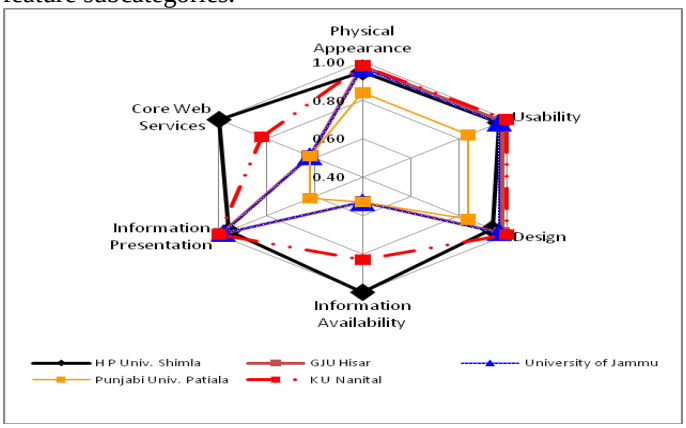


Figure 4. Web Quality Index for Top Scoring UGC-ASCs Functioning in Himachal, Punjab, Haryana, Uttrakhand and Jammu & Kashmir States

7. Conclusion

The web quality instrument was developed from the opinions of the participants of Refresher Course on "IT Awareness" at ASC, Himachal Pradesh University, Shimla and tested in the domain of UGC-ASCs website functioning in five states. Analysis of the results suggests that the web quality instrument has good validity and reliability which can be refined further with larger and varied samples. The real meaning of good website in simple term is appropriate representation of content in the form of text, graphics and images to meet users' requirements. The overall effectiveness of website largely depends on accurate, reliable, authentic and current information.

All ASCs functioning in Jammu and Kashmir, Himachal Pradesh, Punjab, Haryana and Uttrakhand states have their own web presence and providing basic web service to advance level service to their users but still there is a gap in the expectation of the audiences and actually available web features. There is a need to bridge this gap by making improvements in certain areas such as: facility to register online for Refresher/Orientation Programmes and seminars/workshop; online intimation to the candidates regarding their selection; online deposit of registration fee; facility to share and download research papers/articles, projects, question banks and journals; user-friendly navigation within the web pages; attractive and properly aligned text; aesthetic use of images, tables, etc.

The online deposit of registration fee and filling of registration forms will help the Academic Staff Colleges to reduce their administrative costs, physical documents management efforts, etc. Facility to download articles, research projects and papers will provide a common platform for the researchers to contribute and share knowledge at the national level. The other advantage of this online automation is to have in place a centralised database of participant teachers and resource persons which can be shared by all the Academic Staff Colleges at national level for better resource management, administrative support and strategic decision

making. This move will also facilitate the UGC-ASCs to provide good governance and 24X7 days service in a week.

So, a good website in terms of effectiveness should be attractive by nature with catchy domain name, availability of current information, facility to apply online and automatic deposit of fee using electronic banking service. An initiative should also be taken at the end of University's Grants Commission to provide web link facility on its existing website (www.ugc.ac.in) to interconnect all Academic Staff Colleges website on a single web platform. This will facilitate the normal user to get access of any Academic Staff College website without any extra searching efforts on World Wide Web. Further, there is also a need to develop standardise web features for all ASCs website to avoid content navigation and service accessibility problems for new users.

References

[1] Comrey, A., *A First Course in Factor Analysis*, New York: Academic Press, 1973.

[2] Crystalcoastech. (2007). Web Designing Basics. [Online]. Available: http://crystalcoasttech.com/web-design-basic.htm

[3] EUROPA. (2000). Quality of Service Parameters for Internet Service Provision. Final Report. European Commission, DG Information Society. [Online]. Available: http://ec.europa.eu/archives/ISPO/eif/InternetPoliciesSite/QoS/FinalReport.pdf

[4] Lemieux, Martin R. (2006). Creating A Quality Web Site Design!. [Online]. Available: http://www.cyberindian.com/web-designing/creating-quality-website-design.php

[5] Reis Carla. (2005). Your Company's Website -- Part III. Quest Quality Solutions. [Online]. Available: http://www.questqualitysolutions.com/articles/website-design-development-marketing.asp

[6] Singh, Didar. *Electronic Commerce: Issues of Policy and Strategy for India*. Indian Council for Research on International Economic Relations. New Delhi: Working Paper 86. 2002.

[7] Sisson, Derek. (2002). Site Design: Designing in Quality. [Online]. Available: http://www.philosophe.com/

[8] Thio, N. and Karunasekera, S. (2005). Automatic Measurement of a QoS Metric for Web Service Recommendation. Software Engineering Conference. [online]. Available: http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/9629/30439/01402015.pdf?arnumber=1402015

[9] UGC. (2007). National Policy on Education 1986. University Grants Commission. New Delhi. [Online]. Available: http://www.ugc.ac.in/policy/npe86.html

[10] Wikipedia. (2008). Website. Wikipedia Encyclopedia. [Online]. Available: http://en.wikipedia.org/wiki/Website

Author

Dr. Vikas Sharma received his MCA and Doctoral degrees in 1998 and 2008 respectively from Himachal Pradesh University, Summer Hill, Shimla, India. Currently he is serving as Senior Programmer in International Centre for Distance Education and Open Learning (ICDEOL), Himachal Pradesh University, Summer Hill, Shimla and has more than 12 years teaching and software development experience. His research areas include eCommerce, Management Information System, etc.

Advanced Automatic Luminance Management and Real Time Counting System Using High Computing Programmable Gate Arrays: One Step Towards Green IT

[1]Mr. Jitendra S. Edle, [2]Prof. Ajay P. Thakare and [3]Prof. A. M. Agarkar

[1, 2] Department of Electronics and Telecommunication Engineering.
Sipna's College of Engineering and Technology, Amravati, India.
[3]Department of ENTC, S.S.G.M. C. E, Shegaon, India
[1.]jeet.edle@yahoo.com
[2.]apthakare40@rediffmail.com
[3]ajayagarkar@rediffmail.com

Abstract:
The present title discloses a new way to increase energy efficiency and minimize wastage of power at the workstations, mall, libraries and bigger visiting centers, by actively managing and operating the appliances automatically – as a theme of GREEN IT. Green-IT is an attempt to move to more efficient products and procedures so as to manage more active equipments within the given energy foot-print [5].

"The Advanced Automatic Luminance Management and Real Time Counting System" considers different parameters at the front end design level so that total power requirement of the systems can be minimized and hence the carbon footprint can be maintained [5][6].

This system promises to consider different parameters such as, Climatic Conditions, Possible Age Group of the Visitors, Environment to which the application is being used and finally Green Computing.

Keywords :
Green IT, energy footprint, High Speed Integration, Low Power Device, Programmable gate Arrays.

1. Introduction:

Considering the dramatically changed scenario in the field of emergent **S**emiconductor **T**echnologies, **S**urface Mount Devices and High Speed Integration, that has brought down the operating voltage of the devices from ± 12V to +5V and +3.2V and at the present scenario is that most of the high speed computing devices are having only +1.8V operating voltage requirement [1].

But, unfortunately, bringing down the operating voltages of the devices contributes to increase the energy efficiency and further it helps to minimize the total power consumed by the device only, but in real world – real time applications the power consumption can be effectively managed further by intelligently operating the different modules in the appliances as per requirement.

The Advanced Automatic Luminance Management System is basically an implementation that actively manages the available power by intelligently operating the different light sources, in modules. This system is further extended for power management by including the parameters like room temperature as well as wind control and etc. The system can be best designed by using low power and high speed computing device like programmable gate arrays.

2. System Block Diagram:

The ALM-RTS system can be best suitably planned around high performance programmable gate arrays. Fig.1 shows the system build around Field Programmable Gate Arrays (FPGA).

Out of all other intelligent devices, programmable gate arrays are excellent choice because of startling features like 16-bit LUT RAM, In-System Programming (ISP), Boundary scan and Read back ability. Out of available different versions and capacity FPGAs, XC2S100 is the best choice [7].

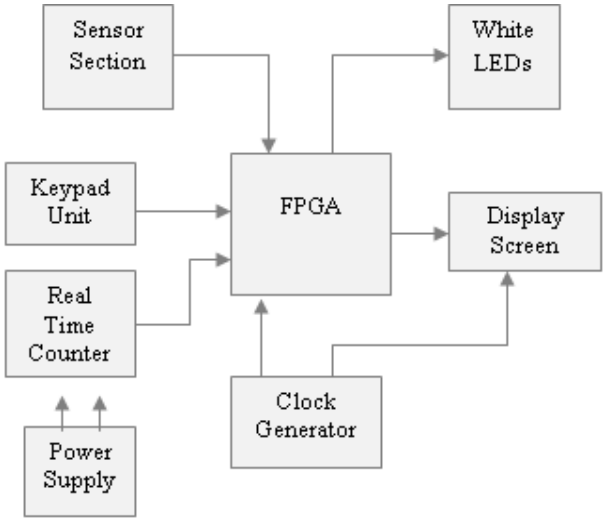


Figure 1. System Block Diagram

3. Programmable Gate Arrays:

Field Programmable Gate Arrays is a flexible, programmable architecture that consist of Configurable Logic Blocks (CLBs), Programmable IO Blocks, Delayed Lock Loops (DLLs), and In-system Programmable Capability etc.[7][8].

CLBs are important blocks in the programmable gate arrays. It is the unit that is configured by the application program created by the designer. Generally the CLBs are configured using some Hardware Description Languages (HDLs) like Very High Speed Integrated Circuit Hardware Description Language (VHDL) and Verilog. Basic element of the CLB is the Look-Up Table (LUT) [3] [4].

3.1 Look-Up Table:

Look up table is basically implemented using universal device i.e. multiplexer as shown in fig.2.

LUT can be suitably implemented using 3-layers of the 2:1 multiplexer device. Layer 1 with control input, 'S0', selects one output of either of the multiplexers in the second layer. With same logic repetition, Layer 2 with control input, 'S1', used to select one out of four outputs from the third layer, and finally Layer 3 with control input, 'S2', it is possible to select one out of two input status lines from the input layer [4].

Status of the input lines (Logic1/Logic0) as well as status of the control inputs (Logic1/Logic0) is provided by the synthesis tool that generates the programmable bits described by the designer using HDLs.

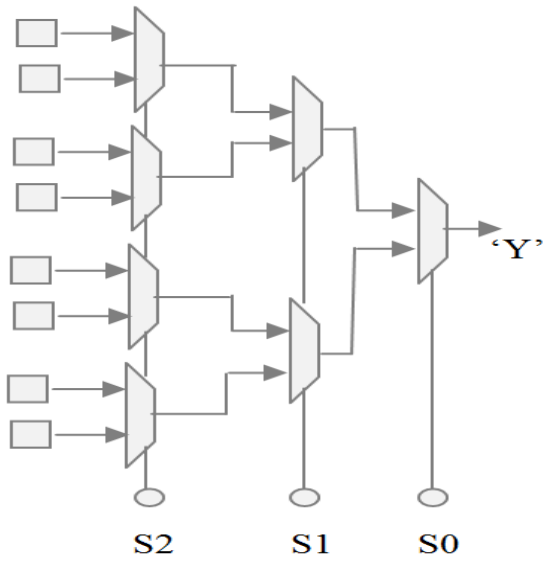


Figure 2. Look-Up Table

Size of the LUT defines the logic holding capacity of the programmable logic devices. In fig 3-stage LUT is shown but in recent Programmable Gate Arrays, more efficient LUTs are available.

3.2 Configurable Logic Blocks:

In digital designing the hardware to be realized may be purely combinational or purely sequential or even combination of both. Configurable Logic Block helps to realize digital module that can be configured as per the hardware to be realized. Fig.3 shows the details.

LUT implements the combinational logic part of the hardware and D- flip-flop takes care of the sequential part of the application hardware. But in real world applications the hardware to be configured may be purely combinational or purely sequential. To implement the module selection facility multiplexers are used.

Multiplexer 'mux1' is used to differentiate between combinational and sequential circuit. With control line S1= '0' the module will be configured to purely combinational mode, with S1= '1' sequential mode is selected. Multiplexer "mux2" is used to configure the module to produce either true or complemented output by setting control input line S2 to '1' and '0' respectively [7][8].

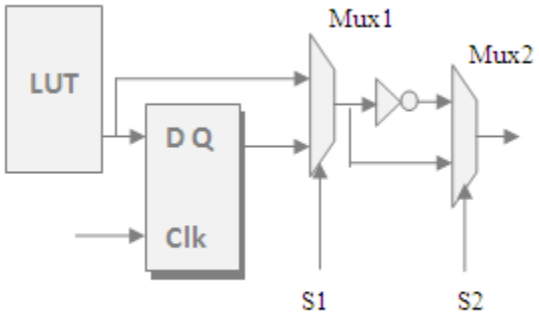


Figure 3. Configurable Logic Block

Configurable Logic Block is the unit that holds the configured hardware by the designer. Same unit can be duplicated to produce higher versions of the configurable logic blocks that will ultimately increase the capability of the programmable logic blocks to implement higher versions of the applications that require large logic capacity devices.

3.2.1 Input – Output Blocks:

These are simply bidirectional buffers used for two special purposes: for signal smoothing and Bounder scan chain execution. In fig.4, control input 'ctrl' provides the facility to select the data transfer mode i.e. either input or output data transfer mode [8].

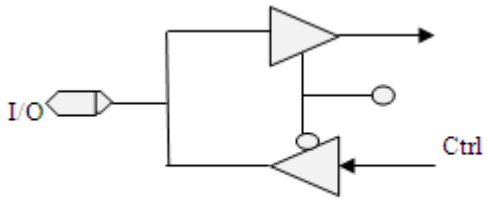


Figure 4. Input-Output Block

These IOBs are placed at the periphery of the device with output of one IOB connected to the input of next one. With such connection the first pin is called as Test Data Input Pin i.e. TDI and the last pin is called as the Test Data Out i.e. TDO [9].

Before describing the hardware inside configurable logic block of the programmable logic device, device test is performed. Some random input data pattern is feed through the input pin TDI if same pattern appears at the output pin TDO then it is concluded that that the device is tested ok else device test fails!

Main element of the device is the LUT, that is implemented using multiplexer, as it is purely combinational logic device, the whole unit becomes 'VOLATILE', hence we can say that Programmable Logic Devices are Volatile devices. To avoid the unnecessary efforts to reload the same program on each power on by the user, Platform Flash device is used [4][10].

3.2.2 In-System Programming:

With this feature it is possible to describe the hardware inside the programmable logic devices without removing them from the circuitry. For the chips that support ISP feature have internal circuitry that generates internal programming voltages from the supply voltage for describing the hardware inside. With this interesting feature manufacturers can integrate the programming and testing into single phase [11].

4. Sensor Bank:

Bank of Passive Infra-Red (PIR) devices can be employed to build a responsive arena. PIR sensors are Pyroelectric Devices that detects motion by measuring changes in the infra-red levels emitted by surrounding objects. This motion can be detected by checking the output at the output of the amplifier.

PIR GH-718 could be the best option. It is made up of crystalline material that generates an electric charge when exposed to an infra-red radiation. The changes in the amount of infrared striking the element change the voltage generated, which can be measured by an on-board available amplifier. This sensor is equipped with the Fresnel lens, which focuses the infrared signals onto the element. As the ambient infrared signal changes rapidly, the output amplifier trips the output to indicate motion.

5. Luminance Control Section:

While working with aspire of increasing the system performance and minimizing the total power wastage, best selection will be definitely Bank of white LEDs. White LEDs are operated at low current and voltage levels but gives efficient luminance when used in group [12].

When white LEDs are used in group, they are operated or activated in an array. With such activation, it is possible to control the luminance more precisely.

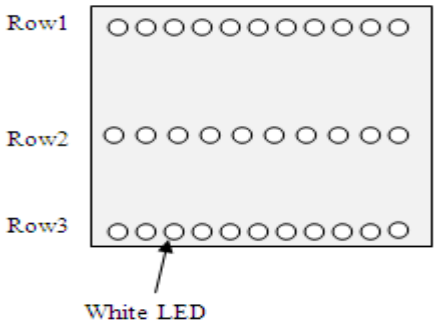


Fig.5: White LEDs arranged in group

Assuming the initial condition at which the natural light is enough, the ALM-RTC system keeps whole bank of LED in off state. As the darkness increases the first row i.e. Row 1 of the bank is operated, in this case other two rows remains in off state. If again, obscurity increases beyond the second threshold, second row i.e. Row 2 is operated by the ALM-RTC system. System keeps track of the amount of obscurity; if it crosses the third threshold whole bank is operated.

6. Real Time Counting Unit:

It is the second part of the system which takes care of counting the number of events occurring. This important part of the system can be build by using simple **I**nfra-**R**ed **(IR)** transmitter and IR receiver. Using pairs of the IR transmitter and IR receiver it is possible to enhance the exactness. Fig.6 depicts the arrangement for the counter section using pairs of IR transmitter and receiver.

Whenever object cuts the link, logic high is generated at the output. With these essential features the ALM-RTC system implements the directional intelligent counting system. Whenever the object moves in forward direction one specific combination is executed which increments the present value, but if the reverse action is performed then the combination generated at the IIR receivers will be different and hence the count value remains unchanged.

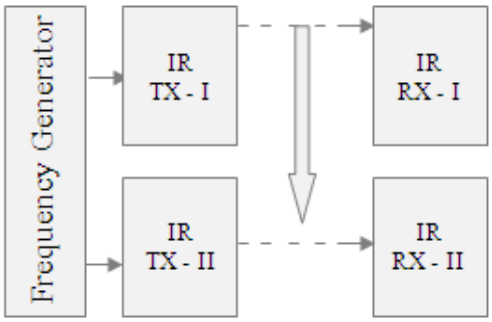


Figure 6. Counting Section

7. System Operation:

Assuming single person is moving inside the arena under surveillance, fig.7 shows basic three sensor assembly of motion sensors which detects the movement of a person.

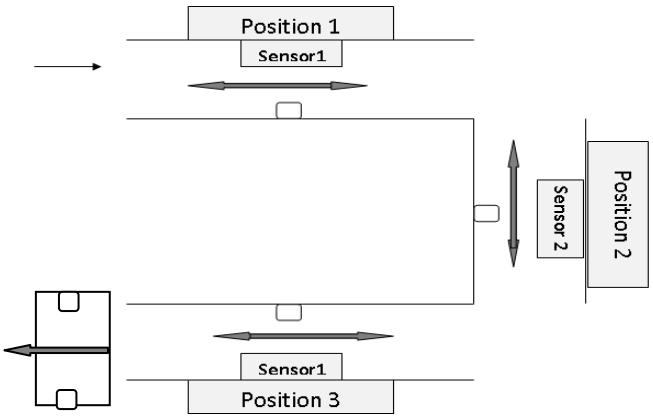


Figure 7. System operation

For the said arrangement, entrance light is kept on round the clock that maintains the necessary illumination at position 1. Person may move to position 2 or even it may come out of the arena under surveillance. If a person moves to position2, sensor2 is activated and respective light is operated.

From this current position person may move to position1 or position3. When a person moves to position3, sensor3 is activated and accordingly light3 is operated.

Out detection mechanism is used to detect final movement of the person. When final out sensor is activated all the lights, except entrance light, are switched off.

8. Observations:

```
Library IEEE;
Use IEEE.std_logic_1164.all;
Entity lmu is
Port(clk, reset: in std_logic;
sensor1, sensor2, sensor3, person_out : in std_logic;
light1, light2, light3: out std_logic);
end lmu;
architecture test of lmu is
signal state: integer range 0 to 2;
begin
process(clk, reset, sensor1, sensor2, sensor3, person_out)
begin
        if reset = '1' then
        state <='0';
        light1<= '0'; light2<='0';light3<='0';
        elsif clk = '1' and clk'event then
                case state is
                when 0 =>
                light1 <= '1';
                        if sensor2 = '1' then
                        state <= 1;
                        end if;
                when 1 =>
                light2 <= '1';
                        if sensor1 = '1' then
                        light2<= '0';
                        state <= 0;
                        elsif sensor3 = '1' then
                        light2 <= '0';
                state <= 2;
                        end if;
                when 2 =>
                light3 <= '1' ;
                        if sensor2 = '1' then
                        light3 <= '0';
                        state <= 1;
                elsif person_out = '1' then
                        light3 <= '0';
                        state <= 0;
                        end if;
                end case;
        end if;
end process;
end test;
```

Fig.8 shows the results for the above considerations. For multi user, moving under the surveillance arena, counter can be used which is incremented at the entrance and decremented by the persons coming out of the arena.

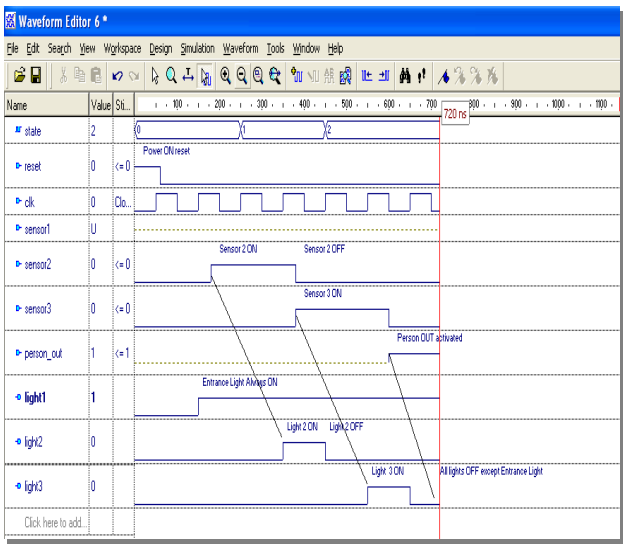


Figure 8. Results of single user system

9. Proposed Algorithm:

For the Advanced Automatic Luminance Management and Real Time Counting System to display the event counted data and other parameters 16x2 character LCD display is used. These LCDs have higher set-up and mode selection timing issue and therefore require to be initialized at very early stages.

Soon after that sensor unit and counter section can be initialized to detect the position of the person. This information is given to the programmable gate arrays. Counter section also gives exact value of the persons entered, in order to operate the bank of white LEDs. PGA drives the bank of white LEDs and also controls the display unit which modifies the count values being displayed. Fig.9 shows the proposed algorithm of the ALM-RTC system.

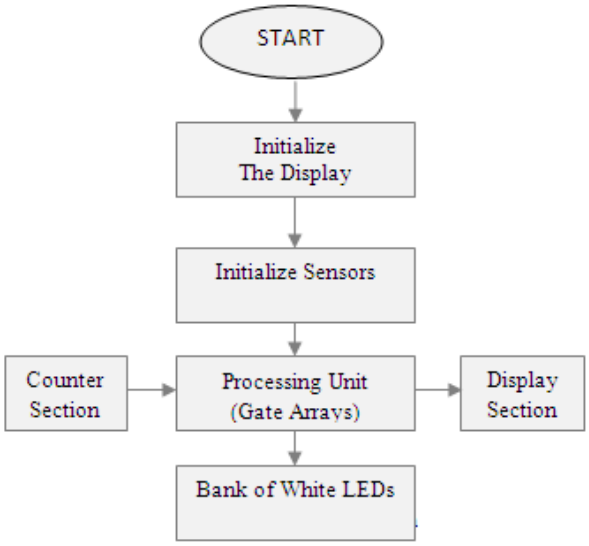


Figure 9. The proposed algorithm of the ALM-RTC system

10. Conclusions:

This system is the novel concept to manage the available power resources. From the results shown in fig.8 and discussion under the section 7and8,iit is evident that, in ideal case, only one light unit is operating and in any pose only two units of the light will be operating, hence it manifest power saving up to 30% directly.

Power can be managed further by considering different parameters like average age of the visitors, external and internal environmental parameters like Natural Light intensity and Natural Air, possible season like winter, summer etc.

By taking benefits of the Back End developments in semiconductor technologies, this system implements the excellent power management system at the Front End – can be considered as the system/application that takes real benefit of the Back End development.

11. References:

[1] Kuan Jen Lin, Yi Tang Chiu and Shan Chief Fang, "Design Optimization and Automation for Secure Cryptographic Circuits", 22nd International conference on VLSI, 5-9 Jan.2009 , New Delhi, India.

[2]FPGA prototyping by VHDL Examples: Xilinx Spartan-3 Version, e-book, 2008.

[3] Arifur Rahman, "FPGA based design and applications", Springer edition, July 2008.

[4] Morris Mano, "Digital Logic and Computer Design-Advanced Digital Design fundamentals and Issues" Prentice Hall , 2003.

[5]http://energypriorities.com/entries/2007/06/what_is_green_it_data_centers.php.

[6]http://timeforchange.org/what-is-a-carbon-footprint.

[7]http://www.xilinx.com Preliminary product specification DS077-1(v1.0).

[8] http://www.xilinx.com DS312.

[9]http://www.xilinx.com, JTAG Programmers Guide.

[10]http://www.alldatasheet.com/datasheet-pdf/pdf/197436/XILINX/XCF01S.html.

[11] http://www.national.com In System Programming Communication Protocol, revision 2..

[12]http://apps1.eere.energy.gov/buildings/publications/pdfs/ssl/led_basics.pdf.

Author Information

Mr. Jitendra S. Edle is currently working as a lecturer in Electronics and Telecommunication Engineering, Sipna College of Engineering, Amravati (India) since 2009. He is also persuing his M.E. in Electronics from the same institute. His areas of interest are Digital System Design, VLSI Design and Green IT.

Prof. Ajay P. Thakare is currently working as Head of Electronics and Telecommunication Engineering Department, Sipna College of Engineering, Amravati (India) since inception. He is also persuing his Ph.D. in the faculty of Electronics Engg. From SGB Amravati University, Amravati (India). His areas of interest are Digital System Design, VLSI Design , Green IT and Embedded Systems.

Prof. A. M. Agarkar is currently working as Assistant Professor in Electronics and Telecommunication Engineering Department, Shri Sant Gajanan College of Engineering , Shegaon (India). He is also persuing his Ph.D. in the faculty of Electronics Engg. From SGB Amravati University, Amravati (India). His areas of interest are Digital System Design, VLSI Design , Green IT , Embedded Systems and Fiber Optics Systems.

A Survey of Energy Efficient Geocast Routing Protocols in Wireless Ad Hoc and Sensor Networks.

[1]**Kaushik Ghosh** and [2]**Pradip K Das**

Faculty of Engineering and Technology
Mody Institute of Technology and Science, Lakshmangarh, India.

[1] kghosh.et@mitsuniversity.ac.in, [2] pkdas@ieee.org

***Abstract*:** Geocast routing protocols are very common in Wireless Ad Hoc Sensor Networks (WASN). The resource constrained environment of WASN demands the protocols to be as energy efficient as possible. Over the years a number of geocast routing protocols have been proposed of which quite a handful are energy efficient. In this paper we attempt to give a detailed study of some of those energy efficient geocast routing protocols. Of the different protocols under discussion, some have tried to reduce the energy expenditure by reducing the number of transmissions while others have focused on reducing the effective transmission distance in order to reduce the energy consumption for the network.

***Keywords*:** Wireless Ad hoc Sensor networks, Geocasting, Beaconless Routing.

1. Introduction

Geocasting, a special variant of multicasting, has become sort of a default choice as the mode of transmission in **Wireless Ad Hoc Sensor Networks** (WASN). Here, unlike multicasting the node id of the traversed nodes are not checked to see whether or not the target nodes are reached. Rather, some specified geographic areas are considered as the *target geocast regions* and the routing protocols thus designed concentrate on the location of the destination for delivering data packets and not on node id. Geocast routing protocols are concerned till the different target regions are reached. What happens to the data packet thereafter is however beyond the scope of many of these protocols. Needless to say that the term energy efficiency is amongst the many performance parameters, which comes into mind while one thinks of WASNs. According to Frii's free space equation,

$$P_r = P_t * (G_t * G_r * \lambda^2) / (16 * \pi^2 * d^2 * L)$$

where P_r and P_t are the power required for receiving and transmitting respectively over a distance of d. In [7] however we get a detailed equation for determining the energy required for transmitting and receiving via the following equations:

$$E_{TX}(m,d) = m * E + m * \epsilon * d^2$$

$$E_{RX}(m) = m * E$$

Where,

E= 50nJ/bit and *ε= 10 pJ/bit /m^2.*
E_{TX}=Energy consumed for transmission.
E_{RX}=Energy consumed for reception.
d=distance between the transmitting and receiving node.
ε=Permittivity of free space.
m=Number of bits.

From the above equations it is clear that energy consumed for both transmission and reception are directly proportional to the total number of bits transmitted. It is clear that the actual payload needs to remain unaffected in the quest of attaining a routing technique which gives a descent performance by saving overall energy of the network and thereby enhancing the lifetime of the network. The only way therefore to curb the number of transmissions is by reducing the number of control packets transmitted. Although its true that traditional geocast routing protocols don't have a complete view of the network topology and in order to keep track of their 1-hop neighbours they rely upon beacon messages; over the years a number of beaconless routing protocols have been proposed for WASNs, which not only has reduced the communication overhead but also has reduced the total energy consumption of the network. The transmitting energy $\mathbf{E_{TX}}$ being a function of both **m** and **d** demands the transmitting distance be reduced by some mechanism or the other for reducing the energy expenditure of the network. After a detailed survey of the literatures, we subdivided all those protocols under two broad heads: i) those attaining energy efficiency by **reducing the number of transmitting bits/packets** without taking a toll on the performance of the network and ii) others by **reducing the transmitting distance**. Genre i) can again be subdivided into a) protocols **reducing redundant transmissions** and b) **beaconless routing protocols**. Genre ii) also is subdivided as a) protocols reducing energy **without considering nodal movement** and b) protocols **considering nodal movement**. There can again be some **hybrid** protocols which may combine both approaches i) and ii) to attain the same goal.

This hybrid protocols are however kept out of the domain of this paper.

In this paper we have discussed some of the works which may not be out and out an energy efficient protocol, but the basic philosophy of those works had nevertheless a positive impact on reducing the energy expenditure of a WASN under consideration. Here we have used the term geographic routing interchangeably with geocast routing since we feel that every geocast routing protocol has to qualify as a geographic routing protocol first. The converse however is not always true.

In section 2 we have discussed about the different energy efficient geocast routing protocols and have concluded in section 3 with a note of future works.

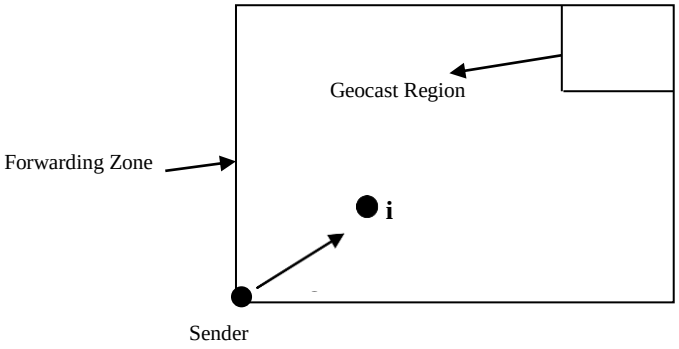
2. ENERGY EFFICIENT ROUTING PROTOCOLS FOR WASN:

As discussed earlier, the geocast routing protocols have been first divided under two broad heads with further two subdivisions under each of them. In section 2.1 we have discussed about the protocols which conserves energy by reducing the number of transmissions and under section 2.2 we have discussed the protocols taking care of energy conservation by reducing the transmission distance.

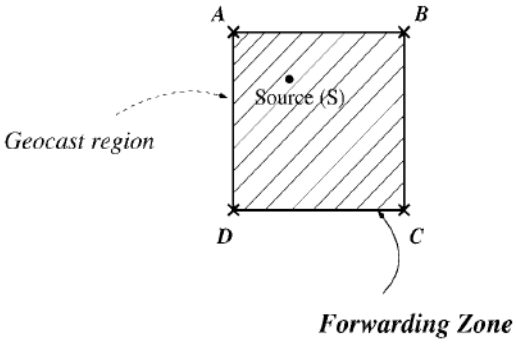
2.1. Transmission Reducing Routing Protocols: As discussed earlier, this category of routing protocols are further subdivided as a)**Redundant Transmission Reducing Routing Protocols** and b)**Beaconless Routing protocols**.

2.1.1. Redundant Transmission Reducing Routing Protocols: One of the very first protocols of this kind was proposed by the authors in **LBM** [8] and **LAR** [9]. The present day geographic and hence geocast routing protocols can well be termed as their successors. In [9] the concept of a *forwarding zone* had actually revolutionized the concept of flooding. It suggests, if a node is to deliver a packet to a certain geocast region, then not all the intermediate nodes take part in the forwarding process (unlike flooding) but the ones within the forwarding zone of a node do (like selective flooding). That is, if an intermediate nodes finds itself to be outside the forwarding zone defined by the sender or its immediate predecessor node, then it refrains itself from forwarding the packet. As a result, a smaller set of nodes take part in the process of forwarding as compared to flooding which eventually helps in reducing the energy expenditure of the network. However, there is a trade off between the accuracy of the protocol and energy consumption. The protocol acts with more accuracy for a larger forwarding region. The forwarding zone talked about may or may not be equal to the geocast region itself depending upon the position of the source node (Figure 1). It may again be either be **static** or **adptive.** If static, the boundary of the zone as defined by the source node doesn't change in course of the packet transmission. Whereas, in case the forwarding zone is considered adaptive, it changes with every intermediate transmission which reduces the number of redundant transmission to an even larger extent. For example, in the static zone scheme, the forwarding region defined as SCBA remains unmodified throughout the packet forwarding from the node S till it reaches the geocast region. But if the scheme is adaptive, then as the packet is forwarded to node I, the new forwarding zone becomes a rectangle of smaller dimension with I as one of the vertex. The forwarding zone goes on becoming smaller with successive forwarding which reduces the number of redundant transmissions. The authors in [8] have emphasized on the accuracy of packet delivery as the performance parameter for the protocol but nevertheless it was a major step forward in reducing energy consumption of the network when compared with flooding.

In [9] the same authors used location information of the nodes to improve the routing performance. They had put forward the notion of *expected zone* and *request zone* whereby the numbers of transmissions are reduced. A node S can determine the expected zone for a node D at time t_1 based on the location information of D at time t_0 ($t_1 > t_0$) and its velocity v. In fact it is the region where S expects D at time t_1. In case node S do not have the location information then the entire network region is considered as expected zone. In this case the algorithm reduces to flooding.



(a)



(b)

Figure 1 (a & b). Relationship between forwarding zone and geocast region.

The request zone is the region where a node searches for a route to the destination during the route request phase. A node forwards a route request packet only when it is within the request zone. In order to ensure that the route request packet will reach the destination, the expected zone is included within the request zone (Figure 1c).

2.1.2. Beaconless Routing Protocols: The second kind of transmission reducing protocols is the beaconless routing protocols. As the name goes, this kind of protocols does not

transmit beacon messages. By doing so, they assure energy consumption because the number of transmissions are reduced. Many of the beaconless routing protocols talk of a *forwarding area*. This forwarding area is a *reuleaux triangle* within the *greedy area* of a node S currently holding the data packet (Figure 2). Nodes within the forwarding region are the only ones to reply back to node S in its search for the next hop.

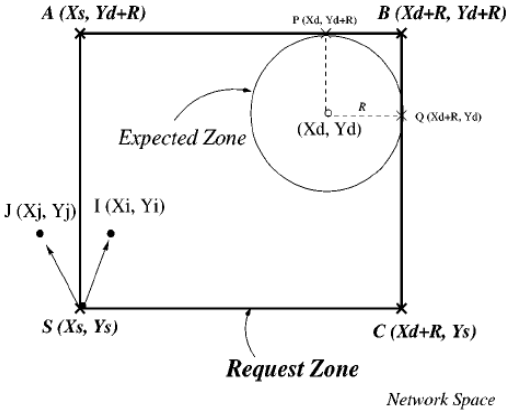


Figure 1c. Request Zone

Implicit Geographic Forwarding (IGF) [21] is one of the oldest beaconless routing protocols to have used this concept of forwarding area.

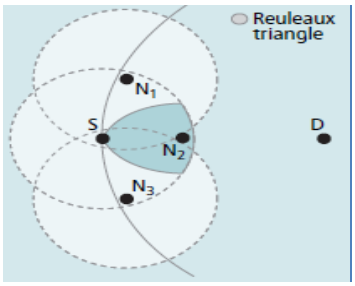


Figure 2. Forwarding area.

In IGF while replying, the nodes within the forwarding region set a timer and actually reply when the timer has expired. The value of the timer thus set depends upon the distance of the node from the destination D. The farther the node from D, greater is the value of the timer set. This in a way ensures that the node within the forwarding area with least distance from D becomes the one to reply first. The other candidate nodes within the forwarding area overhearing this reply clear their timers and cancel their reply. This way two basic philosophies of the geographic routing protocols are maintained: advancement in each hop and greedy forwarding [10]. The concept of forwarding area was put forward in order to ensure that all the candidate nodes can overhear each other while replying to the node S in order to suppress redundant transmissions. In [21] the forwarding area is so defined that all candidates within the area are separated by a distance less than the radio range of each node. Inclusion of the nodes within the greedy area as candidates however increases the probability of getting suitable nodes for relaying the message to node D. In **EEBGR** [22] the authors introduce concept of *optimal forwarding distance* ($\mathbf{d_o}$) and thereby marks a circular region around d_o to define it as the *relay search region* $\mathbf{R_u}$ for a forwarding node **u** (Figure 3).

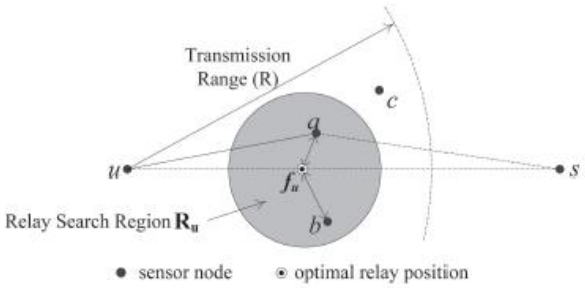


Figure 3. Optimal relay position & relay search region.

