



# IJCSI

## **International Journal of Computer Science Issues**

**Volume 4, Issue 2, September 2009**  
**ISSN (Online): 1694-0784**  
**ISSN (Print): 1694-0814**

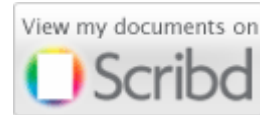
**© IJCSI PUBLICATION**  
**[www.IJCSI.org](http://www.IJCSI.org)**

**IJCSI proceedings are currently indexed by:**



**Cogprints**

**Google scholar**



**SciRate.com**

**CiteSeer<sup>x</sup><sub>beta</sub>**

**© IJCSI PUBLICATION 2009**

**www.IJCSI.org**

## **EDITORIAL**

There are several journals available in the areas of Computer Science having different policies. IJCSI is among the few of those who believe giving free access to scientific results will help in advancing computer science research and help the fellow scientist.

IJCSI pay particular care in ensuring wide dissemination of its authors' works. Apart from being indexed in other databases (Google Scholar, Cornell's University Library, ScientificCommons, CiteSeerX, etc...), IJCSI makes articles available to be downloaded for free to increase the chance of the latter to be cited.

Furthermore, unlike most journals, IJCSI send a printed copy of its issue to the concerned authors free of charge irrespective of geographic location.

IJCSI Editorial Board is pleased to present IJCSI Volume Four Issue Two (IJCSI Vol. 4, No. 2, 2009). The paper acceptance rate for this volume is 29.55%; set after all submitted papers have been received with important comments and recommendations from our reviewers.

We sincerely hope you would find important ideas, concepts, techniques, or results in this special issue.

As final words, PUBLISH, GET CITED and MAKE AN IMPACT.

IJCSI Editorial Board  
September 2009  
[www.ijcsi.org](http://www.ijcsi.org)

## **IJCSI Editorial Board 2009**

### **Dr Tristan Vanrullen**

Chief Editor

LPL, Laboratoire Parole et Langage - CNRS - Aix en Provence, France

LABRI, Laboratoire Bordelais de Recherche en Informatique - INRIA - Bordeaux, France

LEEE, Laboratoire d'Esthétique et Expérimentations de l'Espace - Université d'Auvergne, France

### **Dr Constantino Malagón**

Associate Professor

Nebrija University

Spain

### **Dr Mokhtar Beldjehem**

Professor

Sainte-Anne University

Halifax, NS, Canada

### **Dr Pascal Chatonnay**

Assistant Professor

Maître de Conférences

Université de Franche-Comté (University of French-County)

Laboratoire d'informatique de l'université de Franche-Comté, (Computer Science Laboratory of University of French-County), France

### **Dr Deepak Laxmi Narasimha**

Department of Software Engineering,

Faculty of Computer Science and Information Technology,

University of Malaya,

Kuala Lumpur, Malaysia

### **Prof N. Jaisankar**

School of Computing Sciences,

VIT University

Vellore, Tamilnadu, India

## **IJCSI Reviewers Committee 2009**

- Mr. Markus Schatten, University of Zagreb, Faculty of Organization and Informatics, Croatia
- Mr. Vassilis Papataxiarhis, Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimiopolis, Ilissia, GR-15784, Athens, Greece, Greece
- Dr Modestos Stavrakis, University of the Aegean, Greece
- Dr Fadi KHALIL, LAAS -- CNRS Laboratory, France
- Dr Dimitar Trajanov, Faculty of Electrical Engineering and Information technologies, ss. Cyril and Methodius University - Skopje, Macedonia
- Dr Jinping Yuan, College of Information System and Management, National Univ. of Defense Tech., China
- Dr Alexis Lazanas, Ministry of Education, Greece
- Dr Stavroula Mouggiakakou, University of Bern, ARTORG Center for Biomedical Engineering Research, Switzerland
- Dr DE RUNZ, CRÉSTIC-SIC, IUT de Reims, University of Reims, France
- Mr. Pramodkumar P. Gupta, Dept of Bioinformatics, Dr D Y Patil University, India
- Dr Alireza Fereidunian, School of ECE, University of Tehran, Iran
- Mr. Fred Viezens, Otto-Von-Guericke-University Magdeburg, Germany
- Mr. J. Caleb Goodwin, University of Texas at Houston: Health Science Center, USA
- Dr. Richard G. Bush, Lawrence Technological University, United States
- Dr. Ola Osunkoya, Information Security Architect, USA
- Mr. Kotsokostas N. Antonios, TEI Piraeus, Hellas
- Prof Steven Totosy de Zepetnek, U of Halle-Wittenberg & Purdue U & National Sun Yat-sen U, Germany, USA, Taiwan
- Mr. M Arif Siddiqui, Najran University, Saudi Arabia
- Ms. Ilknur Icke, The Graduate Center, City University of New York, USA
- Prof Miroslav Baca, Associated Professor/Faculty of Organization and Informatics/University of Zagreb, Croatia
- Dr. Elvia Ruiz Beltrán, Instituto Tecnológico de Aguascalientes, Mexico
- Mr. Moustafa Banbouk, Engineer du Telecom, UAE
- Mr. Kevin P. Monaghan, Wayne State University, Detroit, Michigan, USA
- Ms. Moira Stephens, University of Sydney, Australia
- Ms. Maryam Feily, National Advanced IPv6 Centre of Excellence (NAV6) , Universiti Sains Malaysia (USM), Malaysia
- Dr. Constantine YIALOURIS, Informatics Laboratory Agricultural University of Athens, Greece
- Mrs. Angeles Abella, U. de Montreal, Canada
- Dr. Patrizio Arrigo, CNR ISMAC, Italy
- Mr. Anirban Mukhopadhyay, B.P. Poddar Institute of Management & Technology, India
- Mr. Dinesh Kumar, DAV Institute of Engineering & Technology, India
- Mr. Jorge L. Hernandez-Ardieta, INDRA SISTEMAS / University Carlos III of Madrid, Spain

- Mr. AliReza Shahrestani, University of Malaya (UM), National Advanced IPv6 Centre of Excellence (NAV6), Malaysia
- Mr. Blagoj Ristevski, Faculty of Administration and Information Systems Management - Bitola, Republic of Macedonia
- Mr. Mauricio Egidio Cantão, Department of Computer Science / University of São Paulo, Brazil
- Mr. Jules Ruis, Fractal Consultancy, The Netherlands
- Mr. Mohammad Iftekhar Husain, University at Buffalo, USA
- Dr. Deepak Laxmi Narasimha, Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
- Dr. Paola Di Maio, DMEM University of Strathclyde, UK
- Dr. Bhanu Pratap Singh, Institute of Instrumentation Engineering, Kurukshetra University Kurukshetra, India
- Mr. Sana Ullah, Inha University, South Korea
- Mr. Cornelis Pieter Pieters, Condast, The Netherlands
- Dr. Amogh Kavimandan, The MathWorks Inc., USA
- Dr. Zhinan Zhou, Samsung Telecommunications America, USA
- Mr. Alberto de Santos Sierra, Universidad Politécnica de Madrid, Spain
- Dr. Md. Atiqur Rahman Ahad, Department of Applied Physics, Electronics & Communication Engineering (APECE), University of Dhaka, Bangladesh
- Dr. Charalampos Bratsas, Lab of Medical Informatics, Medical Faculty, Aristotle University, Thessaloniki, Greece
- Ms. Alexia Dini Kounoudes, Cyprus University of Technology, Cyprus
- Mr. Anthony Gesase, University of Dar es Salaam Computing Centre, Tanzania
- Dr. Jorge A. Ruiz-Vanoye, Universidad Juárez Autónoma de Tabasco, Mexico
- Dr. Alejandro Fuentes Penna, Universidad Popular Autónoma del Estado de Puebla, México
- Dr. Ocotlán Díaz-Parra, Universidad Juárez Autónoma de Tabasco, México
- Mrs. Nantia Iakovidou, Aristotle University of Thessaloniki, Greece
- Mr. Vinay Chopra, DAV Institute of Engineering & Technology, Jalandhar
- Ms. Carmen Lastres, Universidad Politécnica de Madrid - Centre for Smart Environments, Spain
- Dr. Sanja Lazarova-Molnar, United Arab Emirates University, UAE
- Mr. Srikrishna Nudurumati, Imaging & Printing Group R&D Hub, Hewlett-Packard, India
- Dr. Olivier Nocent, CReSTIC/SIC, University of Reims, France
- Mr. Burak Cizmeci, Isik University, Turkey
- Dr. Carlos Jaime Barrios Hernandez, LIG (Laboratory Of Informatics of Grenoble), France
- Mr. Md. Rabiul Islam, Rajshahi university of Engineering & Technology (RUET), Bangladesh
- Dr. LAKHOUA Mohamed Najeh, ISSAT - Laboratory of Analysis and Control of Systems, Tunisia
- Dr. Alessandro Lavacchi, Department of Chemistry - University of Firenze, Italy
- Mr. Mungwe, University of Oldenburg, Germany
- Mr. Somnath Tagore, Dr D Y Patil University, India
- Ms. Xueqin Wang, ATCS, USA
- Dr. Fondjo Fotou Franklin, Langston University, USA

- Mr. Haytham Mohtasseb, Department of Computing - University of Lincoln, United Kingdom
- Dr. Vishal Goyal, Department of Computer Science, Punjabi University, Patiala, India
- Mr. Thomas J. Clancy, ACM, United States
- Dr. Ahmed Nabih Zaki Rashed, Dr. in Electronic Engineering, Faculty of Electronic Engineering, menouf 32951, Electronics and Electrical Communication Engineering Department, Menoufia university, EGYPT, EGYPT
- Dr. Rashed Kanawati, LIPN, France
- Mr. Koteswar Rao, K G Reddy College Of ENGG.&TECH,CHILKUR, RR DIST.,AP, India
- Mr. M. Nagesh Kumar, Department of Electronics and Communication, J.S.S. research foundation, Mysore University, Mysore-6, India
- Mr. Saqib Saeed, University of Siegen, Germany
- Dr. Ibrahim Noha, Grenoble Informatics Laboratory, France
- Mr. Muhammad Yasir Qadri, University of Essex, UK
- Mr. Annadurai .P, KMCPGS, Lawspet, Pondicherry, India, (Aff. Pondicherry Univeristy, India
- Mr. E Munivel , CEDTI (Govt. of India), India
- Dr. Chitra Ganesh Desai, University of Pune, India
- Mr. Syed, Analytical Services & Materials, Inc., USA
- Dr. Mashud Kabir, Department of Computer Science, University of Tuebingen, Germany
- Mrs. Payal N. Raj, Veer South Gujarat University, India
- Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal, India
- Mr. Mahesh Goyani, S.P. University, India, India
- Mr. Vinay Verma, Defence Avionics Research Establishment, DRDO, India
- Dr. George A. Papakostas, Democritus University of Thrace, Greece
- Mr. Abhijit Sanjiv Kulkarni, DARE, DRDO, India
- Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
- Dr. B. Sivaselvan, Indian Institute of Information Technology, Design & Manufacturing, Kancheepuram, IIT Madras Campus, India
- Dr. Partha Pratim Bhattacharya, Greater Kolkata College of Engineering and Management, West Bengal University of Technology, India
- Mr. Manish Maheshwari, Makhanlal C University of Journalism & Communication, India
- Dr. Siddhartha Kumar Khaitan, Iowa State University, USA
- Dr. Mandhapati Raju, General Motors Inc, USA
- Dr. M.Iqbal Saripan, Universiti Putra Malaysia, Malaysia
- Mr. Ahmad Shukri Mohd Noor, University Malaysia Terengganu, Malaysia
- Mr. Selvakuberan K, TATA Consultancy Services, India
- Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India
- Mr. Rakesh Kachroo, Tata Consultancy Services, India
- Mr. Raman Kumar, National Institute of Technology, Jalandhar, Punjab., India
- Mr. Nitesh Sureja, S.P.University, India
- Dr. M. Emre Celebi, Louisiana State University, Shreveport, USA
- Dr. Aung Kyaw Oo, Defence Services Academy, Myanmar
- Mr. Sanjay P. Patel, Sankalchand Patel College of Engineering, Visnagar, Gujarat, India
- Dr. Pascal Fallavollita, Queens University, Canada

- Mr. Jitendra Agrawal, Rajiv Gandhi Technological University, Bhopal, MP, India
- Mr. Ismael Rafael Ponce Medellín, Cenidet (Centro Nacional de Investigación y Desarrollo Tecnológico), Mexico
- Mr. Supheakmungkol SARIN, Waseda University, Japan
- Mr. Shoukat Ullah, Govt. Post Graduate College Bannu, Pakistan
- Dr. Vivian Augustine, Telecom Zimbabwe, Zimbabwe
- Mrs. Mutalli Vatila, Offshore Business Philipines, Philipines
- Dr. Emanuele Goldoni, University of Pavia, Dept. of Electronics, TLC & Networking Lab, Italy
- Mr. Pankaj Kumar, SAMA, India
- Dr. Himanshu Aggarwal, Punjabi University, Patiala, India
- Dr. Vauvert Guillaume, Europages, France
- Prof Yee Ming Chen, Department of Industrial Engineering and Management, Yuan Ze University, Taiwan
- Dr. Constantino Malagón, Nebrija University, Spain
- Prof Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India
- Mr. Angkoon Phinyomark, Prince of Singkla University, Thailand
- Ms. Nital H. Mistry, Veer Narmad South Gujarat University, Surat, India
- Dr. M.R.Sumalatha, Anna University, India
- Mr. Somesh Kumar Dewangan, Disha Institute of Management and Technology, India
- Mr. Raman Maini, Punjabi University, Patiala(Punjab)-147002, India
- Dr. Abdelkader Outtagarts, Alcatel-Lucent Bell-Labs, France
- Prof Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India
- Mr. Prabu Mohandas, Anna University/Adhiyamaan College of Engineering, india
- Dr. Manish Kumar Jindal, Panjab University Regional Centre, Muktsar, India
- Prof Mydhili K Nair, M S Ramaiah Institute of Technnology, Bangalore, India
- Dr. C. Suresh Gnana Dhas, VelTech MultiTech Dr.Rangarajan Dr.Sagunthala Engineering College, Chennai, Tamilnadu, India
- Prof Akash Rajak, Krishna Institute of Engineering and Technology, Ghaziabad, India
- Dr. Vu Thanh Nguyen, University of Information Technology HoChiMinh City, VietNam

# **TABLE OF CONTENTS**

**1. Password Based a Generalize Robust Security System Design Using Neural Network -**  
(pg 1-9)

Manoj Kumar Singh, Manuro Tech. Research Lab, Bangalore, India

**2. 3D/2D Registration of Mapping Catheter Images for Arrhythmia Interventional Assistance –** (pg 10-19)

Pascal Fallavollita, School of Computing, Queens University, Kingston, Ontario, Canada

**3. Parallel Computation of Finite Element Navier-Stokes codes using MUMPS Solver -**  
(pg 20-24)

Mandhapati P. Raju, Mechanical Engineering, Case Western Reserve University, Cleveland, Ohio 44106

**4. The Folklore of Sorting Algorithms –** (pg 25-30)

Santosh Khamitkar, Parag Bhalchandra, Sakharam Lokhande and Nilesh Deshmukh, School of Computational Sciences, Swami Ramanand Teerth Marathwada University, Nanded (MS ) 431605, India

**5. Color Image Clustering using Block Truncation Algorithm –** (pg 31-35)

Sanjay Silakari, Department of Computer Science and Engineering, Rajiv Gandhi Technical University, Bhopal, M.P., India

Mahesh Motwani, Department of Computer Science and Engineering, Jabalpur Engineering College, Jabalpur, M.P., India

Manish Maheshwari, Department of Computer Science & Application, Makhanlal C. National University of Journalism and Communication, Bhopal, M.P., India

**6. DAMQ-Based Schemes for Efficiently Using the Buffer Spaces of a NoC Router –** (pg 36-41)

Mohammad Ali Jabraeil Jamali and Ahmad Khademzadeh, Department of Computer Science, Islamic Azad University, Science & Research Branch, Tehran, Iran

**7. Distributed Object Medical Imaging Model –** (pg 42-48)

Ahmad Shukri Mohd Noor and Md Yazid Md Saman, Department of Computer Science, Faculty of Science and Technology, University Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia

# PASSWORD BASED A GENERALIZE ROBUST SECURITY SYSTEM DESIGN USING NEURAL NETWORK

Manoj Kumar Singh  
Manuro Tech. Research Lab, Bangalore, India  
[mksingh@manuroresearch.com](mailto:mksingh@manuroresearch.com)

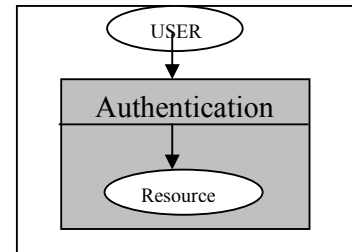
## Abstract

Among the various means of available resource protection including biometrics, password based system is most simple, user friendly, cost effective and commonly used. But this method having high sensitivity with attacks. Most of the advanced methods for authentication based on password encrypt the contents of password before storing or transmitting in physical domain. But all conventional cryptographic based encryption methods are having its own limitations, generally either in terms of complexity or in terms of efficiency. Multi-application usability of password today forcing users to have a proper memory aids. Which itself degrades the level of security. In this paper a method to exploit the artificial neural network to develop the more secure means of authentication, which is more efficient in providing the authentication, at the same time simple in design, has given. Apart from protection, a step toward perfect security has taken by adding the feature of intruder detection along with the protection system. This is possible by analysis of several logical parameters associated with the user activities. A new method of designing the security system centrally based on neural network with intrusion detection capability to handles the challenges available with present solutions, for any kind of resource has presented.

*Keywords:* artificial neural network, security, authentication, intrusion detection.

## 1. Introduction

Security is a broad topic and covers many issues. Malicious people trying to gain some benefit, attention, or to harm someone intentionally cause most security problems. As, Complete security not possible in real life, transition will be long in coming. Conventional cryptographic methods have their own problems. Intrusion Detection System (IDS) is an emerging new technology, being informed is the best weapon in the security analyst's arsenal. "An ounce of prevention is worth a pound of detection". An Intrusion Detection System detects attacks as soon as possible and takes appropriate action. Security is a compulsory need for data operation today. Information or commerce exchanges need security and reliability. Examples are like: Bank Transactions where state-of-the art financial security is mandatory, Protection of Personal Resources etc. password based security system should must posses facilities like: (i) provision to assign complex password without any restriction. (ii) Password must be encrypted up to proper level before storing or transmitting. (iii) Easy and secure means to reset the password. (iv) Facility to record and monitor failed login attempt. (v) Don't let people try to break password in forever. General structure of security system is shown in fig. (1), to access the resource user must verify by authentication environment.



**Figure 1. Security System in Abstract form**

Process of authentication can be defined as developing a unique mapping process from given password to some other unique information in a defined domain. The guarantee of security doesn't only depend on unique mapping but greatly depend upon difficulties associated with getting back the password form the mapped formation. Artificial Neural network characteristics, making it very appropriate for playing an important role in coming era of security design. Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by examples. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well.

## 2. Literature Review

In order for a computer system to perform specific acts on behalf of a specific individual, an identification and authentication step is needed. This process of identification and authentication has been extensively studied and formally described in principle [1], [2]. These formal descriptions include the user as the principal and document all relevant constructs and issues from a single system perspective. In the basic authentication process, the entity desiring authentication presents credentials usually an account ID and some additional information, to prove that the request is coming from a legitimate owner of the ID. This is relatively

straightforward process that has been in use for decades. An example is user ID and password combination, one of the simplest forms of user authentication [3]-[6]. A more complicated example is the smart card system [7], [8], where a user typically has an ID, a password and also a dynamic (time-generating) passkey from the smart card which changes every 60 seconds. The authenticating server has the same time changing numerical sequence as the specific smart-cards assigned to that ID and if the ID, password and card generated number are all correct, authentication is granted. Frequently smart-card are combined with passwords for an account to increase security. This is an example of two-factor authentication and is more secure because it requires more items for authentication. Another form of authentication involves the concept of biometrics. Biometrics can take the form of several measurements, from fingerprints to retinal scans to pupil images. The advantage of biometrics is that always available with user without any extra means and effort and unforgotten. Disadvantages are many, including not being able to change them if needed and complexity of solution. Many modern system have adopted a simple ID/password method of achieving the goals associated with the identification and authentication function, and numerous technical method exists to achieve this [3],[4],[7],[9],[15].the wide Variety of implementation schemes are a result of individual design decisions appropriate to specific circumstances at the time to design of a specific system or class of a system. Important research on human cognitive ability has generated a lot of practical knowledge on the issue of what an individual can remember [16]. The effect of human cognitive ability in the authentication process is a central element, through often overlooked by developers. Remembrance of passwords is one of the cornerstones of the current password based authentication system [17],[18],[24].From system perspective a password should be easily remembered, yet hard for an intruder to guess[19]-[21].Although other system level solutions exist much of the effort to secure password-based system is focus on preventing unauthorized access through better password selection[5],[13],[29],[30].Many known weaknesses exists in password-based system and various fixes have been applied over time[4],[9],[10],[11],[12],[14].

According to the learning ability, artificial neural network has been used in artificial intelligence [31], neural network can be used to model nonlinear statistical data, which can model complex relationship between inputs and outputs. Sometimes, it can generate chaos phenomenon. According to this property, it has complex dynamic action, which can be used to protect data content. For example, the random sequence produced by neural network can be used to encrypt data [38], and the neural networks that generate chaos phenomenon can be used in secret communication [39]. According to complex relation between the nodes of neural network, neural network can produce the sequences with random properties. For the neural network that has chaotic dynamics, its output is often sensitive to the inputs or such control parameter as weight; it is caused by parameter sensitivity of chaos system. Taking the neural network proposed in [37] for example, although there is a slight difference in the initial value, the output changes greatly. This property makes the initial value suitable for the key that

controls the data encryption or decryption. Application of neural network for intrusion detection has been shown in [36]. Intrusion detection is an important technology in network security, which can detect illegal intruders or illegal intrusions. Using neural network's supervised learning, the intrusive operation can be distinguished from normal operation [37]. The one-way property makes neural network a suitable choice for hash function. Hash function is a technique for data integrity authentication. Till now some hash functions based on neural network have been presented [38][39], which were reported to have some advantages compared with existing schemes, such as high time efficiency or flexible extension.

### **3. Risk and complications associated with password authentication**

#### **3.1 Attacks on password**

Because of high sensitivity with attacks, risk factor is very high in password based authentication process. Broadly the types of attack can be divided into three categories: - technical (brute force), discovery and social engineering.

In the brute force attack, two methods can be used, (a) attempting passwords against the system, but this is easily stopped with account lockouts. (b) An offline attack against the password hash file. This is processor intensive search through the entire password key-space, calculating and comparing hash values of potential passwords to the values in the stolen hash file. Various defense exist, including increasing key-space through the use of salts, and physically protecting the password hash file.

Password may also be compromised by discovery. Forms of password discovery may vary and include interception of a script file, an exploit on another system, a Trojan program capturing keystrokes, or the discovery of default passwords associated with other system or programs. The primary defense against discovery is proper system design rules that do not allow discovery of passwords through scripts or default system accounts.

Social engineering represents an attempt by an intruder to elicit password and account information from a user. This attack is exogenous to the computer system in question, coming via phone, fax, email or causal contact. The primary defense against this type of exploit is training and awareness directed at the user with respect to this specific vulnerability.

#### **3.2 Complications with safeguards**

Password rules are either optional or enforced specification about the length of the password and the diversity of the characters that comprise its' the length and diversity contribute to the size of the domain set containing all passwords that increase the difficulty of brute force detection. Prevention of easily guessed passwords reduce discovery. However the same rules that increase password resistance to brute force attack directly reduce discovery. However the same rules that increase password resistance to brute force attack directly reduce the ability of a user to remember a password and increase the need of password aids. System rules relate to the procedural aspects of gaining access and are enabled in a system. For example, the

automatic user lockout after three failed trailed attempts is a system-enforced rule. More complex mechanism includes expiring passwords and the forcing of password changes. System rule can also have the opposite effect through, as they can lead to discovery patterns. There are systems that will email an encrypted password back to user if required presenting an opportunity for discovery. System rule that make the passwords harder to remember can increase the need for user based password memory aids system rules for recovering forgotten password such as emailing forgotten passwords lessen the need for memory aids but increase risk of discovery. In general rules that enforce higher quality passwords do so at the expense of user memory ability. This was an acceptable trade off when the count of passwords per user was low but this trade off today forces users to use memory aids. One of the weakest links in a security system is an untrained user. Formal and informal activities of training and awareness can alleviate a wide Varsity of actions that weaken a system such as choosing poor passwords, writing them down, sharing with others. Training can address issues associated with discovery and social engineering attackers. The primary issue with training is its temporary nature, users forget or become complacent over time and retraining is time consuming and costly.

### 3.3 Complications because of multi-usability

Today users have multiple accounts on multiple systems. users must to remember multiple IDs and multiple passwords for wide range of computer based service they use, this has placed a strain on user memory and users have developed memory aids, such as password lists to assist them in the task of keeping accounts and passwords straight. User password memory aids affect overall system security at the individual application level.

## 4 Neural network's features suitable for security design

### 4.1 Learning ability

Neural network has the possibility of learning. Given a specific task to solve and a class of function, neural network can use a set of observations to solve the task in an optimal sense. Generally, according to learning task, learning ability of neural network can be classified into two main categories, i.e. supervised learning, and unsupervised learning. Supervise learning is the learning with a 'teacher' in the form of a function that provides continuous feedback on the quality of solutions. These tasks include pattern classification, function approximation and speech recognition, etc. Unsupervised learning refers to the learning with old knowledge as the prediction reference. These tasks include estimation problem, clustering, compression or filtering.

### 4.2 One-way property

One-way property means that it is easy to compute the output from the input while difficult or impossible to compute the input from the output. In neural network, there are often more inputs than outputs. Taking a simple neuron model for

example, the input is composed of n elements  $P_0, P_1, P_2 \dots P_{(n-1)}$ , while the output is a unique element C. it is defined as

$$C = f \left( \sum_{j=0}^{n-1} w_j p_j + b \right).$$

As can be seen, it is easy to compute C from  $P=[p_0, p_1, p_2 \dots p_{(n-1)}]$ , while difficult to compute P from C. The difficulty is equal to solve a singular equation. Thus it is one-way process from the input P to the output C. This property is often required by Hash function [35][36] that is used to authenticate data's integrity. This is just one example with one neuron model, complexity of getting C from P with network of neurons can be understand.