As the name implies, the search of relay node(intermediate forwarding node) for node u would thus remain confined within the said region only. Based on the findings in [23], the authors in [22] advocate that, if the following lemma holds good

$$d \leq \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}},$$

then direct transmission is the most energy efficient way to transmit packets from source to destination. Where d= distance between the nodes, k=propagation loss exponent, α_2 is the transmitting amplifier and $\alpha_1=\alpha_{11}+\alpha_{12}$. α_{11} is the energy spent by the transmitter α_{12} is energy spent by receiver. In case the above lemma doesn't satisfy, then the transmitting energy would be minimum when all hop distances are equal to

$$\left\lfloor \frac{d}{d_o} \right\rfloor \text{ or } \left\lceil \frac{d}{d_o} \right\rceil$$

Where

$$d_o = \sqrt[k]{\frac{a_1}{a_2(k-1)}}.$$

From the above lemma, the authors conclude that d_o is the optimal forwarding distance in terms of minimizing the transmitting energy when d is an integral multiple of d_o. Even if d cannot be divided by d_o, the authors argue that d_o is a good approximation of optimal forwarding distance. Although the shape of the area for searching relay nodes is different, both [21] and [22] rely on RTS/CTS mechanism to find the suitable candidate node. A RTS packet is first sent by the forwarding node and it then waits for a CTS response from one of the candidate nodes. The candidate nodes on receiving the RTS message set up a timer which is a function of their distance from the receiving node [21]. This way it is guaranteed that a candidate located nearer to the destination replies with CTS message earlier as compared to the one located farther. Other candidate nodes on over hearing this transmission resets their timer and cancels their respective CTS transmissions. In [22] a *delay function* has been proposed in addition to determine the time each candidate node would take before replying to the RTS transmission. The delay proposed here is dependent upon the distance of a candidate node from the optimal relay position which is situated at a distance d_o from the forwarding node. To find the delay function, the relay search region is subdivided into n concentric coronas of equal areas (Figure 4). If r_1 is the radius of S_1, the width of the i^{th} corona is given by

$(\sqrt{i}-\sqrt{i-1})r_1$

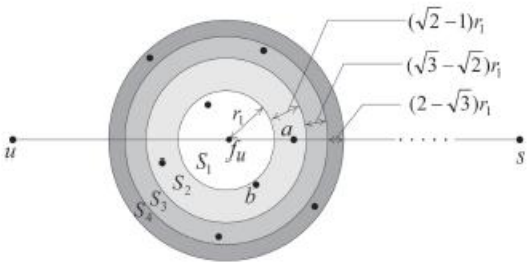


Figure 4. Concentric coronas for determining delay function.

The authors argue that if γ be the delay for transmitting a CTS packet over unit distance, then the delay for transmitting a CTS packet by node v within R_u to node u is given by

$$\frac{\gamma \cdot r_s(u)}{\sqrt{n}} \cdot \left(2\sum_{i=1}^{m}\sqrt{i}-\sqrt{m}-1\right)+\gamma \cdot \left(|vf_u|-\sqrt{\frac{m-1}{n}}r_s(u)\right),$$

where

$$m=\left\lfloor\left(\frac{\sqrt{n}\cdot|vf_u|}{r_s(u)}\right)^2\right\rfloor+1.$$

The deduction and verification of the above expressions are however beyond the scope of this paper.

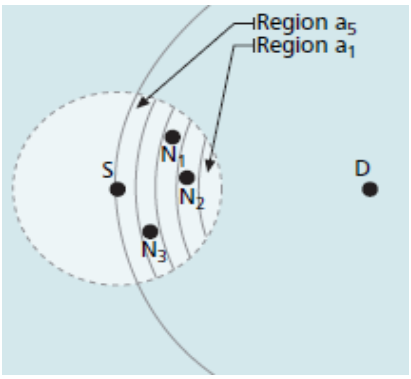


Figure 5. GeRaF

Geographic Random Forwarding (GeRaF) [24] is a beaconless routing protocol that considers all the nodes situated closer to the destination as compared to the forwarding node as the candidates for next hop relay i.e. it considers the entire greedy area for candidate selection and not only the forwarding area. This way it is different from both [21] and [22]. The greedy area is logically divided as shown in Figure 5. Nodes in region a_1 are the ones to reply with a CTS message first as compared to the ones situated in regions a_2 a_3 or a_5. If the transmitter successfully receives a CTS message, it issues a data packet containing the payload and a header indicating the identifier of the neighbour selected as next hop, that is, the one who's CTS was received first. If the transmitter does not correctly receive a CTS frame within the CTS slot, then the data packet issued indicates a collision, and all the nodes in the same region decide whether to send another CTS or not with probability 0.5. If no node answers during the CTS slot, the transmitter indicates in the message that nodes in the next region must answer because there are no neighbours in the first one.

In **Contention Based Forwarding** (CBF) [25] however there are two phases: contention phase and suppression phase. Although in [22] also there are two phases like beaconless greedy forwarding and beaconless recovery phase, qualitatively [22] is different from [25]. In [22] the beaconless recovery phase is for recovering from *local minimum* condition; a topic we consider superfluous for discussing in the present context of energy efficient routing protocols. The phases of [25] on the other hand are concerned with the RTS/CTS packet transmission/reply and suppressing redundant messages respectively-something having direct correlation with energy conservation.

In the contention process, the node currently holding the data packet forwards it and waits for its neighbours to determine for themselves which one will be the next relay in a *distributed contention process*. During the contention process, candidate neighbours compete for becoming the next relay by setting timers related to their actual positions. The neighbour providing the most advance toward the destination waits for the shortest time before forwarding the data packet. All the candidate forwarders cancel their timers when they hear the transmission from the winning neighbour. Here we can see that CBF follows the same philosophy as [21] in determining the next relay node. The suppression phase as said earlier is used to reduce the chance of accidentally selecting more than one node as the next hop, as well as to reduce the overhead of the protocol. Three different suppression schemes are proposed. The basic scheme consists only of cancelling timers after hearing a transmission from another neighbour. A second area-based scheme consists of defining a forwarding area as in IGF. Three different areas are studied in CBF, but the one achieving the best results is the Releaux triangle. Finally, a third suppression mechanism called active suppression is defined, which is in fact the same RTS/CTS approach proposed in IGF. It allows the forwarding node to determine which neighbour to select as next forwarder among the ones whose CTS frames were received.

In **Beacon-Less Routing** (BLR) [26] once again the shape of the forwarding area is again a Releaux triangle. The delay function computing the dynamic forwarding delay (DFD) takes the position of the sending node and and that of the destination into account. The following three delay functions were proposed.

- $Add\_delay = Max\_delay \cdot \left(\frac{r-p}{r}\right)$ (1)
- $Add\_delay = Max\_delay \cdot \left(\frac{p}{r}\right)$ (2)
- $Add\_delay = Max\_delay \cdot \left(\frac{e^{\sqrt{p^2+d^2}}}{e}\right)$ (3)

Function (1) basically implements MFR [27],
function (2) is a modification of NFP [28] and (3) implements a combination of the both to obtain a more advanced DFD function. Here also the other neighbours cancel their scheduled transmission when they overhear the

data packet. BLR also defines a recovery strategy to deal with local minima, which again we are not discussing keeping in mind the demand of the present context.

Of the different beacon less protocols for sensor networks, **BOSS** [29] differs from the rest in the fact that it sends the entire payload with the RTS packet. The authors in [29] argue that since long messages are more error prone, getting a CTS message over a link doesn't necessarily guarantee the delivery of the entire payload. By sending the data packet first, BOSS performs the neighbour selection only among those neighbours that successfully received the data packet. In fact it was designed to take into account packet losses and collisions, which are common in radio communications. It uses a three way hand shake (RTS/CTS/ACK). Along with sending the entire payload with RTS, another novelty of [29] is on defining the DFD in a discrete form called as Discrete DFD (DDFD). DDFD divides the neighbourhood area into sub-areas in a fashion similar to [24] according to the progress toward the destination. Like its predecessors, DDFD function assigns smaller delay times to the neighbors providing the maximum progress toward the destination. To do so, the authors propose to divide the neighborhood in sets of neighbours providing a similar progress . The *Number of Sub Areas* (NSA) have been defined in which the whole coverage area has to be divided and then, taking into account that the maximum difference in progress between two neighbors as *2r*, each neighbor determines in which *Common Sub Area* (CSA) it is placed. To do that it uses the following equation:

$$CSA = \left\lfloor NSA \times \frac{r - P(n,d,c)}{2r} \right\rfloor$$

P(n,d,c) is the progress of a candidate neighbour to be selected as the next forwarder where **c** and **d** are the forwarding and destination nodes respectively. Each neighbour computes its delay **(T)** according to the following equation:

$$T = \left(CSA \times \frac{T_{Max}}{NSA}\right) + random\left(\frac{T_{Max}}{NSA}\right)$$

TMax is a constant representing the maximum delay time that a forwarding node will wait for answers of its neighbors and, *random*(*x*) a function obtaining a random value between 0 and *x*.

2.2. Transmission Distance Reducing (TDR) Routing Protocols: Another set of routing protocols that can directly reduce the transmitting energy are the ones that are capable of reducing the transmitting distance. This is because, various radio models have proved that in free space the transmitting energy of a node is directly proportional to the square of the distance between it and the receiver. This class of routing protocols may again be subdivided as the ones a) which do not consider movement of the nodes to reduce the transmission distance and the others b) which do so and in fact uses mobility prediction of the nodes to reduce the effective transmission distance.

2.2.1. TDR without Node Movement Consideration: Of the different geocast routing protocols of this type, authors in [3] proposed a **geometry driven** *Fermat point* [1] based scheme which had its direct positive impact on reducing the transmitting energy. The authors considered the sender and two of the geocast regions as the three vertex of a triangle and used a pure geometric method to determine the Fermat point of that triangle (Figure 6). Then using greedy forwarding, the packet is sent to the node situated nearest to the Fermat point. From there actually, the packet is transmitted to both the geocast regions. The path thus formed is said to be of minimum possible length since the sum of the vertices of a triangle from the Fermat point is minimum. When the number of geocast regions is more than two, the authors proposes a heuristic method of finding the Fermat point for the region. This method however fails to point out the actual position for the Fermat point of a region which is polygonal. The percentage of error kept on increasing with increasing numbers of geocast regions. This issue was addressed in [4] and a **Global Minima** based scheme was proposed which invariably located the exact position of Fermat point for any number of geocast regions. In [4] the authors gave an alternative approach for finding the Fermat point of a triangular/polygonal region (Figure 7).

In fact they followed the same approach for finding the Fermat point of an enclosed region- be it triangular or polygonal. They initially took any point *P(x,y)* within the boundary of the concerned polygon and defined a function f(x,y) that gives the sum of the distances of the vertex of the polygon from P. Iff P(x, y) happens to be the Fermat point, in that case the value of the function f(x, y) would be minimum. The authors thus found out the minima for the function with the value of the coordinates for P(x, y) as unknown. With all other coordinates for the vertex being known, finding the minima for the function f(x, y) actually locates the Fermat point of the polygon/triangle with zero error. As a result, unlike [3], the error factor for finding the Fermat point don't get increased with increasing number of geocast regions. The results in [4] has shown that when

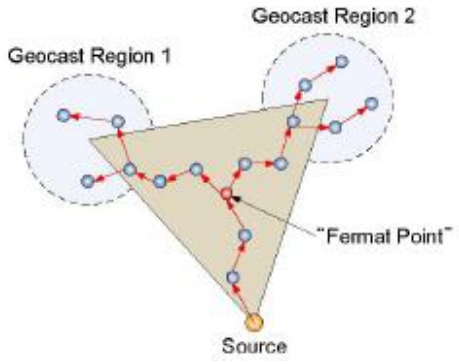


Figure 6. Geometry Driven Scheme.

the number of geocast regions is more than two, their scheme outperforms the one described in [3] when it comes to energy conservation. In fact with increasing number of geocast regions, the energy consumption for a said sensor network reduces exponentially while one uses the scheme in [4] in place of [3]. Based on the findings of [3] in [20] the authors proposed **I-Min** algorithm, which is basically a modification of [3] only. But the I-min algorithm scored over [3] in terms

of energy conservation due to the fact that it was loop free. It there by eliminated the redundant transmissions/receptions and saved the related energy expenditure thereby. Figure 8 gives a comparative result for the schemes [3], [4] and [20].

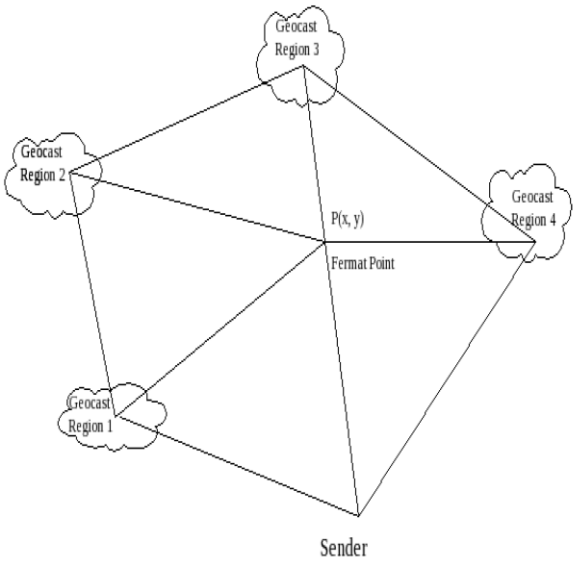


Figure 7. Finding the Fermat point using Global Minima Scheme.

2.2.2. TDR with Node Movement Consideration: In [30] the approach taken for distance reduction is somewhat different. Here, the authors have taken nodal movement in their stride to delay an expected transmission such that the transmission distance can be reduced. They have proposed a general stochastic movement model to capture realistic movements. Following the system model in [32] the authors in [30] have divided a said geographic area into a grid of fixed size cells *(mxn)* and each cell is a square of size *sxs*. The movement model variables are given in Table 1. The model is a quadruple of *network configuration*, *accessibility constraints*, *trip based movement model* and *route selection algorithm.* With this they have derived tight lower bound expected communication distances as a function of average nodal speed and allowable postponement delay. It provides the best achievable communication distance between a mobile node and a receiver.

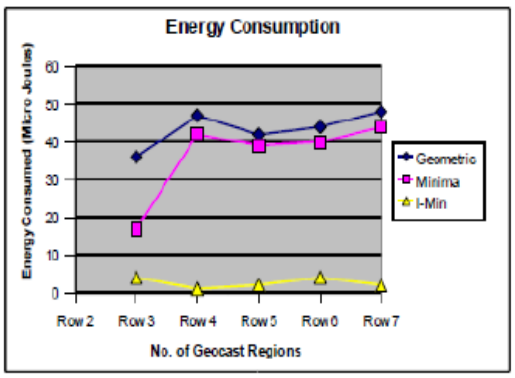


Figure 8. Comparison of Geometry driven, Minima and I-Min schemes.

The authors have actually shown how closely the *least distance* algorithm [31] can match the theoretical optimum. In fact the least distance algorithm works on 37% rule. According to this rule the first 37% candidates are evaluated but not accepted. After that the candidate whose relative rank is best among the candidates seen so far, is selected. It has been shown in [30] that the least distance algorithm can achieve an average reduction in the expected communication distance between 62% to 94% of the optimal. The algorithm in [30] improves its performance with increase in nodal speed or the allowable delay for packet reception. Other than providing the tight lower bound for expected communication distance, the authors in this paper has also proposed a movement prediction algorithm. Strictly speaking, the movement prediction algorithm talked about in [30] is basically the least distance algorithm of [31].

Minima Algorithm

Input: Coordinates of the sender node and that of different geocast regions.
Output: (fx, fy): Coordinates of the Fermat_Point.
Tdist : Total distance traveled by the packet.
Tpow: The sum of power consumed by all the intermediate nodes to forward the packet to m geocast regions

```
1.  max_x=MAX_x (Sx, GRX(N));
2.  max_y=MAX_y (Sy, GRY(N));
3.  min_x=MIN_x (Sx, GRX(N));
4.  min_y=MIN_y (Sx, GRY(N));
5.  dx=0;   /*          Initialize dx */
6.  dy=0     /* Initialize dy */
7.  flag=0;  /* To check the Fermat Point*/
8.  for ( i=min_x; i<max_x; i++ )
9.  {  if (flag==1) break; /*  Fermat point found  */
10. for ( j=min_y; j=max_y; j++ )
11. { x=i;
12. y=j;
13. for ( k=0;k<n;k++ )
14. { dx+= termdx ( x,GRX(k),y,GRY(k) );
15. dy+= termdy ( y,GRY(k),x,GRX(k) );
16. } /* end of for loop (line-11)*/
17. if  ( dx==0 && dy==0 )
18. { flag=1;  /* Fermat point found */
19. break;
20. }
21. dX=0; dY=0;
22. } /* end of for loop (line-8) */
23. }/* end of for loop (line - 6) */
24. if ( flag==1 )
25. { fx=x; fy=y; } /* Fermat point */
26. Tdist= Total_Dist ( Sx, Sy, GRX(N), GRY(N), fx, fy );
27. Tpow= Total_Pow ( Tdist )
```

Figure 9. Minima algorithm for finding the Fermat point.

In [31] a number of several postponement algorithms were proposed of which the least distance has the best practical performance. What authors in [30] have done is to track the *first 37% or more* of the candidate positions in movement history of the mobile node and find the least distance between the sender and the receiver in the movement history. After that in each of the D time units they kept on checking the current distance between the sender and the receiver. If that happens to be less than or equal to the *minimum distance*, communication is made, else communication is done on time D. Here D specifies the maximum allowable delay. At this specific point, there is hardly any difference between the exercise carried out in [31] and [30]. At the end the authors have compared the least distance algorithm of [31] with the theoretical lower bound of the expected

communication distance to provide an absolute measure of performance for the algorithm.

I-Min Algorithm

```
Input: Node_id of the destination.
Output: Boolean.
D_ID: Node_id of the destination.
NN: Number of neighbors of the forwarding
node.
Neigh_ID= Node_id of a neighbor.

1. if (flag==0)
2. next_hop=MAX_DIST(D_ID, NN, neigh[]);
3. else
4. next_hop=D_ID;
5. for(i=0; i<NN; i++)
6. {if (D_ID==Neigh_ID)
7. flag=1;
8. else
9. flag=0; }
```

Figure 10. I-Min Algorithm

Since [31] is the basis of determining the least distance between a mobile node and a receiver using *postponement algorithms* and as it is a basis of the work in [30], we feel it necessary to spend a few words on the different schemes to improve routing protocol performances by using mobility prediction.

Just like [30], here also the entire network space subdivided into virtual grids. Each node is assumed to know its position using GPS. Before a node can actually communicate, its movement history is studied to predict its future position. Then it is decided whether it is sensible to wait for a specific unit of time, say k units, to postpone the communication. Because waiting for that much of time may bring the node closer to its destination and thus save the transmitting energy. In case the node does not come any closer to the destination in k-1 time units, then it would communicate on the k^{th} slot, without waiting any further. It might happen that after k-1 seconds the node is further away from the destination than its initial position. In that case it would end up consuming more energy. This, as put forward by the authors, is the *misprediction penalty*. The authors have then proposed four heuristics based on a nodes movement history to calculate : i) the probability that a node will be within a specific grid within the next k units, ii) the probability that a node will move closer to the destination (termed as *target*) and iii) the weighted average distance between the node and the target. The different heuristics used were i) binary distance heuristic, ii) binary Markov distance heuristic, iii) Markov distance heuristic and iv) average distance heuristic. All the heuristics were used to decide whether to postpone a communication till sometime or not. Informed by the *secretary problem*, [33] the above mentioned heuristics were restructured into the following three heuristics; i) directional and recursive average distance heuristics, ii) least distance heuristics and iii) heuristics in the case of moving target. The authors have shown through experimental results that of the three the least distance is the simplest and most effective. The least distance heuristics is based on the 37% rule, which has been discussed earlier.

[32] is one among the earliest works on mobility prediction of the nodes. The authors defined a function D_t based on the motion parameters of the node along with their present coordinates. D_t gives the amount of time two mobile nodes will stay connected. They termed this time as the *Link Expiration Time* (LET). A node can however, determine LET of its neighbors by periodically measuring the transmission power samples from that neighbor. The time this power drops below a threshold can also be taken as the LET between the nodes. Finding the LET, the authors had used it as a metric for route selection in a number of unicast and multicast routing protocols like FORP, ODMRP and LAR. Although the performance comparison done by the authors never included energy as a metric yet the LET can be used as metric for deciding upon whether or not to defer the transmission as it gives a direct measurement of the distance between a sender and a receiver.

Table 1. Movement model variables.

Variable	Definition	Type
X	width of the network area	input parameter
Y	height of the network area	input parameter
s	cell size	input parameter
V_{max}	maximum nodal speed	input parameter
V_{min}	minimum nodal speed	input parameter
V	nodal speed	random variable
T	travel time	random variable
$E[V]$	expected speed	statistical property
$E[T]$	expected trip time	statistical property

3. CONCLUSION

In this paper we have tried to broadly classify some selected geocast routing protocols for WASNs from the available handful. The protocols may well be classified under smaller heads as well but that may not necessarily affect a paper of the present kind qualitatively. Many other works demands inclusion in discussion of similar kind but had to be left alone as a part of some future work.

References

[1] The Fermat point and generalizations: http://www.cut-the-knot.org/generelization/fermat_point.shtml.

[2] Young-Mi Song, Sung-Hee Lee and Young-Bae Ko"*FERMA: An Efficient Geocasting Protocol for Wireless Sensor Networks with Multiple Target Regions*" T. Enokido et al. (Eds.): EUC Workshops 2005, LNCS 3823, pp. 1138 – 1147, 2005. © IFIP International Federation for Information Processing 2005.

[3] S.H. Lee and Y.B. Ko "*Geometry-driven Scheme for Geocast Routing in Mobile Adhoc Networks*", IEEE Transactions for Wireless Communications, vol.2, 06/ 2006.

[4] Kaushik Ghosh, Sarbani Roy and Pradip K. Das, "*An Alternative Approach to find the Fermat Point of a Polygonal Geographic Region for Energy Efficient Geocast Routing Protocols:Global Minima Scheme"*, AIRCC/IEEE NetCoM 2009.

[5] Lynn Choi, Jae Kyun Jung, Byong-Ha Cho and Hyohyun Choi, "*M-Geocast: Robust and Energy- Efficient Geometric Routing for Mobile Sensor Networks*", SEUS 2008, LNCS 5287,pp. 304–316, IFIP (International Federation for Information Processing) 2008.

[6] W.B. Heinzelman, A.P. Chandrakasan and H. Balakrishnan, "*An application-specific protocol architecture for wireless micro sensor networks*", IEEE Transactions on Wireless Communications, Vol.1, Issue 4, October 2002.

[7] I-Shyan Hwang and Wen-Hsin Pang, "*Energy Efficient Clustering Technique for Multicast Routing Protocol in Wireless Adhoc Networks*", IJCSNS, Vol.7, No.8, August 2007.

[8] Young-Bae Ko and Nitin H. Vaidya "*Geocasting in Mobile Ad Hoc Networks: Location-Based Multicast Algorithms* ", Proceedings of the

Second IEEE Workshop on Mobile Computer Systems and Applications (WMCSA), February 1999.
[9] Y.B. Ko and N. H. Vaidya, "*Location-aided routing (LAR) in Mobile Ad Hoc Networks*", ACM/Baltzer Wireless Networks (WINET) journal, vol. 6, no. 4, 2000, pp. 307-321.
[10] B.Karp and H.Kung, "*GPSR: Greedy perimeter stateless routing for wireless networks*", ACM/IEEE MobiCom, August 2000.
[11] C.Maihofer, "*A survey of geocast routing protocols*", IEEE Communications Survey & Tutorials,vol.6, no.2, Second Quarter 2004.
[12] T Camp, Y Liu, "*An Adaptive Mesh Base Protocol for Geocast routing*", Journal to Parallel and Distributed Computing", Volume 63, Issue 2, February 2003.
[13] T-F Shih and H-C Yen,"*Core Location-Aided ClusterBased Routing Protocol for Mobile Ad hoc Networks*", 10th WSEAS International Conference on Communications.
[14] J. Zhou, "*An Any-cast based Geocasting Protocol for MANET", ISPA 2005, LNCS 3758, pp.915-926, 2005.*
[15] Y.B.Ko and N.H.Vaidya, "*Anycasting based protocol for geocast service in mobile ad hoc networks*", Computer Networks Journals, vol.41, no. 6, 2003.
[16] A.K.Sadek, W.Yu and K.J.Ray Liu, "*On the Energy Efficiency of Cooperative Communications in Wireless Sensor Networks",* ACM transactions on Sensor Networks, vol. 6, N0.1: December 2009.
[17] Y.Dong, W-K Hon, D.K.Y. Yau, J-C Chin, "*Distance Reduction in Mobile Wireless Communication;Lower Bound Analysis & practical Attainment",*IEEE transactions on Mobile Computing, vol. 8 no. 2, February 2009.
[18] M. Zorzi and R R Rao, "*Geographic Random Forwarding(GeRaF) for Adhoc & Sensor Networks:Energy & Latency Performance",*IEEE transactions on Mobile Computing, vol.2 no.4, October-December 2003.
[19] A.G. Ruzzelli, G.'Hare and R. Higgs, "*Directed Broadcast eith Overhearing for Sensor Networks*", ACM Transactions on Sensor Networks, Vol.6, No.1, December 2009.
[20] Kaushik Ghosh, Sarbani Roy & Pradip K. Das *" I-Min: An Intelligent Fermat Point Based Energy Efficient Geographic Packet Forwarding Technique for Wireless Sensor and Ad Hoc Networks*", International journal on applications of graph theory in wireless ad hoc networks and sensor networks (GRAPH-HOC) Vol.2, No.2, June 2010.
[21] B. Blum *et al*., "*IGF: A State-Free Robust Communication Protocol for Wireless Sensor Networks*," *Tech.Rep*., Dept. Comp. Sci., Univ. of VA, 2003.

[22] Haibo Zhang and Hong Shen, "Energy-Efficient Beaconless Geographic Routing in Wireless Sensor Networks", IEEE transactions on Parallel and Distributed Systems, Vol.21 No.6, June 2010.

[23] I. Stojmenovic and X. Lin, "Power-Aware Localized Routing in Wireless Networks," Proc. 14th Int'l Parallel and Distributed Processing Symp. (IPDPS), p. 371, 2000.
[24] M. Zorzi and R. R. Rao, "*Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Energy and Latency Performance*," *IEEE Trans. Mobile Comp*.,vol. 2, no. 4, 2003, pp. 349–65.

[25] H. Füßler *et al*., "*Contention-Based Forwarding forMobile Ad Hoc Networks*," *Ad Hoc Net*., vol. 1, no. 4, 2003, pp. 351–69.

[26] M. Heissenbüttel *et al*., "*BLR: Beacon-Less Routing Algorithm for Mobile Ad-Hoc Networks*," Elsevier J. Comp..Commun., vol. 27, no. 11, July 2004, pp. 1076–86.

[27] H. Takagi and L. Kleinrock, "*Optimal transmission ranges for randomly distributed packet radio terminals*," IEEE Trans. Com-mun., vol. 32, no. 3, pp. 246{257, Mar. 1984.

[28] T.-C. Hou and V. Li, \Transmission range control in multihop packet radio networks," IEEE Trans. Commun., vol. 34, no. 1, pp. 38{44, Jan. 1986.

[29] J. A. Sanchez, R. Marin-Perez, and P. M. Ruiz, "BOSS: Beacon-Less On-Demand Strategy for Geographic Routing in Wireless Sensor Networks," *Proc. 4th IEEE Int'l. Conf. Mobile Ad Hoc and Sensor Sys*., Oct. 2007, pp.1–10.

[30] Y. Dong , W-K Hon, David K.Y. Yau and J-Chit Chin, "*Distance Reduction in Mobile Wireless Communication: Lower Bound Analysis and Practical Attainment*" , IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 8, NO. 2, FEBRUARY 2009.

[31] S. Chakraborty, Y. Dong, D.K.Y. Yau, and J.C.S. Lui, "*On the Effectiveness of Movement Prediction to Reduce Energy Consumption in Wireless Communication*," IEEE Trans. Mobile Computing, vol. 5, no. 2, Feb. 2006.

[32] W. Su, S. J. Lee, M. Gerla. *Mobility Prediction and Routing in Ad Hoc Wireless Networks*. International Journal of Network Management, Wiley & Sons, 2000.

Credits:

Figures:

1. Y.B. Ko and N. H. Vaidya, "*Location-aided routing (LAR) in Mobile Ad Hoc Networks*", ACM/Baltzer Wireless Networks (WINET) journal, vol. 6, no. 4, 2000, pp. 307-321.
2. J. A. Sanchez, P. M. Rafael and R. M. Perez, "*Beacon-Less Geographic Routing Made Practical;Challenges, Design Guidelines and Protocols"*, IEEE communications Magazine, August 2009.
3. Haibo Zhang and Hong Shen, "Energy-Efficient Beaconless Geographic Routing in Wireless Sensor Networks", IEEE transactions on Parallel and Distributed Systems, Vol.21 No.6, June 2010.
4. Do.
5. Same as 2.
6. S.H. Lee and Y.B. Ko *"Geometry-driven Scheme for Geocast Routing in Mobile Adhoc Networks*", IEEE Transactions for Wireless Communications, vol.2, 06/ 2006.

Author Biographies

Kaushik Ghosh is an Assistant Professor in the Department of Computer Science & Engineering, Mody Institute of Technology and Science, Lakshmangargh(Sikar), India. He received his Bachelor of Engineering degree in Electrical & Electronics Engineering from Sikkim Manipal Institute of Technology and has done his M.Tech in Computer Technology from Jadavpur University, India. His area of research includes Mobile Adhoc Networks (MANETs) and Sensor Networks.

Pradip K Das is a Professor and Dean in the Faculty of Engineering & Technology, Mody Institute of Technology & Science, Lakshmangargh(Sikar), India. Earlier, he was a Professor and Head in the Department of Computer Science & Engineering and the founder Director of the School of Mobile Computing & Communication, Jadavpur University, India. He received is B.E. and M.E. degrees in Electronics and Ph.D (Engg.) in Computer Engineering from Jadavpur University.

Evaluation of the Effect of Gen2 Parameters on the UHF RFID Tag Read Rate

Jussi Nummela, Petri Oksa, Leena Ukkonen and Lauri Sydänheimo
Tampere University of Technology
Department of Electronics, Rauma Research Unit
Kalliokatu 2, FI-26100 Rauma
FINLAND
jussi.nummela@tut.fi

Abstract: The UHF RFID technology is nowadays widely spreading in different industries, e.g. in the supply chain and warehouse management of different goods among others. This paper presents passive UHF RFID tag read rate measurements using the EPCglobal Class 1 Gen 2 protocol. Tag encoding schemes such as FM0 and Miller, and Gen2 parameters such as *Tari* and *Backscatter Link Frequency* are varied to find out their effect on the tag's reading rate. Relating to these parameters alterable in the tag reader, this paper provides certain achievements which can be exploited in industrial RFID configurations and setups.

Keywords: *radio frequency identification, RFID, passive tag, read rate, tag encoding, Gen2 parameters*

1. Introduction

The UHF RFID technology has become more and more common in different industries during the last years. For example supply chain management of paper reels is one of the latest applications, as well as small scale positioning, in which promising results have been achieved. [1, 2]

EPCglobal was formed in 2003 to carry on the non-profit standard framework supporting use of RFID in the supply chain and in other applications. As part of this mission, it sought to create a single global standard for the UHF RFID reader-tag air interface. This second-generation (Gen2) standard became publicly in 2005, and today numerous vendors are producing reader and tag models supporting EPCglobal Class 1 Gen2 -standard. [3, 4]

Any passive RFID protocol must provide certain basic functions:

- *Data Modulation:* a set of waveforms that is understood as symbols by tags and readers.
- *Packet Structure:* preambles, training symbols, and timing conventions that enable tags and readers to synchronize and to recognize commands and data.
- *Command Set:* commands and responses that make it possible to read (and write) information stored in tags.

This paper deals with *Data Modulation* issues of Gen2-protocol, and thus it is described in more detail in chapter II. Chapter III presents the measurements, where different Gen2-protocol parameters were varied and read rates observed. Chapter IV gives the measurement results and discussion, and finally Chapter V conclusion and the future work.

2. Data modulation in UHF Class 1 Gen2 - communication

Modulation is the change made in a signal in order to send information. Each RFID standard employs one modulation scheme for *Forward Link* (reader-to-tag) and another for *Reverse Link* (tag-to-reader), because of different roles of reader and tag in backscatter communication. [5, 6]

1.1. Forward Link

In Gen2-protocol reader sends signals by changing its output power level between two states, on state and off state. This is known as *amplitude-shift keying* (ASK). Gen2 uses two data symbols, Data "0" and Data "1", when former specification Class 0 had additional "Null" symbol. Both data symbols begin with an on state, followed by an off state. The length of the on state varies, but the length of the off state is always fixed, and called *Pulse Width* (PW). Symbols Data "1" and Data "0" are distinguished by varying the length of on state, in other words, the interval between off state pulses. Thus, this technique is sometimes referred to as *pulse-interval encoding* (PIE). [4, 5, 6]

The total time of Data "0" is called *Tari*. Data "1" symbol is allowed to between 1,5*Tari or 2*Tari, and the length of the Tari may vary between 6,25 µs and 25 µs. Since the data symbols of Gen2 are not of the same length, there is no fixed Gen2 Forward Link data rate. Instead, there is an "effective data rate assuming equiprobable data", which is, depending on Tari, between 27 and 128 kbps. Fig. 1 clarifies the explanation. [3, 4, 5]

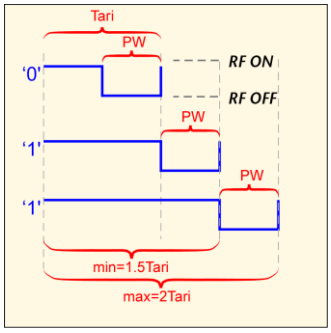


Fig. 1. Gen2 Forward Link symbols [3]

As seen, Gen2 offers a wide selection of allowable Forward Link waveforms. This permits tradeoffs between read rate, read range and transmit bandwidth, and allows convenient adaption to different operating environments.

1.2. Reverse Link

As passive RFID tag does not have its own radio, it sends information back to reader by changing the impedance state of its antenna. The reader senses this as changes in the backscattered power from the tag. The tag encodes its data in the timing of the transitions between backscattering states.

In Gen2-communication, there is a fundamental clock, known as the *Backscatter Link Frequency* (BLF), which specifies the pulse width of the shortest Reverse Link feature. Then there are four different ways to encode data symbols: *FM0 baseband* and three different *Miller modulations* of BLF subcarrier.

The simplest encoding, *FM0*, has a state transition at the beginning of each data symbol. Data "0" symbol has an additional mid-symbol transition. Based on this, a long string of FM0 Data "0" symbols produce a square wave at BLF, and a long string of Data "1" symbols generates a square wave at BLF/2. For FM0 the data rate is same as BLF, and they share the same allowable range, from 40 to 640 kbps. FM0 compliant tags must support the whole range, but readers need not, and in many cases will not, implement all data rates. Figure 2 shows default Gen2 FM0 tag signaling. [5, 6]

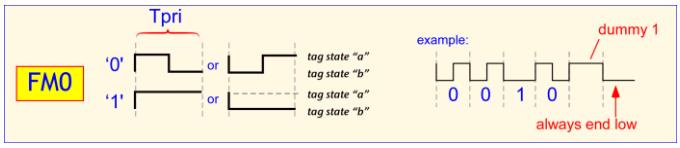


Fig. 2. Default Gen2 FM0 tag signaling [3]

Miller modulated subcarrier (MMS) is more complicated encoding. MMS provides more state transitions per bit and is therefore easier to decode in the presence of interference. However it is slower with the same tag BLF than FM0. MMS exists in three different schemes: Miller-2, Miller-4, and Miller-8. Then number defines how many BLF periods define one data symbol. For example, using the slowest BLF of 40 kHz, the data rate for Miller-8 is the BLF/8 = 5 kbps. Figures 3 and 4 clarify the MMS structure. [5, 6]

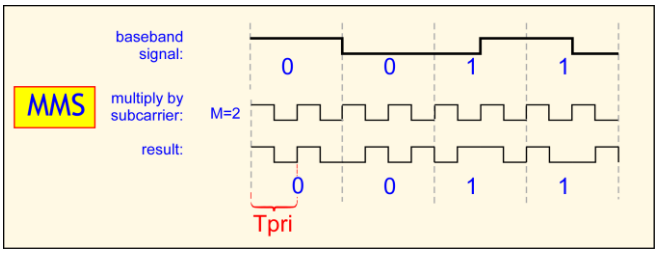


Fig. 3. Miller-modulated subcarrier example with M=2 [3]

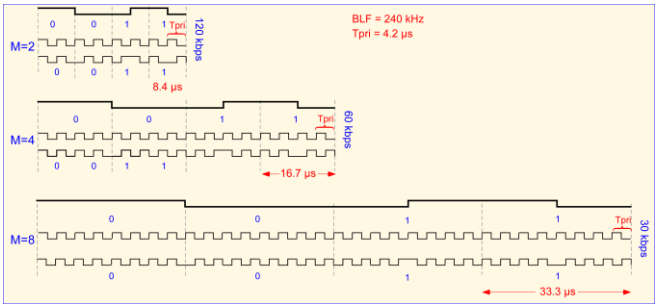


Fig. 4. Example of encoding the same data stream using Miller M=2,M=-4 and M=8 [3]

MMS is supposed to provide a constant performance when a large number of readers are operating in the same facility at the same time. This is due to its narrower Reverse Link spectrum and a tendency to put it into the frequency region between readers' Forward Link channels. [3, 5]

3. Measurements

This paper presents the measurements where UHF RFID tag read rates were studied when varying certain Gen2 parameters as well as reader transmission power. In addition some interference was added with extra reader reading the same tags.

Totally ten individual measurements setups (i.e. circumstances) were conducted. Within each measurement setup the differences between read rates were studied by modifying the Gen2 parameters similarly.

The measurements were conducted with 1, 2 or 4 similar passive UHF tags, and a reader with one antenna (monostatic mode). The taken parameter changes are presented in table I. Each group represents one individual permanent set of parameter values used in this study.

Table 1. The tested protocol parameters

PROTOCOL SETTINGS	**Group A**	**Group B**	**Group C**	**Group D**
Tari (µs)	14,29	14,29	20	20
PIE	2,0:1	2,0:1	2,0:1	2,0:1
Forward Link	PR-ASK	PR-ASK	PR-ASK	PR-ASK
Pulse Width	0,5	0,5	0,5	0,5
BLF (kHz)	436	320	320	320
Reverse Modulation	FM0	Miller M=4	Miller M=4	Miller M=8

Each of ten measurement setups consisted of following adjustments: at first Gen2 parameters were set as presented in Group A –column. The read rate was recorded based on the average values of ten individual 60 second measuring periods. Next *Link Frequency* was decreased to 320 kHz and *Reverse*

Modulation changed from "FM0" to "Miller M=4" (Group B), and read rate again recorded similarly. For third setup the length of *Tari* was increased to 20 µs (Group C), and finally the "Miller M=4" turned to "Miller M=8" (Group D).