## 5 Proposed authentication method based on neural network

As discuss earlier, the two main requirements for developing authentication system are, unique mapping and no reversibility of mapping. These two objectives can be fulfilled by means of applying the suitable learning of password and developing an appropriate architecture. Feed forward architecture has taken, over that the user-defined password can be trained with supervise learning algorithm. This section will contain details about architecture of neural network, learning rule, target definition and process, which will apply for authentication.

### 5.1 Feed forward architecture

The type of Architecture implemented in the project is multiplayer Feed Forward architecture as shown in Fig (2). This neural network is formed in three layers, called the input layer, hidden layer, and output layer. Each layer consists of one or more nodes, represented in this diagram by the small circles. The lines between the nodes indicate the flow of information from one node to the next. In this particular type of neural network, the information flows only from the input to the output.

The nodes of the input layer are passive, meaning they do not modify the data. They receive a single value on their input, and duplicate the value to their multiple outputs. In comparison, the nodes of the hidden and output layer are active. This means they modify the data as shown in Fig (3) each value from the input layer is duplicated and sent to all of the hidden nodes.

The values entering a hidden node are multiplied by weights; a set of predetermined number is stored in the program. The weighted inputs are then added to produce a single number. This is shown in the diagram by the symbol,  $\Sigma$ . Before leaving the node, this number is passed through a nonlinear mathematical function called a Sigmoid. This is an "S" shaped curve that limits the node's output. That is, the input to the sigmoid is a value between  $-\infty$  to  $+\infty$ , while its output can only be between 0 and 1. The outputs from the hidden layer given to output nodes after multiplication with associated weight on that path. The active nodes of the output layer combine and modify the data to produce the output value of the network. In the case of this project the output layer only needs a single node, single hidden layer

contains number of nodes equal to 30 % of the size of the input layer nodes number. The input layer contains the number of nodes according to length of the user identity. The initial weights are generated randomly using uniform distribution within the range of [0 1].

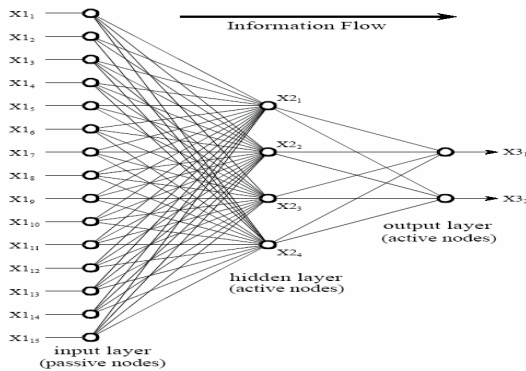


Figure 2. Feed forward architecture

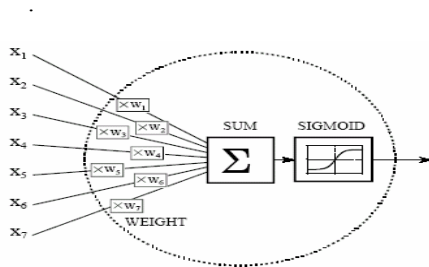


Figure 3. Neural Network Active Node

### 5.2 Transfer function: Sigmoid Function

Sigmoid function (unipolar), mathematically described by the equation  $S(x) = 1/(1+e^{-\lambda x})$ , having several advantages like, (1) Soft limiter, i.e. having sensitivity w.r.t variation in input. (2) Mathematical model of biological neuron, firing phenomenon the characteristic is appearing like sigma function. (3) Its derivative is easily available which is required in learning process,  $S'(x) = s(x) [1 - s(x)]$ .

### 5.3 Physical interpretation of learning rule

Mathematical modeling of any physical phenomena has at least two purposes like Compact form representation and simplification in processing. But sometimes when we observe physical phenomena properly, there may be a chance to get the same solution as it comes from the mathematical modeling.

In ANN, the main goal of learning rule is to minimize the *error*, which is the difference between the target and the observed network output. Weight training is usually formulated as minimization of an error function, such as the mean square error between the target and the actual outputs averaged over all by iteratively adjusting connection weights.

### Why this error exists?

Error exists because initially there is zero co-relation between the connection weights and the applied inputs, since the connection weight values are chosen to be random. So the error value is high initially.

### How this error can be minimized?

In simplified form we can say, if error is high, it indicates that the current weights are very far from the required value and hence a large change in weights required. In other case if error is less, it indicates that the current weights are near to the required value and hence change in weights required for the next iteration is also small. Hence we can say,  $\Delta W \propto \text{Error}$ . Each and every weight is responsible for the output, but in different proportion. So it seems logical to adjust the individual weights in accordance to their affect at the output in order to minimize the error, where the affect is calculated by taking slope, i.e., variation at the output w.r.t the variation in individual weight. Hence we can say,  $\Delta W \propto \text{Affect}$ . Combining with placement of proportional constant  $\eta$ , the final equation of placement of weights adjustment can be expressed as  $\Delta W = \eta(\text{Error})(\text{Affect})$ . Advantage of this method for finding the change in weights are simplicity and very suitable for more complex architecture (like neural network with two or more hidden layers. Author claims to have a method for finding weights adjustment equation for any complex circuit just by visual appearance of network).

### 5.4 Target selection for high sensitivity

Because of supervise learning method; target must be defined at beginning. The output node transfer function is a sigmoid function; hence any value on the curve of sigmoid function can take as target i.e. between 0-1. Authentication system should must be very sensitive w.r.t even a very small variation in the input. The sigmoid characteristic curve can be divided in the three regions namely regions (1)&(3) can be consider as saturation region, because in these regions o/p variation is not very much. Region (2) can be considered as linear region, because changing in o/p with input is approximately linear manner as shown in fig (4). To enhance the sensitivity it's necessary to define the target value in the linear region and the most suited value is equal to 0.5.

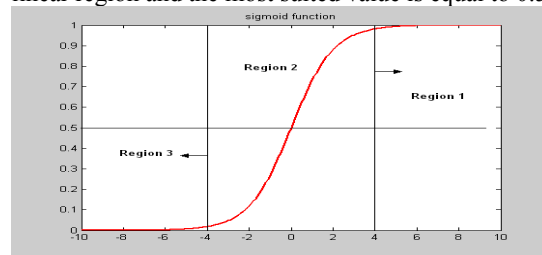


Figure 4. Sigmoid curve with 3 regions

## 6. Authentication process

Unique mapping of a given password can achieve by giving the training of neural network with taking password as an input. Because row information of password may have various characteristics like having alphabet character,

numeric or even some special character or any combination. Hence it's not possible to go for direct training. A preprocessing step required transforming the row data in some suitable format like numeric only. Sensitivity with authentication can be further improved very much if row password is transformed in to corresponding binary format. According to length of this transformed password an architecture with same number of input nodes, one hidden

<b>Correct password: architecture</b>	
<b>Architecture size</b>	
<b>Layer</b>	<b>No. Of Nodes</b>
<b>Input</b>	<b>84</b>
<b>Hidden</b>	<b>25</b>
<b>Output</b>	<b>1</b>

layer containing around 30% nodes of

input layer and one output node created. To get the maximum sensitivity this architecture is fully interconnected. Sigmoid function has taken as transfer function of any active node. Once training completed up to specified error value, the final output of network is the final mapped value of the password. The hidden nodes output also consider as the intermediate mapped values. Even final mapped value is alone enough to provide the authentication but inclusion of intermediate mapped values will give added strength of complete authentication process.

Any test case, first pass through the same preprocessing step as original password has been. If length size of test password is same as original password (otherwise un-authentication declared), same architecture taken as in case of original password, along with trained weights (learning will not take place this time). With the processing of architecture, all mapped values (final & intermediate) compared with the all-original mapped values. Zero difference in result counted as authenticated action otherwise counted as an unauthenticated.

## 7.Experimental results

According to specified description of neural network and its utilization of authentication, several experiments have done. Two examples along with some unauthenticated trail have shown below.

### 7.1 Experiment 1

<b>Correct password: neural</b>	
<b>Architecture size</b>	
<b>Layer</b>	<b>No. Of Nodes</b>
<b>Input</b>	<b>42</b>
<b>Hidden</b>	<b>13</b>
<b>Output</b>	<b>1</b>

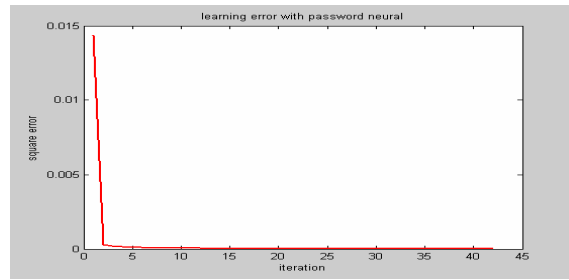


Figure 5. Learning plot

Learning has given corresponding to password 'neural'. It's required very few number of iteration to minimize the error below 0.00001, as shown in fig (5). The available trained weights utilize for authentication process. In the test phase four different passwords 'neural', 'mural', 'neurba', 'signal' has taken to verify the process. Its very clear from the result given system performs as it was expected. Zero difference appears for password 'neural' while for any other case large difference, which is very good enough for declaring as a unauthentication, as shown in fig (6).

### 7.2 Experiment 2

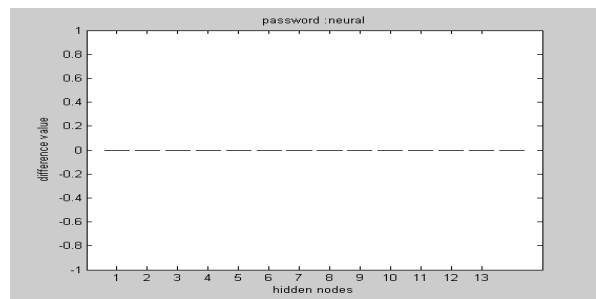


Figure 6(a) test password: neural

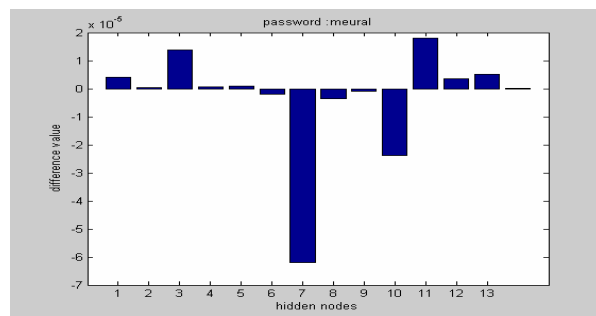


Figure 6(b) test password: meural

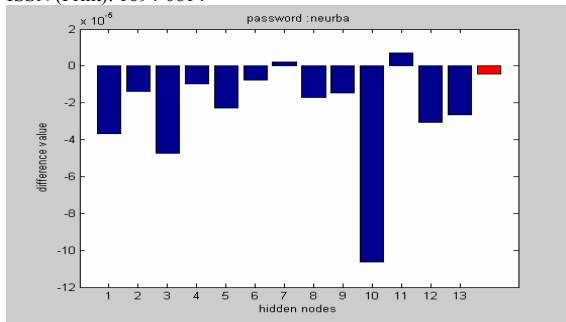


Figure 6(c) test password: neurba

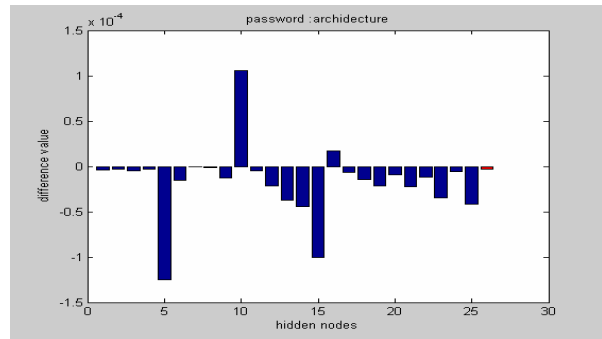


Figure 8(b) test password: architecture

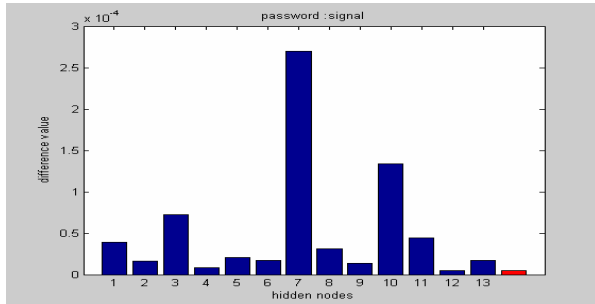


Figure 6(d) test password: signal

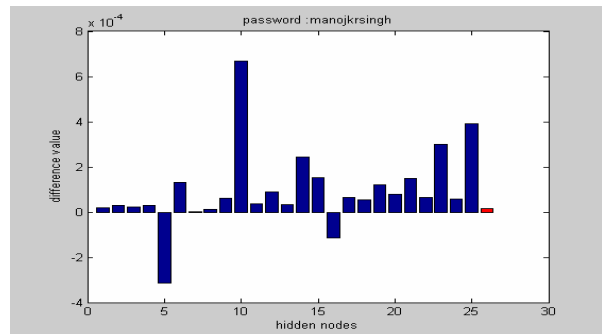


Figure 8(c) test password: manojkrsingh

(\* Last bar with red color represent the final output mapping difference)

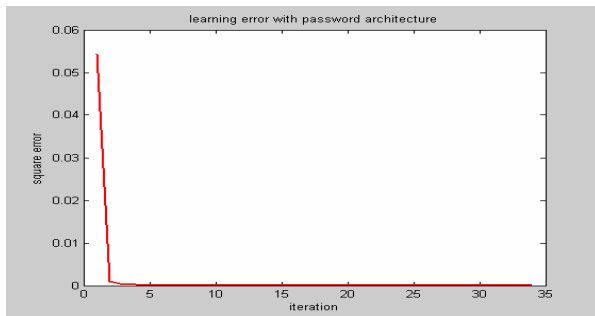


Figure 7. Learning plot

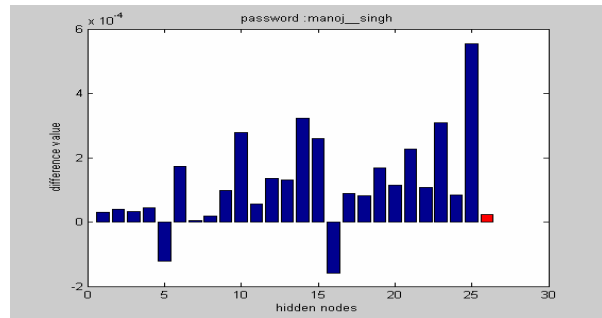


Figure 8(d) test password: manoj\_singh

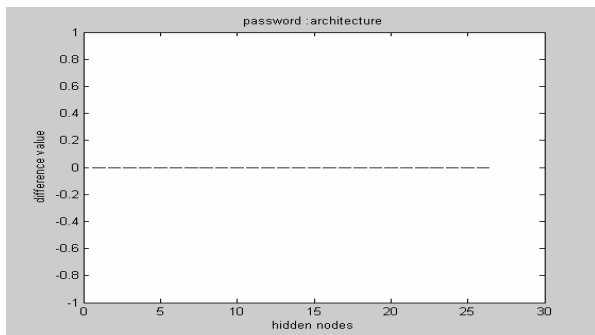


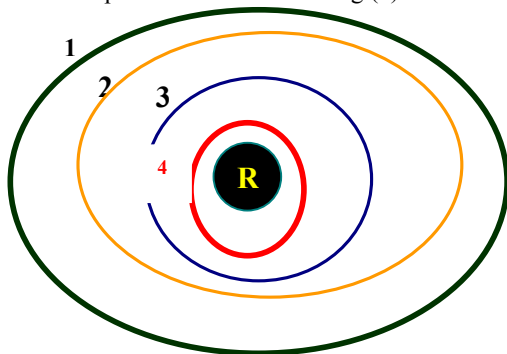
Figure 8(a) test password: architecture

## 8 Proposed security system

### 8.1 Hierarchical structure of security

Based upon above discussion a multiplayer, multiparametric security system has developed. In most of the public oriented services there are service provider along with service users. Each party wanted to allocated a security means to protect their resource. Hence for each case a separate password provision can be allocated (if there is no service provider single password can also be applied). To provide the intruder detection several parameters can taken as measuring consideration likes, (1) number of trails (2) length of password (3) time taken to insert the password. These parameters can surely help to define the activity as normal or intrusion. These parameters are forming a hierarchical structure to enhance the speed of computation and saving the

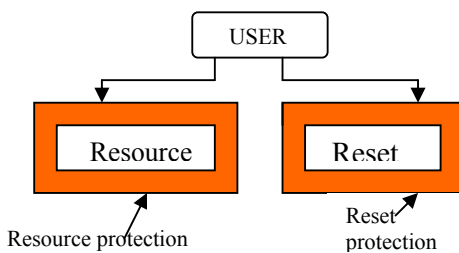
unnecessary involvement of other inner layer modules associated with this security system. These three layers logically define the types of activity without knowing the contents of password. Neural network based authentication is the innermost layer, which is taking care of contents available with the password as shown in Fig (9).



**Figure 9.** 1:Trail layer, 2:length layer, 3:Time layer, and 4:ANN layer, R: Resource.

Any request having permission by all layers only can access the resource otherwise any failure at any layer will be counted as unauthentic activity. Any unauthentic activity will cause to save all the information about the events like time of event, inserted password, and time taken to insert the password. Repeated unauthentic activities (depend upon maximum trail permission allocated for that particular application) result in declaration of intruder in front of system (this situation may also appear if right user have forgotten the password). When intrusion declared, to protect the resource, security environment not allow entering the password further. The other benefit of this facility is, when the right person will try to access the resource, system will not permit to enter the password hence an auto- information mechanism about intrusion available to right user.

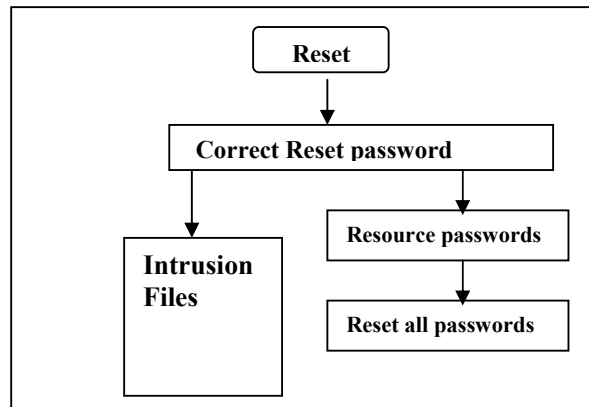
### 8.2 Block diagram of appearance of Security System from User Side.



**Figure 10.** Front view for user

for any user there are two path of progress, (1) accessing of resource (2) reset the old password or intrusion file. If user going for resource access, above defined protocol will follow. Reset mode basically contains two facilities namely intrusion activities file or reset the old passwords either by new password or retaining the existing passwords. Picking up of reset mode is a necessary step for right user if there is seize in resource mode. Because reset mode is open for all,

hence a high level of security provision has given for this by allocating a reset password along with the two same passwords protecting the resource. In summary to reset the passwords three currently existing passwords must be correctly given by the user. The flow of reset mode is given in fig. (11).



**Figure 11.** Reset protocol

Resetting the passwords with same existing ones is having its own advantage in this system. Even for same password the encrypted information will very different compare to previous existing information. Hence instead of changing the password frequently this is much easy means to change the contents of encrypted information.

### 8.3 Two factor security with physical device

With the association of a user-friendly memory device the level of security with given method can be increase very much. The core reason behind this is availability of two different sets of trained weights, hidden layer weights and output layer weights, from neural network. If hidden layer weights transfer to user memory device and output layer weights stored in server or with resource, resistance with brute force trail of attacks can be increase very much because virtually there are infinite possibility exist in the real value domain. Hence even with simple and small password, high level of security can be achieved. Because with each new training of neural network there are different set of weights appear, there is no harm of using the same password with multi-applications. This will remove the burden of memory aids. This proposed method give the level of security comparatively better and cost effective to even smart card.

### 8.4 Advantages of purposed system

In summary the security system given in this paper having advantages like: -

- a similar, simple method to design the one-way hash function.
- with new training of same password encrypted information will be different,. Hence instead to changing the password periodically, with same password the security level can be increased very much.
- free from service provider faith ness circumference.

- Intrusion detection facility.
- Hierarchical protection gives optimum use of security model with high processing speed.
- Multi-user facility from same security environment.
- Encrypted password information always appear in a set of real value.
- can be integrating with any kind of resource environment like soft resource or physical resource.
- with association of external memory device, level of security can achieve a great height and reduce the burden of memory aids.

This project has implemented on MATLAB platform and for experiment it has integrated successfully with various products developed by Manuro tech. research lab.

### Conclusion

**A new method to design the security system using artificial neural network having the intrusion detection capability too, has given. This design is very efficient & robust, at the same time simple. This design having the domain of application numerous irrespective of their nature. With this design, the password identity based security system will get a new direction of development. There is a very good scope to enhance the level of security by splitting the encrypted information in two parts. One part can be stored with security environment whereas other part can embed in to any user-friendly device. As an extension of this design, keystrokes dynamics inclusion can boost the further strength of the solution.**

### Acknowledgments

Author is thanking to Reeta Singh for giving tremendous support, and to lovely Manoree who always inspire and sometimes surprised me to see the real meaning of evolution. Last but not the least I am thankful to all of my associates at Manuro tech. research lab.

### REFERENCES

- [1] M. Burrows, Abadi, M., Needham, R., "A Logic of Authentication," Proceedings of the Royal Society of London, pp. 233-271, 1989.
- [2] B. Lampson, Abadi, M., Burrows, M., Wobber, E., "Authentication in Distributed Systems: Theory and Practice," ACM Transactions Computer Systems, vol. 10, pp. 265-310, 1992.
- [3] J. Anderson, Vaughn, Rayford, "Guide to Understanding Identification and Authentication in Trusted Systems (Light Blue Book)," National Computer Security Center NCSC-TG-017, September 1991 1991.
- [4] U. Manber, "A Simple Scheme to Make Passwords Based on One-Way Functions Much Harder to Crack,," Computers & Security, vol. 15, pp. 171-176, 1996.
- [5] B. Menkus, "Understanding the Use of passwords," Computers & Security, vol. 7, pp. 132- 136, 1988.
- [6] B. L. Riddle, Miron, M. S., Semo, J. A.,

- "Passwords in Use in a University Timesharing Environment," Computers & Security, vol. 8, pp. 569-579, 1989.
- [7] W. Yang, Shieh, S., "Password Authentication Schemes with Smart Cards," Computers & Security, vol. 18, pp. 727-733, 1999.
- [8] M. Abadi, Burrows, M., Kaufman, C., Lampson, B., "Authentication and delegation with smart-cards," Science of Computer Programming, vol. 21, pp. 91-113, 1993.
- [9] K. Dehnad, "A Simple Way of Improving the Login Security," Computers & Security, vol. 8, pp.607-611, 1989.
- [10] M. Bishop, and Klein, D.V., "Improving System Security via Proactive Password Checking," Computers and Security, vol. 14, 1992.
- [11] N. Ahituv, Lapid, Y., Neumann, S., "Verifying the Authentication of an Information System User," Computers & Security, vol. 6, pp. 152-157, 1987.
- [12] D. L. Jobusch, Oldehoeft, A. E., "A Survey of Password Mechanisms: Weakness and Potential Improvements. Part 1," Computers & Security, vol. 8, pp. 587-604, 1989.
- [13] R. Morris, Thompson, K., "Password Security: A Case history," Communications of the ACM, vol. 22, pp. 594-597, 1979.
- [14] M. Peyravian, Zunic, N., "Methods for Protecting Password Transmission," Computers & Security, vol. 19, pp. 466-469, 2000.
- [15] M. Zviran, Haga, William, "A Comparison of Password Techniques for Multilevel Authentication Mechanisms," The Computer Journal, vol. 36, pp. 227- 237, 1993.
- [16] G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," The Psychological Review, vol. 63, pp. 81-97, 1956.
- [17] A. Adams, Sasse, M. A., "Users Are Not The Enemy,," Communications of the ACM, vol. 42, pp. 41-46, 1997.
- [18] R. Pond, Podd, J., Bunnell, J., Henderson, R., "Word Association Computer Passwords: The Effect of Formulation Techniques on Recall and Guessing Rates,," Computers & Security, vol. 19, pp. 645-656, 2000.
- [19] S. N. Porter, "A Password Extension for Improved Human Factors," Computers & Security, vol.1, pp. 54-56, 1982.
- [20] R. Shimonski, "Create effective passwords: strategies for computer systems," vol. 2003: IBM developer Works, 2002.
- [21] S. L. Smith, "Authenticating Users by Word Association," Computers & Security, vol. 6, pp. 464-467, 1987.
- [22] D. Weirich and M. A. Sasse, "Pretty Good Persuasion: A first step towards effective password security in the real world," presented at New security Paradigms, Cloudcroft, New Mexico, 2002.
- [23] J. Yan, Blackwell, A., Anderson, R., Grant, A., "The Memorability and Security of Passwords Some Empirical Results," Cambridge University Computer Laboratory.