As mentioned, the above described measurement arrangement was carried out with totally ten different ways with varying the number of tags, the TX-power of the reader, or adding the extra reader to create interference, as presented in more detail in table II. In measurement setups 7-10 an additional reader was added to create interference. In setups 7-8 this extra reader was using 28 dBm TX power and 4 alternating channels without LBT (listen before talk) feature. The actual reader whose read rate was monitored used "Reader Selects Frequency" option, which made it possible to choose between the same four channels, and LBT. In setups 9-10 the actual reader was forced to act only in one single channel (866,3 MHz) with LBT, and the extra reader was transmitting with 30 dBm first in four channels (#9) and then in the same single channel (#10) again without LBT.

The used RFID equipment was:

- Tags:
  - UPM ShortDipole EPC Class 1 Gen 2 – tags with 240 bit EPC memory
- RFID reader:
  - Impinj Speedway R1000 UHF –reader
  - Firmware version: Octane 3.2.1.240
  - Operating region: ETSI EN302-208 LBT
- Reader Antenna: Huber+Suhner SPA 860/65/9/0/V
- Measurement software: Impinj MultiReader 6.0.0
- Extra reader creating interference:
  - ThingMagic Astra EU –reader, with integrated 6 dBi CP-antenna

Table 2. Different measure arrangements

Measurement setups
Common parametres: - Tag distance: 1,0 m; Height: 1,2 m - Duration: 10 x 60s - RX sensitivity: -70 dBm - Search Mode: 0 - Dual target - Session 1 - Reader Selects Frequency (not in #9 and #10) - LLRP Protocol
Varying parametres: - *Setup #1:* TX power 18 dBm, 1 tag - *Setup #2:* TX power 28 dBm, 1 tag - *Setup #3*: TX power 18 dBm, 2 tags - *Setup #4*: TX power 28 dBm, 2 tags - *Setup #5*: TX power 18 dBm, 4 tags - *Setup #6*: TX power 28 dBm, 4 tags - *Setup #7*: TX power 18 dBm, 4 tags, Extra reader - *Setup #8*: TX power 28 dBm, 4 tags, Extra reader - *Setup #9*: TX power 16 dBm, 4 tags, 1 channel, Extra reader - *Setup #10*: TX power 16 dBm, 4 tags, 1 channel, Extra reader in the same 1 channel

The measurements were conducted in a laboratory, where the wall behind the tags was covered with absorbing material to cut down unintended interference. Otherwise the environment was close to "real-life", not anechoic chamber. In these circumstances, the threshold power level was approximately 15,0 - 15,5 dBm. Figure 5 shows the measurement setup.

Fig. 5. The measurement setup. Tags are attached on the styrofoam pillar and readers and antenna on the rack on right. Impinj reader at the bottom, antenna in the middle, and ThingMagic reader uppermost.

4. Results and Discussion

As the purpose of this paper is to study the effect of adjusting the Gen2-parameters, the achieved read rates are compared to each other's within each measurement setup. This way the differences between each parameter values can be seen, as well as if the measuring circumstances cause any variation for their mutual interrelationships.

The following graphs in figure 6 presents actual amount of successfully read tags in each measurement setup (#1 – #10), and with each Gen2 parameter setup (Group A – D). Graphs #1 – #6 prove clearly the theory that FM0 coding scheme (Group A) achieves higher read rate than MMS coding schemes (Group B – D). In this study the increase in number of tags increased greatly the total read rate per second as well, especially with FM0 coding. The variation in Tari (Group B vs C) had only a minor effect in tested situations. In graphs #7 – #10, presenting setups where interference was created with extra reader, the benefits of MMS coding schemes can clearly be seen. The read rates stayed close or even increased when moved from FM0 (Group A) to MMS (Group B – C).

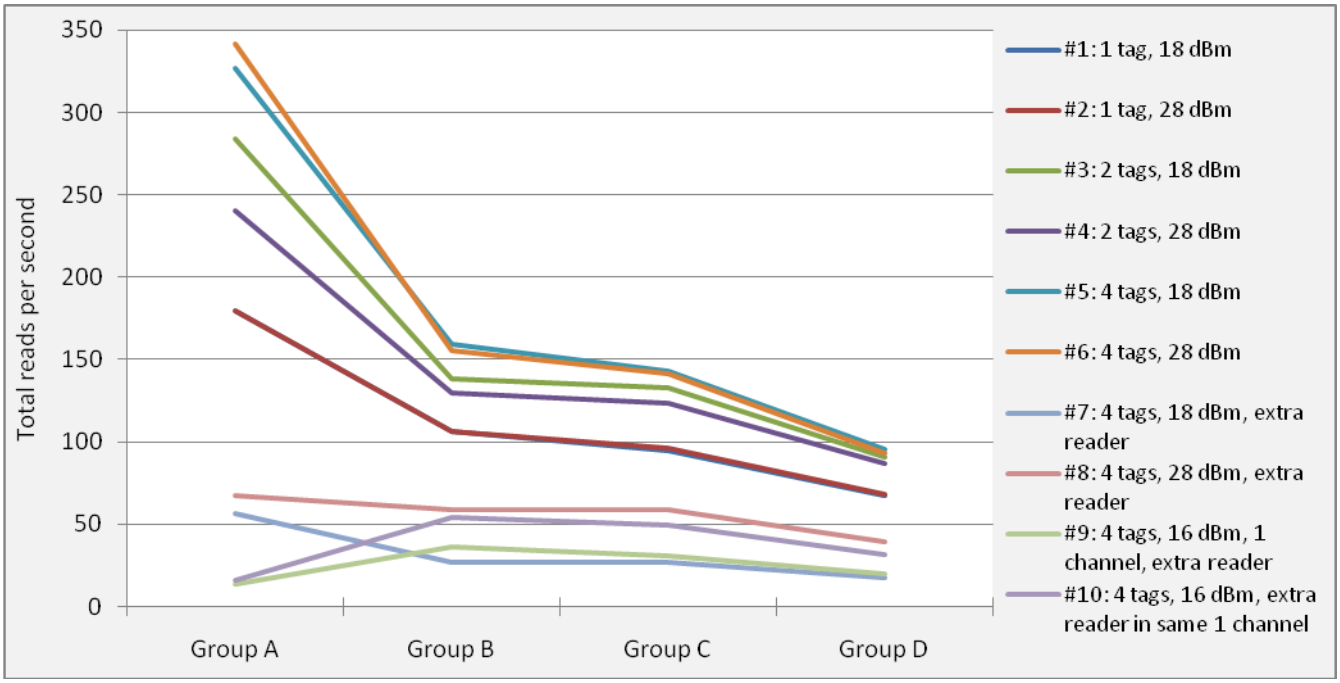


Fig. 6. The actual amount of successfully read tags in each measurement setup (#1 – #10) and with each Gen2 parameter setup (Group A – D)

The figure 7 shows graphs about relative read rates of different tested Gen2 parameter setups (Group A – D), in each of ten tested environments (#1 – #10). Group A, where FM0 coding scheme was used, is individually set as a reference value *within* each of 10 measurement setups. Percentage values *between* different setups (#1 – #10) are not comparable in this figure.

Graphs in figure 7 show the purpose and advantage of the MMS coding scheme. In setups #1 – #6, where no interference was present, the read rates with MMS coding schemes were between 27% and 59% of the FM0 read rate. In setups #7 – #10, where additional reader was present, the MMS coded read rate achieved higher percentages, and finally when channel switching was disabled (#9 and #10), over three times higher values than with FM0 coded reading.

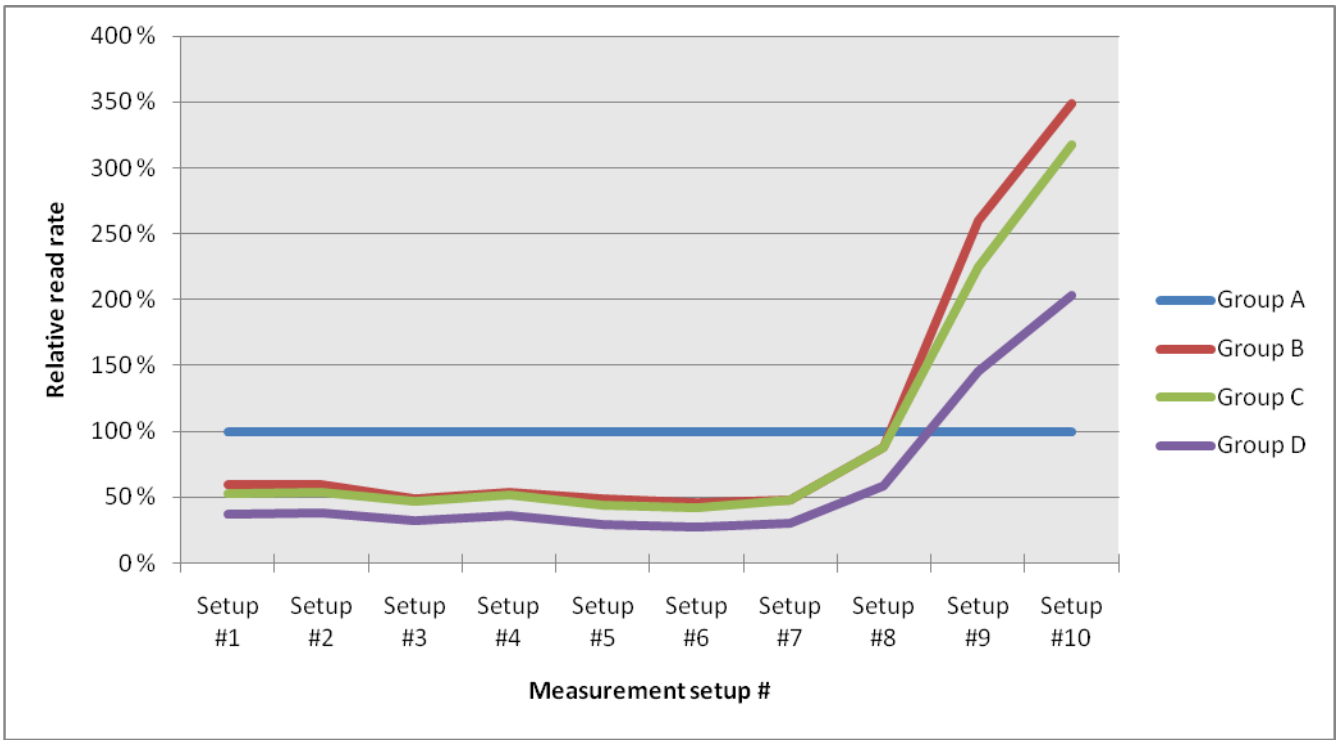


Fig. 7. The relative read rate of different Gen2 parameter setups

5. Conclusions and Future Work

Based on the results presented in this paper can be concluded that changes in UHF Gen2 –parameters have effects on read rate. As long as the operating environment does not have any interference present, FM0 coding gives highest read rates, despite of number of tags (1, 2 or 4) or transmission power level. When one additional reader was brought nearby to read same tags, FM0 still achieved highest read rates, when reader was able to select its channel automatically. When it was forced to act in a one channel, and even further, when additional reader was acting on that same channel, the MMS coding scheme achieved notably higher read rates than FM0. The tested 40% increase in Tari value did not have a relevant effect on read rate, in these circumstances.

If more interference and larger amount tags had been tested, the advantage of MMS coding might have been even more significant. This would be the interesting research topic in the future. Also the effects of varying Tari and PW values could be studied in more detail.

References

[1] Nummela, J., Lehto, A., Ukkonen, L. and Sydänheimo, L. "Automatic reel identification by RFID technology in paper reel supply chains: examples from paper mill and sea port environments", International Journal of RF Technologies: Research and Applications,Vol 1, No 3, pp. 194- 213, September 2009.

[2] Nummela, J., Nikkari, M., Soini, M., Ukkonen, L., Sydänheimo, L. (2009) "A Novel Method for Indoor Positioning with Passive UHF RFID" International Journal of Radio Frequency Identification Technology and Applications (IJRFITA). Inderscience. (ACCEPTED)

[3] D. M. Dobkin, The RF in RFID: passive UHF RFID in practice. Burlington, MA USA: Elsevier Inc., 2008.

[4] EPCglobal, "Specification for RFID air interface: EPC radio-frequency identity protocols Class-1 Generation-2 UHF RFID protocol for communications at 860 MHz-960 MHz, version 1.2.0", 108 Pages, October 2008.

[5] D. M. Dobkin, D. J. Kurtz, " Overview of EPCglobal Class 1 Generation 2 and comparison with 1st Generation EPCglobal standards", 10 Pages, March 2006.

[6] P. V. Nikitin, K. V. S. Rao, "Effect of Gen2 Protocol Parameters on RFID tag Performance", 2009 IEEE International Conference on RFID, 27-28 April 2009.

Fuzzy Rule Based Inference System for Detection and Diagnosis of Lung Cancer

K. Lavanya[1], M.A. Saleem Durai[2], N.Ch. Sriman Narayana Iyengar[3]

School of Computing Sciences and Engineering, VIT University, Vellore, Tamil Nadu, India
[1]lavanya.sendhilvel@gmail.com, [2]masaleemdurai@vit.ac.in, [3]nchsniyengar48@gmail.com

***Abstract*— In this paper we design a fuzzy rule based inference system to determine and identify lung cancer. The proposed system accepts the symptoms as input and provides the confirmed disease and stage as the output it also calculates the membership function for both input as well as the output variable. Domain expert knowledge is gathered to generate rules and stored in the rule base and the rules are fired when there exists appropriate symptoms. The features of fuzzy logic toolbox is used to implement the proposed system and is used as the medical diagnosis model for providing treatments to the patients as well as it can be used to assist the doctors.**

***Keywords*— fuzzy rule base, inference system, symptoms, medical diagnosis, membership function and lung disease.**

I. INTRODUCTION

The advent of computers and information technology in the recent past has brought a drastic change in the fuzzy medical expert system. Information gathered from the domain experts must be transferred to knowledge and must be used at the right time .These Knowledge can be incorporated in the form of fuzzy expert system in the diagnosis of lung cancer in specific.

Lung cancer is the most common cause of death in both men and women throughout the world. The American Cancer Society estimates that 219,440 new cases of lung cancer in the U.S. will be diagnosed and 159,390 deaths due to lung cancer will occur in 2009. According to the U.S. National Cancer Institute, approximately one out of every 14 men and women in the U.S. will be diagnosed with cancer of the lung at some point in their lifetime. The lung cancer is most difficult cancer to treat. Every year ten million people die of lung cancer. [1]

An agent is a goal-directed, computational entity which acts on behalf of another entity (or entities). Agent systems are self-contained software programs possessing domain knowledge and an ability to behave with some degree of independence to carry out actions to achieve specified goals [2] .

II. RELATED WORK

Most of the researchers develop many methods to diagnose lung cancer. An expert system is designed to diagnose the heart disease and is based on fuzzy logic it uses the dataset of V.A.Medical center. All the symptoms are considered and membership function is calculated for both input and the output variable. Finally a rule base is created and rules are fired based on the symptoms for the patient.

This system provides information only about the disease and not about stage and treatment. [3] A fuzzy expert system is developed to determine coronary heart disease risk of patients and gives the user the ratio of the risk for normal life, diet and drug treatment [4]. A Fuzzy rule based lung disease diagnostic system combining the positive and negative knowledge was developed using contexts, facts, rules, modules and strategies of knowledge representation to identify medical entities and relationship between them for diagnosis of lung cancer[5]. An Intelligent system for Lung cancer diagnosis is designed that detects all possible lung nodules from chest radiographs using image processing techniques and feed forward neural networks. It classifies the nodules into cancerous and non-cancerous nodules [6] numerous applications implement computer-assisted evidence-based lung cancer diagnosis by utilizing of Tomosynthesis. It offers high shadow detection sensitivity at a low exposure dose which recreates multi sectional images from a single scan and offers image processing to produce artifacts [7]. In PET/CT Images the exact position of boundary of the tumor was manually identified, five optimal threshold features and two gray level threshold features of the tumors were extracted from the B-mode ultrasonic images, an optimal feature vector was obtained using K-means cluster algorithm and a back propagation artificial neural network, was applied to classify lung tumors [8].Dynamic tumor-tracking treats the lung as an elastic object and analyzes the deformation based on linear Finite Element Method. The doctor planning the radiotherapy can reproduce the movement of the lung tumor by freely adjusting the regions, displacements, and phases of the boundary conditions while comparing the position of the lung tumor in an X-ray photograph [9].

III. PROPOSED SYSTEM

Our proposed system uses the basic idea of fuzzy set theory to create a Fuzzy Inference system for the diagnosis of lung cancer. This proposed system uses the dataset gathered from domain expert including the symptoms, stages and treatment

facilities that provides an efficient and easy method to diagnose if a patient is affected by cancer, If so in which stage and the treatment options for the prognosis of lung cancer. The proposed system is implemented in JADE, MATLAB using the features of Fuzzy Tool box.

A. *System Architecture*

The proposed system architecture (Figure 1) consists of the following Components:

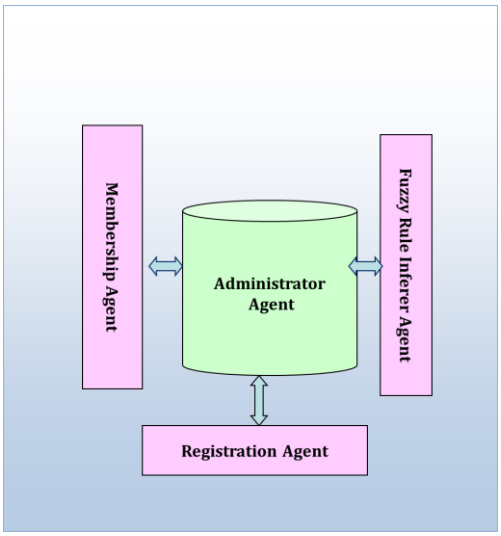


Fig. 1 System Architecture

JADE provides a distributed, fault tolerant, container architecture, internal service Architecture which provides all the services needed for hosting and executing agents. A JADE platform is composed of agent containers that can be distributed over the network. A special container, called the *main container*, is the first container to be started and all other containers must register with the main container to start working. The Registration Agent takes care of authentication and authorization of users. The Fuzzy Rule Inferer Agent and the Membership Agent co-ordinately utilizes the fuzzy rule membership functions to conclude the stage of cancer. The Administrator agent takes control of store/view/update data in database.

B. *Fuzzy Inference System (FIS)*

FIS makes use of fuzzy logic to map the given input in this case the symptoms to an output the possible disease a patient can have. Based on this we can make decision or extract patterns to identify the stage and the treatment for a particular patient. FIS makes use of membership function, If-then rules and logical operators for making these decisions.

FIS performs fuzzification on the inputs and determine the degree to which the input belongs to the fuzzy set. Using fuzzy operator it combines the result of If part of the rule to find the degree of support to the rule and applies to the output function. It uses the degree of support for the entire rule to determine the output fuzzy set. The Then part of the fuzzy rule assigns an entire fuzzy set to the output. Aggregation is used to combine the output of all the rules in to a single fuzzy set and it is defuzzified to make the appropriate decisions.

C. *Calculating membership function*

A membership function defines how each point in the input space is mapped to a degree of membership between 0 and 1 and is denoted by μ. The membership function editor in Fuzzy tool box is used to define the shapes of all membership functions associated with each membership variable. In our system we calculate the membership function for each input variable (symptoms) and output variable (disease). For each of the input variable the membership function is defined as follows

1. *Weight Loss*: If there is more than 10% of weight loss in six months then it corresponds to the symptoms of lung cancer. To calculate the membership function we use four linguistic variables low, medium, high and very high.

$$\mu \text{ low}(x)= \begin{cases} 1 & x < 1 \\ \frac{3-x}{2} & 1 \le x < 3 \end{cases}$$

$$\mu \text{ med}(x)= \begin{cases} \frac{x-2}{1.5} & 2 \le x < 3.5 \\ 1 & x = 3.5 \\ \frac{4.5-x}{1} & 3.5 \le x < 4.5 \end{cases}$$

$$\mu \text{ high}(x)= \begin{cases} \frac{x-4}{1.5} & 4 \le x < 5.5 \\ 1 & x = 5.5 \\ \frac{7-x}{1.5} & 5.5 \le x < 7 \end{cases}$$

$$\mu \text{ veryhigh}(x)= \begin{cases} \frac{x-6.5}{1} & 6.5 \le x, 7.5 \\ 1 & x \ge 7.5 \end{cases}$$

TABLE I-CLASSIFICATION OF WEIGHT LOSS

Input Field	Range	Fuzzy Sets
Weight Loss	<0.2 0.3 – 3.6 2.61 – 4.4 3.5 >	Low Medium High Very high

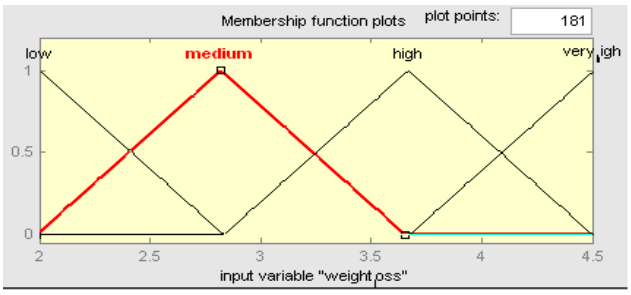


Fig. 2 Membership Function for weight loss

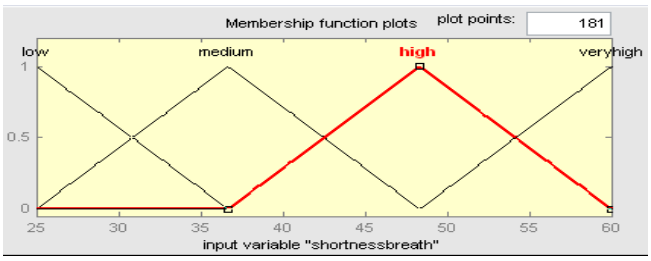


Fig. 3 Membership Function for Shortness of breath

2. *Shortness of breath*: The range 14-16 is normal for this if it is more than this range then we need to check for the occurrence of lung disease. For calculating the membership function (MF) we scale the range to 0-100 and based on the severity we calculate the MF and the range are given as follows

$$\mu \text{ low}(x)= \begin{cases} 1 & x < 10 \\ \frac{14-x}{4} & 10 \le x < 14 \end{cases}$$

$$\mu \text{ med}(x)= \begin{cases} \frac{x-12}{8} & 12 \le x < 20 \\ 1 & x = 20 \\ \frac{30-x}{10} & 20 \le x < 30 \end{cases}$$

$$\mu \text{ high}(x)= \begin{cases} \frac{x-25}{16} & 25 \le x < 41 \\ 1 & x = 41 \\ \frac{60-x}{19} & 41 \le x < 60 \end{cases}$$

$$\mu \text{ veryhigh}(x)= \begin{cases} \frac{x-55}{5} & 55 \le x < 59 \\ 1 & x \ge 59 \end{cases}$$

TABLE II-CLASSIFICATION OF SHORTNESS OF BREATH

Input Field	Range	Fuzzy Sets
Shortness of breath	<14 12 –35 36 – 60 >55	Low Medium High Very high

3. *Chest Pain*: Depending up on the severity, chest pain can be characterized as low, medium, high and very high. For calculating the membership function (MF) we scale the range to 0-100 and based on the severity we calculate the MF and the range are given as follows

$$\mu \text{ low}(x)= \begin{cases} 1 & x < 16 \\ \frac{19-x}{3} & 16 \le x < 19 \end{cases}$$

$$\mu \text{ med}(x)= \begin{cases} \frac{x-17}{9} & 17 \le x < 26 \\ 1 & x = 26 \\ \frac{38-x}{12} & 26 \le x < 38 \end{cases}$$

$$\mu \text{ high}(x)= \begin{cases} \frac{x-35}{13} & 35 \le x < 48 \\ 1 & x = 48 \\ \frac{72-x}{24} & 48 \le x < 70 \end{cases}$$

$$\mu \text{veryhigh}(x)= \begin{cases} \frac{x-69}{2} & 69 \le x < 71 \\ 1 & x \ge 71 \end{cases}$$

TABLE III-CLASSIFICATION OF CHEST PAIN

Input Field	Range	Fuzzy Sets
Chest pain	<19 17 – 38 35 – 59 >60	Low Medium High Very high

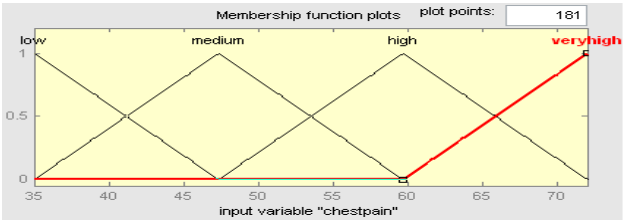


Fig. 4 Membership Function for Chest pain

4. *Persistence Cough:* If there exist persistence cough for more than eight weeks then there exists symptoms of lung cancer. Based on the severity experienced by the patient we scale the range from 0-100 and calculate MF.

$$\mu low(x)=\begin{cases}1 & x<18\\ \frac{20-x}{2} & 18\le x<20\end{cases}$$

$$\mu med(x)=\begin{cases}\frac{x-18}{5} & 18\le x<23\\ 1 & x=23\\ \frac{29-x}{6} & 23\le x<29\end{cases}$$

$$\mu high(x)=\begin{cases}\frac{x-26}{16} & 26\le x<42\\ 1 & x=42\\ \frac{60-x}{18} & 42\le x<60\end{cases}$$

$$\mu veryhigh(x)=\begin{cases}\frac{x-58}{1} & 58\le x<59\\ 1 & x\ge 59\end{cases}$$

TABLE IV-CLASSIFICATION OF PERSISTENCE COUGH

Input Field	Range	Fuzzy Sets
Persistence Cough	<20 18 – 59 60 – 78 78 >	Low Medium High Very high

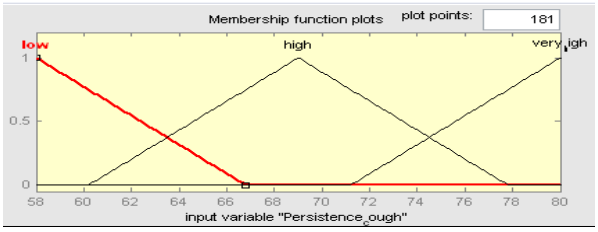


Fig. 5 Membership Function for Persistence cough

5. *Blood in Sputum:* If the blood in sputum is brownish in color then it is one of the symptoms of lung cancer. We scale the range from 0-100 to calculate the MF.

$$\mu low(x)=\begin{cases}1 & x<9\\ \frac{11-x}{2} & 9\le x<11\end{cases}$$

$$\mu med(x)=\begin{cases}(x-9) & 9\le x<22\\ 1 & x=22\\ \frac{36-x}{14} & 22\le x<36\end{cases}$$

$$\mu high(x)=\begin{cases}(x-\frac{31}{18} & 31\le x<49\\ 1 & x=49\\ \frac{68-x}{19} & 49\le x<68\end{cases}$$

$$\mu veryhigh(x)=\begin{cases}\frac{x-66}{1} & 66\le x<67\\ 1 & x\ge 67\end{cases}$$

TABLE V-CLASSIFICATION OF BLOOD SPUTUM

Input Field	Range	Fuzzy Sets
Blood in Sputum	<30 31– 55 55.1– 68 68.1 >	Low Medium High Very high

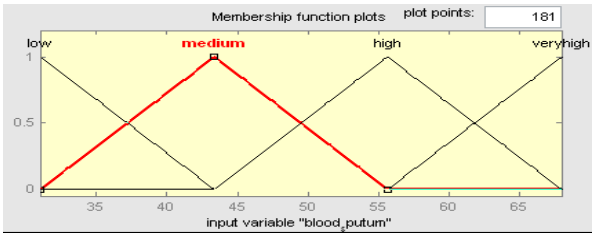


Fig 6 Membership Function for Blood Sputum

Output variable:

The decision field refers to the presence of lung cancer. It is an integer value which represents no cancer or any one of the stage. In the proposed system the output variable is divided into the following fuzzy sets stage1, stage2, stage3 and stage4 if the patient is affected with lung cancer. For the output variable the membership function is defined as follows

TABLE VI-CLASSIFICATION OF OUTPUT

Input Field	Range	Fuzzy Sets
Decision	<3.9 3.91–6.1 5.6– 6.4 6.41>	Stage-1 Stage-2 Stage-3 Stage-4

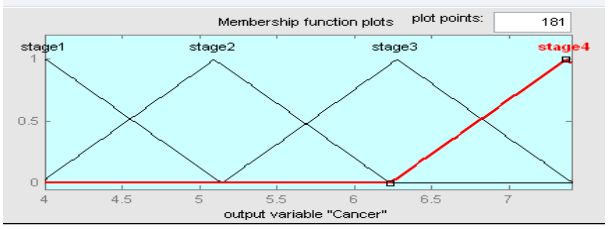


Fig 7 Membership Function for Output variable

D. *Fuzzy Rule Base*

FIS makes use of fuzzy rule base to make decision for the diagnosis of disease and the efficiency of the system depends on the fuzzy rules generated. Rule based systems are mainly used in medical diagnosis. In this the knowledge base stores all information about the symptoms and disease in the form of rules in the Rule base. The rules are generated as per the data gathered from domain expert. All possibility of rules for lung cancer diagnosis is generated and some of it are given below.

Rule 1: If (persistence cough is very high) and (sputum in blood is very high) and (chest pain is very high) and (shortness of breath is very high) and (weight loss is very high) then disease =lung cancer and stage= stage 4.

Rule 2: If (persistence cough is very high) and (sputum in blood is very high) and (chest pain is high) and (shortness of breath is very high) and (weight loss is high) then
disease =lung cancer and stage= stage 4.

Rule 3: If (persistence cough is high) and (sputum in blood is very high) and (chest pain is high) and (shortness of breath is very high) and (weight loss is very high) then
disease =lung cancer and stage= stage 4.

Rule 4: If (persistence cough is very high) and (sputum in blood is high) and (chest pain is high) and (shortness of breath is high) and (weight loss is very high) then
disease =lung cancer and stage= stage 4.

Rule 5: If (persistence cough is very high) and (sputum in blood is very high) and (chest pain is high) and (shortness of breath is high) and (weight loss is very high) then
disease =lung cancer and stage= stage 4.

Rule 6: If (persistence cough is very high) and (sputum in blood is medium) and (chest pain is medium) and (shortness of breath is high) and (weight loss is high) then disease =lung cancer and stage= stage 3

Rule 7: If (persistence cough is high) and (sputum in blood is high) and (chest pain is medium) and (shortness of breath is high) and (weight loss is) then disease =lung cancer and stage= stage 3.

Rule 8: If (persistence cough is medium) and (sputum in blood is high) and (chest pain is medium) and (shortness of breath is high) and (weight loss is medium) then
disease =lung cancer and stage= stage3.

Rule 9: If (persistence cough is medium) and (sputum in blood is medium) and (chest pain is medium) and (shortness of breath is high) and (weight loss is medium) then disease =lung cancer and stage= stage 2.

Rule 10: If (persistence cough is high) and (sputum in blood is medium) and (chest pain is medium) and (shortness of breath is high) and (weight loss is medium) then
disease =lung cancer and stage= stage 2.

Rule11: If (persistence cough is low) and (sputum in blood is low) and (chest pain is medium) and (shortness of breath is low) and (weight loss is low) then
disease = no cancer and stage= nil.

The Rules generated using Rule Editor is given in fig 8.

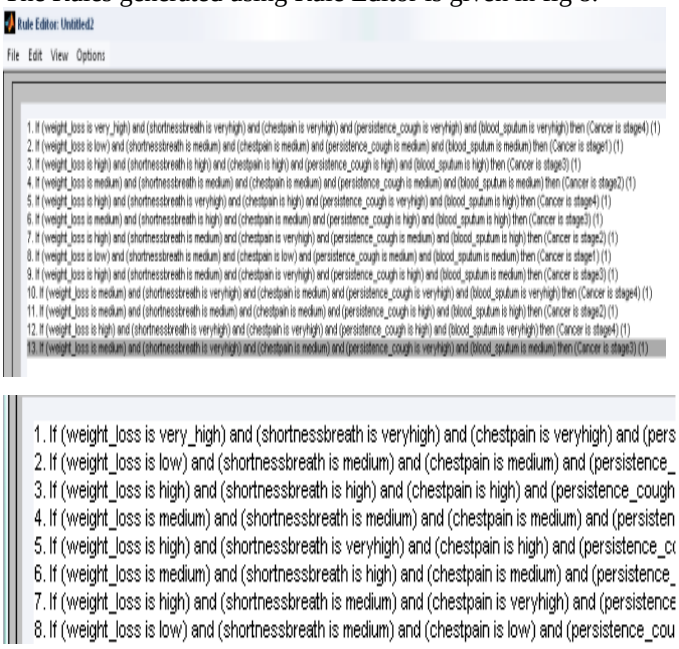


Fig. 8 Rules Based System

E. *Fuzzification And Defuzzification:*

The proposed system uses Mamdani-type inference system in which the output membership function is a fuzzy set. Centroid method is used for defuzzification which takes fuzzy set as input and output is crisp set.

Surface Viewer is a three dimensional curve that represents the mapping from the symptoms to cancer stage. A graph is plotted from the rules and the symptoms. The curve has two inputs and one output.

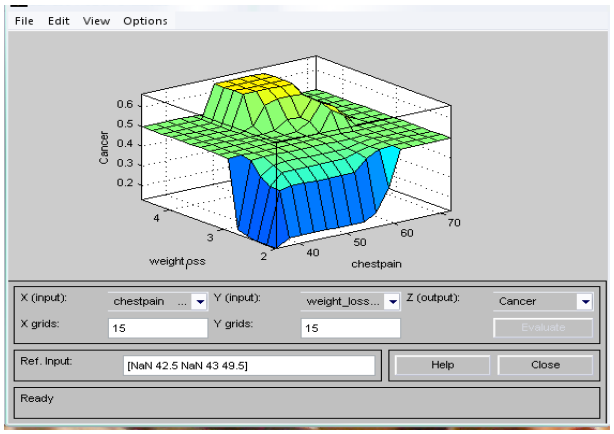


Fig 9 Surface Viewer for Chest pain and Weight loss

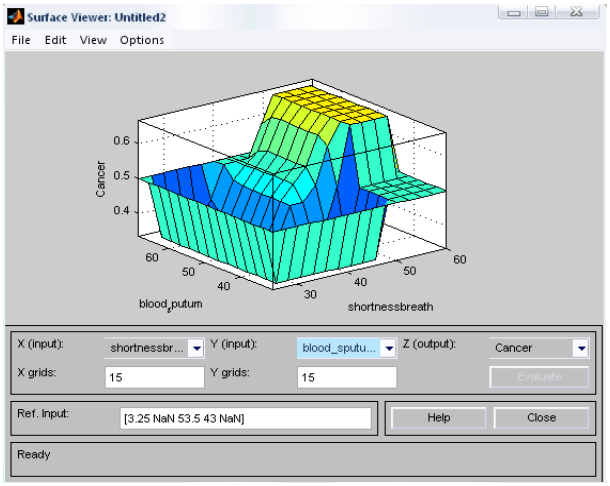


Fig. 10 Surface Viewer for Shortness of Breath and Blood in Sputum

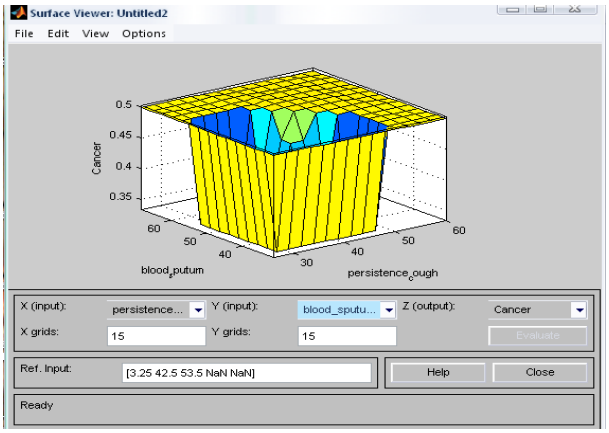


Fig.11 Surface Viewer for Persistence cough and Blood in Sputum

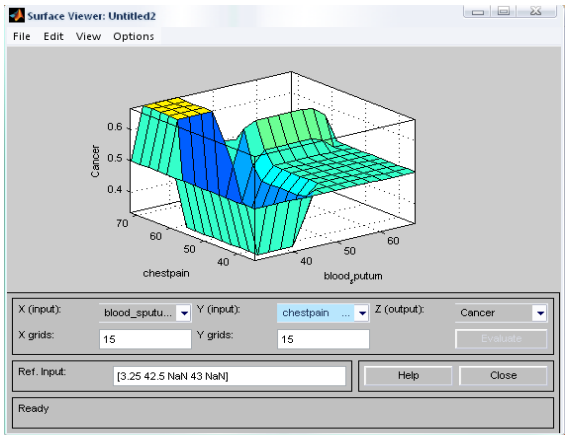


Fig. 12 Surface Viewer for Blood in Sputum and Chest pain

The Rule Viewer displays the execution of the Fuzzy Inference process and also represents the antecedent and consequent of the generated rules and interprets the entire fuzzy inference process.

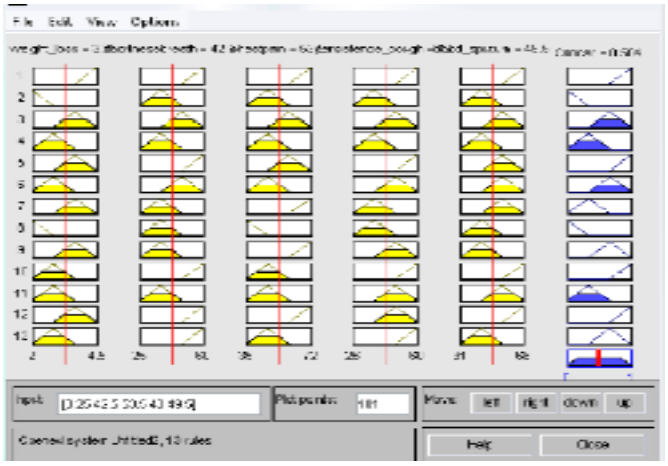


Fig. 13 Rule Viewer for generated rules

IV. Conclusion and Future Work

The developed system yields a good result and serves as an expert medical diagnosis system even for ordinary users since it is simple and easy to implement. The proposed system is implemented in JADE and MATLAB using the features of Fuzzy Logic Toolbox. This system is compatible with other techniques and can be easily incorporated with case based reasoning, Decision tree and other data mining techniques. To increase the system efficiency we can make use of neuro-fuzzy system and rough sets.

References

[1] American cancer society, Lung cancer facts and figures 2009-2010, Atlanta: American cancer society, Inc.

[2] F. Bellifemine, G. Caire, A. Poggi, G. Rimassa, "JADE-A WHITE PAPER" Volume 3, No.3, September 2003. (jade.tilab.cpm/papers/WHITEPAPEREXP.pdf)

[3] Ali.Adeli and Mehdi.Neshat "A Fuzzy expert system for heart disease diagnosis",Proceeding of the IMECS 2010 Vol 1.