- [24] M. Zviran, Haga, William, "Password Security: An Empirical Study," *Journal of Management Information Systems*, vol. 15, pp. 161- 185, 1999.
- [25] T. Jones, "Too many secrets? Password proliferation leads to user fatigue," in *Columbia News Service - Columbia University Graduate School of Journalism*. New York, 2002.
- [26] S. Swanson, "Way too many passwords, not enough protection," in *Chicago Tribune*, online edition ed. Chicago, 2003, pp. 1.
- [27] I. Sommerville, *Software Engineering*, 6th ed. Essex, England: Pearson Educational Limited, 2001.
- [28] J. J. Whitmore, "A method for designing secure solutions," *IBM Systems Journal*, vol. 40, pp.747-768, 2001.
- [29] E. Spafford, "Opus: Preventing weak password choices," *Computers & Security*, vol. 11, pp.273-278, 1992.
- [30] E. Spafford, "Observing Reusable Password Choices," presented at *Proceedings of the 3rd Security Symposium*, Usenix, 1992.
- [31] Bezdek j.c. On the relationship between neural networks, pattern recognition and intelligence. *The international journal of approximate reasoning*, 1992,6(2): 85-107.
- [32] Grossberg S. *Neural network and neural intelligence*. Cambridge, Mass: MIT press, 1988
- [33] Ohira T, *Neural network model with delay toward encryption*. proceeding of the IEEE-INNS-ENNS International joint conference on neural network (IJCNN), Vol.5, pp417-4221, 2000.
- [34] Tao Yang, C.M. Yang application of neural network to unmasking chaotic secure communication. *PHYSICA D* 124(1-3):248-257 Dec 1 1998.
- [35] *Secure Hash Standard*. FIPS 180-2, 2002.
- [36] Stephen Northcutt, Judy Novak, *Network intrusion detection: An analyst's Handbook*, 2<sup>nd</sup> edition New Riders publishing 2000.
- [37] Bivens A, Palagiri C, Smith R, et al. *Network-based intrusion detection using neural networks*. Proceeding of ANNIE-2002. New york: ASME press, 2002, 579-584.
- [38] Lian S, Sun J, Wang Z. *Secure Hash function based on neural network*. *Neurocomputing*, Vol.69 No.16-18, 2006.
- [39] Lian S, Liu Z, Ren z, Wang H, *Hash function based on chaotic neural networks*. 2006 IEEE International symposium on circuit and system (ISCAS2006), May 21-24, 2006, Greece. pp221-224.



**Manoj kr. Singh**, Currently holding the post of director in manuro tech. research lab. He is actively associated with industry as technology consultant & guiding several research scholars. He is having background of R&D in Nanotechnology, VLSI CAD, Evolutionary computation, Neural network, Advanced signal processing etc.

## 3D/2D Registration of Mapping Catheter Images for Arrhythmia Interventional Assistance

Pascal Fallavollita  
School of Computing, Queen's University,  
Kingston, Ontario, Canada  
*pascal@cs.queensu.ca*

### Abstract

Radiofrequency (RF) catheter ablation has transformed treatment for tachyarrhythmias and has become first-line therapy for some tachycardias. The precise localization of the arrhythmogenic site and the positioning of the RF catheter over that site are problematic: they can impair the efficiency of the procedure and are time consuming (several hours). Electroanatomic mapping technologies are available that enable the display of the cardiac chambers and the relative position of ablation lesions. However, these are expensive and use custom-made catheters. The proposed methodology makes use of standard catheters and inexpensive technology in order to create a 3D volume of the heart chamber affected by the arrhythmia. Further, we propose a novel method that uses *a priori* 3D information of the mapping catheter in order to estimate the 3D locations of multiple electrodes across single view C-arm images. The monoplane algorithm is tested for feasibility on computer simulations and initial canine data.

**Key words:** 3D reconstruction, monoplane imaging, navigation system, catheter ablation, cardiac arrhythmias.

### 1. Introduction

Severe disorders of the heart rhythm that can cause sudden cardiac death or morbidity, can be treated by radiofrequency (RF) catheter ablation, which consists of inserting a catheter inside the heart, near the area from which originates the abnormal cardiac electrical activity, and delivering RF currents through the catheter tip so as to ablate this arrhythmogenic area. The precise localization of the arrhythmogenic site and positioning of the RF catheter at that site are problematic: they can impair the efficacy of the procedure and the procedure can last many hours, especially for complex arrhythmias. To shorten the duration of RF catheter ablation and increase its efficiency, commercial systems that provide a 3D color display of the cardiac electrical activation sequence during the arrhythmia have been proposed.

These systems incorporate basket electrode arrays (**Constellation**, EPT Inc.), catheters with a balloon electrode array (**Ensite 3000**, Endocardial Solutions Inc.) and catheters with magnetic position detectors

(**CARTO™**, Biosense Webster Inc.). A complete navigation and registration framework is also available (**CartoMerge**, Biosense Webster Inc.). All these systems including purchase of system-specific catheters are costly for hospitals. The first two technologies can map the cardiac activation sequence using data recorded during a single beat whereas the **CARTO™** system relies on data recorded point-by-point during numerous beats, which implies that the arrhythmia must remain stable during the procedure.

Other approaches have also been proposed to guide RF ablation therapy, such as the visualization of an optically tracked catheter by making use of magnetic resonance imaging (MRI) [1-2], the combination of MRI and fluoroscopy [3], ultrasound imaging of the ablation catheter [4], the combination of ultrasound and pre-operative computer tomography (CT) [5], or preoperative imaging (CT/MRI) for ablation planning [6-7]. These approaches omit incorporating the all important electrophysiological data which allows the interventionist to determine the origin of the arrhythmia. Inverse electrocardiography, an established formulation is the imaging of the activation time map on the entire surface of the heart from ECG mapping data. A few examples computed from body surface potential measurements have been incorporated to segmented MRI images by Tilg et al. [8-9] and to biplane reconstructions of the cardiac geometry by Ghanem et al. [10-11]. DeBuck et al. [12] have constructed a patient-specific 3D anatomical model from MRI and merged it with fluoroscopic images in an augmented reality environment that enables the transfer of cardiac activation times onto the model. Cristoforetti et al. [13] have developed a strategy for the spatial registration of the coarse electroanatomic map obtained by the **CARTO™** system, and the detailed geometrical reconstruction of the left atria and pulmonary veins retrieved from CT images. Methods based on real-time integration of electroanatomic and tomographic modalities are being developed and offer an intermodal fusion based on semi-automatic registration procedures initialized by fiducial point pairing [14-15]. The above methods make use of preoperative data which do not reflect the real-time movement of the heart at the time of intervention, and introduce a logistical problem as these technologies are not present in the electrophysiological intervention room.

Recently, we have proposed a more affordable fluoroscopic navigation, emulating CARTO™, by obtaining local activation times from a roving catheter whose positions are computed point-by-point from biplane fluoroscopic projections. We also proposed a novel concept by superimposing the isochronal map depicting the cardiac electrical activation sequence directly over the 2D fluoroscopic image of the heart [16], a method that the CARTO™ technology lacks. However, biplane fluoroscopy systems are rare in the clinical setting which in turn emphasizes the importance of developing single-view 3D reconstruction algorithms.

The focus of this paper is a continuation of our previous work, but with two significant modifications. To begin with, we consider implementing a full perspective camera model instead of the parallel projection concept so as to create a more precise 3D geometry of the heart chamber affected by the arrhythmia. Second, our aim is to answer the following question: is it possible to estimate precisely the 3D locations of the mapping catheter across all C-arm image frames in a cardiac cycle using only a single view? We propose to use *a priori* 3D information of the mapping ablation catheter positions in order to achieve this. If this is possible, then the interventionist could possibly reduce procedure time accordingly as a monoplane C-arm fluoroscope is the primary imaging modality used in clinic and can provide real-time data images as well. To our knowledge this is the first work reported on estimating 3D locations using single view sequences in order to assist catheter ablation procedures. Therefore, we tag this work as a feasibility study by presenting a practical implementation and experimental analysis using computer simulations and initial results on canine data.

## 2. Methodology

### 2.1 Central Intuition

Clinical accuracy is determined by 3D reconstruction precision in the order of 2 mm or less and we don't expect to achieve this using a monoplane sequence. As showed in [16], using only a single 2D C-arm image the recovered catheter depth had an error of about 10 mm. We believe that by incorporating temporal knowledge of the catheter coordinates may help improve depth estimation results.

Initial *a priori* 3D information of the mapping catheter can be obtained using pre-operative data from CT/MRI or via two-view reconstruction by rotating the monoplane C-arm fluoroscope at two perpendicular angulations and acquiring monoplane sequences. In this paper, we focus on the latter since the principal imaging modality used to guide cardiac ablation procedures is the C-arm fluoroscope. Like most work in the cardiac field, the 3D *a*

*priori* coordinates are reconstructed in the diastolic cardiac phase as heart motion is considered to be minimal at this instant. We then proceed to select one of the monoplane sequences acquired and estimate the depth of the mapping catheter at each image frame. The idea is to determine if or when the 3D reconstruction becomes inaccurate. In the end, what will be required is providing a 3D visual aid to the interventionist of the heart chamber and fusing the isochronal times on it in order to help them position correctly the mapping ablation catheter on the arrhythmogenic site.

### 2.2 C-arm Fluoroscopy Calibration

**Full Perspective Camera Model:** Figure 1 shows the full perspective camera model that will be used for the 3D reconstruction problem. If we define a three dimensional point  $P_{world} = [X \ Y \ Z \ 1]^T$  in the world coordinate system, then its 2D projection in an image,  $m = [u \ v \ 1]^T$ , is achieved by constructing a projection matrix:

$$P_{mat} = \begin{bmatrix} kf & 0 & u_o \\ 0 & kf & v_o \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}$$

$$m = P_{mat} P_{world} \quad (1)$$

The intrinsic matrix of size [3x3], contains the pixel coordinates of the image center, also known as the principal point  $(u_o, v_o)$ , the scaling factor  $k$ , which defines the number of pixels per unit distance in image coordinates, and the focal length  $f$  of the camera (in meters). The extrinsic matrix of size [3x4] is identified by the transformation needed to align the world coordinate system to the camera coordinate system. This means that a translation vector,  $t$ , and a rotation matrix,  $R$ , need to be found in order to align the corresponding axis of the two reference frames. Lastly, image resolution (usually mm/pixel) is calculated from the imaging intensifier size divided by the actual size of the image in pixels.

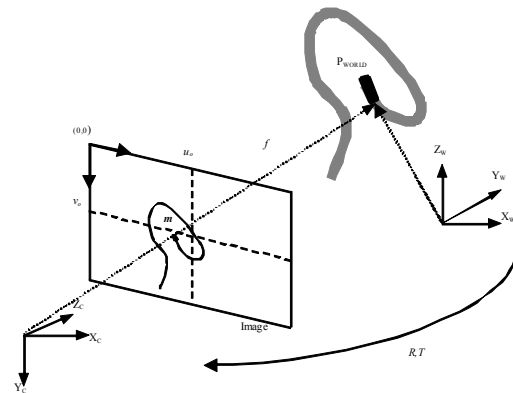


Fig 1. The perspective camera model. Any 3D world point can be projected onto a 2D plane and its coordinates would be  $(u, v)$

pixels. The camera model is taken from the Epipolar Geometry Toolbox [17]. Image resolution would be equal to C-arm intensifier size divided by image size in pixels.

**Parallel & Weak Perspective Camera Models:** An orthographic camera is one that uses parallel projection to generate a two dimensional image of a three dimensional object. The image plane is perpendicular to the viewing direction. Parallel projections are less realistic than full perspective projections, however they have the advantage that parallel lines remain parallel in the projection, and distances are not distorted by perspective foreshortening. The parallel projection matrix is given by:

$$P_{\text{affine}} = \begin{bmatrix} k * r_{11} & k * r_{12} & k * r_{13} & (k * t_x) + u_o \\ k * r_{21} & k * r_{22} & k * r_{23} & (k * t_y) + v_o \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

The weak perspective camera is an approximation of the full perspective camera, with individual depth points  $Z_i$  replaced by an average depth  $Z_{\text{avg}}$ . We define the average depth,  $Z_{\text{avg}}$  as being located at the centroid of the cloud of 3D points in the world coordinate system. The weak perspective projection matrix is given by:

$$P_{\text{weak}} = \begin{bmatrix} f * k * r_{11} & f * k * r_{12} & f * k * r_{13} & (f * k * t_x) + (u_o * Z_{\text{avg}}) \\ f * k * r_{21} & f * k * r_{22} & f * k * r_{23} & (f * k * t_y) + (v_o * Z_{\text{avg}}) \\ 0 & 0 & 0 & Z_{\text{avg}} \end{bmatrix}$$

where  $Z_{\text{avg}} = ([r_{31} \ r_{32} \ r_{33}]^T \times \text{centroid}) + t_z$  (3)

### 2.3 Mapping Catheter Segmentation

We have recently developed a four-step filter in order to enhance coronary arteries visible in C-arm fluoroscopy images [18]. Optimal filter parameters are presented there and omitted in this paper for brevity. We applied the same filter to enhance the catheters and extract electrode coordinates. Following is a brief description of these filters.

**Homomorphic Filtering:** A homomorphic filter is first used to denoise the fluoroscopic image. The illumination component of an image is generally characterized by slow spatial variation. The reflectance component of an image tends to vary abruptly. The homomorphic filter tends to decrease the contribution made by the low frequencies and amplify the contribution of high frequencies. The result is simultaneous dynamic range compression and contrast enhancement. The homomorphic filter is given by:

$$H(u, v) = (\gamma_H - \gamma_L) (1 - e^{-c(D^2(u, v)/D_o^2)}) + \gamma_L \quad (4)$$

with  $\gamma_L < 1$  and  $\gamma_H > 1$ . The coefficient  $c$  controls the sharpness of the slope at the transition between high and

low frequencies, whereas  $D_o$  is a constant that controls the shape of the filter and  $D(u, v)$  is the distance in pixels from the origin of the filter.

**Perona-Malik Filtering:** The Perona-Malik filter is implemented here in order to reduce and remove both noise and texture from the image, as well as, to preserve and enhance structures. The diffusion equation is given by

$$\frac{\partial I}{\partial t} = \text{div}(c(x, y, t) \nabla I) \quad (5)$$

Where  $I$  is the input image and  $c(x, y, t)$ , the diffusion coefficient, will control the degree of smoothing at each pixel point in the image. The diffusion coefficient is a monotonically decreasing function of the image gradient magnitude. It allows for locally adaptive diffusion strengths; edges are selectively smoothed or enhanced based on the evaluation of the diffusion function. Although any monotonically decreasing continuous function of the gradient would suffice as a diffusion function, two functions have been suggested:

$$c(x, y, t) = \exp\left(-\frac{|\nabla I|}{K}\right)^2 \quad (6)$$

$K$  is referred to as the diffusion constant or the flow constant. The greatest flow is produced when the image gradient magnitude is close to the value of  $K$ . Therefore, by choosing  $K$  to correspond to gradient magnitudes produced by noise, the diffusion process can be used to reduce noise in images.

**Complex Shock Filtering:** The complex shock filter couples shock and linear diffusion in the discrete domain, showing that the process converges to a trivial constant steady state. To regularize the shock filter, the authors suggest adding a complex diffusion term and using the imaginary value as the controller for the direction of the flow instead of the second derivative. The complex shock filter is given by

$$I_t = -\frac{2}{\pi} \arctan\left(a \text{Im}\left(\frac{I}{\theta}\right)\right) |\nabla I| + \lambda I_{\eta\eta} + \tilde{\lambda} I_{\zeta\zeta} \quad (7)$$

where  $a$  is a parameter that controls the sharpness of the slope,  $\lambda = r^{ei\theta}$  is a complex scalar,  $\tilde{\lambda}$  is a real scalar,  $\zeta$  is the direction perpendicular to the gradient and  $\eta$  is the direction of the gradient. As fluoroscopy images have a low signal to noise ratio they tend to be noisy with artifacts present in them. Hence, we believe that applying a complex filter will result in a robust and stable deblurring process for the images as the filter is effective in very noisy environments.

**Morphological Operation:** Morphological filtering was applied as a final image processing step in order to eliminate background elements around the object of interest. The structuring element consists of a pattern specified as the coordinates of a number of discrete points relative to a defined origin. We chose a disk structuring element that has a radius of a few pixels, since the contours of the catheters and electrodes can be modeled as a disk.

**Electrode Segmentation & Convex Hull:** Using MatLab, the 4-step filter was implemented and applied on both monoplane C-arm datasets, beginning with the diastolic image. The smoothed images were then automatically thresholded using Otsu's method. We then labeled connected components in this image using the *bwlabel* function in MatLab. The centroid of the electrodes in each labeled region was then calculated using the *regionprops* function so as to obtain the coordinates required in the two C-arm views. The algorithm outputs the centroid automatically and in case of failure, due to electrodes or catheters overlapping, we manually segment the image to obtain the desired coordinates. The convex hull algorithm is a classical and popular method used to reconstruct a 3D object from an unorganized set of points. The 2D version of this algorithm (which is easier to understand) is described as follows. Let  $S$  be a finite set of two-dimensional points in the plane. The convex hull of  $S$  is the smallest convex set that contains  $S$ . This means that the boundary of the convex hull is a convex polygon, whose vertices are points of  $S$ , and whose edges are line segments joining pairs of points of  $S$ .

#### 2.4 *A Priori* 3D Monoplane Algorithm

We suppose that we have at our disposal a set of  $n$  3D ablation catheter electrode coordinates  $(X0_n, Y0_n, Z0_n)$  at time  $t = 0$  obtained from two-view fluoroscopic data. These *a priori* coordinates are expressed in the camera reference frame in order to have a Z-direction corresponding to catheter electrode depth. Secondly, we have at our disposal the C-arm fluoroscope gantry parameters which can be extracted from the image header DICOM files. These parameters enable us to construct a projection matrix for a specific viewing angle. By selecting one of the two acquired monoplane datasets, we can solve for the 3D displacements  $(dx_{i,i+1}, dy_{i,i+1}, dz_{i,i+1})$  between consecutive C-arm image frames beginning with the first image  $i=1$ . Expanding equation (1) and using an additional C-arm image  $i = 2$ , we obtain our first two equations as follows

$$u_2 = \frac{m_1(X0 + dx_{12}) + m_2(Y0 + dy_{12}) + m_3(Z0 + dz_{12}) + m_4}{m_9(X0 + dx_{12}) + m_{10}(Y0 + dy_{12}) + m_{11}(Z0 + dz_{12}) + m_{12}}$$

$$v_2 = \frac{m_5(X0 + dx_{12}) + m_6(Y0 + dy_{12}) + m_7(Z0 + dz_{12}) + m_8}{m_9(X0 + dx_{12}) + m_{10}(Y0 + dy_{12}) + m_{11}(Z0 + dz_{12}) + m_{12}}$$
(8)

Both equations describe the pixel coordinates in the second image  $(u_2, v_2)$  and the twelve coefficients  $m_i=1:12$  are the values of the projection matrix. By adding an additional C-arm image frame  $i = 3$  we obtain two new equations with three additional unknowns in 3D

$$u_3 = \frac{m_1(X0 + dx_{12} + dx_{23}) + m_2(Y0 + dy_{12} + dy_{23}) + m_3(Z0 + dz_{12} + dz_{23}) + m_4}{m_9(X0 + dx_{12} + dx_{23}) + m_{10}(Y0 + dy_{12} + dy_{23}) + m_{11}(Z0 + dz_{12} + dz_{23}) + m_{12}}$$

$$v_3 = \frac{m_5(X0 + dx_{12} + dx_{23}) + m_6(Y0 + dy_{12} + dy_{23}) + m_7(Z0 + dz_{12} + dz_{23}) + m_8}{m_9(X0 + dx_{12} + dx_{23}) + m_{10}(Y0 + dy_{12} + dy_{23}) + m_{11}(Z0 + dz_{12} + dz_{23}) + m_{12}}$$
(9)

These four equations take into account the spatial positions of a projected 3D world point on the acquired C-arm images. As we have four equations with six unknowns we can extract two additional equations based on the fact that the Euclidean distances in pixels,  $d$ , between catheter electrode points in two consecutive images are known. It is to note that the distance between 3D points is not the same as the distance between their projected image points. Thus, we consider orthogonal projection estimations in this case and we deem that this approximation is suitable enough for the proposed analysis. We arrive at the following two equations

$$d_{12}^2 = \sqrt{(u_2 - u_1)^2 + (v_2 - v_1)^2}$$

$$d_{23}^2 = \sqrt{(u_3 - u_2)^2 + (v_3 - v_2)^2}$$
(10)

We can now solve for the three dimensional displacements. A Levenberg-Marquardt optimization scheme [19] can be used here in order to solve for the unknown displacements. For the optimization scheme, initial approximations are a requirement to initialize the process. Hence, a suitable approximation for the displacements  $dx$  and  $dy$  can be obtained if we consider a parallel back projection of the 2D image points into the world coordinate system. As for the displacements  $dz$ , if we assume that the average depth of the catheter electrodes remains relatively constant in consecutive time frames (i.e. weak perspective camera model), then we can calculate the average depth of the 3D points  $(X0, Y0, Z0)$ . This average depth should be relatively the same at future time instants  $t = 2, 3$ , etc., signifying that depths  $dz$  will be equal to zero for the optimization scheme. However, for the sake of a more exhaustive analysis, we also consider depth displacements  $dz \in [1-5]$  millimeters as well. To justify the range of approximations, we note that the acquisition frame rate of the fluoroscopy images was such that the movement between catheter electrodes in two adjacent images

resulted in displacements less than ten pixels. Finally, we note that the proposed equations do not incorporate nonrigid constraints implying that the monoplane 3D reconstruction of rigid moving objects (i.e. catheter electrodes) might be feasible.

## 2.5 Clinical Data Acquisition

A mongrel dog was anesthetized and laid on its right side on a fluoroscopy table (Integris Allura, Philips Inc.). A reference catheter and a pacing catheter were inserted into the right ventricle, close to the septal wall. The role of the reference catheter was to define an origin for our 3D coordinate system. This was important as motion artifacts are ever present during the experiments (heart beat, respiration, etc.), hence we deemed it appropriate to position it near a rigid landmark so that it experiences less movement due to artifacts. The role of the pacing catheter was to produce a simple electrical activation sequence so as to validate the isochronal maps. Finally, a standard RF ablation catheter was inserted from the femoral vein into the left ventricle (LV) of the dog. During the course of the experiment, this mapping catheter was moved to 20 different sites (i.e. point-by-point as the CARTO™ system) within the ventricle in order to obtain electrical and geometrical data from sufficient sites to map the activation sequence. The 20 landmarks were selected to reflect as closely as possible the entire volume of the ventricle. Electrograms were recorded using the CardioMap software system (*Research Center at Sacré Coeur Hospital, Montreal, Canada*). Specifications for the acquisition software were: 1000 samples/second, 0.05 Hz high-pass and 450 Hz low-pass frequencies. Local activation time was measured (ms) as the difference in the times of the fastest negative deflections (dV/dt) seen in the two electrograms recorded with the reference catheter and the mapping catheter. The fluoroscopic image acquisition rate was set to 60 fps so as to minimize motion artifacts as interframe 2D images are thus closer to each other and yield a smaller displacement between the objects in consecutive images. Images were recorded during approximately 2 seconds at the end of the expiration. The monoplane fluoroscopic C-arm was rotated by 90° to acquire two images for each mapping site: a left lateral view (the C-arm in a vertical position) and a posterior view (the C-arm in a horizontal position). Images were recorded with a 512 x 512 pixel resolution. We selected the diastolic frame for each mapping site so as to minimize motion blur by identifying the frame having the smallest root-mean-square difference with the preceding frame.

## 2.6 Evaluation

**Computer Simulations:** In order to validate our method, we first tested our proposed procedure on synthetic experimentation. We created a 3D helix containing 30

coordinate points so as to model the shape of a catheter. Then we created two C-arm gantry setups that represented the posterior/anterior (PA) and left lateral (LAT) views of the heart. The focal length of the fluoroscopic X-ray system was equal to 1 meter and the helix location was set to 50 cm along the focal axis. The primary angles were equal to (90°, 0°) respectively for the PA and LAT views, whereas the secondary angles were equal to (0°, 0°) for both views respectively. The image sizes were set to [512x512] pixels and the intensifier size was chosen to be [178x178] millimeters. This allowed for a resolution 0.347mm/pixel. We could now calculate two projection matrices and we projected the 3D helix points on a first set of biplane images. These two-view images represent a first time instant at  $t=0$ . Subsequent biplane images are calculated by applying rigid movement on the 3D helix coordinates using the following rigid motion equation:

$$X_{t=2} = R_{\theta\phi\psi} \times X_{t=1} + [T_x, T_y, T_z] \quad (11)$$

The 3D angles were set to  $[\theta, \Phi, \varphi] \in [0.5^\circ, 0.5^\circ, 0.5^\circ]$  and the 3D translations were set to  $[T_x, T_y, T_z] \in [1, 1, 1]$  millimeters each. This accounted for 2D interframe displacements in the images of (-6.82, -6.03) pixels in the left lateral view and (4.97, 7.12) pixels for the posterior/anterior views. These values are similar to what was perceived using the 60 fps acquisition rate in the clinical images. A new set of biplane images and 3D helix points are obtained at  $t=1$ . In a similar fashion, equation (11) is reapplied to produce subsequent biplane data. Excluding the *a priori* biplane set at  $t=0$ , a total of five biplane datasets are generated. For good measure we added errors of up to 2 mm to the coordinates.

**Clinical Experiment:** Regarding the clinical data, as we had a total of 20 biplane datasets, we calculated the 3D catheter points using a conventional biplane reconstruction algorithm (triangulation method). Details of the biplane algorithm are omitted here as it can be deduced that by using the two projection matrices for the X-ray gantry settings and the 2D electrode coordinates, we can easily estimate the 3D world points by solving a set of 4 projection equations with 3 unknowns (X, Y, Z) when considering a single point. A complete description is available in [20]. For a specific dataset, we first extracted the diastolic images and used our 4-step filter to extract the electrodes. Then we performed two-view reconstruction. This 3D *a priori* model was then used to initialize our monoplane equations. Since we are supposing that the C-arm fluoroscope gantry parameters remain fixed in a monoplane analysis, we can then calculate projection matrices that will allow the 3D points to be projected in the images temporally. We optimized the equations using a total of 12 visible electrodes including the actual mapping tip-electrode; the one that is in contact with the ventricle wall and directs heat to the arrhythmogenic site. With

respect to the initial 3D displacements for each electrode, we used the parallel projection approximation for  $(dx, dy)$  and depth estimates were set in the interval  $dz \in [0-5]$  mm.