[4] Allahverdi Novruz, Torun Serhat and Saritas Ismail "Design of a Fuzzy expert system for determination of coronary heart disease risk", International conference on Computer Systems and technologies-CompSysTech'07.

[5] Phuong Nguyen Hoang, Nghi Le Huu, Cho Hune and Kwak Yun Sik "LDDS-A Fuzzy rule based lung diseases diagnostic system combining positive and negative knowledge".

[6] Khanna Nehemiah.H and Kannan .A "An intelligent system for lung cancer diagnosis from chest radiographs" International Journal of Soft Computing 1(2):133-136,2006.

[7] Rachid Sammouda, Jamal Abu Hassan, Mohamed Sammouda , Abdulridha Al-Zuhairy , Hatem Abou Elabbas. "Computer Aided Diagnosis System for Early Detection of Cancer Using Chest Computer Tomography Images", GVIP '05 Conference, 19-21 December 2005, CICC, Cairo, Egypt. www.igst.com.

[8] I. Jafar, Hao Ying , Shields A.F, O.Muzik . "Computerized Detection of Lung Tumors in PET/CT Images". EMBS 2006, 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006.

[9] Nakao M, Kawashima A, Kokubo M, Minato K. "Simulating Lung Tumor Motion for Dynamic Tumor-Tracking Irradiation". Nuclear Science Symposium Conference Record, 2007. NSS 2007.

REP Models versus OntoAidedRE - A Parameters Based Study

Shilpa Sharma[1] and Maya Ingle[2]

[1]Sanghvi Institute of Management and Science, Indore, India
[2]Indore Institute of Computer Application, Indore, India
shilpa_1819@yahoo.com maya_ingle@rediffmail.com

Abstract: *Presently, the software requirement specifications are developed to accomplish the objectives of a concerned domain only. The traditional approaches in requirement engineering have been restricted to limited time span. In such a context, the successful construction of generic, flexible, reusable and extensible software proves to be difficult. The ontology may be proved as a substantial assurance in aiding the elicitation and elaboration of requirements. Ontology here refers to the basic existential pool of knowledge in the world that are of interest to the discipline. In this view, the explicit treatment of knowledge and emphasizing on the category of requirement in requirement engineering practices suggests a fundamental shift in the domain oriented underpinnings of requirement engineering process. The paper attempts to review the existing requirement engineering process models to reconcile and unify the domain-specific concepts, approaches and knowledge in order to illustrate the OntoAidedRE (Ontology Aided Requirement Engineering). Thus, in this perspective, we have suggested some parameters related with project to verify the ability of OntoAidedRE against the conventional requirement engineering models.*

1. Introduction

With the advances in software development, the role of requirement engineering is to introduce engineering principles with the aim of amenable to analysis, communication, and subsequent implementation. Therefore, a systematic and disciplined process should be followed. The traditional requirements engineering is an iterative process which continues iteratively until the project is complete. This process involves the activities such as defining the terms in the domain of discourse, stating, clarifying and agreeing on assumptions and constraints, discussing and negotiating the needs and the objectives for a software development [1] [2]. There are many conventional requirements engineering process models are used to define these activities. These models includes linear sequential model, linear iterative processes model, iterative process model and spiral model. However, these models are different and sometimes conflicting in their nature, ranging from linear and incremental to cyclical and iterative in structure [3].

The requirement engineering phase of a software project is vital to its success. Consequently, the ability to identify the problems and suggestions for improvements in the requirement engineering process opens up significant potential for increasing the success of software projects. To deal with the increasing competencies of project parameters such as Project Type, Project Size, Project Team, Project Effort, Project Quality, Project prioritized element and Project key element the software systems demand a knowledge intensive requirement engineering process. Also, this process is required to promote the cohesiveness among the information gathered and to provide a coherent view between the stakeholders. Therefore, one possible solution may be encapsulating the ontology for invention of generalized requirement set that may cater various applications from different domains. Ontology is a formal and rigorous approach to the representation of knowledge to provide an unambiguous and precise terminology that can jointly explicable and functional across various realms [4].

Illustrating and understanding the project parameters affecting requirement engineering process is an important step towards improving RE practice and therefore increasing the success of software projects. In this context, Section 2 defines various project parameters. Section 3 exemplifies existing descriptive RE process models that provide an indication of common RE activities and their sequence. Section 4 aims to confer insight into the gap between the conventional RE process models in practice by rising the OntoAidedRE emphasizing on the category of requirement. Section 5 is the comparative study of conventional RE process models versus OntoAidedRE on the basis of project parameters. Finally, we conclude with the benefits of an OntoAidedRE.

2. Project Parameters

The project attributes such as project type, project size, project team, project effort, project quality, project prioritized element and project key element play a very significant role during project development. By defining, articulating, and managing these project attributes in a following way will add a new dimension to requirement engineering process and thus contributing to project success.

- **Project Type:** It defines the statement of work that must be completed during the development. Typically it includes building the product prototype [7]. The project type illustrates the classification of projects according to acquired distinctiveness such as operations support (includes transaction processing, process control and office automation), management

support (includes management information, decision support) and others (includes expert systems).

- **Project Size:** It is defined as number of lines of source code (LOC) in the software. There may exist various kinds of size based upon LOC such as small, medium and large. It determines the scalability of the methodology. The project size estimates the effort hours [8]. The effort hours for categorizing projects are as follows:

Table 1. Project Size

Size	Effort hours
Small	1-250 hours
Medium	251 – 5000 hours
Large	over 5000 hours

- **Project Team:** The project team is a group of individuals with appropriate and complementary professional, technical or specialist skills that aims to provide technical expertise in support of project objectives. It also contributes to understand the use project management standards specified in a project and maintain the project documentation in line with the project quality plan. The project team can consist of human resources within one functional organization, or it can consist of members from many different functional organizations such as analyst, change control board, client project manger, designer, project manager and team leader.

- **Project Effort:** The project effort is computed as a function of program size. Program size is expressed in estimated thousands of lines of code (KLOC).

Effort Applied = a_b (KLOC)$^{b}{}_{b}$ **[man-months]**

The coefficients a_b, b_b, c_b and d_b are given in the following table.

Table 2. Project Effort

Project Size	a_b	b_b	c_b	d_b
Small	2.4	1.05	2.5	0.38
Medium	3.0	1.12	2.5	0.35
Large	3.6	1.20	2.5	0.32

- **Project Quality:** The project quality is the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied requirements. There are several indicators are used in a range of measuring the project quality, or the perception of quality [9].
  - Engagement Measures [EM]: Internal Customer involvement in key project activities; expected vs. actual.
  - Planned vs. Actual Cumulative Review Count [P vs. A CRC]
  - Assessment Measures [AM]: Customer satisfaction surveys; stakeholder expectations evaluation

Table3. Project Quality

EM	Pvs.A CRC	AM	Quality Status	EM	Pvs.A CRC	AM	Quality Status	EM	Pvs.A CRC	AM	Quality Status
High	High	High	~100%	Med	High	High	75-99%	Low	High	High	65-74%
High	High	Med	75-99%	Med	High	Med	51-64%	Low	High	Med	41-49%
High	High	Low	65-74%	Med	High	Low	41-49%	Low	High	Low	31-39%
High	Med	High	75-99%	Med	Med	High	51-64%	Low	Med	High	41-49%
High	Med	Med	51-64%	Med	Med	Med	~50%	Low	Med	Med	~40%
High	Med	Low	41-49%	Med	Med	Low	~40%	Low	Med	Low	26-30%
High	Low	High	65-74%	Med	Low	High	41-49%	Low	Low	High	31-39%
High	Low	Med	41-49%	Med	Low	Med	~40%	Low	Low	Med	26-30%
High	Low	Low	31-39%	Med	Low	Low	26-30%	Low	Low	Low	0-25%

- **Project prioritized element:** The project prioritized element measures the project characteristics on three aspects such as cost effectiveness, time effectiveness and the project functionality. It will begin with the project cost factor comprised of acquisition cost and the operation cost. It then describes the project time factors which includes the overall time required to complete the project Finally, the project functionality will characterized by usability, serviceability and compatibility and the modes of operation.

- **Project key element:** It is a well-defined knowledge artifact for every project that will express the value of the product and its competitive benefit when the project type such as process control, decision support and expert systems etc. are to be developed. It also provide the guidelines to project team for making the knowledge support systems.

3. Conventional Requirements Engineering Process Models

The requirement engineering process is often depicted with a linear, incremental, cyclical and iterative model. Within these models, common RE activities, such as elicitation, analysis and negotiation to documentation and validation are combined under different label, which follow varied organization formats [5].

3.1 Linear Requirements Engineering Process Model

The linear requirements engineering process model was proposed by Linda Macaulay in 1996. There are five activities arranged sequentially in this model namely concept, problem analysis, feasibility and choice of options, analysis and modeling and requirement documentation. This is a simple model and mostly used for small projects with some less amount of complexity but this model is not good for some large and huge projects to get their requirements.

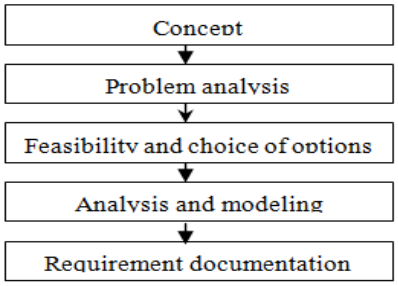


Figure 1. Linear RE Model

3.2 Linear Iterative Requirements Engineering Process Model

The linear iterative requirements engineering process model was proposed Kotonya and Sommervile in 1998 with some of the iterations used for validation the requirements engineering again and again. This iteration continues until the stakeholders are agreed and the final system specification is achieved. This model is useful for the system where the specifications should be pin point accurate and should be validated multiple numbers of times through the potential stakeholders.

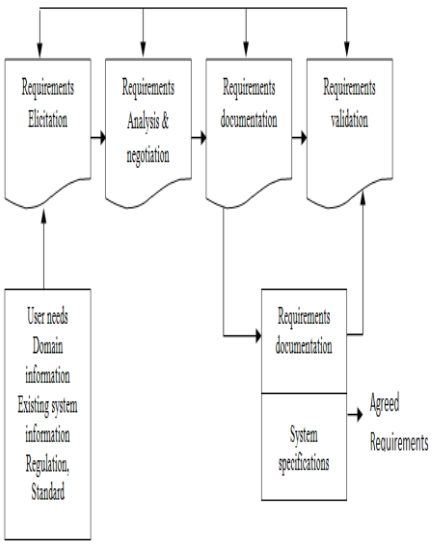


Figure 2. Linear Iterative RE Model

3.3. Iterative Requirements Engineering Process Model

The linear iterative requirement engineering model was proposed by Loucopoulos and Karakostas. This model is used to perform the requirements engineering in multiple iterations and hence is better for those software development which are launched versions by versions in the market. There are three simple phases of this model called elicitation, specification and validations. The requirements are obtained from the user and problem domain of the software systems to be developed.

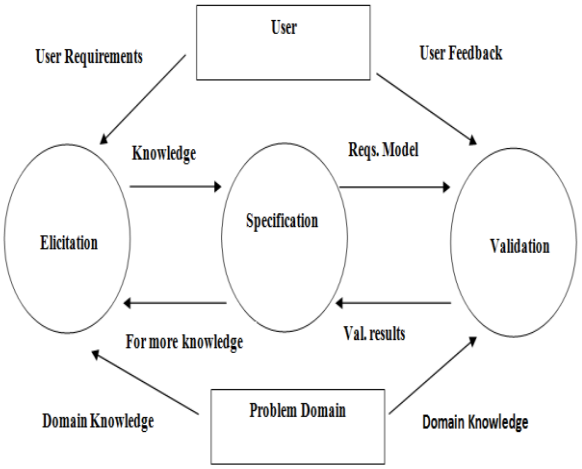


Figure 3. Iterative RE Model

3.4. Spiral Requirements Engineering Process Model

The spiral model is suggested by Kotonya and Sommerville in 1998. This model is performed in spirals. One spiral represents the complete version of the requirements on the basis of which the system has to be developed. Each spiral is divided into four quadrants called specification elicitation, requirements analysis and negotiation, requirements documentations and requirements validations. The main characteristic of this model is to handle the unwanted consequences called risks such as speciation delay, requirements change, low ROI etc which can badly affect the cost schedule and quality of the project.

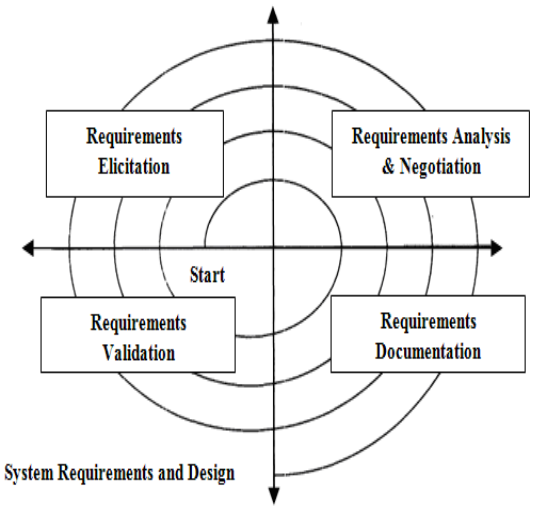


Figure 4. Spiral RE Model

4. OntoAidedRE

The model presents a framework indicating the kind of applications can or should be built and the interrelation. Consequently, an eccentric approach to provide guidelines and commendation for the application of simulated requirement set to ensure pragmatism and control, referred to as an OntoAidedRE (Ontology Aided Requirements Engineering) model. It follows an encrusted approach and intended to be a generic paradigm to enable knowledge driven requirements engineering. It is not only supported by ontology but also driven by requirement type depending on relevant changes of knowledge. It is anticipated that OntoAidedRE will enhance the traditional RE process, using this layered approach. There are mainly four layers devised on the basis of requirement type as shown in Figure 5. The first layer, OntoPre Requirements is dedicated for the use of system. The second layer, OntoInput Requirements refers to initial system qualifying terms. The third layer is divided into three sub layers specifically OntoSystem Operational Requirements (for system access procedures), OntoSystem Control Requirements (for system control procedures) and OntoSystem Parameter Requirements (for system parameterize procedure). Finally, the fourth layer called OntoOutput Requirements reflects the system eventual presentation [6].

4.1 Layer1. OntoPre-requisites

- System activation by User verification/identification such as Login-Password or any security checks services, User registration facility
- System defined constraints verification

4.2 Layer2. OntoInput Requirements

- Initial inputs or the system required specifications to accelerate the system such as system-defined/access details stacking (or initial data values)

4.3 Layer3. OntoSystem Requirements

To create a chain of commands, services or concerns associated with the system and the environment, the layer 3 is divided into three sub layers.

4.3.1 Sublayer 3. a. OntoSystem Operational Requirements

- System modification and updation competencies such as addition, deletion etc.
- Appropriate user responsive messages where applicable

4.3.2 Sublayer 3. b. OntoSystem Control Requirements

- Advanced electronic Control provisions for accomplishment of system goals such as automation of variety of appliances and managing wide range of disparate technology from single point.
- Knowledge engineering environment

4.3.3 Sublayer 3. c. OntoSystem Parameter Requirements

- Allow to capture the system decision information as intelligent records such as conventional databases and data dictionaries

4.4 Layer4. OntoOutput Requirements

- View the system output in the form of reports, transaction receipt, bills or invoices
- Adding the final information updation to any kind of communication portal such as to mail services or on mobile phones

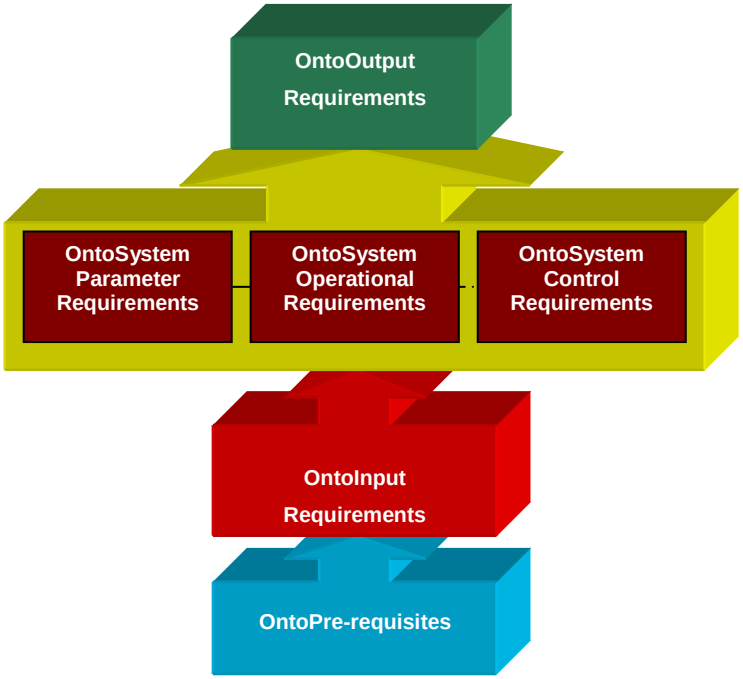


Figure 5. Ontology Aided Requirements Engineering

5. REP Models versus OntoAidedRE

The various parameters related to the project attributes are presented, on the basis of which the different conventional requirement engineering models versus OntoAidedRE can be compared. To understand the impact of various parameters during the project development by using different models as mentioned in Table 4, we map these project parameters with different models. For a Project Type such as TPS or MIS to implement the existing *Sales and Promotion System*, the Linear RE model is most suitable, as relatively straightforward requirements are specified and the model follows sequential life cycle. A medium level of RE process awareness in the larger or medium size projects with many RE activities performed iteratively, as in the Linear Iterative RE model. Developing a *customized website for institutional clients*, with contents such as interactive performance data, that

handles requirements invalidation and requirement freezing. In Project Team, the role defined for RE changed on each project between business systems analyst, project manager and no defined role etc. Consequently, the roles are influenced by the size of the project, since the larger project allowed for greater separation of tasks between the decisions since the new work needs to integrate into a version that already exists such as to replace an existing customer relationship management (CRM) system with a new version from the same vendor. Thus, the Spiral RE model is convenient.

Effective teams identify the factors that affect both the defect rate in results, and the perception of Project Quality. Effective Program and Project Managers establish the prerequisites of Quality from individual roles of business analyst and project manager as in Iterative RE model. Since, the model is appropriate for complicated software development that takes multiple cycles to completely develop the user feedback intensive application. The Project Effort for modifying existing software version that assumes many of the design assignments to overall project results. The differences in the type of RE models appeared to affect the role of Project prioritized element such as cost, time and functionality. Also, Project key element that includes well defined knowledge artifact to establish the open-ended progression of RE process, OntoAidedRE have been assessed showing a major shift in the way that requirements engineering is practiced from process-driven to knowledge driven requirements engineering.

Table 4: Comparison of Requirement Engineering Models vs OntoAidedRE

Parameter	Linear RE Model	Linear Iterative RE Model	Iterative RE Model	Spiral RE Model	OntoAided RE
Project Type	1. TPS 2. MIS	1. PCS	1. OAS	1. DSS 2. ES	All
Project Size	Small/ Medium	Medium	Medium/ Large	Medium/ Large	Small/ Medium/ Large
Project Team	1-10 [in no's]/ 11-25 [in no's]	11-25 [in no's]	11-25 [in no's]/ 26-100[in no's]	11-25[in no's]/ 26-100 [in no's]	Only 1[in no]
Project Effort	8PM	25PM	540PM	100PM	<8PM
Project Quality	[EM: Low P vs. A CRC: High AM: High] Quality Status: upto 65%-74%	[EM: High P vs. A CRC: Medium AM: Medium] Quality Status: Upto 51% to 64%	[EM: Medium P vs. A CRC: Medium AM: Medium] Quality status: Upto 50%	[EM: Medium P vs. A CRC: Medium AM: Medium] Quality Status: Upto 50%	[EM: High P vs. A CRC: High AM: High] Quality Status: Upto 100%
Project prioritized element	Functionality	Time/ Functionality	Cost/ Functionality	Time/ functionality	Time/ Cost/ functionality
Project key element	Not required	Required	Not required	Required	Required

6. Conclusion

The conventional RE process models were reviewed and compared to the proposed OntoAidedRE on the basis of different project parameters. It has been observed that the RE process varies between projects. Consequently, the RE process models were examined on a project-by project basis. While RE occurred in all projects, the attainment of project parameters to its optimal level has been seen as one aspect of the entire RE process.

The Linear, Linear iterative, Iterative and Spiral models were compared to the RE process in each project. It was determined that none of these models were a good representation of every process, but each model represented some RE processes better than others. The Linear iterative model represented processes that were generally linear but involved some iteration of activities. The Linear model characterized purely linear processes, and did not represent iteration of activities. The Iterative and Spiral model represented highly iterative and

enhancive processes, such as prototyping and ad hoc processes. Also, it has been established that one model can not accomplish each and every project parameter. This, in turn, often severely affects the successful completion of projects. OntoAidedRE is one example for how a knowledge-driven as opposed to process-driven approach to requirements engineering could be put into practice in the future to help overcome some of the problems mentioned above and consequently the project attributes can be optimally contrived by adapting the OntoAidedRE.

References

[1] Saiediana H , Daleb R, Requirements engineering: making the connection between the software, developer and customer, 2000

[2] Carina Alves, Anthony Finkelstein, Workshop on software engineering decision support: components: Challenges in COTS decision-making: a goal-driven requirements engineering perspective, Proc. 14th int. conf. on Software engineering and knowledge engineering SEKE '02, July 2002

[3] Kauppinen Marjo, Vartiainen Matti, Kontio Jyrki, Kujala Sari, Sulonen Reijo Implementing requirements engineering processes throughout organizations: success factors and challenges, 2004

[4] Ivan J. Jureta, John Mylopoulos, St´ephane Faulkner, Revisiting the Core Ontology and Problem in Requirements Engineering, 16th IEEE International Requirements Engineering Conference, 2008

[5] Hoffmann, M., Kuhn, N., Weber, M.; Bittner, M., Requirements for requirements management tools, Requirements Engineering Conference, Proc. 12th IEEE International, Page(s):301 – 308, 2004

[6] Shilpa Sharma, Maya Ingle, An Ontology Aided Requirement Engineering Framework, International Journal of Advanced Research in Computer Science, 2011

[7] Aaron Shenhar, Michael Poli, Thomas Lechler, A New Framework for Strategic Project Management, 2006

[8] http://www.lifecyclestep.com/open/406.0HomeProjSize.htm

[9] Stacy Goff, Measuring and Managing Project Quality, 2008

[10] Lubars, M., Potts, C. and Richter, C, A Review of the State of the Practice in Requirements Modeling, *Proceedings of the IEEE International Symposium on Requirements Engineering*, San Diego, USA, pp. 2-14, IEEE Computer Society, 1993

[11] Biffl S., Aurum A., Boehm B., Erdogmus H., Grünbacher P. (eds.), Value-Based Software Engineering, Springer Verlag, ISBN:3-540-25993-7, 2005

Imperative Pathway Analysis to Identify the Potential Drug Target for Aspergillus Infection

V. K. Morya[1]* and Shalini Kumari[2], Eunki Kim[1]
[1]Department of Biological Engineering
Inha University, Incheon, Republic of Korea, 402-751
[2]Department of molecular and cellular engineering
SHIATS, Allahabad, 211007, India

Corresponding address
Dr. V. K. Morya,
Department of Biological Engineering
Inha University, Incheon, Republic of Korea, 402-751
moryavivek@gmail.com

Abstract: *The genus Aspergillus consorting with a prevalent air born filamentous fungal pathogen causing severe and fatal invasive infection in immunecompromised patients. To identify and prioritize antifungal drug targets, against these pathogens is required to develop new pharmaceuticals. We have analyzed various enzymes from different biochemical pathways of pathogen using KEGG database server and compared with human protein. We have identified eight enzymes as the potential target for drug design and are Anthranilate phosphoribosyl transferase / 2.4.2.18, Anthranilate synthase/4.1.3.27, Urate oxidase /1.7.3.3, Keto-acid-reductoisomerase /1.1.1.86, Chitin synthase /2.4.1.16, 3,4-dihydroxy-2-butanone 4-phosphate synthase /4.1.99.12, alpha-alpha-trehalose-phosphate synthase / 2.4.1.15 and Dihydroxy-acid dehydratase/ 4.2.1.9 respectively. The validations of these targets as potential targets were performed by various computational approaches.*

***Keywords*:** *Aspergilosis, Drug target, Antifungal, Keto-acid-reductoisomerase (KARI).*

1. Introduction

It has been suggested that Aspergillus pathogenesis is based on its saprophytic lifestyle in combination with the immunosuppressal state of the host, (Stevens et al. 2000) rather than from genuine fungal virulence factors. The frequency of serious mycotic infection is rising, and this trend has been attributed to such factors as the increasing use of cytotoxic and immunosuppressive drugs to treat both malignant and non-malignant diseases. The increasing prevalence of infection due to HIV (Gupte et al. 2002) and the world –spread use of the newer and more powerful antimicrobial drugs and current treatment options for Aspergillosis are limited to three classes of antifungal therapeutic, Choline, Azole and Echinocandans, fungin and anidulafungin (Petrikkos & Skiada 2007, Hu et al. 2007). Aspergillus Despite the recent introduction of new antifungal agents with promising anti aspergillus activity, mortality associated with invasive Aspergillosis remain prominent, approaching 80-90 % in the high risk leukemia patient and allogenic bone marrow transplant recipient (De Backer et al. 2001, Firon et al. 2003, Osmani et al. 2006,). The genus Aspergillus, is a member of the phylum Ascomycota, includes over 185 known species. Until a date, around 20 of them have been reported to cause harmful infections in humans and animals termed as Aspergillosis (Baba et al. 2006, Giaever et al. 2002, Andersen et al. 2008). Although new generation of antifungal drugs is appearing reviewed by (Baker 2006), development of more effective drugs for treatment of Aspergillosis and other fungal infections is required. Computational method of drug designing and target validation helps in the development of a putative drug molecule. Target identification, target validation and drug targets are three different aspect of drug designing. Computational approach of drug targeting uses various databases and data bank for antifungal drug target, Insilico pathway analysis and drug designing to utilize the information from KEGG, PubChem, ChEBI, RSCB, Swiss-Prot, drug bank, NCBI, DEG, PASUB and variety of structure viewing applets. An approach for antifungal drug design was recently proposed through identification of essential enzyme in Aspergillus. Combination of antifungal design and gene essentiality in Aspergillus has previously been introduced but at that point the main limitation was the generation of gene deletion strain collection.

2. Material and Methods

Pathway analysis was made on Kyoto encyclopedia of gene and genome server (http://www.genome.jp/kegg) (Nierman et al. 2005, Koonin et al. 2004). On the source of Aspergillus flavus open reading frame analysis associated with the reactions. Identification of orthologous candidate's essential genes in A. niger, A. nidulans, A. fumigatus and A. oryzae was made. The use of metabolic models to predict essential reactions and pathways in Aspergillus spp. It has guaranteed to inform reverse genetic studies of gene essentiality and identify potential targets for antifungal development have achieved from above server. Essential reaction pathways, gene and enzyme of Aspergillus species were identified by the help KEGG and essential gene database. The metabolic pathways of A. flavus, A. fumigatus, A. niger, A. nidulans and A. oryzae were searched and compared with the human metabolic pathways online. And the vital enzyme was searched out via comparative study of the metabolic pathway of Aspergillus and human. The essentiality of reaction was predicted by comparing the information to what reported earlier and through study the Insilico pathway analysis. Finally, the

targets were examined for homology search by BLASTp (http://www.blast.ncbi.nlm.nih.gov/) (Ballance and Turner 1985, van Hartingsveldt et al. 1987, de Ruiter-Jacobs et al. 1989) against the human genome. Search for essential gene of Aspergillus was made on database of essential gene (http://tubic.tju.edu.cn/deg/) (Gouka et al. 1995). Afterward the subcellular site of the enzymes was facilitated through PASUB (http://pasub.cs.ualberta.ca:8080/pa/Subcellular) (D'Enfert 1996) database, gives an idea about the existence of the enzyme in cell organelle the statistical term, i.e. pi diagram. It shows the percentage of enzyme in cell organelle. In addition the lists of inhibitors were reported from Helmward Zollner (Weidner 1999). These enzymes were also subjected to Database of essential genes, to find the genes which are essential for the survival of the Aspergilli, although this data bank does not have data related to Aspergilli but still the search was made by subjected the sequence for Blast in the database. For biological significance and distribution of these essential targets were analysed by PA-SUB (Proteome Analyst Specialized Sub Cellular Localization) Server v2.5. (Lu et al. 2004). This is required to find out the surface membrane proteins which could be probable vaccine targets.

3. Results and discussion

All vital reactions and pathways were recognized in Aspergillus by study of the metabolic pathways for this organism and by comparative analysis of the metabolic pathways of the Aspergilli with the human. Comparative study of the metabolic pathways of Aspergillus with human gives the idea of antifungal drug target. Since in the metabolism various enzymes are engaged and these enzymes may be the key for antifungal drug targeting. Those enzymes which are the absence in human can be a potential target for Aspergillosis. However, not all the enzymes which are the absence in the human can be targeted as antifungal drug because they also contain a large number of the pathway and enzyme which are not essential for the viability of the Aspergillus (D'Enfert et al. 1996). The common enzymes from the five species of Aspergillus, i.e. A. flavus, A. fumigatus, A. niger, A. nidulance and A. oryzae but absence in human have been analyzed. These enzymes were involving in one or more metabolic pathway/s. The metabolic model based on pathway analysis were characterized as being essential and as illustrated in these reactions grouped into 15 essential pathways covering the following categories: amino acid biosynthesis and metabolism, Arachidonic acid metabolism, Base excision Repair, cell wall biosynthesis and disaccharide metabolism, Glyoxylate and dicarboxylate metabolism, Inositol phosphate metabolism, purine metabolism, mannitol biosynthesis, nitrogen metabolism, Nicotinate and nicotinamide metabolism, propionate metabolism, Pantothenate and CoA biosynthesis, Riboflavin metabolism, Starch/Sucrose metabolism, and Sulfur metabolism. The Insilico approach for identification of potential antifungal targets, the Aspergillus genes identified through pathway analysis were all compared by pBLAST to the human genome (Pel et al. 2007). 40 enzymes from Aspergilli with scores (bits) 50 were categorized as they are absent in human and these were classified as possible antifungal targets and a potential drug can be design based on these data (Table 1) this study is also supported by earlier reports (Liebmann et al. 2004, Maerker et al. 2005, Ibrahim-Granet et al. 2008).

Most of the enzyme which was found as an antifungal drug target is the enzyme of the amino acid biosynthesis pathways, and these enzymes can be a valid target for the drug targeting because these pathways are already validated. In these 40 enzymes, some of the enzymes are present in one pathway while some of them are present in more than one pathway. DEG analysis shows that, out of 40 enzymes only eight enzymes were found, out of which 4 are putative and 3 are those enzymes of which not positive result was found (Table-2). These enzymes are listed along with their gi number it also shows abundance of individual enzyme in cellular compartment. After PASUB analysis was carried out to get the information about the enzyme that which enzyme is present in which organelle of the cell. The pi diagram of essential enzymes for their distribution in the cellular component was also evaluated, and the targets were distributed in cytoplasm and mitochondria only while for the rest no data was available (Figure 1). Inhibitors for enzyme were identified from the previously reported database and are listed in table 3. This inhibitor can exploit for construction of Zinc library for docking and other analysis. Eight potential and promising targets namely Anthranilate phosphoribosyltransferase (2.4.2.18), Anthranilate synthase (4.1.3.27), Urate oxidase (1.7.3.3), Keto-acid-reductoisomerase (1.1.1.86), Chitin synthase (2.4.1.16), 3,4-dihydroxy-2-butanone4-phosphate synthase (4.1.99.12), alpha,alpha-trehalose-phosphate synthase (2.4.1.15), Dihydroxy-acid dehydratase (4.2.1.9) could be exploiting further rational drug design for opportunistic fungal pathogen like Aspergillus. The brief of the selected eight enzymes was given below.

In enzymology, an anthranilate phosphoribosyltransferase (EC 2.4.2.18) is an enzyme that catalyzes the chemical reaction N-(5-phospho-D-ribosyl)-anthranilate + diphosphate anthranilate + 5-phospho-alpha-D-ribose 1-diphosphate. Thus, the two substrates of this enzyme are N-(5-phospho-D-ribosyl)-anthranilate and diphosphate, whereas its two products are anthranilate and 5-phospho-alpha-D-ribose 1-diphosphate. This enzyme belongs to the family of glycosyltransferases enzyme, specifically the pentosyltransferases. This enzyme participates in the phenylalanine, tyrosine and tryptophan biosynthesis and two-component system – general [Kima et al., 2002].

Anthranilate is a common precursor of all these compounds and anthranilate synthase (AS) catalyzes the synthesis of anthranilate from chorismate and Gln. Although the synthesis of anthranilate from chorismate and Gln requires - and -subunits of AS, the -subunit can catalyze the synthesis of anthranilate from chorismate with ammonia as an amino donor in the presence of high concentrations of ammonium. Specific amino acids in the -subunit of bacterial AS are important for feedback inhibition of the enzyme by Trp [Matsui et al., 1987]. Thus inhibition of this enzyme may suppress the normal growth and proliferation of pathogen.

Table 1. Potential antifungal targets and the pathway/s in which they are involved with their function/s, known inhibitors.

EC No. /Essential enzyme	No. of pathways	Potential function	Known Inhibitors
2.4.2.18 / Anthranilate phosphorib-Osyltransferase	Phenylalanine, tyrosine and tryptophan biosynthesis Metabolic pathways	Anthranilate to N- (5-phosphoB-D- ribosyl) Anthranilate	anthrnilic acid; 3-hydroxyanthranilic acid; n-(5-phospho-d-ribosyl) anthrnilaiccid; pyrophosphate
1.1.1.86 / Ketol acid reductoisomerase	Valine, leucine and isoleucine biosynthesis Pantothenate and CoA biosynthesis Metabolic pathways	2- Acetolactate to 3 – Hydroxy, 3- methyl2-oxobutanoate	P-chloromercuribenzoic acid N-hydroxy-n-isopropy loxamate 2-methyllactic acid 2-0x0-p-hy droxy isovalerate P-chloromercuribenzoic acid N-hydroxy-n-isopropy loxamate 2-methyllactic acid 2-0x0-p-hydroxy isovalerate
2.4.1.16 / Chitien synthase	Amino sugar and nucleotide sugar metabolism	UDP-GlcNAc to chitin	N-acetyl, d-glucosamine; Amphotericin b; Calcofluor Captan; 8-0-demethylpseurotin a; Edifenphos; Glycerol; Hexachlorophene; Leu-polyoxin b Linoleic acid; 2.2- methylenebis [3,4,6-trichlorophenol] Nikkomycin; Nystatin; N-p-octanoyllysyl-uraciplo lyoxin c; Oleic acid; Pentachlorophenol; Polyixin
4.1.3.27 / Anthranilate synthase	Phenylalanine, tyrosineand tryptophan biosynthesis	Chorismate to Anthranilate	Acivicin; **(**4r*,5s*,6s*)4-amino-5-[(1-carboxyethenyl)oxy]-6-hydroxycyclohex-1-ene carboxylic acid; l-2-amino-4-oxo-5-chloropentanaociicd; anthranilic acid; azaserine (lr-trans)-2-[(bromo-6-hydroxy-2,4-cyclohe~dien-l-yl)-oxyi-2-propenoic acid; bromopyruvic acid; 0-carbamoylserine; (s)-3-[(1-carboxyethenyl)oxy]-1,6-cycloheptadiene-l- carboxylic acid; Chanoclavine; p-chloromercuribenzoic acid; 5-chloro-4-oxonorvaline; 6-diazo-5-0x0-l-norleucine dimethy 1amine; Elymoclavine; n-ethylmaleimide d,l-5-fluorotryptophan l-glutamic acid
1.7.3.3/ Urate oxidase	Purine metabolism Caffeine metabolism Metabolic pathways	Urate to 5-hydroxyisourate	Amelide; 5-azaorotic acid; 8-azaxanthine; 2-chloro 6-amino 8-hydroxypurine; Cyanurate; dicyandiamide; dithionite; n-ethylmaleimide; glycerol; guanine; 2-hydroxypurine; oxonic acid; sucrose; metal ion heavy
2.4.1.15/ alpha,alpha-trehalose-phosphate	Starch and sucrose metabolism	UDP glucose to alpha alpha' terhalose-6p	Not available
4.1.99.12 / 3,4-dihydroxy-2-butanone4-phosphate synthase	Riboflavin metabolism	Ribulose 5- phosphate to 3,4 Dihydroxy 2-butanone 4- phosphate	Not available
4.2.1.9 / Dihydroxy-acid dehydratase	Valine, leucine and isoleucine biosynthesis	2,3 – Hydroxy, 3-methybutanoate to 3-methyl-2- oxobutanoate	Not available

Uric acid is catabolized to allantoin by urate oxidase, or uricase (E.C. 1.7.3.3.), in most vertebrates except humans, some primates, birds, and some species of reptiles. Urate oxidase has attracted considerable interest for several reasons: (i) It has a unique evolutionary feature in that the enzyme has been lost during primate evolution with no

obvious explanation. (ii) The development of mice with the complete hypoxanthine-guanine phosphoribosyltransferase deficiencies that does not display any of the symptoms of Lesch-Nyhan syndrome has raised the possibility that the presence of urate oxidase in these mice may somehow protect them from neurological damage. (iii) The enzyme has been used as a peroxisomal marker and is potentially a good system for studying protein sorting into peroxisomes. Therefore, targeting this enzyme may serve as the better drug target having low cross reactivity [Wu et al., 1989]. The biosynthesis of the branched-chain amino acids Val, Leu, and Ile in microorganisms and plants uses the common enzymes ALS (EC 4.1.3.18), KARI (EC 1.1.1.86), and dihydroxyacid dehydratase (EC 4.2.1.9). **KARI (also referred to as acetohydroxy-acid reductoisomerase)** is the second enzyme in the biosynthetic pathway. For the biosynthesis of Val and Leu, the enzyme catalyzes the reductive isomerization (alkyl migration) of 2-acetolactate to 2,3-dihydroxyisovalerate; for catalysis, the enzyme requires Mgz+ and NADPH as cofactors. For the corresponding formation of Ile, KARI converts 2-aceto-2- hydroxybutyrate to 2,3-dihydroxy-3-methylvalerate [Durner et al., 1993].