### 3. Results & Discussion

#### 3.1 Computer simulation results

Table 1 shows the simulation results for rigid motion analysis on a helix. We observe that as the number of C-arm images used to optimize our monoplane equations increases, the overall reconstruction results deteriorate. This is expected as the uncertainty of landmark positions increases temporally in a single view framework. The initial depth approximation  $dz$  plays a role in the convergence process. If we select an initial depth approximation of  $dz = 1$  mm, which is equal to the true simulated displacements from our computer simulations, then we obtain lower 3D root mean square errors (RMS) between optimized and true three dimensional coordinates. The RMS values are accumulated values across the number of images used for the monoplane algorithm. Depth initial estimates  $dz \in [1,2,3]$  mm produce RMS values less than 3 mm when using three consecutive monoplane images. The PA view simulations produce better results probably due to the helix points being projected with no coplanarity in the 2D images. As expected, if the initial guess is unrealistic (i.e.  $dz \in [4, 5]$  mm) reconstruction results deteriorate as RMS errors are larger than 10mm. Interestingly, the temporal 3D RMS values for all 30 helix points are 8.1mm and 7.2 mm for the LAT and PA views respectively when using six consecutive images and an initial guess of  $dz=3$  mm. This is less than the 10 mm estimated electrode depth using only a single image in [16].

#### 3.2 Clinical results

Image contrast in the clinical experiment affects the accuracy of the segmentation algorithm and depth estimates: visual analysis revealed that the LAT images always had a better contrast than the PA images because of the oblong section of the canine torso. First, we were able to automatically segment the reference and mapping catheter electrodes in 16 of 20 datasets using the proposed 4-step filter and thresholding algorithm. The remaining four images had to be manually segmented. Regarding the PA images, the filtering and thresholding algorithm failed in the majority of images and therefore we relied on a manual segmentation of the electrodes. Figure 2 shows an example of an C-arm image segmentation using the proposed filter on a LAT and PA image. Once the electrodes were extracted from the twenty datasets, we

selected the mapping catheter tip electrode coordinates from the diastolic images and performed two-view triangulation and constructed a convex hull as shown

Table 1. Computer simulation results for various depths  $dz$  (in mm).

	Left Lateral View				Antero-Posterior View			
	mean	min	max	3D RMS	mean	min	max	3D RMS
$dz=0$								
# Images								
3	2.188	0.148	3.840	3.976	1.033	0.096	1.917	1.911
4	1.752	0.094	3.837	4.527	0.976	0.003	2.067	2.613
5	1.463	0.003	3.951	5.152	0.929	0.007	2.128	3.204
6	1.150	0.015	3.373	5.324	0.869	0.004	2.160	3.674
$dz=1$								
# Images								
3	1.491	0.043	2.877	2.763	0.578	0.005	0.929	0.983
4	1.272	0.039	2.876	3.122	0.651	0.006	1.080	1.541
5	1.247	0.101	2.998	3.802	0.657	0.010	1.138	2.004
6	1.230	0.011	2.413	4.411	0.641	0.002	1.165	2.398
$dz=2$								
# Images								
3	1.254	0.005	1.913	2.048	0.915	0.060	1.830	1.742
4	1.223	0.098	2.628	2.813	0.977	0.002	1.996	2.592
5	1.363	0.006	3.052	4.095	1.045	0.003	2.070	3.443
6	1.522	0.021	3.327	5.645	1.114	0.001	2.109	4.314
$dz=3$								
# Images								
3	1.190	0.059	2.860	2.341	1.890	1.050	2.794	3.142
4	1.418	0.008	3.643	3.873	1.942	0.896	2.954	4.458
5	1.730	0.009	4.072	5.787	2.004	0.842	3.023	5.815
6	2.074	0.012	4.351	8.097	2.074	0.825	3.057	7.220
$dz=4$								
# Images								
3	1.698	0.014	3.862	3.390	2.866	2.040	3.758	4.635
4	2.197	0.007	4.659	5.567	2.913	1.885	3.912	6.471
5	2.568	0.064	5.091	8.037	2.974	1.832	3.976	8.363
6	3.021	0.465	5.375	10.979	3.042	1.820	4.006	10.314
$dz=5$								
# Images								
3	2.668	0.978	4.862	4.714	3.842	3.030	4.723	6.154
4	3.173	0.969	5.673	7.476	3.885	2.873	4.872	8.528
5	3.544	0.812	6.111	10.492	3.944	2.822	4.930	10.965
6	4.004	1.424	6.399	14.028	4.011	2.814	4.955	13.468

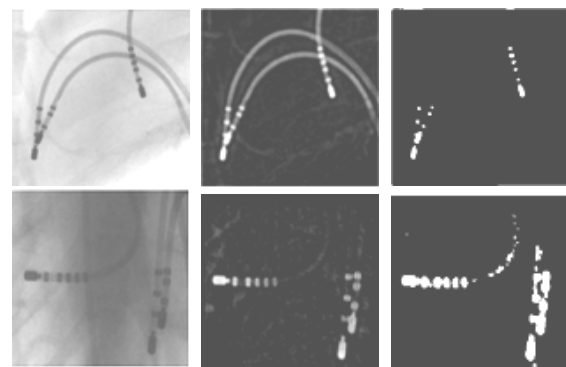


Fig 2. Example of segmentation algorithm using (top row) left lateral view and (bottom row) the posterior anterior view. The

reference and pacing catheters are adjacent to each other in the right ventricle; the third catheter is the ablation catheter in the left ventricle. (*Left column*): original cropped image; (*Center column*): the electrodes are enhanced and background

suppressed using a 4-step filter; (*Right column*): segmentation of the filtered image using Otsu's method and labeling. Manual segmentation would be necessary for the PA image.

Table 2. Various depth estimations of catheter electrodes using 3 and 6 consecutive X-ray images over the 20 datasets. ( in mm).

	Left Lateral View								Posterior-Antero View							
	Mapping electrode				Twelve electrodes				Mapping electrode				Twelve electrodes			
	mean	min	max	3D RMS	mean	min	max	3D RMS	mean	min	max	3D RMS	mean	min	max	3D RMS
$dz=0$																
# Images																
3	0.498	0.295	0.725	0.791	0.470	0.131	0.887	0.755	0.977	0.654	1.300	1.435	0.931	0.322	1.674	1.418
6	0.533	0.121	1.075	1.397	0.553	0.044	1.311	1.449	0.958	0.249	1.814	2.642	0.994	0.055	2.295	2.762
$dz=1$																
# Images																
3	1.021	0.797	1.240	1.615	1.031	0.628	1.411	1.624	1.596	1.157	2.020	2.567	1.580	0.815	2.450	2.566
6	1.039	0.494	1.721	3.554	1.050	0.332	1.948	3.609	1.578	0.634	2.585	5.190	1.626	0.386	3.060	5.323
$dz=2$																
# Images																
3	1.902	1.641	2.164	3.042	1.942	1.485	2.360	3.095	2.581	2.105	3.056	4.138	2.560	1.763	3.436	4.109
6	1.968	1.331	2.683	6.730	1.993	1.143	2.935	6.846	2.509	1.522	3.557	8.399	2.467	1.193	3.855	8.355
$dz=3$																
# Images																
3	2.876	2.616	3.136	4.578	2.918	2.460	3.334	4.635	3.566	3.083	4.040	5.693	3.531	2.731	4.403	5.638
6	2.933	2.297	3.645	9.922	2.957	2.113	3.898	10.029	3.461	2.394	4.533	11.605	3.501	2.040	5.000	11.718
$dz=4$																
# Images																
3	3.848	3.592	4.103	6.113	3.893	3.434	4.310	6.177	4.550	4.067	5.033	7.250	4.528	3.734	5.387	7.215
6	3.912	3.269	4.630	13.173	3.923	3.082	4.859	13.201	4.432	3.359	5.507	14.824	4.469	2.988	5.970	14.932
$dz=5$																
# Images																
3	4.795	4.534	5.057	7.609	4.871	4.406	5.291	7.724	5.536	5.049	6.022	8.808	5.519	4.710	6.376	8.767
6	4.885	4.242	5.597	16.399	4.899	4.063	5.827	16.468	5.405	4.330	6.482	18.051	5.440	3.959	6.939	18.149

in Figure 3. This 3D representation of the left ventricle combined with the electrophysiological data was obtained using a full perspective camera model.

Table 2 presents the values for the 3D monoplane analysis using three and six consecutive images. Once again we validated our algorithm using various initial depth estimates  $dz \in [0-5]$  mm. We performed two-view reconstruction at each image frame beginning with the diastolic image allowing us to have ground truth for the true 3D displacements and 3D reconstruction between consecutive images. We also determined that the minimum average 3D displacement between consecutive images for the twenty datasets was 0.10 mm and that the maximum average 3D displacement of the catheter

electrodes was equal to 1.63 mm. This useful information allows us to hypothesize that the best results for our monoplane analysis should come from initial depth approximations in the range of  $dz \in [1, 2]$  mm.

From Table 2 we observe that the LAT views used for monoplane reconstruction provided better recovered displacements and lower 3D accumulated RMS values when compared to the PA view. The 3D RMS values increased as we doubled the number of consecutive images used in the algorithm. The algorithm produced 3D RMS errors lower than 4.1 mm using  $dz \in [0, 1, 2]$  mm initial approximations when considering both the LAT and PA views and three successive images. When considering six consecutive images, a maximum RMS

error equal to 8.36 mm was obtained. Depending on the initial guess used to solve our equations our results are up to five times better than in [16] where we obtained 15 mm (posterior/anterior view) and 10 mm (left lateral

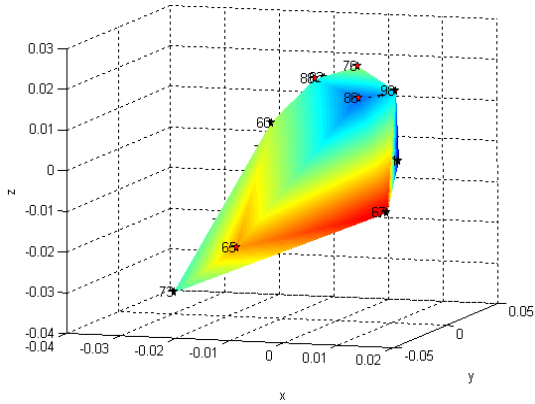


Fig 3. Three dimensional depiction of the left ventricle using a full perspective model and convex hull. The local activation times measured with the mapping catheter and colored regions having the same local activation times are superimposed on the model. The star (\*) symbol represents the tip electrode of the mapping catheter which was moved to 20 different locations.

view) depth errors for the ablation catheter tip electrode using a single C-arm image. Even when using six consecutive images and an initial depth error of  $dz = 3$  mm we obtain 3D accumulated RMS errors of about 10 mm in LAT and 12 mm in PA views respectively which is again more accurate than [16]. Nevertheless, our method relies on initial *a priori* 3D information of the catheter, this implies more preprocessing time. Also, we could not attain our goal of estimating accurately the 3D coordinates of the electrodes across all C-arm images in a cardiac cycle. Depending on the acquisition rate, the total number of images making up a cardiac cycle can be at most 15. We thus conclude that we can recover adequately the rigid movement of the electrodes up to six consecutive images and that nonrigid equations should be incorporated in our proposed monoplane algorithm.

The clinical strategy is to provide the electrophysiologist with a real time representation of the activation times during catheter ablation procedures, as well as the 3D volume of the heart chamber obtained by a mapping catheter. As an addition to the CARTO™ technology, we register 3D isochronal information directly on the 2D C-arm images. Figure 4 and Figure 5 show the fusion of the 3D activation model with the fluoroscopic images. For both figures, the image depicts the mapping catheter tip electrode with a local activation time of 73 milliseconds. Figure 4 shows two local

activation times that are inaccurate (56 ms and 61 ms); similarly in Figure 5, we observe two local activation times (79 ms and 94 ms) which are no longer part of the convex hull before projection. The smallest activation times are positioned

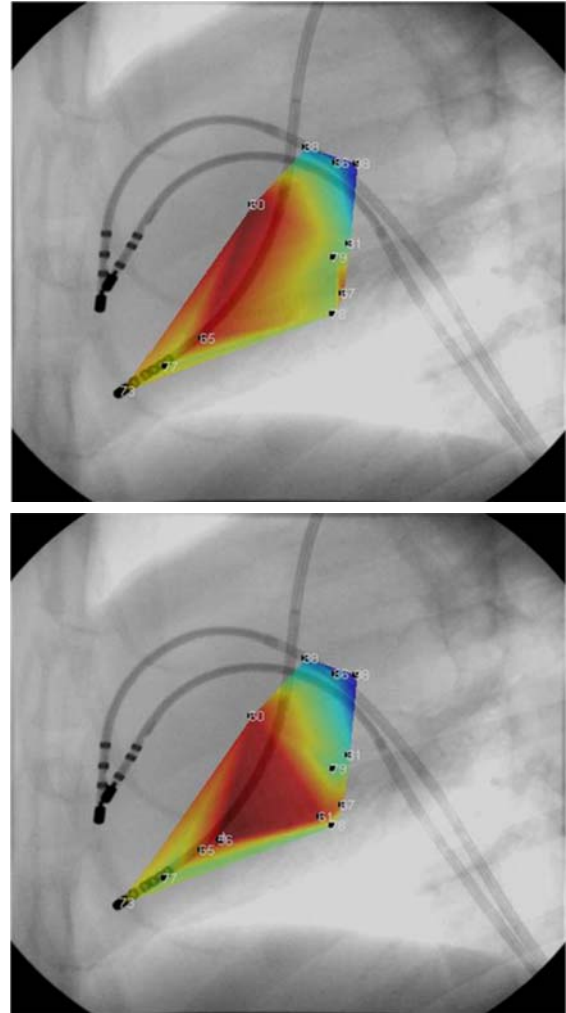


Fig 4. Fusion of electrophysiology on left lateral images of a mongrel dog. (Top) Visible electrophysiological data on exact depth coordinates of the ablation catheter. (Bottom) Erroneous electrophysiological data on estimated depth coordinates using our monoplane three view analysis.

closer to the pacing catheter in the left ventricle, as would be expected.