Chitin synthase (CHS) (EC 2.4.1.16) catalyzes the polymerization of chitin from UDP-N-acetylglucosamine (UDP-GlcNAc) monomers. CHSs are large enzymes located in the plasma membrane, enabling the newly synthesized chitin to be extruded from the cell into extracellular environment. CHSs belong to the GT2 family of glycosyltransferases, which also includes cellulose synthases.

CHSs have been extensively studied in fungi. Fungal CHSs have been found to have different roles in cell wall and septum biosynthesis and are expressed at different developmental stages [Hogenkamp et al., 2005].

L-3,4-dihydroxy-2-butanone-4-phosphate synthase (DS) is involves in catalysis of the ribulose 5-phosphate to L-3,4-dihydroxy-2-butanone-4-phosphate (DHBP) and format. 3,4-Dihydroxy-2-butanone 4-phosphate synthase supplies the building blocks for the assembly of the xylene ring of the vitamin. In fact, all eight carbon atoms of the xylene moiety are derived from the product of the enzyme. The riboflavin pathway is absent in human host there for inhibitor design for enzymes involve in this pathway would be safer and effective too [Liao et al., 2001].

Trehalose 6-phosphate synthase (TPS: E.C.2.4.1.15) catalyzes the formation of trehalose 6-phosphate from UDP-glucose (UDPG) and glucose 6-phosphate (G-6-P). Trehalose 6-phosphate is subsequently dephosphorylated to trehalose by a specific trehalose 6-phosphate phosphatase. Trehalose biosynthesis has been extensively studied in microorganisms.

Dihydroxy acid dehydratase (2,3-dihydroxy acid hydrolyase, E.C. 4.2.1.9) catalyzes the third step in branched-chain amino acid biosynthesis which involves the dehydration and tautomerization of two naturally occurring2 ,3-dihydroxycarboxylic acids to the corresponding 2-kaectoid s. Its substrates and products (intermediates in the biosynthesis of valine, leucine, **CoA,** and isoleucine).

Targeting the mentioned enzyme as the drug target, may serve as better and safer for developing new type therapeutics against Aspergillus infection. These enzymes were not reported from the human host. Therefore, the adverse effect of inhibitors on the host will minimum.

4. Conclusion

The pathway analysis of Aspergilli can be moved forward to get better footnote of the metablome in the designing and targeting of the antifungal drug. In addition we find the metabolic model establishment of functional links between genes, enabling the upgrade of the information content of transcriptome data. Comparative analyses of pathogen and host pathways and molecular research on genes related with fungal viability, and virulence has led to the detection of many targets and putative targets for novel antifungal agents on the basis KEGG information. Although when the essential enzymes' search for Aspergillus was made on the DEG, essential enzyme for Aspergillus was only 8 because DEG does not contain essential gene data for Aspergillus it may be one region for fewer essential genes. So far, the rational approach to antifungal discovery, in which compounds are optimized against being targeted then progressed to efficacy against the integrated fungal pathogens and ultimately to infected humans has delivered no new agents. However, the approach continues to hold promise for the future.

5. References

1. Stevens DA, Kan VL, Judson MA 2000. Practice guidelines for diseases caused by Aspergillus. *Clin. Infect. Dis. 30:* 696-709.
2. Gupte M, Kulkarni P, Ganguli BN 2002. Antifungal antibiotics. *Appl. Microbiol. Biotechnol. 58:* 46-57.
3. Petrikkos G, Skiada A 2007. Recent advances in antifungal chemotherapy. *Internat. J. Antimicrob. Agents 30:* 108-117.
4. Hu W, Sillaots S, Lemieux S, Davison J, Kauffman S 2007. Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. *PLoS Pathog. 3(3):* e24. doi:10.1371/journal.ppat.0030024
5. De Backer MD, Nelissen B, Logghe M 2001. An antisense-based functional genomics approach for identification of genes critical for growth of Candida albicans. *Nature Biotechnol. 19:* 235-241.
6. Firon A, Villalba F, Beffa R, d'Enfert C 2003. Identification of essential genes in the human fungal pathogen *Aspergillus fumigatus* by transposon mutagenesis. Eukary. *Cell 2:* 247-255.
7. Osmani AH, Oakley BR, Osmani SA 2006. Identification and analysis of essential *Aspergillus nidulans* genes using the heterokaryon rescue technique. *Nature Protocols 1:* 2517-2526.
8. Baba T, Ara T, Hasegawa M 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 0008 doi:10.1038/msb4100050.
9. Giaever G, Chu AM, Ni L 2002. Functional profiling of the Saccharomyces cerevisiae genome. *Nature 418:* 387-391.
1. 10.Andersen MR, Nielsen ML, Nielsen J 2008. Metabolic model integration of the bibliome, genome, metabolome, and reactome of Aspergillus niger. *Mol. Syst. Biol. 4:* 178.
10. Baker SE 2006. *Aspergillus niger* genomics: past, present and into the future. *Med. Mycol. 44:* S17-S21.
11. Nierman WC, Pain A, Anderson MJ 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature 438:* 1151-1156.

12. Koonin EV, Fedorova ND, Jackson JD 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol. 5:* R7.
13. Ballance DJ and Turner G 1985. Development of a high-frequency transforming vector for *Aspergillus nidulans*. *Gene 36:* 321-331.
14. van Hartingsveldt W, Mattern IE, van Zeijl C, M., Pouwels PH, van den Hondel CA 1987. Development of a homologous transformation system for *Aspergillus niger* based on the pyrG gene. *Mol. Gen. Genet. 206:* 71-75.
15. de Ruiter-Jacobs YM, Broekhuijsen M, Unkles SE 1989. A gene transfer system based on the homologous pyrG gene and efficient expression of bacterial genes in Aspergillus oryzae. *Curr. Genet. 16:* 156-163.
16. Gouka RJ, Hessing JG, Stam H, MustersW, van den Hondel CA 1995. A novel strategy for the isolation of defined pyrG mutants and the development of a site-specific integration system for *Aspergillus awamori*. *Curr. Genet. 27:* 536-540.
17. D'Enfert C 1996. Selection of multiple disruption events in Aspergillus fumigatus using the orotidine-5?-decarboxylase gene, pyrG, as a unique transformation marker. *Cur. Genet. 30:* 76-82.
18. Weidner G, d'Enfert C, Koch A, Mol PC, Brakhage AA 1998. Development of a homologous transformation system for the human pathogenic fungus Aspergillus fumigatus based on the pyrG gene encoding orotidine 5'-monophosphate decarboxylase. *Curr. Genet. 33:* 378-385.
19. Lu Z, Szafron D, Greiner R, Lu P, Wishart DS 2004. Predicting Subcellular Localization of Proteins using Machine- Learned Classifiers. *Bioinformatics 20:* 547-556.
20. D'Enfert C, Diaquin M, Delit A 1996. Attenuated virulence of uridine-uracil auxotrophs of *Aspergillus fumigatus*. *Inf. Immun. 64:* 4401-4405.
21. Pel HJ, de Winde JH, Archer DB 2007. Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nature Biotechnol 25:* 221-231.
22. Liebmann B, Muhleisen TW, Muller M 2004. Deletion of the *Aspergillus fumigatus* lysine biosynthesis gene lysF encoding homoaconitase leads to attenuated virulence in a low-dose mouse infection model of invasive aspergillosis. *Arch. Microbiol. 181:* 378-383.
23. Maerker C, Rohde M, Brakhage AA, Brock M 2005. Methylcitrate synthase from *Aspergillus fumigatus*-Propionyl-CoA affects polyketide synthesis, growth and morphology of conidia. *FEBS J 272:* 3615-3630.
24. Ibrahim-Granet O, Dubourdeau M, Latge´ JP 2008. Methylcitrate synthase from *Aspergillus fumigatus* is essential for manifestation of invasive aspergillosis. *Cell Microbio. 10:* 134-148.
25. Kima C., Xuong N. H., Edwards S., Madhusudan, Yee M.C., Spraggon G. Mills S.E., (2002). The crystal structure of anthranilate phosphoribosyltransferase from the enterobacterium Pectobacterium carotovorum. FEBS Letters 523 (2002) 239-246
26. Matsui K., Miwa K, and Sano K. (1987). Two single-base-pair substitutions causing desensitization to tryptophan feedback inhibition of anthranilate synthase and enhanced expression of tryptophan gene of *Brevibacterium lactofermentum. J. Bacteriol.* 169, 5330-5332.
27. Wu X, Leeo CC, Muznyt DM, and Caskey CT. (1989). Urate oxidase: primary structure and evolutionary implications. Proc. Nati. Acad. Sci. Usa Vol. 86, pp. 9412-9416,
28. Durner J, Knorzer O. C., and Boger P., (1993). Ketol-Acid Reductoisomerase from Barley *(Hordeum vulgare):* Purification, Properties, and Specific lnhibition. Plant Physiol. (1993) 103: 903-910
29. Hogenkamp D G, Arakane Y, Zimoch L, Merzendorfer H, Kramer KJ, Beeman RW, Kanost MR, Specht CA, Muthukrishnan S, (2005), Chitin synthase genes in Manduca sexta: characterization of a gut-specific transcript and differential tissue expression of alternately spliced mRNAs during development. Insect Biochemistry and Molecular Biology 35 (2005) 529–540
30. Liao Der-Ing, Calabrese Joseph C., Wawrzak Zdzislaw, Viitanen Paul V., and Jordan Douglas B., (2001) Crystal Structure of 3,4-Dihydroxy-2-Butanone 4-Phosphate Synthase of Riboflavin Biosynthesis. Structure, Vol. 9, 11–18,

In-Vitro Analysis of RM Stent for Marginal Coronary Artery

Sanjay pujai[a], and Dr.V.R.udupi[a]
[a]Department of Electronics and Communication research center, G.I.T, Belgaum,
Belgaum – 591 008, India.
e-mail: sapujari@rediffmail.com, vishw_u@yahoo.com

Abstract: *The production of stent designs poses difficult problems to clinicians, who have to learn the relative merits of all stents to ensure optimal selection for each lesion, and also to regulatory authorities who have the dilemma of preventing the inappropriate marketing of substandard stents while not denying patients the benefits of advanced technology. Of the major factors influencing long-term results, those of patency and restenosis are being actively studied where as the mechanical characteristics of devices influencing the technical results of stenting remain under-investigated. Each different stent design has its own particular features. A robust method for the independent objective comparison of the mechanical performance of each design is required. To do this by experimental measurement alone may be prohibitively expensive. A less costly option is to combine computer analysis, employing the standard numerical technique of the finite element method (FEM), with targeted experimental measurements of the specific mechanical behaviour of stents for specific coronary artery i.e marginal artery. In this paper the FEM technique is used to investigate the structural behaviour of sap stent in marginal artery. The stress, strain and increasing the RM stent diameter, performance were investigated when the internal presuure is applied. The experimental scheme presented in this paper provides invitro stent expansion, process by finite element analysis using ANSYS 10.0 analysis software.*

Keywords: *Finite element analysis; RM Stents; Mechanical properties; ANSYS.*

1. Introduction

Background and motivation the current global stent use and research is enthusiastic in progress, domestic use of stents to treat coronary artery stenosis ratio also year by year increase, but so far the domestic use of vascular stent products is foreign imports. The design of cardiovascular stents, currently there is no relevant research papers in domestic and self designed products. Manufacture of a complete addition to the structural frame design, they also need a lot of research such as material selection, surface treatment and processing of the planning process and so on. We must first have sufficient strength to support the vessel wall the future, but also to maintain sufficient flexibility, and whether the uniform stent expansion and will not produce changes in the length of stent placement is an important factor. This mechanism design approach to change the adverse cardiovascular stent patent overseas on both the structural design to the main design goals, the future conduct of new stent design. To design a change defective structure, structural features in the expansion process, does not have the axial length change, and expansion after the formation of structures with strong the strength to resist vessel wall stress, reduce restenosis opportunities.

1.1 Literature Review

Stent term was first in the UK by a dentist Charles Stent and the future; he uses a rack for future improvement of the teeth, then the future stent because of its similar design to the next name to his name. The earliest animal testing is carried out Zhijia 1954 AD by Charles Dotter conducted; the success of this experiment will be woven mesh stent position in experimental bile duct in the dog. AD 1986, Puel team conducted the first human stent clinical trials; the use of status is also woven mesh stent (WallStent), the open order to test the use of stents in the first such body, after Scaffold to flourish. Years 1995 to 1997 AD, Cardon team, Rudnick team, and Robert team in the United States Patent Office, respectively, we propose a novel stent design. Cardon team has published composite design of the stent, also includes tubular and net position woven structure; Rudnick team made changes Leung-type toroidal stent structure; Robert team for the first time to design easy on the placement of the stent BETH (Butterfly Expandable to Honeycomb). AD 1997, Balcon, Beyar and Chierchia and other researchers published the first level of management on the entire stent design, manufacture and application of relevant norms and recommendations.

Meanwhile, there have been efforts to model and design stents computationally, as it had already been recognized that stent design affects restenosis (Kastrati et al., 2001; Rogers and Edelman, 1995; Rogers et al., 1998). Linear elastic models by Rogers et al. (1998) modeled balloon expansion with stent and artery contact using a 2-dimensional model. Investigators such as Migliavacca and colleagues (Petrini et al., 2004; Migliavacca et al., 2002; Migliavacca et al., 2005) have focused mostly on the characterization of mechanical properties of stents. Prendergast and colleagues (Lally et al., 2005) modeled the stent-artery interaction of commercially available stents (NIR – Boston Scientific; S7 – Medtronic AVE) on an idealized stenosed artery. Furthermore, they created a simplified restenosis algorithm that would simulate the process of neointimal hyperplasia and restenosis. Holzapfel et al. (2002) modeled the balloon expansion of a full 3-dimensional anisotropic diseased artery Because of rapid development of computer techniques it is possible to simulate the stent implantation and utilization of the stent. It is possible to create the RM stent model, on the basis of physical and mathematical model and FEM allows to determine the implant behaviour in conditions nearing real ones and searching optimal material properties and constructions at the some time.

2.0 Methodology

2.1.1 Introduction to FEA

Finite element analysis (FEA) is a numerical technique that allows users to calculate how a field (e.g. displacement, temperature, velocity) can vary in space within a particular geometry.Since the computer efficiency, power and memory capacity have improved over the last few decades, the level of complexity of problems has increased, particularly concerning the geometrical presentation and material behaviour of the physical problems [5].A complete Finite Element Analysis consists of three stages: preprocessing, simulation, and postprocessing (see Fig 1).

2.1.2 Preprocessing:

The first steps of any finite element simulation are to represent the physical problem by a mathematical model or idealization. That means that the material properties in the geometry and the environment around the geometry (boundary conditions, applied loads) must be expressed mathematically [5]. The actual geometry of the structure is then discretize using a collection of finite elements. Each finite element represents a discrete portion of the physical structure. The finite elements are joined by shared nodes. The collection of nodes and finite elements is called the mesh.

2.1.3Simulation:

In a stress analysis the displacements of the nodes are the fundamental variables that an FE software program calculates. Once the nodal displacements are known, the stresses and strains in each finite element can be determined. The simulation is the stage in which the Finite Element Package solves the numerical problem defined in the model and stores the output files ready for postprocessing.

2.1.4 Postprocessing:

Once the solution is completed, the next step is to analyse the results with a postprocessing module of the FEM software package. The values of the field variables at all nodes or other physical parameter calculated on the basis of the field variables are listed or graphically displayed.

The visualisation of the results severs to understand and to interpret the results. An error analysis serves to analyse the computed results critically and to make decisions of their Accuracy

The main advantage of FEM is that the performance of cardiovascular prostheses can be predicted without considerable costs associated with laboratory experimentation. FEM analyses can be carried out before prototype are manufactured, and therefore, reduces the development costs. However, performing a simulation, certain simplifying assumptions have to be made. Thus, simulation is an approximation to describe real problems and it is considered as complement to experimental tests [1-2].

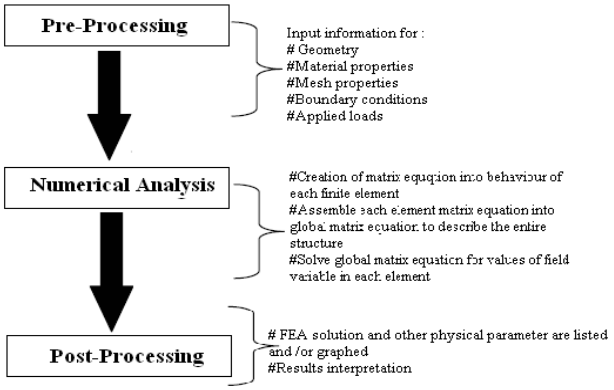


Fig.1 Schematically description of complete Finite Element Analysis.Representing the three stages: preprocessing, simulation, and postprocessing.The main tasks of each stage are summarised.

2.2 Stent materials

The correct choice of the stent material is an essential step of stent design process as well as for every other medical device. In this chapter, the current stent metal options available are provided and the influence of their features on this choice is discussed. The stent material choice is a symbiosis between the required mechanical properties (flexibility, ability to support the duct wall, ability to expand etc.) and the biocompatibility.

2.2.1Biocompatibility

It is not easy to provide a unique definition of biocompatibility. In fact, the interaction of a material with the body involves many factors and a single test is not able to define whether a material is biocompatible or not. Here, the biocompatibility of a stent is considered. This device is a long-term implanted device, so its biocompatibility can be defined as the property of the device to perform its task, not providing any undesirable local or systemic effects in the host body. A biocompatible material may not lead to a toxic or harmful immunological response. The way to assess the biocompatibility or the tissue reaction is to test the substances firstly in vitro then in vivo. The usual response of the host to an implant includes:

- Trauma;
- Inflammation;
- The immune system reaction;
- Eventual healing or scarring.

The biocompatibility of the metals is directly related to their ability to resist to the aggression of biological fluids. In fact the biological fluids have a well-known corrosive action. The corrosion has consequences on the mechanical properties of the metal implant and a toxicological action since it leads the release of metallic ions in the body. The anti-corrosion features of the metals can be improved applying ad hoc technologies acting on the chemical stability of the device surface (i.e. passivation) and on the surface geometry (i.e. electrochemical polishing).

The evaluation of the biocompatibility requires a well define procedure which starts investigating the chemical behavior of the sole metal and concludes with the implant set in the body host. In fact the in vitro tests involve the cytotoxicity (i.e. the quality of being toxic to cells; examples of toxic agents are a chemical substance or an immune cell) of the material but they are not able to assess the cell interaction with the surface properties of the material. For this reason, it is necessary to perform in vivo tests on animals (i.e. pigs, dogs and sheep) having features comparable to the human body. The final phase of the testing is the implant of the device into the human body during clinical trials[1].

The stent market is a very interesting research field also from the point of view of material science, the discovery of a new material can lead to a sudden wide and rapid revolution. Furthermore, the rapid influx of the new stent material requires well understanding of the relative strengths and weaknesses of the various materials. The mechanical and chemical properties, highlighting their role in the stent design; the material properties used in this paper *is* ***Stainless steel (SS)*** Steel is an alloy of iron. It includes 2% of carbonium (C) (at most). The stainless steels are specific types of steel, which well resists, to the chemical agents. It is includes chromium (Cr) (more than 10%) and nickel (Ni). By the crystal microstructure, the SSs can be classified as:

- martensitic SS;
- austenitic SS;
- ferritic SS;

Here, the focus will be on the austenitic SSs which are the most important class of materials for medical device applications and most BX stents, which are based on the metal plastic deformation, are stainless steel made (SS). These types of SSs are interesting because they propose a

good compromise among corrosion resistance, strength, formability, weld ability, and cost. They are also non magnetic allowing magnetic resonance imaging (MRI) of the device. The first austenitic SS used in the medical application was the Type 302 but it became obsolete since its scarce corrosion resistance. Nowadays the most widely used SS is the Type 316L, a variation of Type 316 including molybdenum, which provides more resistance to the salt water: the Type 316L contains less carbon than Type 316 as shown in the Table1.

Table.1: Composition of austenitic stainless steels Type 316 and 316L.

Type	C	Cr	Ni	Other alloying elements
316	0.1% (max)	16-18%	10-14%	Mo:2-3%
3i6L	0.03% (max)	16-18%	10-14%	Mo:2-3%

2.2.2 Ideal Characteristics of Stent

Coronary stenting was introduced into clinical practice in 1987, just ten years following the introduction of percutaneous angioplasty. Stent performance depends on both the material properties and the physical attributes of the stent.

The ideal stent material will contain the following characteristics:

1. Good biocompatibility – To ensure no adverse reaction occurs when the stent is implanted in the body.
2. Fatigue resistant – Fatigue failure can occur in stents due to the cyclic stresses created by the blood flow.
3. Good radiopacity – To enable the visibility of the stent under standard Xray and MRI.
4. Sufficient radial strength – Low yield strength is required to allow for sufficient stent expansion and high tensile properties are required after expansion to achieve sufficient radial strength to maintain patency.
5. Low recoil – both radial and longitudinal.
6. Good axial and radial flexibility – To enable navigation through the tortuous vessel.
7. Good deliverability – To enable access to smaller vessels.
8. Low profile – Higher tensile properties enable the use of thinner stent struts and therefore an overall lower profile. This also improves the flexibility and deliverability of the stent.

2.3 Cardiovascular stent design methods and procedures

The cardiovascular stent design, and analysis steps is divided into four categories, the first stent designed structure, the second for graphics rendering drawn by the SoildWork

Stent the same appearance of the geometry and finite element analysis using ANSYS 10.0 software to conduct analysis, the third stent is designed to come out of the actual processing, and finally the fourth is the future of the stent on the processing carried out a variety of analysis and testing and the with the first three steps in the simulation results can be compared.

2.3.1 RM Stent Model

To serve as a basis for comparison, a series of "typical" stents were designed, dimensioned, and modeled as described in this section. The stents analyzed here, within the class of arterial stents, lay within the industry norm with regard to dimensions, slenderness ratios, and other geometric and material parameters Representative crown-patterns/links of each geometry are shown in Fig. 1. Note that in point of fact the geometry of stent designs varies widely between manufacturers and models; however, it was found that the trends and conclusions drawn here with regard to numerical modeling were valid for all of the analyzed stents. Due to this, the remaining sections summarize the results for the bottom right stent of Fig. 1 as a typical example.

2.3.2 Geometry Model

Commercially available stent geometries are often subjected to very strict patent claims. For this reason, manufacturer's specific and detailed information regarding the stent geometry is usually not available in the public domain. After discussing with the cardiologist, depending upon the environment and structure of a plaque in marginal coronary artery. In the present analysis, RM types of stents have been considered for the specife marginal coronary artery.

Regular Meshed (Rm) Model in unexpanded configuration this model illustrated in figure is having a length of 8mm and external diameter of 1.5 mm. It comprises of 540 slots in different rows with an internal diameter of 0.7386 mm. This configuration provides highest rigidity with least amount of deformation. The models are designed using SolidWorks 2004 is 3-D geometric modeling software. This software permits accurate creation of a 3-D component with given appropriate dimensions. The balloon and stent are created into 3-D models in SolidWorks 2004. These models are exports as parasolid files to ANSYS for creation of the FEM. The stent geometry was initially modeled in the three-dimensional Cartesian coordinate system, representing the stent in an opened-out planar configuration. To generate the planar geometry of the stent, line profiles that represented the skeleton of the stent were created and a circular area measuring in diameter was subsequently extruded along the line profiles to generate the stent volume.

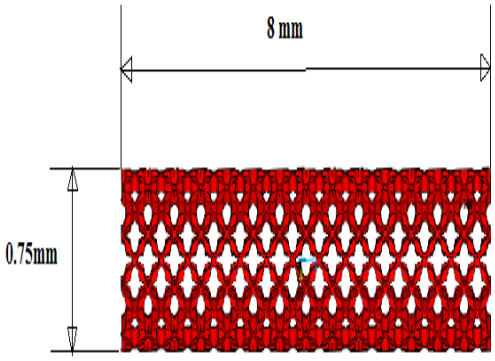


Fig. 2 The Geometerical RM model.

The volumes of the stent were then discretised using twenty-node isoparametric, three dimensional elements. A mesh convergence study was carried out which specified an optimum mesh density.

The integrity of the elements was checked for their shape to ensure that no distorted elements were generated. Adaptive meshing was used to ensure minimal penetration occurred when the stent came into contact with the artery by subdividing the elements on the artery which were contacted by the stent. This adaptive meshing and subsequent mesh refinement results in an improvement in the accuracy of the solution as well as reducing the computational time that would be required if the entire artery were to have a refined mesh.

The nodal coordinates of the meshed model were transferred from a Cartesian coordinate system into a cylindrical coordinate system, using a procedure reported by Lally *et al*[4]. Whereby the planar configuration was wrapped to represent the cylindrical structure of the stents. The elements, nodes and their connectivity were then transferred into the finite element code ANSYS and large deformation analyses were solved to simulate the expansion of the stent.

2.3.3 Material

The two primary assumptions used in the material modeling are initial homogeneity and isotropy. The former implies that the material properties are initially constant throughout the device, while the latter implies that material response is not a function of its orientation. The specific material under consideration is 316L stainless steel. The material properties are described in Table 2. The properties listed were assumed identical under tension and compression. The inelastic response is that of classical linear isotropic hardening plasticity . Given that the three modeling strategies considered required different levels of dimensionality in the constitutive response i.e., beams, shells, and solids, the constitution in each case was derived from the same basic formulation.

Tabel 2: Material Properties of SS 316.

Parameters	
Young's modulus E	**200000 Mpa**
Poisson's ratio v	**0.33**
Yield stress	**690Mpa**

2.3.4 Boundary Conditions.

In the analysis of stent deployment, boundary conditions are critical for obtaining an approximate solution which accurately captures the behavior of the physical system. Incorrect imposition of boundary conditions can lead to systems which are either numerically stiff, lack the correct response, or have spurious rigid body modes[2] .

The boundary conditions applied to the boundary value problem included displacement boundary conditions, pressure, and contact. The vessel was stretched in the axial direction by 59% simulating the axial tethering that was measured in vivo, it was necessary to apply boundary conditions on the arterial wall at the 0^o and 360^o positions such that when inflated, the artery would deform uniformly while the wall at those positions would remain in its original plane. The artery was then inflated by applying a pressure in the range of.8 to 1.6Mpa. These pressure was determined by numerical experiments and it was found that this value dilated the artery enough such that the 10% oversized stent could be "implanted". The stent was originally positioned outside the artery and then translated in the axial direction such that the stent and artery mid-points along that direction coincided. In the Windows version of Ansys it is possible to simply alter the contacts table such that during inflation load step, the artery is allowed to pass through the stent without making contact. In subsequent load steps the contact table can be modified and reactivated so that contact may occur during systole and diastole. The boundary conditions on the stent beyond the translation step, included in plane deformation for the struts identical to those applied to the artery, and an analytical contact boundary condition.

3.0 Result

Results of the finite element method with ANSYS are nodal values by default. The resulting table of nodal values can be plotted as a color map of the model for qualitative analysis. The table can also be evaluated by manipulating the quantitative outputs. Both approaches are used herein to provide a more complete conception of the impact of stent design on stresses in the artery wall [6].

The symmetry boundary conditions necessary to take advantage of the reduced computational load, can cause edge effects due to the nature of the contact boundary condition is not a symmetric one, and therefore it can occasionally cause anomaliesin the results, such as those detected in RM stent model simulations as seen in below table 3.

The stent models employing distinct variations of the stent parameters outlined above were developed. After solution is done the reasult are obtained to calculate displacements, stresses, and strain Residual stresses and strains, inherent in many biologic tissues, result from adaptation mechanisms. The considerations have a strong influence on the stress and strain distributions at physiological condition across the stent wall, and on the global pressure/radius response of arteries [1].

The following tables summarize all the data pertaining to stent under different pressures, which concludes that there is linear variation of displacement, stress and strain with applied pressure

Table 3: data pertaining to stent under different pressures

	Pressures in MPa				
DISCRIPTON	**0.8**	**1**	**1.2**	**0.8**	**1.6**
Sum dimeter mm	0.88348	1.105	1.326	1.547	1.768
Stress in XY plane MPa	4.282	5.352	6.422	7.493	8.563
Stress in YZ plane MPa	10.759	13.49	16.139	18.829	21.518
Stress in XZ plane MPa	11.039	13.798	16.558	19.318	22.077
Principal Stress MPa	20.121	25.151	30.182	35.212	40.242
Vomeises Stress MPa	21.677	27.096	32.515	37.935	43.354
Strain in XY plane	0.54375	0.67969	0.81563	0.95157	1.088
Strain in YZ plane	1.366	1.708	2.05	2.391	2.733
Strain in XZ	1.402	1.752	2.103	2.453	2.804
Principal Strain	1.043	1.304	1.565	1.826	2.086
vomeises Strain	1.462	1.828	2.193	2.559	2.924

.

Data from the table 5.1 indicates that the maximum displacement in X, Y and Z directions increase with the increase in pressure and exhibits linearity which is shown in figure 3. Also the displacement in z direction is very less compared to those in x and y directions

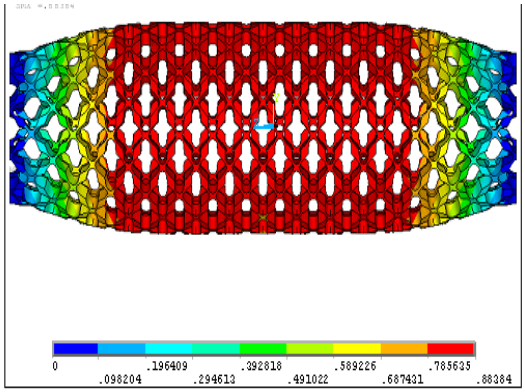


Fig. 3: Sum Displacements for RM model is 0.88384 for pressure 0.8 MPa

The Von Mises strain intensity for RM model is shown in figure 4.explain the variations of Von Mises strain intensity stress in the stentThe is ranging from 0.11111 to 1.09 when the pressure is .8Mpa.

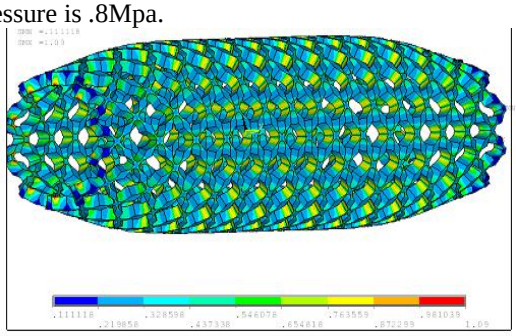


Fig. 4 The Von Mises strain intensity for RM model

The Von Mises stress for RM model is deplicated in the figure 5 shows the Von Mises stress is ranging from 2.557MPa to 32.515 Mpa when the pressure is1.2 Mpa.

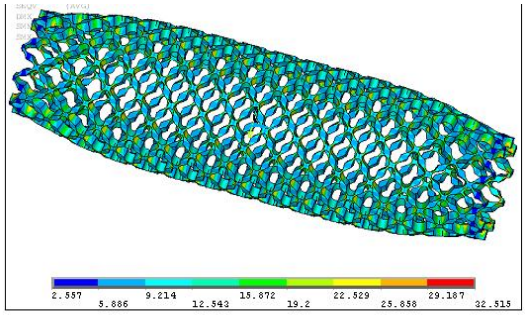


Fig. 5: The Von Mises stresses for RM model at pressure 1.2 MPa

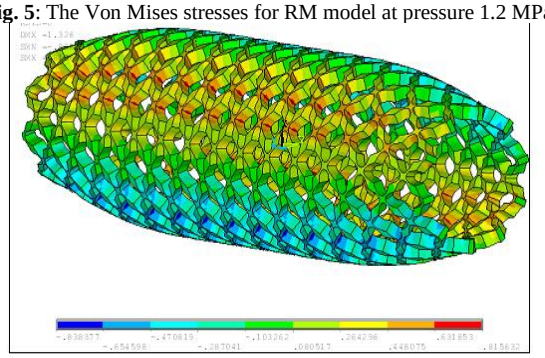


Fig. 6: The shear strain variation in XY plane for RM model at pressure 1.2 Mpa.

The Shear strain in XY plane for RM model is shown as above .The shear strain are ranging from -0.838377 to 0.815632.

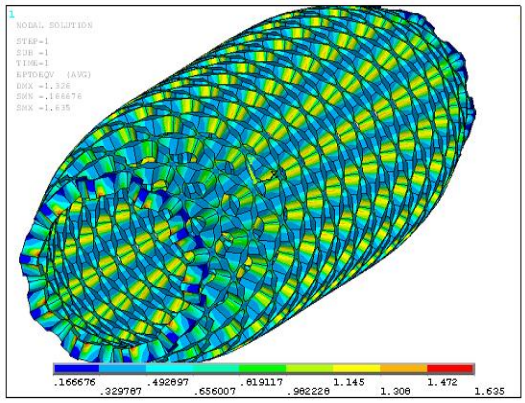


Fig.7: The Von Mises strain for RM model at pressure 1.2 Mpa.

The Von Mises strain for RM model is shown as above .The Von Mises strain intensity is ranging from 0.166676 to 1.635.

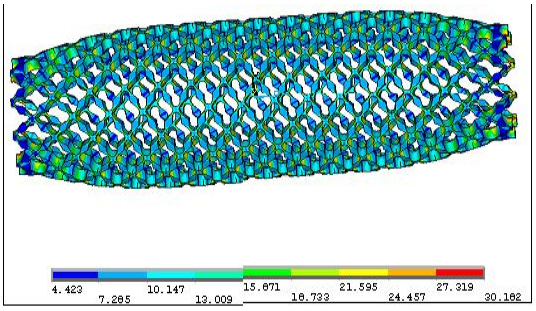


Fig. 8: The First principal stress variation in RM model at pressure 1.2 MPa The principal stress are ranging from 4.423MPa to 30.182Mpa..

Data from the table 2 indicates that the maximum total displacement in x, y and z directions increase with the increase in pressure and exhibits linearity which is shown in figure 9.

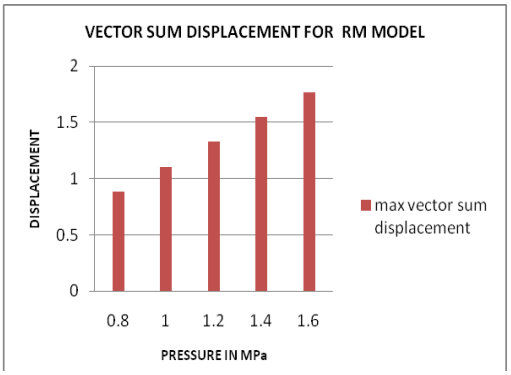


Fig. 9: vector sum displacement in RM model stent.

The shear stress variation in xy, yz and xz plane vs. various pressures in the artery indicate that, there is linear variation of the shear stress[5] with respect to pressure as demonstrated in the figure 10.

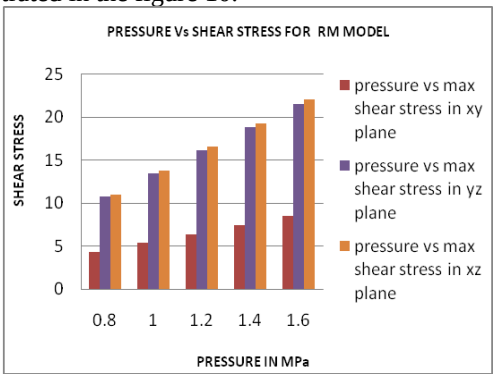


Figure 10.Shear stress in RM model stent.

The von Mises stress vary linearly with respect to pressures as exhibited in the Figure 11 below

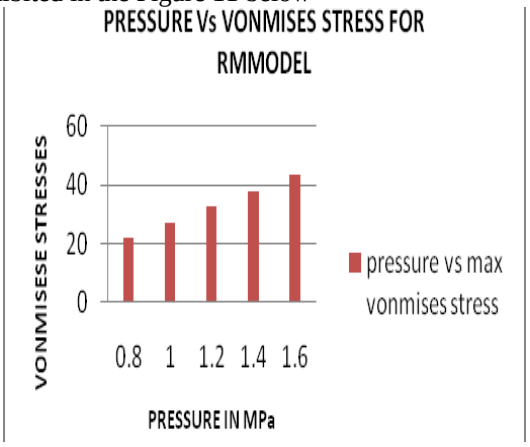


Fig. 11: Principal strain in RM model stent.

The others parameters like strains in x,y and z directions, principal strains in x , y and z directions vary linearly with respect to various pressures[4]. The graph of strains vs various pressures is demonstrated in figure 12.

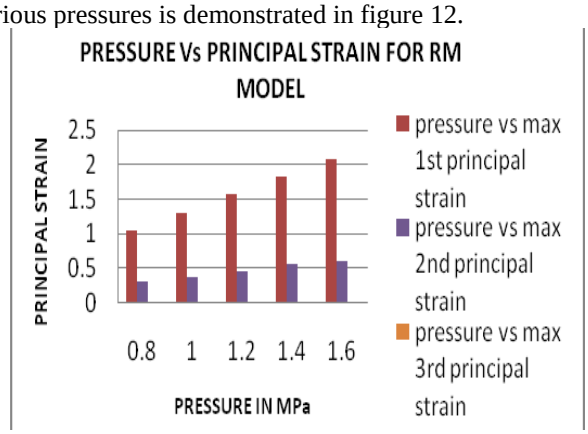


Fig. 12.Principal strain in RM model stent.

4. Conclusions

This study describes the development of a Finite Element analysis tool to simulate and predict the mechanical behaviour of Regular Model (RM) Marginal stent for the use in cardiovascular prostheses. [4] The FE analysis of RM stent was focused on stress, strain and diameter (displacement) analysis to determine the mechanical behaviour in the stentstructures.The stress analysis of the numerical results indicated stress values smaller than the ultimate tensile stress in the SS 316 structures.shows the liner and the prostheses structures. The applied Finite Element Analyses allowed to determinate biomechanical characteristics of the proposed coronary stent. The obtained results may be the base for optimization of geometrcal features of the implant as well as its mechanical properties [4-6].

References

[1] Sanjay Pujari, V.R.Udupi, "Biomechanics in Marginal coronary Stenatic Arteries," International Journal of Recent Trend in Computer Engineering ISSN: 1797-9617, Vo2, pp.90-92, Year. May2009.

[2] Sanjay Pujari, V.R.Udupi, "Stenotic Artery Revascularization through Finite-element Analysis" ,national Conference at G.I.T college of Engineering, Belgaum, pp.12-17, Year. Nov 2008.

[3] Sanjay Pujari, V.R.Udupi, "rehabilitation of Artery Revascularizatin through stent Insertions," International Journal of Recent Trend in Mechanical Engineering ISSN: 1797-9617, Vo2, pp.142-147, Year. May2009.

[4] Lally, C., Dolan, F., Prendergast, P.J., Biomechanics, Cardiovascular stent design and vessel stresses: a finite element analysis", Vol. 38, pp. 1574-1581.

[5] Rogers, C., Tseng, D.Y., Squire, J.C., Edelman, E.R., "Balloon Artery Interactions during Stent Placement: A Finite Element Analysis Approach to Pressure, Compliance, and Stent Design as Contributors to Vascular Injury", Circulation, Vol. 84, pp. 378-383.

[6] Manak J.J., (1980) "The two-dimensional in vitro passive stress-strain elasticity relationships for the steer thoracic aorta blood vessel tissue", Journal of Biomechanics, Vol. 13. pp. 637-646.

Author Biographies

Sanjay A. Pujari was born on 1st June 1973, in Belgaum, Karnataka, India. He studied his diploma in Electronics and Communication Engineering, bachelor degree in Instrumentation Engineering from Karnatak University, Dharwad, Karnataka, India in 1994 and master degree in biomedical instrumentation from Mysore University in 1997. He is currently pursuing his doctoral in biomedical engineering field under Visvesvaraya Technological University, Belgaum, Karnataka, India. His research interests include biomedical sensors, biocompatible materials, biomechanics and medical electronics. He has presented several papers in national and international conferences and published 03 papers in international journals.

Dr. V. R. Udupi has completed bachelor degree in Electronics and Communication Engineering from Mysore University, Karnataka in 1983 and did his masters in electronics from Shivaji University, India in 1989. He pursued his Ph.D. from Shivaji University, Kolhapur, Maharashtra,India in 2003. His major area of interest medical electronics, Computer Vision and Intelligent systems. He has presented several papers in national and international conferences and published 12 papers in international journals.He is the life member of several technical societies.

RM Analysis On µCOS for Embedded Systems

R.R.Maggavi, D.A.Torse

Department of Electronics and Communication Engineering
Gogte Institute of Technology,
Belgaum, India
raghu_maggavi@indiatimes.com , datorse@git.edu

***Abstract*:** Real-time operating system (RTOS) is a very useful tool for developing the application on embedded boards with least software development effort. Though number of RTOS products are available in the market, µC/OS-II is a freeware with minimum facility and more popular among the hobbyist, researchers and small embedded system developers. The µC/OS-II supports preemptive scheduling which is not efficient with respect to processor utilization. As a result, this may lead to missing deadline of the task assigned and hence may cause system failure. In this paper, a Rate Monotonic scheduling (RM), which is a better scheduling method when compared to preemptive technique, is implemented on µC/OS-II and its operation in terms of task execution and processor utilization are discussed. This paper presents the RM Analysis (RMA) on µC/OS-II with hardware i.e 8051 microcontroller based system with four tasks one task is on LCD display and three tasks on LED ON/OFF are taken to demonstrate the RMA. The software tools Keil IDE, Phillips Flash Magic are used for implementation of tasks on 8051 embedded development board. The scaled-version of µC/OS-II with multiple tasks uses 4 kB of flash and 512 bytes of RAM in 8051 board. The results obtained indicate optimum utilization of processor with RMA scheduler for realizing low cost software for developing the application on embedded boards with least software development effort.

***Keywords*:** Embedded systems, Keil IDE, Flash Magic, µC/OS-II RTOS and Microcontroller.

1. Introduction

Real-time operating system (RTOS) is useful for developing the application on embedded boards with least software development effort. The human effort needed for implementing µC/OS-II is less compared to other RTOS's. Among the available RTOS's µC/OS-II is suitable for various controllers and processors, since it is low cost and easily available. The µC/OS-II supports preemptive scheduling and not efficient with respect to processor utilization.

Processor utilization is a measure of CPU loading. Improving the CPU loading is difficult since embedded processor need to complete the tasks within the deadlines. Rate Monotonic Analysis (RMA) helps to achieve optimum CPU loading for running real-time tasks with time deadlines. In RMA, most frequently used task gets highest priority. Optimum utilization of resources by meeting the deadlines is discussed in this paper.

This paper presents implementation of RMA on µc/os-ii. For this, µC/OS-II is simulated into Keil IDE [2] and RMA code is appended to it. ROM image file is ported into 8051 based microcontroller flash and tested for scheduling with RMA. And the performance parameters are discussed.

The scaled-version of µC/OS-II with multiple tasks uses 4 kB of flash and 512 bytes of RAM in 8051 board. The results indicate optimum utilization of processor with RMA scheduler for realizing low cost software and hardware for developing embedded system.

A micro-kernel operating system (µC/OS-II) written by John J Lebrosse is a freeware for peaceful research purpose. The source code is written in C language which is compliant to ANSI C format. It has about 10,000 lines of source code with well-documented comments in the source code. It is a well-tested source code and has been ported to thousands of devices. It has a built-in preemptive scheduler. But, it does not support a rate-monotonic scheduler. In this paper, it is proposed to implement a rate-monotonic scheduler in µC/OS-II and will be discussed in this paper.

The microkernel approach is based on the idea of only placing essential core real-time operating system functions in the kernel, and other functionality is designed in modules that communicate through the kernel via minimal well-defined interfaces. The microkernel approach results in easy reconfigurable systems without the need to rebuild the kernel.

The µC/OS-II kernel is implemented completely in software and is a well-used real-time operating system for embedded systems. It is written as a monolithic kernel in the language of ANSI C with a minor part in assembler for context switching.

The paper is organized as follows. Section 2 describes Configurable parameters for µC/OS-II. Section 3 describes the implementation of scheduler on low-end hardware board. Section 4 describes the analysis of results. Section 5 describes the conclusion of this work.

2. Configurable Parameters in µC/OS-II

It is necessary to configure the parameters so that only the required features are built into the target system. The code size is reduced from 10,000 lines to 5,500 lines by suitable configuration of parameters. This is needed to reduce the memory requirements and to achieve optimum utilization of resources for the required functions in 8051 hardware board.

Figure 1 shows the architecture of µC/OS-II.The files of µC/OS-II is shown in figure 1. The application software is the code written by the user in the main function to perform several tasks. Processor independent codes are the files which can be used in any hardware without the need to configure it.

Processor-specific and Application-specific code can be modified to suit the application requirements.

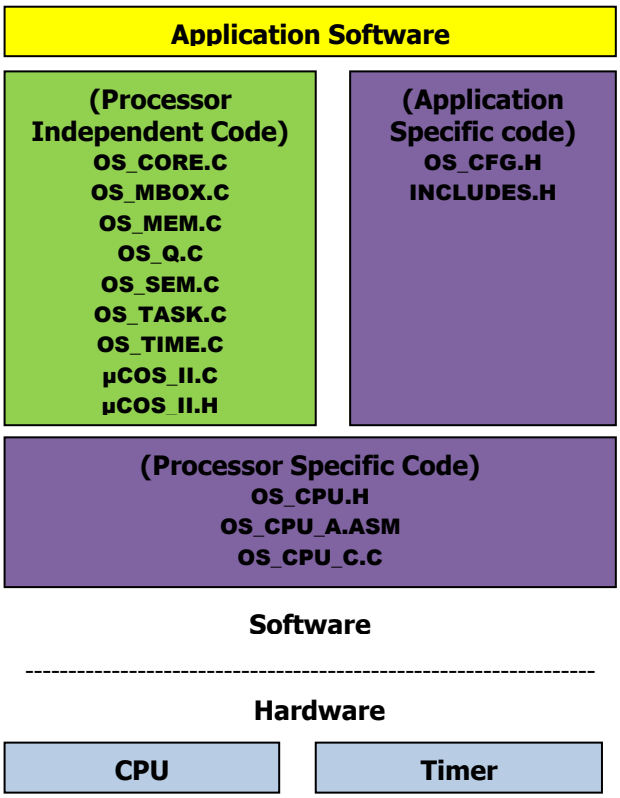


Fig. 1 µC/OS-IIarchitecture

The parameters that are configured in **µC/OS-II** are in **file os_cfg_r and os_cpu_a sm.** The parameters that are configured **in os_cfg_r** with its explanation are given below

```
#define OS_TICK_STEP_EN          1
/* Set to 1/0 to Enable/Disable tick stepping feature respect ively for uC/OS-View */
#define OS_MAX_TASKS             3   /* Max. Number of tasks in your application, MUST be >= 2      */
#define OS_LOWEST_PRIO           5   /* Defines the lowest priority that can be assigned ...        */
#define OS_TMR_CFG_TICKS_PER_SEC 10  /*  Rate at which timer management task runs (Hz)          */
```

The other features like semaphores, timers, queues and memory management are not needed. Hence, they are set to '0' to disable these features.

The size of the OSStack and timer delay is configured in file os_cpu_asm. The routines are written in assembly language. The pseudo source code of main () function is given below.

```
void main(void)
{
  OSInit();/*Initialize OSStack and memory blocks */
TargetInit(); /*  Initialize the target hardware */
OSTaskCreate(Task0,(void    *)0,&Task0Stack[MaxStkSize-1],0);/*Create Task0 with priority 0*/
OSTaskCreate(Task1,(void    *)0,&Task1Stack[MaxStkSize-1],1);/*Create Task1 with priority1*/
OSTaskCreate(Task2,(void    *)0,&Task1Stack[MaxStkSize-1],2);/*Create Task2 with priority2*/
OSStart();// start muti-tasking
}/*end of main*/
```

The above code consists of OS initialization, target initialization, Creation of tasks and start multi-tasking functions. The priorities are assigned to tasks so that a task with a lowest priority is executed most frequently. This is the scheduling approach which is known as rate-monotonic scheduler approach.

3. Implementation of Rate-Monotonic Scheduler on a Hardware Board

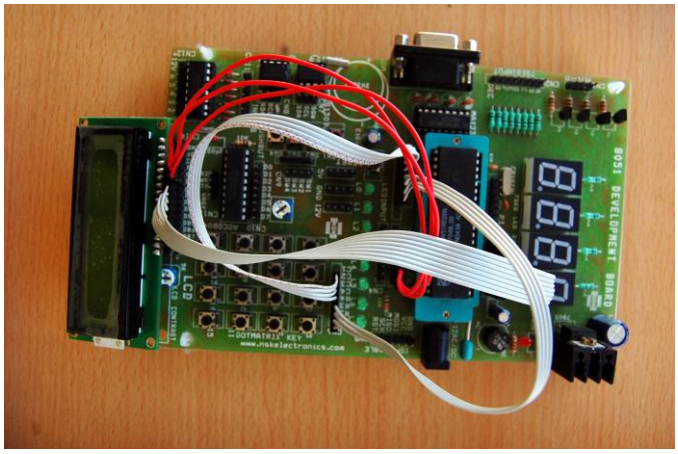


Fig. 2 Hardware used for testing rate-monotonic scheduler

A rate-monotonic scheduler is designed and implemented with a low-end system. The microcontroller 8051 used in this work will be discussed in this paper.

The hardware implementation has been developed in Microcontroller (8051) Embedded Development Kit which consists of a 4 x 4 keypad, 2 x 16 LCD, 8 LEDs and an RS-232 port for serial communication. An external crystal of clock frequency 11.0592 MHz is used in the board. The board details are shown in the figure 2.

The software implementation has been developed in an IDE (Integrated Development Environment) with keil compiler and Phillips Flash Magic tool [3] for 8051 hardware board.

The source code for the tasks described in table1 is written in C language and compiled with keil compiler. The hex file generated by the compiler is used for programming the micro-controller (P89V51RD2) with Phillips Flash Magic tool. The user-defined application is written in file **main.c.** It consists of user-defined tasks to be performed by assigning the priorities in an infinite loop. The memory requirements projected after compilation of code in 8051 are discussed in table 2 and table 3 respectively.

The CPU utilization or time-loading factor U is a measure of the percentage of non idle processing. A system is said to be time-overloaded if U > 100%. Systems that are too highly utilized are undesirable because changes or additions cannot be made to the system without risk of time-overloading. The target for U as 69% in rate-monotonic systems gives a very useful result in the theory of real-time systems.

U is calculated by summing the contribution of utilization factors for each (periodic or aperiodic) task. Suppose a system has $n \geq 1$ periodic tasks, each with an execution period of pi, and hence execution frequency, $f_i = 1/p_i$. If task i is known to have a maximum (worst case) execution time of e_i, then the utilization factor, u_i, for each task e_i is $u_i = e_i/p_i$. Then the overall system utilization is

$$U = \sum_{i=1}^{n} u_i = \sum_{i=1}^{n} e_i/p_i$$

Table 1: Execution Time for Various Tasks in 8051

Using above equation and substituting n=4.The theoretical results observed are close to that of experimental results. The

Task number	Task name	e_i (execution time of a task)	p_i (execution period of a task)
1	***LCD_display0***	**1 s**	**8 s**
2	***LED_display1***	**2.13 s**	**8s**
3	***LED_display2***	**1.42 s**	**8 s**
4	***LED_display3***	**1.065 s**	**8 s**

results are computed for four different tasks and are shown in table1. The value of U obtained is 70.18% for the four tasks shown in table1.
Theoretical value of CPU utilization = **75.68%**
Obtained value of CPU utilization = **70.18%**

LCD_display: The data is displayed in the LCD by initializing the values in the command register. The function is executed within 8s.
LED_display1: LED is turned ON/OFF to indicate completion of a task. To perform this task, data has to be written to o/p port of microcontroller. The function is executed within 8 s.
LED_display2 and LED_display3: It is similar to the LED_display1 task.

The memory requirement of the above tasks in 8051 controller is shown in the table 2 below. For multiple tasks are discussed in table 3.

Table 2: MEMORY REQUIREMENT OF 8051CONTROLLER

Serial no:-	Types of Memory	Memory size (Bytes)
1	***Code memory***	**4897 bytes**
2	***Data memory***	**410 bytes**

Comparison of results

The results discussed above are compared with respect to CPU utilization and are discussed in the table below.

Table 3: Comparison of Results With Respect To Optimum CPU Utilization

Number of Tasks	Theoretical value	Practical value in 8051 controller
4	**75.68%**	**70.18%**

The results in the above table indicate that optimum CPU utilization is not only dependent on the scheduling technique employed in an embedded system. But, it is also dependent on the processing power of controller and the complexity of the application code. A qualitative analysis of the results obtained with respect to 8051 controller

4. Analysis of Results

The execution time for various tasks is shown in table 1. The system can be operated with 69% CPU utilization but in this work we could operate up to 70.18% in 8051 hardware board. This is due to the fact that the interfaces in the board are limited. Also, there are other constraints like limited ON-chip memory, limited number of ports and lower processing power. If the tasks like displaying the data on LCD, indicating the status of task by turning ON/OFF LEDs, scanning and detecting the key pressed in keypad has to be done continuously for an infinite duration of time. This causes a significant delay in the RTOS.

5. Conclusions

µC/OS-II was ported into Keil IDE and an RMA analysis result was discussed in this paper. ROM image file was ported into 8051 based microcontroller flash and tested for scheduling with RMA. The entire µC/OS-II with multiple tasks uses 4 kB of flash and 512 bytes of RAM. The results have indicated optimum utilization of processor with RMA scheduler for realizing low cost software and hardware for developing embedded system. The overall CPU utilization was 70.18% with RMA approach in 8051. The result obtained indicates optimum utilization of processor with RMA scheduler for realizing low cost software for developing the application on embedded boards with least software development effort.

References

[1] Micrium, Inc, www.micrium.com, (2005)
[2] Keil, Inc, www.keil.com, (2005)
[3]Phillips, Inc, www.phillips.com
[4] F. Engel, G. Heiser, I. KuZ, S. M. Petters and S. Ruocco, *"Operating Systems on SoCs: A Good Idea?,"* in ERTSI in conjunction with 25th IEEE RTSS04, Lisbon, Portugal, December 2004.
[5] Furunas, J. *"Benchmarking of a Real-Time System that utilises a booster."* In International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA200), June, 2000.
[6] H. Hartig, M. Hohmuth, J. Liedkte, S. Schonberg and J. Wolter, "The performance of µ-Kernel-Based Systems", in proceedings of the 16th ACM symposium on Operating Systems Principles, p 66-77, *Saint Malo, France*, 1997.
[7] L. Johansson and T. Samuelsson, *"Integration of an Ultra-fast Real-Time Accelerator in the Real-Time Operating System µC/OS-II",* Master Thesis report, Malardalen University, Vasteras, Sweden, October 2004.
[8] P. Kohout, B. Ganesh and B. Jacob, *"Hardware Support for Real-Time Operating Systems"*, in Conference on Hardware/Software codesign and system synthesis of contents, p.45-51, Newport Beach, USA, 2003.
[9] P. Kuacharoen, M. A. Shalan and V. J. Mooney III, *"A Configurable Hardware Scheduler for Real-Time Systems,"* in Precedings of the International Conference on Engineering

of Reconfigurable Systems and Algorithms (ERSA´03), p 96-101, Las Vegas, USA, June 2003.
[10] J. Liedkte, *"Toward Real Microkernels,"* in Communicatons of the ACM, vol. 39, No 9, September 1996.
[11] L. Lindh, and F. Stanischewski, "FASTCHART – A Fast Time Deterministic CPU and Hardware Based Real-Time-Kernel." *In IEEE, Euromicro workshop on Real-Time Systems, June 1991.*
[9] L. Lindh, T. Klevin, and J. Furunas, "Scalable Architecture for Real- Time Applications – SARA*". Swedish National Real-Time Conference SNART99 Linkoping, Sweden, August, 1999.*
[10] T. Nakano, Y. Komatsudaira, A. Shiomi and M. Imai. VLSI Implementation of a Real-time Operating System. Proc. of ASPDAC '97, pp. 679-680, January, 1997.

Author Biographies

R. R. Maggavi was born on 15th July 1979, in Belgaum, Karnataka, India. He studied his bachelor degree in Electronics and Communication Engineering from Visvesvaraya Technological University, Belgaum, Karnataka, India in 2004 and currently pursuing his masters in digital electronics under Visvesvaraya Technological University, Belgaum, Karnataka, India. His research interests include embedded systems and real time operating systems design.

D. A. Torse was born on 12th April 1979, in Belgaum, Karnataka, India. He studied his bachelor degree in Electronics Engineering from Dr. Babasaheb Ambedkar Marathwad University, Aurangabad, Maharashtra, India in 2001 and the masters in digital electronics from Amaravati University, India in 2005.His major area of research is mcirowaves and antenna design and embedded systems design.

Performance Analysis of Conventional Crypto-coding

Rajashri Khanai[1], Dr. G. H. Kulkarni[2],

[1]Department of Electronics & Communication Engineering,
Jain College of Engineering, Visvesvaraya Technological University, Belgaum
rajashri.khanai@gmail.com
[2]Department of Electrical and Electronics Engineering,
Jain College of Engineering, Visvesvaraya Technological University, Belgaum
ghkulkarni1@rediffmail.com

***Abstract*:** New communications paradigms such as wireless communications for handheld devices have increased requirements for operating speed of relevant security mechanisms, such as encryption for confidentiality, hashing for message integrity and combined hashing and public key encryption for digital signature. Also, a major concern of designing digital data transmission and storage systems is the control of errors, so that reliable reproduction of data can be obtained. A great deal of effort has been expended on the problem of devising efficient encoding and decoding methods for error control in a noisy environment. In this paper, we introduce a novel method which combines cryptography and error correction together i.e. we utilize the methods for data security as a tool for the detection and correction of transmission errors. To support our research, we provide simulation results obtained using this method, which we call as conventional method of crypto-coding.

***Keywords*:** cryptography, coding theory, channel coding, crypto code, convolution code, security.

1. Introduction

The need to minimize the effect of noise in our increasingly digital world on the entire digital communication system (i.e. transmission from source to destination) is becomingly increasingly pertinent. This effect can be mitigated by the use of error control coding [1]; the addition of redundancy that is utilized to correct or detect errors but to the extent delineated by the theoretical limit known as Shannon limit. Cryptography on the other hand is primarily used for secure communications. In modern society, exchange and storage of information in an efficient, reliable and secure manner is of fundamental importance. There is an increasing amount of transactions using communications over the Internet and over mobile communication channels. Therefore secure communication will be essential for the exploitation of mobile communication and the Internet to its full potential, such as for the transfer of sensitive data in, for example, payment systems, e-commerce, m-commerce, health systems, oil and gas exploration monitoring, control system communications and in numerous other applications. For many of these applications, systems for authentication will also be necessary [2]. Cryptology comprises the interrelated areas of cryptography and cryptanalysis. Cryptography is fundamental in order to protect information against wiretapping, unauthorized changes and other misuse of information. A cryptanalyst studies vulnerabilities of ciphers and other cryptographic techniques.

The paper is organized as follows. Section 2 describes a brief overview of S-DES. Section 3 presents the convolutional encoder and decoder using Viterbi algorithm. Section 4 describes the concept of conventional crypto-coding approach. Experimental results are presented in Section 5 and finally the conclusion and future work are presented in Section 6.

2. Cryptography

The aim of cryptography is to provide secure transmission of messages [3], in the sense that two or more persons can communicate in a way that guarantees that the desired subset of the following four primitives is met:
(i). Confidentiality. This primitive is usually perceived to be the main focus of cryptography, providing a way such that the information can only be viewed by those people authorized to see it.
(ii). Data integrity. This service will provide a means to check if the transmitted information was altered in any way, including but not limited to things like insertion, deletion and substitution of messages.
(iii). Authentication. This service will establish some identity pertaining to the message. Thus, this primitive can (among others) be used to guarantee the identity of the sender, guarantee the identity of the receiver, or guarantee the time the message was sent.
(iv). Non-repudiation. This serves to prevent someone from denying previous commitments.

It is needed in cases where disputes might have to be resolved, for instance in E-commerce. While cryptography is often thought of as a tool to provide confidentiality, the other three primitives are actually much more important in daily life [4]. In this paper we have implemented S-DES to secure information over the channel during transmission.

2.1 S-DES

S-DES encryption (decryption) algorithm takes 8-bit block of plaintext (ciphertext) and a 10-bit key, and produces 8-bit ciphertext (plaintext) block [5]. Encryption algorithm involves 5 functions: an initial permutation (IP); a complex function f_k, which involves both permutation and substitution and depends on a key input; a simple permutation function that switches (SW) the 2 halves of the data; the function f_k again; and finally, a permutation

function that is the inverse of the initial permutation (IP^{-1}). Decryption process is similar.

The function f_k takes 8-bit key which is obtained from the 10-bit initial one two times. The key is first subjected to a permutation $P10$. Then a shift operation is performed. The output of the shift operation then passes through a permutation function that produces an 8-bit output ($P8$) for the first subkey (K_1). The output of the shift operation also feeds into another shift and another instance of P8 to produce the 2nd subkey f_k.

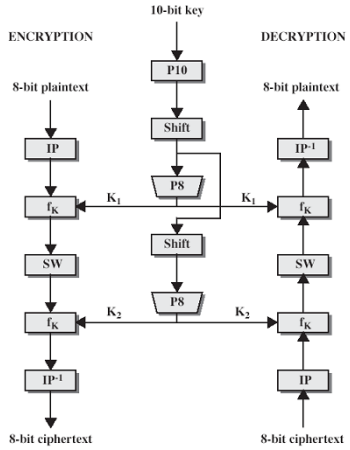


Figure 1. Simplified DES Scheme

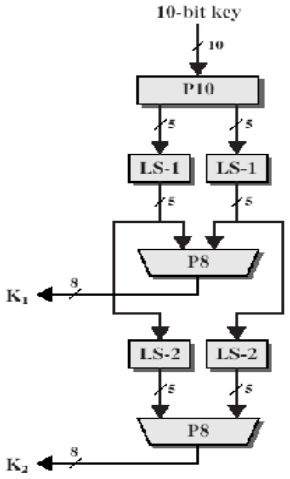


Figure 2. Key generation

We can express encryption algorithm by using following functions:

$$IP^{-1} \circ fk_2 \circ SW \circ fk_1 \circ IP$$

$$Ciphertext = IP^{-1}(fk_2(SW(fk_1(IP(Plaintext)))))$$

Where

$$K_1 = P8(Shift(P10(Key)))$$

$$K_2 = P8(Shift(Shift(P10(Key))))$$

Decryption is the reverse of encryption:

$$Plaintext = IP^{-1}(fk_1(SW(fk_2(IP(Plaintext)))))$$

3. Coding Theory

The aim of coding theory is to provide secure transmission of messages, in the sense that (up to a certain number of) errors that occurred during the transmission can be corrected. However, for this capability a price must be paid, in the form of redundancy of the transmitted data [6].

Coding theory has developed methods of protecting information against noise. Without coding theory and error correcting codes, there would be no deep space pictures, no satellite TV, no CD, no DVD and many more...

Since transmission line and storage devices are not 100% reliable device, it has become necessary to develop ways of detecting when an error has occurred and, ideally, correcting it. The theory of error-correcting codes originated with Claude Shannon's famous 1948 paper "A Mathematical Theory of Communication" and has grown to connect too many areas of mathematics, including algebra and combinatorics. The cleverness of the error-correcting schemes that have been developed since 1948 is responsible for the great reliability that we now enjoy in our modern communications networks, computer systems, and even compact disk players.

3.1 Convolutional Encoder

The encoder for a convolutional code also accepts k-bit blocks of the information sequence u and produces an encoded sequence (code word) v of n-symbol blocks. Each encoded block depends not only on the corresponding k-bit message block at the same time unit, but also on m previous message blocks.

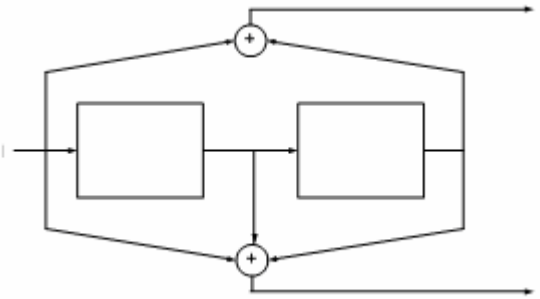


Figure 3. Convolutional encoder $r = 1/2$, $m = 2$

Hence, the encoder has a memory order of m. The set of encoded sequences produced by a k-input, n-output encoder of memory order m is called an (n, k, m) convolutional code. The ratio $R = \frac{k}{n}$ is called the code rate. Since the encoder contains memory, it must be implemented with a sequential logic circuit.

3.2 Viterbi decoder:

Viterbi Algorithm for Decoding of Convolutional Codes

Convolutional codes are frequently used to correct errors in noisy channels. They have rather good correcting capability and perform well even on very bad channels (with error probabilities of about 10−3). Convolutional codes are extensively used in satellite communications. Although convolutional encoding is a simple procedure, decoding of a convolutional code is much more complex task. Viterbi

algorithm is a well-known maximum-likelihood algorithm for decoding of convolutional codes and is an optimal (in a maximum-likelihood sense) algorithm for decoding of a convolutional code using trellis. Its main drawback is that the decoding complexity grows exponentially with the code length. So, it can be utilized only for relatively short codes [10].

A Viterbi algorithm consists of the following three major parts:

1. *Branch metric calculation* – calculation of a distance between the input pair of bits and the four possible "ideal" pairs ("00", "01", "10", "11").
2. *Path metric calculation* – for every encoder state, calculate a metric for the survivor path ending in this state (a survivor path is a path with the minimum metric).
3. *Traceback* – this step is necessary for hardware implementations that do not store full information about the survivor paths, but store only one-bit decision every time when one survivor path is selected from the two.

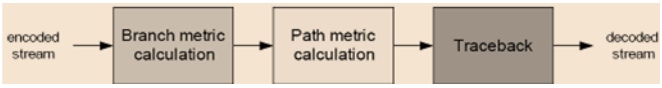


Figure 4. Viterbi decoder data flow

3.2.1 Branch Metric Calculation

Methods of branch metric calculation are different for hard decision and soft decision decoders. For a hard decision decoder, a branch metric is a Hamming distance between the received pair of bits and the "ideal" pair. Therefore, a branch metric can take values of 0, 1 and 2. Thus for every input pair we have 4 branch metrics (one for each pair of "ideal" values). For a soft decision decoder, a branch metric is measured using the Euclidean distance. Let x be the first received bit in the pair, y – the second, x_0 and y_0 – the "ideal" values. Then branch metric is,

$$M_b = (x - x_0)^2 + (y - y_0)^2$$

Furthermore, when we calculate 4-branch metric for a soft decision decoder, we don't actually need to know absolute metric values – only the difference between them makes sense. So, nothing will change if we subtract one value from the all four branch metrics:

$$M_b = (x^2 - 2xx_0 + x_0{}^2) + (y^2 - 2yy_0 + y_0{}^2)$$

$$M_b^* = (M_b - x^2 - y^2) = (x_0{}^2 - 2xx_0) + (y_0{}^2 - 2yy_0)$$

Note that the second formula, M_b^* can be calculated without hardware multiplication: $x_0{}^2$ and $y_0{}^2$ can be pre-calculated, and multiplication of x by x_0 and y by y_0 can be done very easily in hardware given that x_0 and y_0 are constants. It should be also noted that M_b^*, is a signed variable and should be calculated in 2's complement format.

3.2.2 *Path Metric Calculation*

Path metrics are calculated using a procedure called *ACS (Add-Compare-Select)*. This procedure is repeated for every encoder state.

1. *Add* – for a given state, we know two states on the previous step which can move to this state, and the output bit pairs that correspond to these transitions. To calculate new path metrics, we add the previous path metrics with the corresponding branch metrics.
2. *Compare, select* – we now have two paths, ending in a given state. One of them (with greater metric) is dropped.

As there are $2^{(K+1)}$ encoder states, we have $2^{(K+1)}$ survivor paths at any given time.

It is important that the difference between two survivor path metrics cannot exceed $g(K-1)$, where '∂' is a difference between maximum and minimum possible branch metrics.

The problem with path metrics is that they tend to grow constantly and will eventually overflow. But, since the absolute values of path metric don't actually matter, and the difference between them is limited, a data type with a certain number of bits will be sufficient.

There are two ways of dealing with this problem:

1. Since the absolute values of path metric don't actually matter, we can at any time subtract an identical value from the metric of every path. It is usually done when *all* path metrics exceed a chosen threshold (in this case the threshold value is subtracted from every path metric).

This method is simple, but not very efficient when implemented in hardware.

2. The second approach allows overflow, but uses a sufficient number of bits to be able to detect whether the overflow took place or not. The *compare* procedure must be modified in this case.

The whole range of the data type's capacity is divided into 4 equal parts. If one path metric is in the 3-rd quarter should be selected. In other cases an ordinary compare procedure is

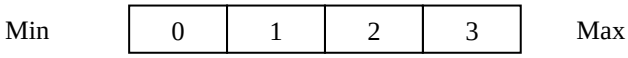


Figure 5. A modulo-normalization approach

applied. This works, because a difference between path metrics cannot exceed a threshold value, and the range of path variable is selected such that it is at least two times greater than the threshold.

3.2.3 Traceback

It has been proven that all survivor paths merge after decoding a sufficiently large block of data (D on Figure 5), i.e. they differ only in their endings and have the common beginning. If we decode a continuous stream of data, we want our decoder to have finite latency. It is obvious that when some part of path at the beginning of the graph belongs to every survivor path, the decoded bits corresponding to this part can be sent to the output. Given the above statement, we can perform the decoding as follows:

1. Find the survivor paths for $N + D$ input pairs of bits.
2. *Trace back* from the end of any survivor paths to the beginning.
3. Send *N* bits to the output.
4. Find the survivor paths for another *N* pairs of input bits.
5. Go to step 2.

In these procedures D is an important parameter called *decoding depth*. A decoding depth should be considerably large for quality decoding, no less than $5K$. Increasing D decreases the probability of a decoding error, but also increases latency.

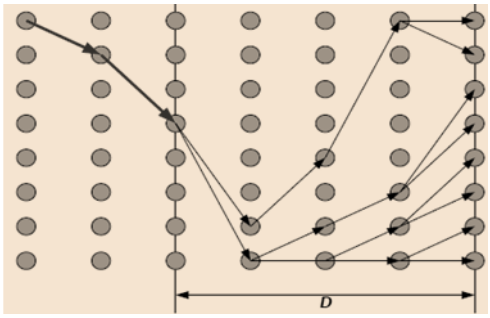


Figure 6. Survivor paths graph example.

Blue circles denote encoder states. It can be seen that all survivor paths have a common beginning (dark) and differ only intheir endings.

As for N, it specifies how many bits we are sending to the output after each traceback. For example, if N=1, the latency is minimal, but the decoder needs to trace the whole tree every step. It is computationally ineffective. In hardware implementations N usually equals D.

4. Conventional Crypto-coding

In view of designing digital data transmission and storage systems it is of interest to re-open a 25 years old question of joint encryption and error-correction coding, named here as Crypto-coding. Crypto-coding is a procedure in which encryption/decryption and error-correction coding/ decoding are performed in a single step. We give simulation results for conventional crypto-coding in which reliable, secure data transmission is employed. Contemporary cryptography and coding theory have more than 60 years of a successful history. Since the publication of the seminal works of Shannon in 1949 [9] and Hamming [7] these two scientific fields evolved with mutual borrowing of the cross fertilizing ideas, but kept a clear distinction of the functionality of the developed methods and algorithms. That may partly have been because cryptographic and error-correcting algorithms have opposite intentions with respect to the type of information processing they perform. While cryptographic algorithms in order to provide information security, in the process of decryption need an errorless input, error-correcting algorithms are meant to handle certain amount of errors in the input data, but they are not designed to provide any security of the data they process. However, there are many situations when both information security and error-correction are needed or required. In that case a combination of cryptographic algorithms and error correcting codes, which basically means encryption, decryption, encoding and decoding, is realized by a sequential execution of two separate algorithms, one for encryption/decryption and the other one for coding/decoding [8].

5. Experimental Results

In this paper, for secure and reliable communication a cryptographic algorithm S-DES is implemented followed by channel coding technique as convolutional coding is implemented. Similarly, at the decoder decryption algorithm and channel errors corrected using Viterbi decoder.

We give simulation results (table 1) for conventional crypto-coding in which reliable, secure data transmission is employed.

Table 1. Experimental results

$P_b = 0.01$		
No. Of Trails	BER	No. Of channel errors
1	0.1793	17
2	0.3259	27
3	0	19
4	0	22
5	0	25
6	0.0767	26
7	0	26
8	0	23
9	0	18
10	0	25
11	0	18

$P_b = 0.05$		
No. Of Trails	BER	No. Of channel errors
1	0.2655	119
2	0.3543	117
3	0.2371	101
4	0.4862	118
5	0.4310	115
6	0.3319	140
7	0.1317	109
8	0.4560	129
9	0.2336	113
10	0.2629	88
11	0.3431	113

The simulation results are plotted in fig. 7.

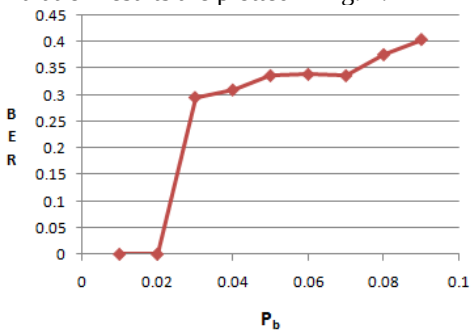


Figure 7. BER v/s P_b

6. Conclusion and Future Work

This paper analyzes conventional crypto-coding for correction of channel errors results of messages which are protected by security mechanisms. The cryptocoding has not attracted more attention is the fact that error correction introduces data redundancy (data expansion in turn more bandwidth) which is not desirable in the design of communication system as a whole. However, data has to be considered for transmissions in a network, then redundancy have to be introduced. In that situation, the computational costs for encryption/encoding and decryption/decoding are bigger.

For secure and reliable data transmission, encryption/encoding and decryption/decoding can be implemented in a single algorithm, called as cryptocoding, reducing computational costs.

References

[1] Svein Johan Knapskog “New Cryptographic Primitives (Plenary Lecture)”, 7th Computer Information Systems and Industrial Management Applications, Center of Excellence for Quantifiable Quality of Service Norwegian University of Science and Technology, Trondheim, Norway.

[2] Nataša Živić, Christoph Ruland “Parallel Joint Channel Coding and Cryptography”, International Journal of Electrical, Computer, and Systems Engineering 4:2 2010.

[3] William Stallings “Cryptography and Network Security

Principles and Practices" Pearson Education, 2004.
[4] Ayyaz Mahmood "Method to Improve Channel Coding Using Cryptography", World Academy of Science, Engineering and Technology, 41, 2008.
[5] Neal Koblitz "A course in Number Theory and Cryptography" Springer International Edition, 2008.
[6] Nataša ŽIVIĆ, Christoph RULAND and Obaid Ur REHMAN "Error correction over wireless channels using symmetric cryptography", Institute for Data Communications Systems University of Siegen, Germany.
[7] S.K.Pal "Fast, Reliable & Secure Digital Communication Using Hadamard Matrices", Scientific Analysis Group DRDO, Metcalfe House Complex, Delhi, INDIA.
[8] J. L. Massey "Some Applications of Coding Theory in Cryptography", Signal and Information Processing Laboratory, Swiss Federal Institute of Technology (ETH), Zürich.
[9] Shannon C.E., Communication theory of secrecy systems, Bell System Tech. J. 28, pp 656–715, 1949.
[10] S. Lin, D.J. Costello: *Error Control Coding*, Pearson Prentice Hall, USA, 2004.

Author Biographies

Rajashri Khanai was born on 15th July 1977 in Belgaum. She completed her B.E. in Electronics and Communication engineering in 2000 from Karnatak University, Dharwad, Karnataka, India and M.Tech. in Digital Communication and Networking in 2007 from Visvesvaraya Technological University, Belgaum, Karnataka, India. She is currently pursuing her Ph.D. in Visvesvaraya Technological University, Belgaum. Her area of interest is error control coding and cryptography.

Prof. G. H. Kulkarni was born in 18th May 1962 in Belgaum, Karnataka, India. He completed his B.E. in Electrical Engineering from Karnataka University, Dharwad in 1985, M.Tech in Power System Engineering from University of Mysore, Mysore, India in 1992 and PhD in from Jawaharlal Nehru Technological University, Hyderabad, India in 2007. His major area of research is ANN Applications to Power System.

Analysis of Thin Smart Antisymmetric Laminated Composite Cylinder Integrated with Distributed ACLD Patches

Ashok M. Hulagabali[1], Dr. J. Shivkumar[2] and Rajesh Maji[3]
[1]Faculty, Mechanical Engg. Dept, M.M. Engg. College, Belgaum, VTU, Karnataka, India
[2]Faculty, Mechanical Engg. Dept, BIET, Davangere, VTU, Karnataka, India
[3] Scientist , DRDL, Hyderabad, Andhra Pradesh, India
ashokmh@gmail.com, jogadashivkumar@indiatimes.com , rajesh_maji@yahoo.co.in

***Abstract*:** In this paper, an attempt has been made to analyze and assess the dynamic performance of thin smart antisymmetric laminated composite (Graphite/Epoxy) cylinder integrated with patches of Active Constrained Layer Damping (ACLD) treatment. The active material used for the ACLD is Piezoelectric Fiber Reinforced Composite (PFRC) with PZT5H and epoxy. To predict the behavior of the shell, a first-order shear-deformation theory (FSDT) and the linear displacement field are employed. A finite element model is developed to predict the performance under dynamic loading case. Numerical results are obtained for an antisymmetric laminated cylindrical shell composed of an epoxy as the matrix, and graphite as fiber. The numerical results demonstrate an effective shape control of the antisymmetric laminated composite cylinder when the patches of PFRC are activated.

***Keywords*:** FSDT, Smart Material, Antisymmetric Laminated Composite, ACLD, PFRC

1. Introduction

To make efficient and cost effective light weight structure the composite material is most suitable materials because of its high specific strength to weight and high specific stiffness to weight ratios. The term" smart materials" sometimes also called intelligent materials or active materials describe a group of material system with unique properties. Generally speaking smart materials respond with change in shape upon application of externally applied driving forces. Typically this shape change is reflected in an elongation of the sample, thus allowing it to be used as vibration controller in aircraft structure. In the present study piezoelectric smart material (PZT5H) in the form of reinforcement along with epoxy as matrix i.e. PFRC is used for the analysis purpose.

Expediently, the use of piezoelectric materials as the distributed actuators and sensors being integrated with the flexible structures was found to provide the self-controlling and self-monitoring capabilities. Such a structure integrated with piezoelectric actuators/sensors is called a 'smart structure'. Point force actuators as control transducers are cumbersome and require some form of restraining support. To overcome this limitation, actuators bonded to the structure were developed using either PCLD (Passive constrained layer damping) or ACLD (Active constrained layer damping)[1-2]. Normally an adaptive or smart structure contains one or more active or smart materials. Advanced smart structures with integrated sensors, actuators and control electronics are becoming increasingly important in high-performance space structures and mechanical systems.

M C Ray and Mallik [4] have investigated the effective elastic and piezoelectric properties of unidirectional piezoelectric fiber reinforced composite (PFRC) materials. The investigations revealed that the effective piezoelectric co-efficient denoted by e_{31} of these PFRC materials which quantifies the induced normal stress in the fiber direction due to the applied electric field in the direction transverse to the fiber direction becomes significantly larger than the corresponding co-efficient of the piezoelectric material of the fibers within the useful range (0.4–0.8) of fiber volume fraction.

M C Ray and Mallik [7] investigated the exact solutions for static analysis of simply supported symmetric and anti-symmetric cross-ply laminated plates integrated with a layer of PFRC material. The results suggest the potential use of PFRC materials for the distributed actuators of smart structures with both thick and thin substrate composite plates. M C Ray and J.N.Reddy [2] demonstrated the effectiveness of ACLD using PFRC material for the constraining layer, in enhancing the damping characteristics of thin laminated composite shells. M C Ray and J.N.Reddy [1] optimized the size of ACLD patches for vibration damping of cantilever cylindrical shell. M. C. Ray, J. Oh and A. Baz[9] developed a finite element model (FEM) to describe the dynamic interaction between the shells and the ACLD treatments. Experiments were performed to verify the numerical predictions. The obtained results suggest the potential of the ACLD treatments in controlling the vibration of cylindrical shells. The performance of this PFRC material as the material for distributed actuators of simply supported composite cylindrical panels was investigated [8]. The effect of variation of the piezoelectric fiber orientation in the constraining PFRC layer and the shallowness angle of the panels on the control authority of the patches was studied.

In this paper an endeavor has been made to investigate the performance of ACLD treatment to improve the damping characteristics of thin antisymmetric laminated circular cylindrical shells made of composite materials which are the building blocks of many engineering structures. The shells studied here are antisymmetric cross-ply and angle-ply laminated shells and assumed to be simply supported.

2. Theoretical formulation

In this section, the equations of motion governing the open loop behavior of antisymmetric laminated cylindrical shells integrated with the patches of ACLD treatment are derived. Schematic diagram of the antisymmetric laminated composite cylindrical shell is shown in Fige 1, which is made of N

number of orthotropic layers. The length, thickness and average radius of the shell are denoted by a, h and R, respectively. The top surface of the shell is integrated with the rectangular patches of ACLD treatment. The constraining layer of the ACLD treatment is made of the PFRC material in which the fibers are unidirectionally aligned and parallel to the plane of the shell and its constructional feature is also schematically demonstrated in Fig 1. The thickness of the PFRC layer is h_p and that of the viscoelastic constrained layer of the ACLD treatment is h_v. The mid-plane of the substrate shell is considered as the reference plane. Four points of both ends of the shell are considered as clamped. The vibration analysis has been done using finite element model of the cylindrical shell with ACLD patches of PFRC. The symmetric laminated cylindrical shell is assumed to be thin with individual ply made up of homogeneous, orthotropic and linearly elastic material. The layers are assumed to be perfectly bonded together. Although the nature of the excitation of actual shells is stochastic, the composite cylindrical shell has been modeled with a harmonic point force excitation. It is also assumed that PFRC patches are perfectly bonded with the viscoelastic layers where the viscoelastic material is isotropic and there is perfect bond between viscoelastic patches and composite shell and PFRC layers.

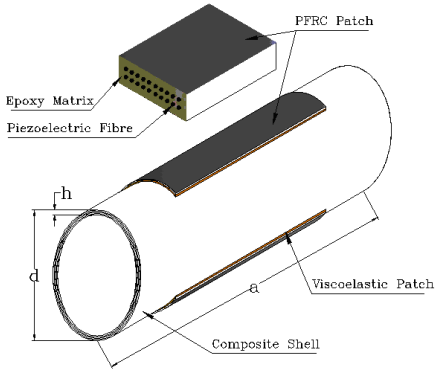


Fig 1: Schematic representation of antisymmetric laminated composite shell integrated with the patches of ACLD treatment

3. Displacement field:

Figure 2 describes a schematic representation of the kinematics of deformation based on FSDT and linear displacement theories. In the figure u_0 and v_0 are the generalized translational displacements of a reference point (x, y) on the mid-plane (z = 0) of the substrate shell along x- and y-axes, respectively. θ_x, $\emptyset_x$, and γ_x are the generalized rotations of the normal to the middle planes of the substrate, viscoelastic layer and the PFRC layer, respectively about y-axis, while the generalized rotations of these normal about x-axis are denoted respectively by θ_y, $\emptyset_y$, and γ_y.

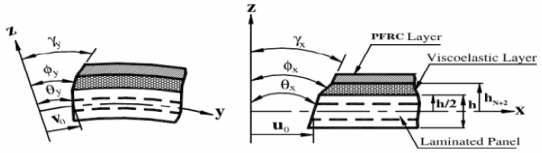


Fig 2: Kinematics of deformation

According to the kinematics of deformation shown in figure 2, the displacements u, v and w at any point in any layer of the overall shell along x, y and z-directions, respectively can be expressed as

$$u(x,y,z,t)=u_0(x,y,t)+\left(z-\left\langle z-\frac{h}{2}\right\rangle\right)\theta_x(x,y,t)+\left(\left\langle z-\frac{h}{2}\right\rangle-\left\langle z-h_{N+2}\right\rangle\right)\phi_x(x,y,t)+\left\langle z-h_{N+2}\right\rangle\gamma_x(x,y,t) \quad (1)$$

$$v(x,y,z,t)=v_0(x,y,t)+\left(z-\left\langle z-\frac{h}{2}\right\rangle\right)\theta_y(x,y,t)+\left(\left\langle z-\frac{h}{2}\right\rangle-\left\langle z-h_{N+2}\right\rangle\right)\phi_y(x,y,t)+\left\langle z-h_{N+2}\right\rangle\gamma_y(x,y,t) \quad (2)$$

and , $$w(x,y,z,t)=w_0(x,y,t) \quad (3)$$

In which, a function within the bracket $\langle\ \rangle$ represents the appropriate singularity functions. For convenience, the generalized displacements are separated into generalized translational $\{d_t\}$ and rotational $\{d_r\}$ displacements as follows:

$$\{d\}=\{d_t\}+[z]\{d_r\} \quad \text{in which}$$

$$\{d\}=[u\ v\ w]^T \qquad \{d_t\}=[u_0\ v_0\ w_0]^T$$

and $$\{d_r\}=[\theta_x\ \theta_y\ \phi_x\ \phi_y\ \gamma_x\ \gamma_y]^T \quad (4)$$

In order to compute the element stiffness matrices corresponding to the transverse shear deformations, the state of strain at any point in the overall shell is divided into the following two strain vectors $\{\varepsilon_b\}$ and $\{\varepsilon_s\}$

$$\{\varepsilon_b\}=[\varepsilon_x\ \varepsilon_y\ \varepsilon_{xy}]^T \text{ and } \{\varepsilon_s\}=[\varepsilon_{xz}\ \varepsilon_{yz}]^T \quad (5)$$

In which ε_x and ε_y are the normal strains along x and y directions, respectively; ε_{xy} is the in-plane shear strain; and ε_{xz} , ε_{yz} are the transverse shear strains. The state of strain in cylindrical coordinate is given by

$$\{\varepsilon_x\ \varepsilon_y\ \varepsilon_{xy}\ \varepsilon_{xz}\ \varepsilon_{yz}\}=\{\frac{\partial u}{\partial x}\ \frac{\partial v}{\partial x}+\frac{w}{R}\ \frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\ \frac{\partial w}{\partial x}+\frac{\partial u}{\partial z}\ \frac{\partial w}{\partial y}+\frac{\partial v}{\partial z}-\frac{v}{R}\} \quad (6)$$

By using the displacement fields (Eqs. 1-3) and the linear strain displacement relations (Eq. 5), the vectors $\{\varepsilon_b\}_c$, $\{\varepsilon_b\}_v$ and $\{\varepsilon_b\}_p$ defining the state of in-plane strains, the vectors $\{\varepsilon_s\}_c$, $\{\varepsilon_s\}_v$ and $\{\varepsilon_s\}_p$ defining the state of transverse shear strains at any point in the substrate composite shell, viscoelastic layer and the active constraining layer (i.e. PFRC) respectively, can be expressed as

$$\{\varepsilon_b\}_c=\{\varepsilon_{bt}\}+[z_1]\{\varepsilon_{br}\} \quad (7)$$

$$\{\varepsilon_b\}_v=\{\varepsilon_{bt}\}+[z_2]\{\varepsilon_{br}\} \quad (8)$$

$$\{\varepsilon_b\}_p=\{\varepsilon_{bt}\}+[z_3]\{\varepsilon_{br}\} \quad (9)$$

$$\{\varepsilon_s\}_c=\{\varepsilon_{st}\}+[z_4]\{\varepsilon_{sr}\} \quad (10)$$

$$\{\varepsilon_s\}_v=\{\varepsilon_{st}\}+[z_5]\{\varepsilon_{sr}\} \quad (11)$$

and $$\{\varepsilon_s\}_p=\{\varepsilon_{st}\}+[z_6]\{\varepsilon_{sr}\} \quad (12)$$

Where $\{\varepsilon_{bt}\}=\left[\frac{\partial u_o}{\partial x}\ \frac{\partial v_o}{\partial y}+\frac{w}{R}\ \frac{\partial u_o}{\partial y}+\frac{\partial v_o}{\partial x}\right]^T$,

$$\{\varepsilon_{sl}\}=\left[\frac{\partial w_o}{\partial x}\quad \frac{\partial w_o}{\partial y}-\frac{v}{R}\right]^T$$

$$\{\varepsilon_{sr}\}=\left[\frac{\partial \theta_x}{\partial x}\quad \frac{\partial \theta_y}{\partial y}\quad \frac{\partial \theta_x}{\partial y}+\frac{\partial \theta_y}{\partial x}\quad \frac{\partial \phi_x}{\partial x}\quad \frac{\partial \phi_y}{\partial y}\quad \frac{\partial \phi_x}{\partial y}+\frac{\partial \phi_y}{\partial x}\quad \frac{\partial \gamma_x}{\partial x}\quad \frac{\partial \gamma_y}{\partial y}\quad \frac{\partial \gamma_x}{\partial y}+\frac{\partial \gamma_y}{\partial x}\right]^T$$

$$\{\varepsilon_{sr}\}=\left[\theta_x\ \ \theta_y\ \ \phi_x\ \ \phi_y\ \ \gamma_x\ \ \gamma_y\right]^T \qquad (13)$$

And the forms of the matrices Z_1-Z_6 are given in the appendix.

Similarly the state of stresses at any point in the overall shell is described by the following stress vectors:

$$\{\sigma_b\}=[\sigma_x\ \ \sigma_y\ \ \sigma_{xy}]^T \text{ and } \{\sigma_s\}=[\sigma_{xz}\ \ \sigma_{yz}]^T \qquad (14)$$

Where σ_x, σ_y are the normal stresses along x, y directions, respectively; σ_{xy} is the in plane shear stresses; σ_{xz} and σ_{yz} are the transverse shear stresses.

The constitutive relations for the material of any orthotropic layer of the substrate shell are given by

$$\left\{\sigma_b^k\right\}=\left[\bar{C}_b^k\right]\left\{\varepsilon_b^k\right\} \text{ and } \left\{\sigma_s^k\right\}=\left[\bar{C}_s^k\right]\left\{\varepsilon_s^k\right\}$$

for k=1,2,3,……N (15)

The constraining PFRC layer will be subjected to the applied electric field E_z acting across its thickness (i.e. along the z-direction) only. Thus the constitutive relations for the material of the PFRC layer can be expressed as

$$\left\{\sigma_b^k\right\}=\left[\bar{C}_b^k\right]\left\{\varepsilon_b^k\right\}-\left\{\bar{e}_b^k\right\}E_z \text{ and } \left\{\sigma_s^k\right\}=\left[\bar{C}_s^k\right]\left\{\varepsilon_s^k\right\} \qquad (16)$$

$$\text{And, } D_z=\left\{\bar{e}_b^k\right\}\left\{\varepsilon_b^k\right\}+\bar{\in}_{33}E_z \qquad \text{for k=N+2} \qquad (17)$$

In which Dz is the electric field in the z-direction and $\bar{\in}_{33}$ is the transformed dielectric constant. The transformed elastic coefficient matrices $\left[\bar{C}_b^k\right]$, $\left[\bar{C}_s^k\right]$ and the transformed piezoelectric coefficient matrix $\left\{\bar{e}_b^k\right\}$ appearing in Eqs. 15 and 16, referred to the laminate coordinate (x, y, z) system are given by

$$\left[\bar{C}_b^k\right]=\begin{bmatrix}\bar{C}_{11}^k & \bar{C}_{12}^k & \bar{C}_{16}^k\\ \bar{C}_{12}^k & \bar{C}_{22}^k & \bar{C}_{26}^k\\ \bar{C}_{16}^k & \bar{C}_{26}^k & \bar{C}_{66}^k\end{bmatrix},\left[\bar{C}_s^k\right]=\begin{bmatrix}\bar{C}_{55}^k & \bar{C}_{45}^k\\ \bar{C}_{45}^k & \bar{C}_{44}^k\end{bmatrix}$$

$$\text{and } \left\{\bar{e}_b\right\}=\begin{Bmatrix}\bar{e}_{31}\\ \bar{e}_{32}\\ \bar{e}_{36}\end{Bmatrix} \qquad (18)$$

$\bar{C}_{ij}^k$ (i, j = 1, 2, 3 ... 6) are the transformed elastic coefficients with respect to the reference coordinate system. The material of the viscoelastic layer is assumed to be linearly viscoelastic and is modeled by using the complex modulus approach in which, the shear modulus G and the Young's modulus E of the viscoelastic material are given by

$$G=G'(1+i\eta) \text{ and } E=2G(1+\upsilon) \qquad (19)$$

In which G' is the storage modulus υ is the Poisson's ratio and η is the loss factor at a particular operating temperature and frequency. Employing the complex modulus approach the constitutive relations for the material of the viscoelastic layer (k = N + 1) can also be represented by Eq. 15 with $\bar{C}_{ij}^{N+1}$ (i, j = 1, 2, 3 ... 6) being the complex elastic constants.

The energy functional T_p of the shell integrated with the patches of ACLD treatment describing the total energy comprising of the strain energy, electrical energy and the work done by the external load can be written as

$$T_P=\frac{1}{2}\left[\sum_{k=1}^{N+2}\int_v\left(\{\varepsilon_b{}^k\}^T\{\sigma_b{}^k\}+\{\varepsilon_s{}^k\}^T\{\sigma_s{}^k\}\right)dv\right. \qquad (20)$$

$$\left.-\int_v E_z D_z dv\right]-\int_A\{d\}^T\{f_w\}dA$$

And total kinetic energy $\boldsymbol{T_k}$ of the overall shell can be expressed as

$$T_k=\frac{1}{2}\left[\sum_{k=1}^{N+2}\int_v\rho^k\left(\dot{u}^2+\dot{v}^2+\dot{w}^2\right)dv\right] \qquad (21)$$

In which, $\boldsymbol{\{f\}}$ is the externally applied surface traction vector acting over a surface area $\boldsymbol{A}$, $\boldsymbol{V}$ represents the volume of the k^{th} layer in concern and $\boldsymbol{\rho^k}$ is the mass density of the concerned layer.

4. Finite Element formulation

The overall shell has been discretized by eight noded isoparametric quadrilateral elements. Following Eqs. 4, the generalized displacement vectors for the i^{th} (i = 1, 2, 3, ..., 8) node of an element can be represented as

$$\left\{d_{ti}\right\}=\left[u_{0i}\ \ v_{0i}\ \ w_{0i}\right]^T \text{ and}$$

$$\left\{d_{ri}\right\}=\left[\theta_{xi}\ \ \theta_{yi}\ \ \phi_{xi}\ \ \phi_{yi}\ \ \gamma_{xi}\ \ \gamma_{yi}\right]^T \qquad (22)$$

Hence the generalized displacement vectors at any point within the element can be expressed in terms of the nodal generalized displacement vectors $\left\{d_t^e\right\}$ and $\left\{d_t^e\right\}$ as:

$$\left\{\mathrm{d_t}\right\}=\left[\mathrm{N_t}\right]\left\{\mathrm{d_t}^e\right\} \text{ and } \left\{\mathrm{d_r}\right\}=\left[\mathrm{N_r}\right]\left\{\mathrm{d_r}^e\right\} \qquad (23)$$

Wherein, $\left\{d_t^e\right\}$ is the nodal generalized translational displacement vector, $\left\{d_t^e\right\}$ is the nodal generalized rotational displacement vector, $[N_t]$ and $[N_r]$ are the shape function matrices are given by

$$\left\{d_t^e\right\}=\left[\left\{d_{t1}^e\right\}^T\left\{d_{t2}^e\right\}^T\left\{d_{t3}^e\right\}^T....\left\{d_{t8}^e\right\}^T\right]^T$$

$$\left\{d_r^e\right\}=\left[\left\{d_{r1}^e\right\}^T\left\{d_{r2}^e\right\}^T\left\{d_{r3}^e\right\}^T....\left\{d_{r8}^e\right\}^T\right]^T$$

$$\left[N_t\right]=\left[N_{t1}\ \ N_{t2}\ \ N_{t3}....N_{t8}\right]$$

$$\left[N_r\right]=\left[N_{r1}\ \ N_{r2}\ \ N_{r3}....N_{r8}\right]$$

$$N_{ti}=n_i I_t \text{ and } N_{ri}=n_i I_r \qquad (24)$$

While, I_t and I_r are the (3 × 3) and (9 × 9) identity matrices,

respectively and n_i is the shape function of natural coordinates associated with the i^{th} node. Making use of the relations given by Eqs. 7-11 and 24, the strain vectors at any point within the element can be expressed in terms of the nodal generalized displacement vectors as follows

$$\{\varepsilon_b\}_c=[B_{tb}]\{d_t^e\}+[z_1][B_{rb}]\{d_r^e\}$$
$$\{\varepsilon_b\}_v=[B_{tb}]\{d_t^e\}+[z_2][B_{rb}]\{d_r^e\}$$
$$\{\varepsilon_b\}_p=[B_{tb}]\{d_t^e\}+[z_3][B_{rb}]\{d_r^e\}$$
$$\{\varepsilon_s\}_c=[B_{ts}]\{d_t^e\}+[z_4][B_{rs}]\{d_r^e\}$$
$$\{\varepsilon_s\}_v=[B_{ts}]\{d_t^e\}+[z_5][B_{rs}]\{d_r^e\}$$
and $$\{\varepsilon_s\}_p=[B_{ts}]\{d_t^e\}+[z_6][B_{rs}]\{d_r^e\} \quad (25)$$

In which the nodal strain-displacement matrices $[B_{tb}]$, $[B_{rb}]$, $[B_{ts}]$ and $[B_{rs}]$ are given by

$$[B_{tb}]=[B_{tb1}\ B_{tb2}\ B_{tb3}\ldots.B_{tb8}]$$
$$[B_{rb}]=[B_{rb1}\ B_{rb2}\ B_{rb3}\ldots.B_{rb8}]$$
$$[B_{ts}]=[B_{ts1}\ B_{ts2}\ B_{ts3}\ldots.B_{ts8}]$$
$$[B_{rs}]=[B_{rs1}\ B_{rs2}\ B_{rs3}\ldots.B_{rs8}] \quad (26)$$

The submatrices of [Btb], [Brb], [Bts] and [Brs] as shown in Eqs.26, have been presented in the Appendix. After substitution of Eqs.15 and 25 into Eq. 20, the potential energy T_p^e of a typical element augmented with the ACLD treatment can be expressed as

$$T_P{}^e=\frac{1}{2}[\{d_t^e\}^T[k_{tt}^e]\{d_t^e\}+\{d_t^e\}^T[k_{tr}^e]\{d_r^e\}+\{d_r^e\}^T[k_{tr}^e]^T\{d_t^e\}+\{d_r^e\}^T[k_{rr}^e]\{d_r^e\}-2\{d_t^e\}^T[F_{\varphi}^e]V-2\{d_r^e\}^T[F_{rp}^e]V-2\{d_t^e\}^T[F^e]]-\bar{\epsilon}_{33}\bar{E}^2V^2 \quad (27)$$

In which $\bar{E}=-\frac{1}{h_p}$ and $\boldsymbol{V}$ are the potential difference across the thickness of the PFRC layer. The elemental stiffness matrices ($[K_{tt}^e]$, $[K_{tr}^e]$ and $[K_{rr}^e]$), the elemental electro-elastic coupling vectors ($\{F_{tp}^e\}$ and $\{F_{rp}^e\}$) and the elemental load vector $\{F^e\}$ appearing in Eq. 27 are given by

$$[k_{tt}^e]=[k_{tb}^e]+[k_{ts}^e]\quad [k_{tr}^e]=[k_{trb}^e]+[k_{trs}^e]$$
$$[k_{tr}^e]=[k_{rt}^e]^T\quad [k_{rr}^e]=[k_{rrb}^e]+[k_{rrs}^e]$$
$$\{F_{tp}^e\}=\{F_{tb}^e\}_p+\{F_{ts}^e\}_p\quad \{F_{rp}^e\}=\{F_{rb}^e\}_p+\{F_{rs}^e\}_p$$
$$\{F^e\}=\int_0^{be}\int_0^{ae}[N_t]^T\{f\}dxdy$$

Where,

$$[K_{tb}^e]=\int_0^{b_e}\int_0^{a_e}[B_{tb}]^T\left([D_{tb}]+[D_{tb}]_v+[D_{tb}]_p\right)[B_{tb}]dxdy$$
$$[K_{ts}^e]=\int_0^{b_e}\int_0^{a_e}[B_{ts}]^T\left([D_{ts}]+[D_{ts}]_v+[D_{ts}]_p\right)[B_{ts}]dxdy$$
$$[K_{trb}^e]=\int_0^{b_e}\int_0^{a_e}[B_{tb}]^T\left([D_{trb}]+[D_{trb}]_v+[D_{trb}]_p\right)[B_{rb}]dxdy$$
$$[K_{trs}^e]=\int_0^{b_e}\int_0^{a_e}[B_{ts}]^T\left([D_{trs}]+[D_{trs}]_v+[D_{trs}]_p\right)[B_{rs}]dxdy$$
$$[K_{rrb}^e]=\int_0^{b_e}\int_0^{a_e}[B_{rb}]^T\left([D_{rrb}]+[D_{rrb}]_v+[D_{rrb}]_p\right)[B_{rb}]dxdy$$
$$[K_{rrs}^e]=\int_0^{b_e}\int_0^{a_e}[B_{rs}]^T\left([D_{rrs}]+[D_{rrs}]_v+[D_{rrs}]_p\right)[B_{rs}]dxdy$$
$$\{F_{tb}^e\}_p=\int_0^{be}\int_0^{ae}[B_{tb}]^T\{D_{tb}\}_p\,dxdy$$
$$\{F_{rb}^e\}_p=\int_0^{be}\int_0^{ae}[B_{rb}]^T\{D_{rb}\}_p\,dxdy\ ,\{F_{ts}^e\}_p=\int_0^{be}\int_0^{ae}[B_{ts}]^T\{D_{ts}\}_p\,dxdy$$
$$\{F_{rs}^e\}_p=\int_0^{be}\int_0^{ae}[B_{rs}]^T\{D_{rs}\}_p\,dxdy$$

Wherein, a_e and b_e are the length and circumferential width of the element in consideration and the various rigidity matrices originated in the above elemental matrices are given by

$$[D_{tb}]=\sum_{k=1}^{N}\int_{h_k}^{h_{k+1}}[\bar{C}_b^k]dz \qquad [D_{trb}]=\sum_{k=1}^{N}\int_{h_k}^{h_{k+1}}[\bar{C}_b^k][Z_1]dz$$
$$[D_{rrb}]=\sum_{k=1}^{N}\int_{h_k}^{h_{k+1}}[Z_1]^T[\bar{C}_b^k][Z_1]dz \qquad [D_{ts}]=\sum_{k=1}^{N}\int_{h_k}^{h_{k+1}}[\bar{C}_s^k]dz$$
$$[D_{trs}]=\sum_{k=1}^{N}\int_{h_k}^{h_{k+1}}[\bar{C}_s^k][Z_4]dz \qquad [D_{rrs}]=\sum_{k=1}^{N}\int_{h_k}^{h_{k+1}}[Z_4]^T[\bar{C}_s^k][Z_4]dz$$
$$[D_{tb}]_v=h_v[\bar{C}_b^{N+1}] \qquad [D_{trb}]_v=\int_{h_{N+1}}^{h_{N+2}}[\bar{C}_b^{N+1}][Z_2]dz$$
$$[D_{rrb}]_v=\int_{h_{N+1}}^{h_{N+2}}[Z_2]^T[\bar{C}_b^{N+1}][Z_2]dz \qquad [D_{ts}]_v=h_v[\bar{C}_s^{N+1}]$$
$$[D_{trs}]_v=\int_{h_{N+1}}^{h_{k+2}}[\bar{C}_s^{N+1}][Z_5]dz \qquad [D_{rrs}]_v=\int_{h_{N+1}}^{h_{k+2}}[Z_5]^T[\bar{C}_s^{N+1}][Z_5]dz$$
$$[D_{tb}]_p=h_p[\bar{C}_b^{N+2}] \qquad [D_{trb}]_p=\int_{h_{N+2}}^{h_{N+3}}[\bar{C}_b^{N+2}][Z_3]dz$$
$$[D_{rrb}]_p=\int_{h_{N+2}}^{h_{N+3}}[Z_3]^T[\bar{C}_b^{N+2}][Z_3]dz \qquad [D_{ts}]_p=h_p[\bar{C}_s^{N+2}]$$
$$[D_{trs}]_p=\int_{h_{N+2}}^{h_{N+3}}[\bar{C}_s^{N+2}][Z_6]dz \qquad [D_{rrs}]_p=\int_{h_{N+2}}^{h_{N+3}}[Z_6]^T[\bar{C}_s^{N+2}][Z_6]dz$$
$$[D_{tb}]_p=\int_{h_{N+2}}^{h_{N+3}}\{\bar{e}_b\}\bar{E}dz \qquad [D_{rb}]_p=\int_{h_{N+2}}^{h_{N+3}}[Z_3]^T\{\bar{e}_b\}\bar{E}dz$$
$$[D_{ts}]_p=\int_{h_{N+2}}^{h_{N+3}}\{\bar{e}_s\}\bar{E}dz \qquad [D_{rs}]_p=\int_{h_{N+2}}^{h_{N+3}}[Z_6]^T\{\bar{e}_b\}\bar{E}dz$$

Substituting Eq.23 into Eq. 21 and neglecting the rotary inertia of the overall shell as the host shell is very thin; the expression for kinetic energy T_k^e of the element can be obtained as

$$T_k^e = \frac{1}{2}\left\{\dot{d}_t^e\right\}^T [M]\left\{\dot{d}_t^e\right\} \quad (28)$$

In which the elemental mass matrix **[M^e]** is given by

$$[M^e] = \int_0^{b_e}\int_0^{a_e}\left(\sum_{k=1}^{N}\rho^k\left(h_{k+1}-h_k\right)+\rho^{N+1}h_v+\rho^{N+2}h_p\right)[N_t]^T[N_t]\,dxdy$$

Using the principle of virtual work the governing equations of an element as follows:

$$\left[M^e\right]\left\{\ddot{d}_t^e\right\}+\left[K_{tt}^e\right]\left\{d_t^e\right\}+\left[K_{tr}^e\right]\left\{d_r^e\right\}=\left\{F_{tp}^e\right\}V+\left\{F^e\right\} \quad (29)$$

$$\left[K_{tr}^e\right]\left\{d_t^e\right\}+\left[K_{rr}^e\right]\left\{d_r^e\right\}=\left\{F_{rp}^e\right\}V \quad (30)$$

It is to be noted now that in case of an element without integrated with the ACLD treatment, the electro-elastic coupling matrices $\left\{F_{tp}^e\right\}$, $\left\{F_{rp}^e\right\}$ turn out to be the null matrices and the other elemental stiffness matrices in Eqs. 29 and 30 become real.

The elemental equations of motion are assembled to form the global equations of motion in such a manner that each patch can be activated separately as follows:

$$[M]\left\{\ddot{X}\right\}+\left[K_{tt}\right]\{X\}+\left[K_{tr}\right]\{X_r\}=\sum_{j=1}^{m}\left\{F_{tp}^j\right\}V^j+\{F\} \quad (31)$$

$$\left[K_{tr}\right]\{X\}+\left[K_{rr}\right]\{X_r\}=\sum_{j=1}^{m}\left\{F_{rp}^j\right\}V^j \quad (32)$$

Where, **[M]** is the global mass matrix; **[K_{tt}]**, **[K_{tr}]**, **[K_{rr}]** are the global stiffness matrices; **{X}** and **{X_r}** are the global nodal translational and rotational degrees of freedom; $\left\{F_{tp}^j\right\}$ and $\left\{F_{rp}^j\right\}$ are the global electro-elastic coupling matrices corresponding to the j[th] patch, V^j is the voltage applied to this patch, m is the number of patches and {F} is the global nodal force vector. After invoking the boundary conditions, the global rotational degrees of freedom can be eliminated to derive the global open loop equations of motion in terms of the global translational degrees of freedom onl0y as follows:

$$[M]\left\{\ddot{X}\right\}+\left[K^*\right]\{X\}=\sum_{j=1}^{m}\left\{F_{tp}^j\right\}V^j+\{F\} \quad (33)$$

In which $\left[K^*\right]=\left[K_{tt}\right]-\left[K_{tr}\right]\left[K_{rr}\right]^{-1}\left[K_{tr}\right]^T$ and $\left\{F_p^j\right\}=\left\{F_{tp}^j\right\}-\left[K_{tr}\right]\left[K_{rr}\right]^{-1}\left\{F_{rp}^j\right\}$

Since the elemental stiffness matrices of an element augmented with the ACLD treatment are complex, the global stiffness matrices become complex and the energy dissipation characteristics of the overall shell is attributed to the imaginary part of these matrices. Hence, the global equations of motion as derived above also represent the passive (uncontrolled) constrained layer damping of the substrate shell when the constraining layer is not subjected to any control voltage following a derivative control law.

5. Closed loop model

The control voltage necessary for activating the patches of the ACLD treatment, a simple velocity feedback control law has been employed. According to this law, the control voltage supplied to each patch can be expressed in terms of the derivatives of the global nodal degrees of freedom as follows:

$$V^j = -K_d^j\,\dot{w} = -K_d^j\left[N^j\right]\left\{\dot{X}\right\} \quad (34)$$

In which K_d^j is the control gain for the j[th] patch and **[N^j]** is a row vector defining the location of the concerned point for sensing the velocity signal that will be fed back to this patch. Finally, substituting Eq. 33 into Eq. 32, the equations of motion governing the closed loop dynamics of the substrate shells activated by the patches of ACLD treatments can be derived as follows:

$$[M]\left\{\ddot{X}\right\}+\left[C_d\right]\left\{\dot{X}\right\}+\left[K^*\right]\{X\}=\{F_t\}+\{F\} \quad (35)$$

It is obvious from Eq. 35 that the implementation of the control strategy yields the active damping matrix [C_d] that is given by

$$\left[C_d\right]=\sum_{j=1}^{m}K_d^j\left\{F_p^j\right\}\left[N^j\right]$$

Since the stiffness matrices for an element augmented with the ACLD treatment are complex, the global stiffness matrix [K*] becomes complex and its imaginary part is responsible for contribution to the dissipation of energy. Hence, in the absence of applied control voltage the equations of motion given by Eq. 33 govern the passive (uncontrolled) constrained layer damping of the substrate shells. In order to study the performance of the shells in frequency domain, it is considered that the shells are subjected to time-harmonic excitation force and the motion is harmonic. Thus it can be written that

$$\{X\}=\left\{\overline{X}\right\}e^{-i\omega t} \text{ and } \{F\}=\left\{\overline{F}\right\}e^{-i\omega t} \quad (36)$$

Where, $\left\{\overline{X}\right\}$ and $\left\{\overline{F}\right\}$ are the vectors of amplitudes of the nodal displacements and excitation forces and ω is the frequency of excitation. Substitution of Eq. 35 into Eq. 34 leads to the following algebraic equation:

$$\left\{\overline{X}\right\}=\left[K_{eq}\right]^{-1}\left\{\overline{F}\right\} \quad (37)$$

In which $\left[K_{eq}\right]=-\omega^2[M]+i\omega\left[C_d\right]+\left[K^*\right]$

6. Numerical Results:

In this section, the numerical results are evaluated using the finite element model derived in the previous sections. Performance of antisymmetric cross-ply and angle-ply thin circular cylindrical shell integrated with two rectangular patches of ACLD treatment are investigated.

Geometry:

The patches are placed on the outer surfaces of the shell as shown in Fig 1. The length and width of the patch are assumed to be 3/5 of the longitudinal length and 1/6 of

circumferential length of the shell respectively. The piezoelectric fiber orientation in the constraining PFRC layer is 0^0. The thickness of the PFRC layer, the viscoelastic layer and the laminated shells are considered as 150 µm, 200 µm and 3 mm, respectively. The axial length (a) and the radius (R) of the cylindrical shell are 1m and 150 mm, respectively.

Material Properties:

The shell is constructed with 4 layer of graphite/epoxy laminated composite. The ACLD patches made with PFRC and viscoelastic material. The materials of the piezoelectric fiber and the matrix of the PFRC layer are considered as PZT5H and epoxy, respectively. Considering 40% fiber volume fraction, the following elastic and piezoelectric properties of the PFRC layer with respect to its material coordinate system are obtained from the existing micromechanics model and are used for evaluating the numerical results

$\boldsymbol{C_{11}}$ = 32.6 GPa; $\boldsymbol{C_{12}}$ = 4.3 GPa; $\boldsymbol{C_{22}}$ = 7.2 GPa;

$\boldsymbol{C_{44}}$ = 1.05 GPa; $\boldsymbol{C_{55}} = \boldsymbol{C_{66}}$ = 1.29 GPa; ρ = 3640 kg/m^3

The applied electric field acts only in the z direction. Hence, the piezoelectric co-efficient e_{15} and e_{24} are not required for evaluating the numerical results. Also, the magnitudes of the piezoelectric co-efficient e_{32} and e_{33} of the PFRC material considered in this study are negligibly small in comparison to that of its piezoelectric co-efficient e_{31}.

e_{31} = -6.76 C/m^2;

The material properties considered for the orthotropic graphite/epoxy layers of the substrate shells are considered as follows:

$\boldsymbol{E_L}$ = 172.9 GPa; $\boldsymbol{E_L} / \boldsymbol{E_T}$ = 25; $\boldsymbol{G_{LT}} = 0.5\boldsymbol{E_T}$;

$\boldsymbol{G_{TT}} = 0.2\boldsymbol{E_T}$; $\boldsymbol{v_{LT}} = \boldsymbol{v_{TT}}$ = 0.25; $\boldsymbol{\rho}$ =1600 kg/m^3

Complex shear modulus, Poisson's ratio and density of the viscoelastic layer are 20(1 + i) MPa, 0.49 and 1140 kg/m^3, respectively. The loss factor of this viscoelastic material remains invariant within a frequency range (0-600Hz) of interest.

Boundary Conditions and Loads:

The model is meshed with eight noded isoparametric elements. Only four points of the both ends of the shell are clamped, so at that points boundary conditions are considered for evaluating the numerical results are given by

$$v_0 = w = \theta_y = \phi_y = \gamma_y = 0 \text{ and}$$
$$u_0 = w = \theta_x = \phi_x = \gamma_x = 0$$

On the top surface of the cylindrical shell a time harmonic point force (1N) is considered to act at a point (a/2, 0, h/2) to excite the first few modes of the shells. The control voltage supplied to each patch is negatively proportional to the velocity of the point located on the shell outer surface which corresponds to the midpoint of the free length of the patch.

Results:

Figure 3 illustrates the frequency response functions of a four layered antisymmetric cross-ply ($0^o/90^o/0^o/90^o$) shell and the variation of the required control voltage applied to each patch with the frequency of excitation has been shown in figure 4 when the value of control gain k_d is zero, 2000 and 4000. Displayed in figure 3 is the response of the cylindrical shell when the patches are passive and active with different control gains. It is evident from these figures that the active patches significantly improve the damping characteristics of the shell for both the modes over the passive damping with very low control voltage. As the gain increases, attenuation of the amplitudes of vibration also increases. The deformed shape of the shell in first fundamental mode is shown in figure 5.

It has been shown that the patches efficiently attenuate the amplitudes of vibrations by enhancing the damping characteristics of the cylindrical shell. For further reference, the magnitudes of the uncontrolled and controlled transverse displacement of the shells subjected to active constrained layer damping and undergoing fundamental mode of vibration are presented in table 1 for cross ply and angle ply arrangements.

Table 1: Amplitudes of the uncontrolled and controlled transverse displacement of the shells undergoing fundamental mode of vibration

Shell Type	Frequency of 1st Fundamental Mode (Hz)	Control gain Kd	Amplitude (w) at (a/2,s/2,h/2) (m)	Control voltage (v)
$0^o/90^o/0^o/90^o$	254	0	3.581×10^{-5}	0
		2000	4.177×10^{-6}	42.95
		4000	2.174×10^{-6}	45.6
$-45^o/45^o/-45^o/45^o$	251	0	3.68×10^{-5}	0
		2000	7.392×10^{-6}	16.88
		4000	7.13×10^{-6}	44.78

By comparing the present model with modifications for antisymmetric configuration and boundary conditions with R/h=50 and reference [2] the comparative observations were presented in table 2 which shows the close agreement of the present model with Ref [2].

Next, the effect of variation of fiber orientation in the constraining PFRC layer on the damping of first mode of vibration of antisymmetric cross-ply and angle ply cylindrical shell has been studied. The plot for the attenuating capability of the patches is shown in figure 8. It has been observed that the attenuation enhanced in the shell with cross-ply substrate when the piezoelectric fiber orientation angle (ψ) in PFRC layer is 30^0 and 90^0 but for shell with angle-ply substrate amplitude of the vibration enhanced when the piezoelectric fiber orientation angle (ψ) in PFRC layer is 30^0. Maximum attenuation is achieved for 15^0 piezoelectric fiber orientation angle (ψ) in PFRC layer in the cylindrical shell with angle-ply substrate.

Table 2: Comparison of present model with Ref [2]

Shell Type	Frequency of 1st Fundamental Mode (Hz)	Control gain K_d	Amplitude (w) at (a/2,s/2,h/2) (m) Present FEM	Amplitude (w) at (a/2,s/2,h/2) (m) from ref [2]
$0^o/90^o/0^o/90^o$ Antisymmetric Cross Ply	153 (from ref [2] 156)	0	1.48×10^{-4}	1.56×10^{-4}
		600	0.42×10^{-4}	0.40×10^{-4}
		1500	0.20×10^{-4}	0.22×10^{-4}
$-45^o/45^o/-45^o/45^o$ Antisymmetric Angle Ply	100 (from ref [2] 102)	0	4.04×10^{-4}	4.2×10^{-4}
		600	0.6×10^{-4}	0.75×10^{-4}
		1500	0.25×10^{-4}	0.28×10^{-4}

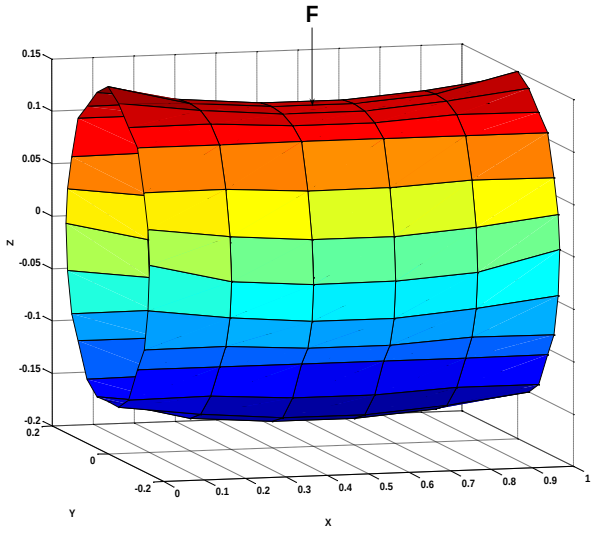


Fig 5: Deformed shape in z direction in 1st fundamental mode

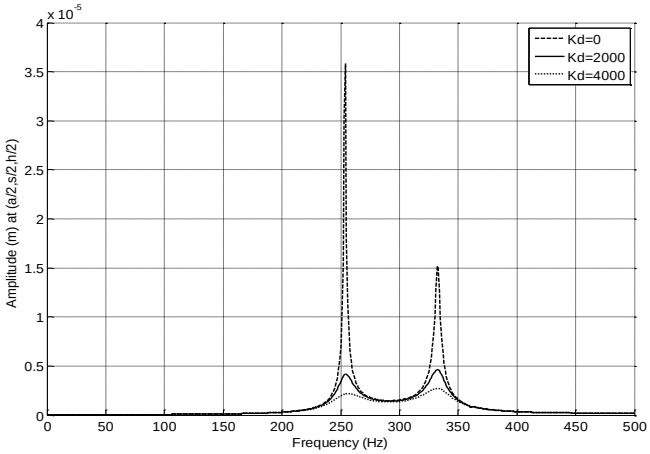


Fig 3: Frequency response of antisymmetric cross-ply ($0^o/90^o/0^o/90^o$) laminated shell

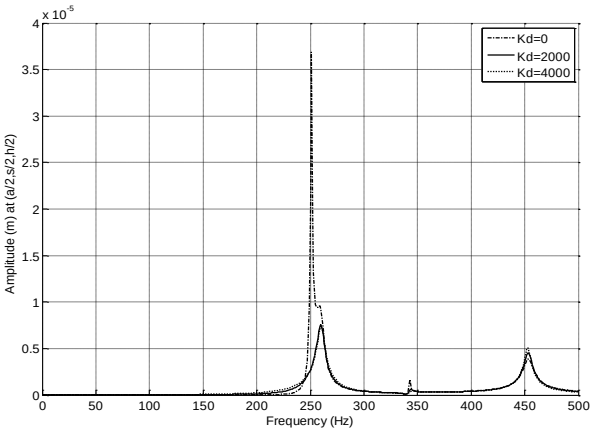


Fig 6: Frequency response of antisymmetric angle-ply ($-45^o/45^o/-45^o/45^o$) laminated shell

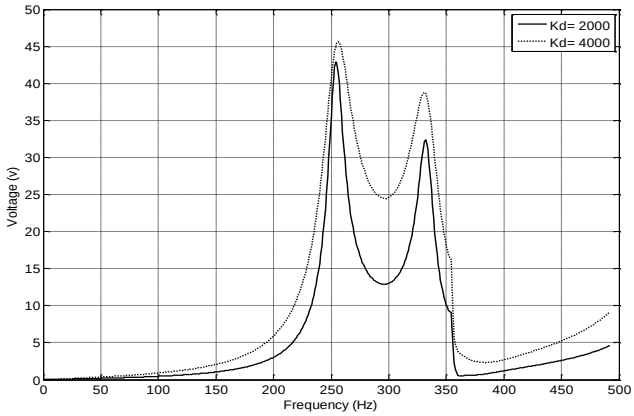


Fig 4: Control voltage for antisymmetric cross-ply ($0^o/90^o/0^o/90^o$) laminated shell with different gains

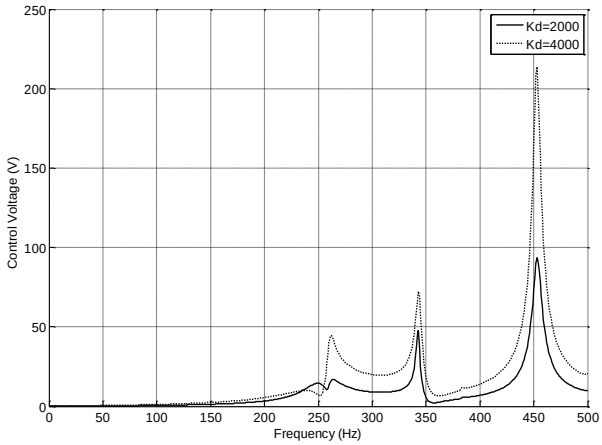


Fig 7: Control voltage for antisymmetric angle-ply ($-45^o/45^o/-45^o/45^o$) laminated shell with different gain

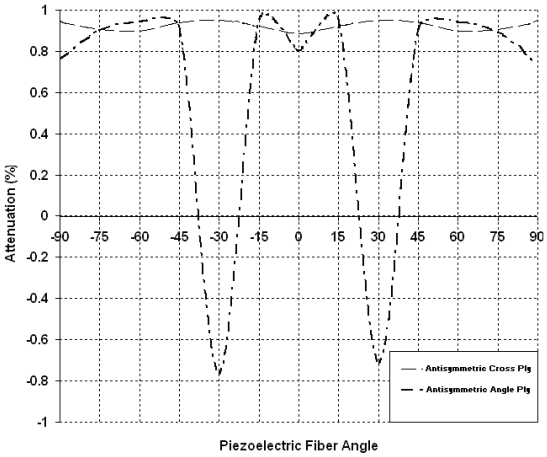


Fig 8: Effect of the fiber orientation in the constraining PFRC layer on the control authority of the ACLD patches for controlling the first mode of vibration of antisymmetric angle ply and cross ply shell

7. Conclusions:

A finite element analysis of active constrained layer damping (ACLD) of thin laminated circular cylindrical composite shell has been carried out to demonstrate the performance of the piezoelectric fiber reinforced composite (PFRC) materials as the material for the constraining layer of ACLD treatment. Both cross-ply and angle-ply laminated composite shell under antisymmetric case are considered for evaluation of the numerical results. Two patches of ACLD treatment are used which are placed on the outer surface of the shell such that the fundamental modes are efficiently controlled. The results illustrate the significant enhancement of damping characteristics of the shell over the passive damping.

The effect of variation of the fiber orientation in the constraining PFRC layer of the patches of the ACLD treatment on the attenuating capability of the patches has been investigated. If the piezoelectric fiber angle in the constraining PFRC layer of the ACLD treatment be 30^0 and 90^0 then the attenuating capability of the patches become maximum for controlling the first mode of vibrations of the cross ply composite cylindrical shell. Maximum attenuation achieved in cross ply composite cylindrical shell for 15^0 piezoelectric fiber angle in the constraining PFRC layer of the ACLD treatment.

8. References

[1] M C Ray and J N Reddy, “Optimal control of thin circular cylindrical laminated composite shells using active constrained layer damping treatment”, Smart Mater. Struct. 13 (2004) 64–72.

[2] M C Ray and J.N.Reddy, “Active Control of Laminated cylindrical shells using piezoelectric fiber reinforced composites”, Composites Science and Technology, vol.65, (2005), pp. 1226-1236.

[3] D.Chakravorty and P.K.Sinha, “Free vibration of point supported laminated composite doubly curved shells - A Finite Element Approach”, Computers and Structures, vol. 54, (1995), pp. 191-198.

[4] M. C. Ray and N. Mallik, “Effective coefficients of piezoelectric fiber reinforced composites”, AIAA journal, vol. 41, no A, pp. 704-710, 2003.

[5] M. C. Ray and N. Mallik, “Active control of laminated composite beams using a piezoelectric fiber reinforced composite layer”. Smart Materials and Structures 13 (1), 146–152, 2004.

[6] M. C. Ray and N. Mallik, “Finite element analysis of smart structures containing piezoelectric fiber reinforced composite actuator”. AIAA Journal 42 (7), 1398–1405, 2004.

[7] M. C. Ray and N. Mallik, “Exact solutions for the analysis of piezoelectric fiber reinforced composites as distributed actuators for smart composite plates”. International Journal of Mechanics and Materials in Design (2005) 2: 81–97.

[8] M. C. Ray, “Smart damping of laminated thin cylindrical panels using piezoelectric fiber reinforced composites”. International Journal of Solids and Structures 44 (2007) 587–602.

[10] M. C. Ray, J. Oh and A. Baz, “Active constrained layer damping of thin cylindrical shells”. Journal of sound and vibration (2001) 240(5), 921.935.

[11] Ray, M. C. and Sachade, H. M.,. "Exact Solutions for the functionally graded plates integrated with a layer of piezoelectric fiber-reinforced composite," accepted for publication in the ASME Journal of Applied Mechanics, 2005.

[12] Ootao, Y. and Tanigawa, Y.,2000,”Three dimensional transient thremoelasticity in functionally graded rectangular plate bonded to a piezoelectric plate”, International Journal of Solids and Structures, 37, pp. 4377-4401.

[13] J N Reddy and Z. Q. Cheng, 2001,” Three dimensional solutions of smart functionally graded plate”, ASME Journal of Applied Mechanics, 68, pp.234-241.

[14] Ray, M. C. and Sachade, H. M., 2005. "Finite Element Analysis of Smart functionally graded plates," accepted for publication in International Journal of Solids and Structures.

[15] Yang, J, Kitipomchi and Liew K M., 2003, "Large amplitude vibration of thermoelectro- mechanically stressed FGM laminated plates," Computer Methods in Applied Mechanics and Engineering, 192 (35), pp. 3861-3885.

9. Appendix:

$$[z_1]=\begin{bmatrix} z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & z & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$[z_2]=\begin{bmatrix} h/2 & 0 & 0 & (z-h/2) & 0 & 0 & 0 & 0 & 0 \\ 0 & h/2 & 0 & 0 & (z-h/2) & 0 & 0 & 0 & 0 \\ 0 & 0 & h/2 & 0 & 0 & (z-h/2) & 0 & 0 & 0 \end{bmatrix}$$

$$[z_3]=\begin{bmatrix} h/2 & 0 & 0 & (h_{N+1}-h/2) & 0 & 0 & (z-h_{N+2}) & 0 & 0 \\ 0 & h/2 & 0 & 0 & (h_{N+1}-h/2) & 0 & 0 & (z-h_{N+2}) & 0 \\ 0 & 0 & h/2 & 0 & 0 & (h_{N+1}-h/2) & 0 & 0 & (z-h_{N+2}) \end{bmatrix}$$

$$[z_4]=\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\\ 0 & (1-z/R) & 0 & 0 & 0 & 0\end{bmatrix}$$

$$[z_5]=\begin{bmatrix}0 & 0 & 1 & 0 & 0 & 0\\ 0 & -z/2R & 0 & \left(1-(z-h/2)/R\right) & 0 & 0\end{bmatrix}$$

$$[z_6]=\begin{bmatrix}0 & 0 & 0 & 0 & 1 & 0\\ 0 & -z/2R & 0 & -(z_{N+1}-h/2)/R & 0 & \left(1-(z-h_{N+2})/R\right)\end{bmatrix}$$

$$[B_{rbi}]=\begin{bmatrix}\frac{\partial n_i}{\partial x} & 0 & 0 & 0 & 0 & 0\\ 0 & \frac{\partial n_i}{\partial y} & 0 & 0 & 0 & 0\\ \frac{\partial n_i}{\partial y} & \frac{\partial n_i}{\partial x} & 0 & 0 & 0 & 0\\ 0 & 0 & \frac{\partial n_i}{\partial x} & 0 & 0 & 0\\ 0 & 0 & 0 & \frac{\partial n_i}{\partial y} & 0 & 0\\ 0 & 0 & \frac{\partial n_i}{\partial y} & \frac{\partial n_i}{\partial x} & 0 & 0\\ 0 & 0 & 0 & 0 & \frac{\partial n_i}{\partial x} & 0\\ 0 & 0 & 0 & 0 & 0 & \frac{\partial n_i}{\partial y}\\ 0 & 0 & 0 & 0 & \frac{\partial n_i}{\partial y} & \frac{\partial n_i}{\partial x}\end{bmatrix}$$

$$[B_{rsi}]=\begin{bmatrix}n_i & 0 & 0 & 0 & 0 & 0\\ 0 & n_i & 0 & 0 & 0 & 0\\ 0 & 0 & n_i & 0 & 0 & 0\\ 0 & 0 & 0 & n_i & 0 & 0\\ 0 & 0 & 0 & 0 & n_i & 0\\ 0 & 0 & 0 & 0 & 0 & n_i\end{bmatrix}.$$

$$[B_{tbi}]=\begin{bmatrix}\frac{\partial n_i}{\partial x} & 0 & 0\\ 0 & \frac{\partial n_i}{\partial y} & \frac{n_i}{R}\\ \frac{\partial n_i}{\partial y} & \frac{\partial n_i}{\partial x} & 0\end{bmatrix}$$

$$[B_{tsi}]=\begin{bmatrix}0 & 0 & \frac{\partial n_i}{\partial x}\\ 0 & -\frac{n_i}{R} & \frac{\partial n_i}{\partial y}\end{bmatrix}$$

Author Biographies

Ashok M Hulagabali born in Belgaum on 8th of January 1980. He completed his Bachelors degree in Engineering in Mechanical from MM Engg. College, Belgaum affiliated to Karnataka University, Dharwar in 2001, obtained his Masters degree in Mecanical Systems Design from IIT Kharagapur in 2008, and presently he is pursiing Ph.D at Vishweshwarayya Technological University, Belgaum. His areas of interest are Composite Mechanics, Smart materials, Dynamic analysis of composite structures etc.

Dr. J.Shivkumar born on 11th of October 1966 completed his Bachelors degree in Engineering in Mechanical from BDT college of Engineering, Davangere affiliated to Mysore University in 1989, obtained his Masters degree in Machine Design from BMS college of Engg. in 1996, and pursued Ph.D from IIT Kharagapur in 2008. His areas of interest are Applied Mechanics, Smart materials. He has published several research papers in International Journals, presented several papers in International and National conferences.

Rajesh Maji born on 12th of December 1980 obtained his Bachelors degree in Engineering from REC Durgapur, West Bengal, and obtained his Masters degree in Mecanical Systems Design from IIT Kharagapur in 2008 His areas of interest are Composite Materials and Smart materials

A Theoretical Study of Fast Motion Estimation Search Algorithms

S. H. D. S. Jitvinder, S. S. S. Ranjit, K. C. Lim S. I. MD Salim and A. J. Salim

Universiti Teknikal Malaysia Melaka, Faculty of Electronic and Computer Engineering,
Hang Tuah Jaya, 76100 Durian Tunggal,
Melaka, Malaysia.
jit_1986@yahoo.com, ranjit.singh@utem.edu.my, ranjit82@gmail.com

Abstract: A review and theoretical study about all the existing superior fast motion estimation algorithms has been carried out. Each algorithm has been study theoretically and is summarized to differentiate each algorithm in terms of their performances. Hence, the advantages and drawbacks of each algorithm are also presented for comparison purposes. In addition to that, the basic search function of each algorithm is discussed in this paper which includes the computational complexity, modeling of search points and elapsed processing time. The computational complexity, modeling of search points and elapsed processing time will describe performance analyze of each algorithm.

***Keywords*:** Motion Estimation, Search Points, Computational Complexity, Block-Based, Block Matching, Block Matching Motion.

1. Introduction

Motion estimation is important technique in video compression with an ability to reduce the temporal redundancy which exists in a sequence of video frames [1]. Temporal redundancy extracts the information from the video frame sequences during the motion estimation process whereby the motion information is represented by a motion vector coordinate [2]. There are few types of motion estimation techniques, block matching algorithm (BMA) technique has been widely adopted into the video coding standards (ITU-T, H.261, H.263, MPEG-1, MPEG-2 and MPEG-4) [1, 3, 4]. It is adopted due to the simplicity and easy hardware implementation [1, 3, 4]. BMA technique is apply to reduce the computational complexity during the video compression process is being executed. During the video compression process, each frame will be divided into small square size blocks of 16 × 16 pixels or 8 × 8 pixels [5]. The pixels in each small square size blocks of the current frame will be compared with the previous frame to search the matching motion vector coordinate in a particular block [6, 7]. There are numerous block matching algorithm has been developed which is based on block matching technique. Full Search Algorithm (FSA) is the high end and exhaustive search algorithm but requires high computational complexity during performing the motion vector search [1, 7]. Due to this drawback, various research have been conducted to develop fast block matching algorithms [1].

In this paper, all the superior fast block matching algorithms are reviewed. That superior block matching algorithms are Three Step Search (TSS), New Three Step Search (NTSS), Four Step Search (FSS) and Diamond Search (DS). This paper will explain on the algorithm methodology and algorithm process during motion estimation video compression is being conduct. The advantages and drawback of each algorithm will also be presented.

2. Motion Estimation Algorithm Review

a) Three Step Search Algorithm (TSS)

Three Step Search (TSS) algorithms were introduced by Koga et. al. in the year 1981 [8]. TSS is mainly developed for stationary video sequences. These videos are known as low bit rate video such as video conferencing and videophone [9-11]. TSS applies a center-biased motion vector technique to determine the global minimum motion vector[11]. This algorithm methodology applies a course-to-fine search approach which means the logarithm moves to the decreasing order in each search step. In each search step there is a step size which usually equal to or slightly larger than half of the maximum search range [9]. When an initial center point is picked up, there are eight search points surrounding the center point during the search methodology. Hence, the center search point and other eight search points are evaluated and during each step search the step size is reduced by half of each step size. This methodology is process until the search motion vector search stops at step size of one pixel [9]. This process also known as Unimodal Error Surface, where the search for matching block will moves away from the global minimum error [12].

The TSS algorithm is mainly used because of the simplicity and effectiveness (uniformly distributed search pattern) [1, 12]. TSS can be easily extended to n-steps depending on the search window (d) using the same strategy[1]. TSS is more efficient to find for those sequences with large motion[1]. The drawback for this algorithm is that it's inefficient for estimating small motions in the local minimum, thus reducing the accuracy dramatically of the motion estimation [1, 13]. This drawback is due to the pattern of search which is uniformly allocated which is inefficient for capturing small motion appearing in stationary or qusi-stationary blocks [10]. TSS also has many checking points in a situation where the motion is small for the block, thus, it is inefficient in terms of computation for images with low motion [14].

Figure 1 illustrates the search methodology of TSS algorithm. TSS algorithm starts with a 9 × 9 search window with the centre search point corresponding as zero Motion Vector

(MV). The initial step size when the search start is equal to 4 [15]. There will be 8 (dotted black) search point's surrounding the centre point [1, 10].

During the initial search methodology, each point is evaluated and the point that measured as the minimum distortion is at coordinate (-4,-4). A new search window will take place with a step size 2 [15]. Hence, the search window size is reduced to 5 × 5. Again each search point at new search window will be evaluated to detect the minimum distortion which is coordinated at (-4,-6). The search window will then reduced to 3 × 3 which is 1 step size for the final evaluation search point [15]. The final motion vector yield in the 3 × 3 search window at coordinated at (-5,-7) [10, 12].

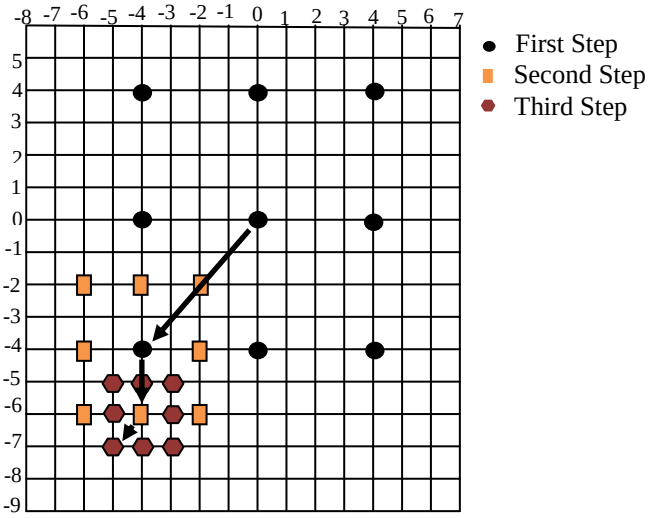


Figure 1. TSS Motion Vector Search Method

b) New Three Step Search Algorithm (NTSS)

Li, Zeng and Liou [10] proposed modified New Three Step Search algorithm which was modified from TSS algorithm [14]. The NTSS are highly centre-biased technique and applies the same characteristics search as TSS with two minor modifications in the search methodology. NTSS algorithm applies the halfway-stop technique with adding eight central search points during the first step of search [1, 10]. The halfway-stop technique is applied to improve the stationary and qusi-stationary block search [1, 10].

Figure 2 explains the search methodology of NTSS algorithm. NTSS algorithm consists nine initial TSS search points and eight additional search points are added (dark circles)[10]. The eight additional search points only will take place if the minimum distortion is determined at the centre of the search window [1, 10, 14]. If the minimum distortion point is evaluated at one of the eight neighbors search points in the first step, 3 or 5 search points will be evaluated (triangle) for minimum distortion search and then stop search [10, 14]. NTSS algorithm requires total of 33 search points (worst case) for a ±7 search window for motion displacement of while TSS algorithm only consists 25 search points [16].

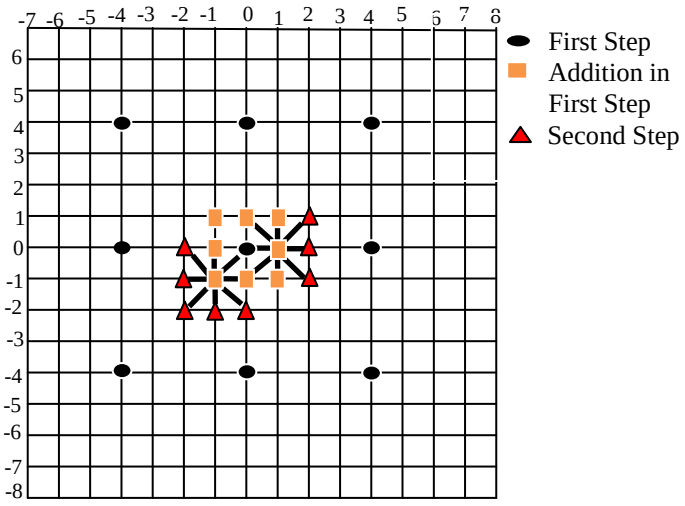


Figure 2. NTSS Motion Vector Search Method

NTSS algorithm improvement has resulted better performance compare with TSS algorithm in terms of number of search points and mean square error [1, 10, 17]. NTSS also perform search for small motions in an image, hence creates a drawback to check for unnecessary search point for big motion in an image [17]. The first-step-stop technique in NTSS reduces the computational complexity in terms of mathematical calculation for search points. The motion vector captured in the first-step-search will be the final motion vector and the search will discard to save the second-step-search points [10]. NTSS also produces almost the same result as Full Search (FS) algorithm and produce robust search compared with TSS algorithm during sub-sampling motion field [10]. NTSS is less accurate during global minimum search at central area search for halfway-stop technique [1].

c) Four Step Search Algorithm (FSS)

In the year 1996, L. M. Po and W. C. Ma proposed Four Step Search (FSS) algorithm which applies centre-biased motion vector distribution characteristic [12, 16, 17]. FSS algorithm is developed to exploit the real world video sequences characteristics. In initial step of FSS algorithm, a smaller step size of 5 × 5 in a search window [1, 12]. Thus, for ±7 search window FSS algorithm requires four steps to perform a search for motion vector coordinate [1]. FSS algorithm has similar technique as TSS algorithm but differs in terms of search steps and step size [18].

FSS algorithm has nine search points during its entire search [18] as illustrated in Figure 3. In the first step search 5 × 5 search window size is applied while reducing the search window size to 3 × 3 during fourth step search [12, 18]. Search window has a centre coordinate at location (0, 0) while each search point is evaluated and compared using the minimum Block Distortion Measure (BDM) [12]. The BDM measurement starts at the centre of search window during the first step search. The first step search will be reduced to 3 × 3 search step during the final search for motion vector distribution [12]. During the first search methodology, if the BDM is not located at the centre search window, then the search will proceed towards the minimum BDM point at

second and third step search [18]. According to Figure 3, centre point in second search step is located at the bottom left corner at coordinate (-2, -2). The second step search continues till the motion vector coordinate is determined at the centre of the search window. In continuing the search step for motion vector the motion vector is located at (-2, -4) before proceeding towards the final fourth search step. In the fourth search step the final motion vector coordinate is determine at coordinate (-4, -6). When the search take place for the final search step, the search window size is reduced and minimum BDM is located at coordinate (-5, -7).

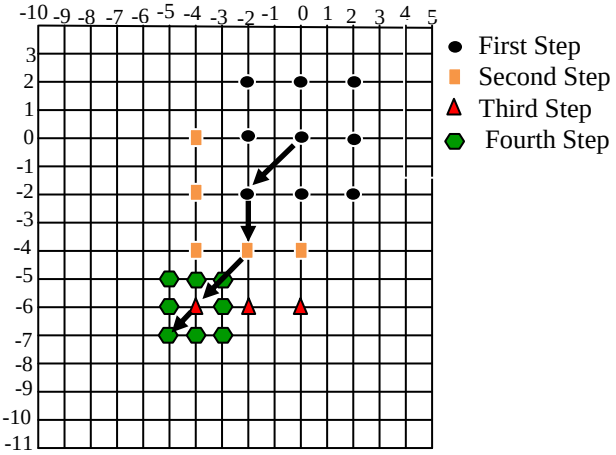


Figure 3. FSS Motion Vector Search Method

Experimental result for FSS algorithm has presented improvement in terms of performance compared with TSS algorithm. FSS algorithm produces similar performance compared with NTSS while improving the computational complexity in terms of fewer search points [1, 12, 17]. FSS algorithm performs checking on 27 search points (worst case) in a block instead of 33 search points in a block in NTSS algorithm [12, 17]. FSS algorithm reduces the computational complexity compared with TSS algorithm because of the overlapping search points [16, 19] during the motion vector coordinate search. FSS algorithm produces robust performances compared with TSS and NTSS because FSS algorithm is used for complex movement in image sequences [18]. FSS algorithm reduces the search points when half-way stop technique is applied in the second and third search step. The half-way search step helps to reduce the search points while improves the FSS algorithm performances [18, 19]. FSS algorithm uses a small search window which performs the search precisely to estimate the motion vector in small motion region [18].

d) Diamond Search Algorithm (DS)

In year 2000, Zhu Shan and K. K. Ma proposed Diamond Search (DS) algorithm [19]. This algorithm applies the center biased ideology which models the search shape as diamond shape search pattern [20]. The diamond search shape pattern methodology is implemented as an unrestricted search step [21]. In the search implementation, DS algorithm models two different sizes of diamond pattern. The two different diamond patterns modeling are categories to large diamond search pattern (LDSP) and small diamond search pattern (SDSP) as depicted in figure 4. The LDSP consist one centre searching point which surrounded with eight searching points [9, 19]. The SDSP model consists of four searching point surrounded with one searching point at the centre location [9, 19].

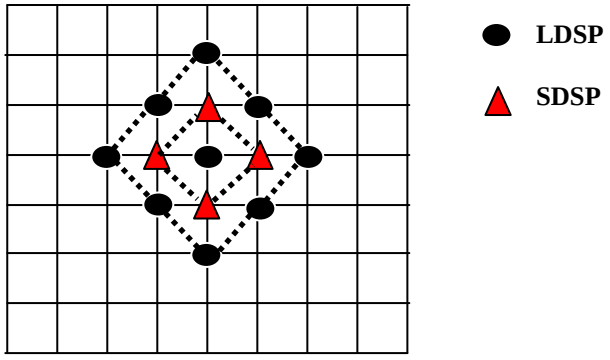


Figure 4. LDSP and SDSP

LDSP performs a repeated throughout search till the minimum BDM is found to be at the centre location of LDSP [19, 20, 22]. During the BDM at the center, SDSP search will take place and yield the MDB for final motion vector coordinate [19].

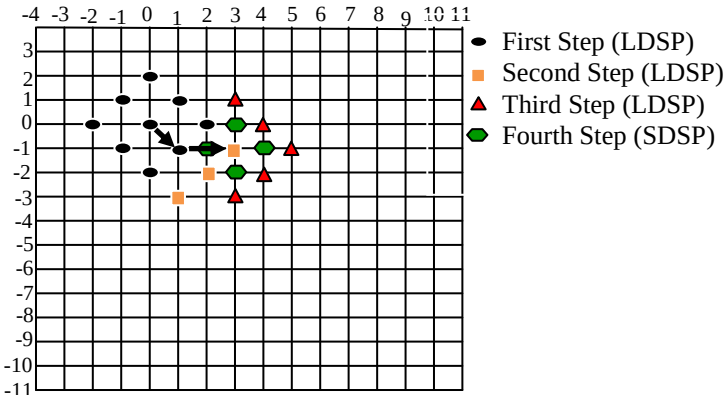


Figure 5. DS Motion Vector Search Method

Figure 5 illustrates the search methodology of DS algorithm. DS algorithm starts it search with the LDSP at location (0, 0) [19]. The DS algorithm search modeling will continue to search for motion vector till the new center search point located at (1, 1). It will then continue to the third search point with the center search point is at coordinate (3, -1) because the minimum BDM is located at the center point of the third search. Once the minimum BDM is found at center point of LDSP, SDSP search will take place to yield the five search points for the final motion vector [9, 19].

DS algorithm has an advantage during performing LDSP modeling. The overlapping search methodology among the nine search points in LDSP helps to reduce the computational complexity in search for motion vector[19]. DS algorithm have two one pel at four diagonal direction and two pels in horizontal and vertical direction in the LDSP modeling. The diagonal, horizontal and vertical pels allows DS algorithm not be easily trapped into a local minimum point for large motion in a block [19]. The compact LDSP search in DS

algorithm increases the possibilities of locating the global minimum BDM in the LDSP search step[19]. The compact LDSP search also help to produce a similar or smaller minimum BDM compared with other fast block matching algorithm [19].

3. Discussion

This paper discusses about fast motion estimation search algorithms that are used for low bit rate video application. Most of the fast motion estimation search algorithms studied is developed for slow, moderate and fast motion estimation. Each algorithm development process associated with improvement and all the processes increase the efficiency of each algorithm. Studies conducted shows TSS algorithm is inefficient for estimating small motion. Due to this drawback NTSS algorithm was developed to improve the motion estimation search for small motion in images. Even though NTSS algorithm improved for small motion estimation search but NTSS algorithm performs unnecessary search point checking on fast motion images. FSS algorithm produces good search for small motion due to its small window search size. It also has the similar performances to NTSS algorithm. The DS algorithm is developed to reduce the computational complexity while applying the overlapping search point technique.

4. Conclusion

In this paper, the fast motion estimation search algorithm which is the TSS, NTSS, FSS and DS algorithms are discussed. The methodology of each algorithm is described step by step to estimate the motion vector for minimum Block Distortion Measure. Based on each algorithm description, the advantages and drawbacks of each algorithm are discussed briefly.

References

[1] L. P. C., X.J., *An Efficient Three-Step Search Algorithm for Block Motion Estimation.* IEEE Transactions on Multimedia, June 2004. Vol. 6, No. 3.

[2] Phadtare, M., *Motion Estimation Techniques in Video Processing.* Electronic Engineering Times India, August 2007.

[3] C. B. D. Xu, R.S., *An Improved Three-Step Search Block-Matching Algorithm for Low Bit-Rate Video Coding Applications.* URSI International Symposium on Signals, Systems and Electronics (ISSSE), 1998.

[4] Y. S.-y. Luo Tao, S.Z.-f., Gao Peng, *An Improved Three-Step Search Algorithm with Zero Detection and Vector Filter for Motion Estimation.* International Conference on Computer Science and Software Engineering, 2008.

[5] W. Z. Ishfaq Ahmad, J.L., Ming Liou, *A Fast Adaptive Motion Estimation Algorithm.* IEEE Transactions On Circuits and Systems for Video Technology, March 2006. Vol.16.

[6] M. Ezhilarasan, P.T., *Simplified Block Matching Algorithm for Fast Motion Estimation.* Journal of Computer Science 4 (4), pp. 282-289, 2008.

[7] S. S. V.S.Kumar Reddy, Y.M., *A Fast and Efficient Predictive Block Matching Motion Estimation.* IJCSNS International Journal of Computer Science and Network Security, December 2007. Vol. 7.

[8] T. Koga, K.I., A. Hirano, Y. Iijima and T. Ishiguro, *Motion Compensated Interframe Coding For Video Conferencing.* Pro. Nat. Telecommun. Conf., New Orleans, , Nov. 1981: p. pp. G5.3.1- 5.3.5.

[9] A. Samet, N.S., W. Zouch, M. A. Ben Ayed, N. Masmoudi, *New Horizontal Diamond Search Motion Estimation Algorithm For H.264/AVC.* UBVideo Inc., Vancouver, Canada, October 27, 2005.

[10] Renxiang Li, B.Z., Ming L. Liou, *A New Three-Step Search Algorithm for Block Motion Estimation.* IEEE Transactions On Circuits And Systems For Video Technology, August 1994. Vol. 4, No. 4.

[11] Donglai Xu, C.B., Reza Sotudeh, *An Improved Three-Step Search Block-Matching Algorithm for Low Bit-Rate Video Coding Applications.* Computer Architecture Research Unit, School of Science and Technology University of Teesside, Middlesbrough, Cleveland, TS 1 3BA, United Kingdom.

[12] Lai-Man Po, W.-C.M., *A Novel Four-Step Search Algorithm for Fast Block Motion Estimation.* IEEE Transactions On Circuits And Systems For Video Technology, June 1996. Vol. 6, No. 3.

[13] Her-Ming Jong, L.G.C., Tzi-Dar Chiueh, *Modifications and Performance Improvements of 3-Step Search Block-Matching Algorithm for Video Coding.* International Symposium on Speech, Image Processing and Neural Networks,Hong Kong, April 1994.

[14] Jianhua Lu, M.L.L., *A Simple and Efficient Search Algorithm for Block-Matching Motion Estimation.* IEEE Transactions on Circuits and Systems for Video Technology, April 1997. Vol. 7, No. 2.

[15] Luo Tao, Y.S.-y., Shi Zai-feng, Gao peng, *An Improved Three-Step Search Algorithm with Zero Detection and Vector Filter for Motion Estimation.* International Conference on Computer Science and Software Engineering, 2008.

[16] Duanmu, C.J., *Fast Scheme for the Four-Step Search Algorithm in Video Coding.* IEEE International Conference on Systems, Man, and Cybernetics , Taipei, Taiwan, October 8-11, 2006.

[17] Jong-Nam Kim, T.-S.C., *A Fast Three-Step Search Algorithm with Minimum Checking Points Using Unimodal Error Surface Assumption.* IEEE Transactions on Consumer Electronics, August 1998. Vol. 44, No.3.

[18] Angus Wu, S.S., *VLSI Implementation of Genetic Four-Step Search for Block Matching Algorithm.* IEEE Transactions on Consumer Electronics, November 2003. Vol. 49, No. 4.

[19] Shan Zhu, K.-K.M., *A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation.* IEEE Transactions on Image Processing, February 2000. Vol. 9, No. 2.

[20] Xiaoquan Yi, N.L., *Rapid Block-Matching Motion Estimation Using Modified Diamond Search Algorithm.* IEEE International Symposium in Circuits and Systems (ISCAS), May 2005. Vol. 6: p. pg 5489 - 5492

[21] Chi-Wai Lam, L.-M.P., Chun Ho Cheung, *A New Cross-Diamond Search Algorithm for Fast Block Matching Motion Estimation.* Department of Electronic

Engineering City University of Hong Kong and Department of Information Technology, Hong Kong Institute of Technology, Hong Kong SAR.

[22] A.Ahmadi, M.M.A., *Implementation of Fast Motion Estimation Algorithms and Comparison with Full Search Method in H.264*. International Journal of Computer Science and Network Security (IJCSNS), March 2008. Vol.8 No.3.

ijltc.excelingtech.co.uk