Even though monoplane reconstruction yielded depth errors, the isochronal maps were close enough to ground truth which raises promise for our monoplane technique. We are optimistic that in the near future this method can be validated on actual patient data. When compared to [1-13], our method uses readily available intraoperative information in order to calculate the 3D coordinates of the mapping catheters in real time. We use this

information to estimate catheter electrode positions in a single view sequence and determine the feasibility of

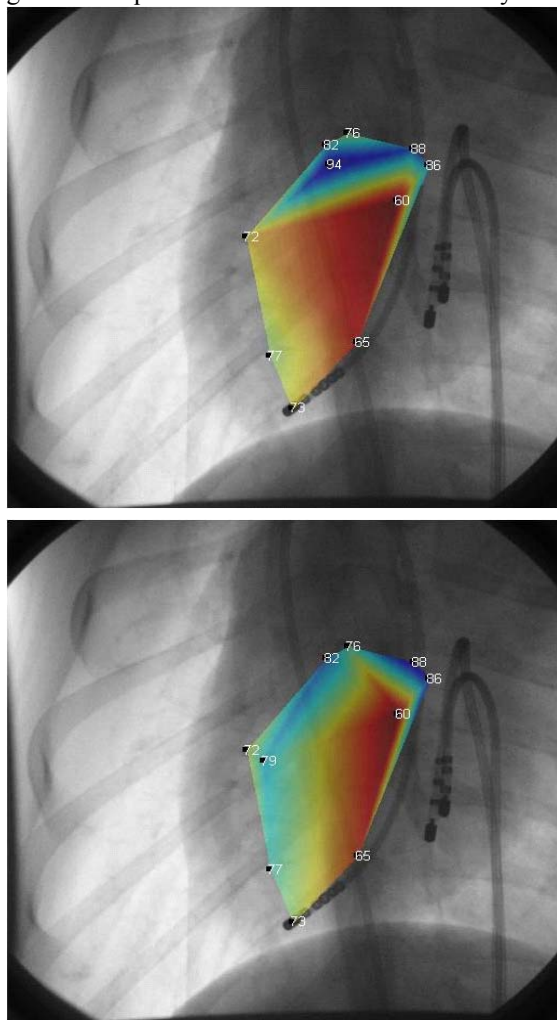


Fig 5. Fusion of electrophysiology on posterior/anterior images of a mongrel dog. (Top) Visible electrophysiological data on exact depth coordinates of the ablation catheter. (Bottom) Erroneous electrophysiological data on estimated depth coordinates using our monoplane three view analysis.

monoplane 3D reconstruction as adequate when analyzing rigid moving objects. When compared to the CARTO™ system we are cost effective by making use of standard catheters and existing technology (i.e. monoplane C-arm fluoroscope), we attempt to detect more than one electrode across monoplane C-arm sequences, and we fuse 2D isochronal maps directly in the fluoroscopy images.

#### 4. Conclusion

A novel method to estimate the depth of the mapping catheter was presented. By assuming that *a priori* 3D

model is present at a first time instant, we can estimate the depth of the catheter by making use of consecutive single view fluoroscopy images. Our feasibility study provided results that were an improvement of up to 50% when compared to the two-view projection method developed in [16] which use only a single image. Activation times were fused directly on the C-arm images in order to illustrate the evolution of cardiac electrical propagation. By exploiting spatial and projective information using only single plane sequences, we aim to decrease overall intervention time and still maintain high level accuracy when predicting the depth position of the mapping catheter. Future work will focus on adding additional nonrigid constraints to our monoplane equation which capture the inherent movement of the heart. With some additional work, we are confident that our proposed analysis can be adapted to meet the demands of MRI/C-arm or CT/C-arm registration in providing a navigational aid when treating any sustained arrhythmia.

#### Acknowledgments

The author would like to thank Dr. Pierre Savard, from École Polytechnique, Canada, for providing the image datasets.

#### References

- [1] Lardo AC, Halperin H, Yeung C, Jumrussirikul P, Atalar E, McVeigh E (1998) Magnetic resonance guided radiofrequency ablation: Creation and visualization of cardiac lesions. *Med Image Comput Assist Interv Int Conf* 1496:189-196
- [2] Lardo A, McVeigh E, Jumrussirikul P et al (2000) Visualization and temporal/spatial characterization of cardiac radiofrequency ablation lesions using magnetic resonance imaging. *Circulation* 102(6): 698-705
- [3] Chu E, Fitzpatrick A, Chin M et al (1994) Radiofrequency catheter ablation guided by intracardiac echocardiography. *Circulation* 89(3):1301-1305
- [4] Sun Y, Kadoury S, Li Y et al (2007) Image guidance of intracardiac ultrasound with fusion of pre-operative images. *MICCAI* 1: 60-67
- [5] Razavi R, Hill DLG, Keevil SF et al (2003) Cardiac catheterization guided by MRI in children and adults with congenital heart disease. *Lancet* 362(9399): 1877-1882
- [6] Kato R (2003) Pulmonary vein anatomy in patients undergoing catheter ablation of atrial fibrillation: lessons learned by use of magnetic resonance imaging. *Circulation* 107(15): 2004-2010
- [7] Lacomis JM, Wigginton W, Fuhrman C et al (2003) Multi-detector row CT of the left atrium and pulmonary veins before radio-frequency catheter ablation for atrial fibrillation. *Radiographics* 23: S35-S48

- [8] Tilg B, Fischer G, Modre R et al (2003) Electrocardiographic imaging of atrial and ventricular electrical activation. *Med Image Anal* 7(3): 391-398
- [9] Modre R, Tilg B, Fischer G, Wach P (2001) An iterative algorithm for myocardial activation time imaging. *Comput Methods Programs Biomed* 64(1): 1-7
- [10] Ghanem R, Jia P, Rudy Y (2003) Heart-surface reconstruction and ECG electrodes localization using fluoroscopy, epipolar geometry and stereovision: Application to noninvasive imaging of cardiac electrical activity. *IEEE Trans Med Imag* 22(10):1307-1318
- [11] Ramanathan C, Ghanem R, Jia P, Rudy K (2004) Noninvasive electrocardiographic imaging for cardiac electrophysiology and arrhythmia. *Nature Med* 10(4): 422-428
- [12] De Buck S, Maes F, Ector J et al (2005) An augmented reality system for patient-specific guidance of cardiac catheter ablation procedures. *IEEE Trans Med Imaging* 24:1512-1524
- [13] Cristoforetti A, Masè M, Faes L et al (2007) A stochastic approach for automatic registration and fusion of left atrial electroanatomic maps with 3D CT anatomical images. *Phys Med Biol.* 52(20):6323-6337
- [14] Tops LF, Bax JJ, Zeppenfeld K et al (2005) Fusion of multislice computed tomography imaging with three-dimensional electroanatomic mapping to guide radiofrequency catheter ablation procedures. *Heart Rhythm* 2: 1076-81
- [15] Sra J (2005) Registration of three dimensional left atrial images with interventional systems. *Heart* 91:1098-1104
- [16] Fallavollita P, Savard P, Sierra G (2004) Fluoroscopic navigation to guide RF catheter ablation of cardiac arrhythmias. *IEEE 26th Annual International Conference of the Engineering in Medicine and Biology Society* 1:1929-1932
- [17] Mariottini GL, Prattichizzo D (2005) EGT for Multiple View Geometry and Visual Servoing. *Robotics and Vision with Pinhole and Panoramic Cameras. IEEE Robotics and Automation Magazine* 12(4):26-39
- [18] Fallavollita P, Cheriet F (2006) Towards an automatic coronary artery segmentation algorithm. *IEEE 28th Annual International Conference of the Engineering in Medicine and Biology Society* 3037-3040
- [19] Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math* 11(2):431-441
- [20] Fallavollita P, Cheriet F (2008) Optimal 3D Reconstruction of Coronary Arteries for 3D Clinical Assessment. *Comput Med Imaging Graph*, 32(6):476-487

**Pascal Fallavollita** received his B.Eng. degree in Mechanical Engineering at McGill University in 2002 and Ph.D degree in Biomedical Engineering at École Polytechnique de Montréal in 2008. He is currently a postdoctoral fellow at the School of Computing at Queen's University, Canada. His main research interests lie in the areas of medical image processing and computer aided surgery. His clinical focus lies in cardiac arrhythmia ablation, coronary angiography and prostate brachytherapy procedures.

# Parallel Computation of Finite Element Navier-Stokes codes using MUMPS Solver

Mandhapati P. Raju

<sup>1</sup> Mechanical Engineering, Case Western Reserve University,  
Cleveland, Ohio 44106  
[raju192@gmail.com](mailto:raju192@gmail.com)

## Abstract

The study deals with the parallelization of 2D and 3D finite element based Navier-Stokes codes using direct solvers. Development of sparse direct solvers using multifrontal solvers has significantly reduced the computational time of direct solution methods. Although limited by its stringent memory requirements, multifrontal solvers can be computationally efficient. First the performance of MULTifrontal Massively Parallel Solver (MUMPS) is evaluated for both 2D and 3D codes in terms of memory requirements and CPU times. The scalability of both Newton and modified Newton algorithms is tested.

**Key words:** *finite element, MUMPS solver, distributed computing, Newton method.*

## 1. Introduction

Discretization of Navier-Stokes equations involves a large set of non-linear equations, requiring high computing power in terms of speed and memory. The resulting set of weak form (algebraic) equations in such problems may be solved either using a direct solver or an iterative solver. The direct solvers are known for their generality and robustness. The direct solution methods generally involve the use of frontal algorithms [1] in finite element applications. The advent of multifrontal solvers [2] has greatly increased the efficiency of direct solvers for sparse systems. They make full use of the high computer architecture by invoking level 3 Basic Linear Algebra Subprograms (BLAS) library. Thus the memory requirement is greatly reduced and the computing speed greatly enhanced. Multifrontal solvers have been successfully used both in the context of finite volume problems [3-5], finite element problems [6] and in power system simulations [7-9]. The disadvantage of using direct solvers is that the memory size increases much more rapidly than the problem size itself [6]. To circumvent this problem, out-of-core multifrontal solvers [10] have been developed which has the capability of storing the factors on the disk during factorization. Another viable alternative

is to use direct solvers in a distributed computing environment.

The system of non-linear equations obtained from the discretization of Navier-Stokes equations is usually solved using a Newton or a Picard algorithm. Newton algorithms are known for their quadratic convergence behavior. When the initial guess is close to the final solution, Newton achieves quadratic convergence. In this paper, only the Newton algorithm is used. In using direct solvers, factorization of the left hand side matrix is the most time consuming step. To avoid factorization during every iteration, a modified Newton is used in which the factorization is done only during the first iteration. The left side matrix evaluated for the first iteration is retained and is not changed during the subsequent iterations. Only the right hand side matrix is updated during each iteration step. The right hand side vector is appropriately modified to give the final converged solution. Since the factorization is done only during the first iteration, the subsequent iterations are extremely cheap. It usually requires more number of iterations to obtain the overall convergence. So there is a tradeoff between the computational time per iteration and the number of iterations to obtain the final convergence. Although the convergence rate is lower compared to the Newton iteration, the savings in computational time per iteration is so high that it can more than compensate the decrease in the convergence rate.

MUMPS [11-13] and SUPERLU [14] are amongst the fastest parallel general sparse direct solvers that are available under public domain software. A detailed description of the various features and algorithms employed in these packages can be found in [15]. MUMPS is found to be much faster compared to SUPERLU, although its scalability is low compared to that of SUPERLU. In this paper, parallelization is achieved using a MULTifrontal Massively Parallel Solver (MUMPS) on a distributed environment using MPI. The linear system of equations is evaluated on different processors

corresponding to the local grid assigned to the processor. The right hand side vector is assembled on the host processor and is input to the MUMPS solver. On exit from the MUMPS solver, the solution is assembled centrally on the host processor. This solution is then broadcast to all the processors. In the context of modified Newton algorithm, the LU factors evaluated during the first iteration are reused and the solution of the linear system with the new right hand side vector is solved. The performance of the solver in terms of scalability and memory issues for both two-dimensional and three-dimensional problems are discussed in detail.

## 2. Mathematical Formulation

The governing equations for laminar flow through a two-dimensional rectangular duct are presented below in the non-dimensional form.

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \quad (1)$$

$$\frac{\partial}{\partial x}(u^2) + \frac{\partial}{\partial y}(uv) = -\frac{\partial p}{\partial x} + \frac{\partial}{\partial x}\left(\frac{2}{Re}\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{Re}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\right), \quad (2)$$

and

$$\frac{\partial}{\partial x}(uv) + \frac{\partial}{\partial y}(v^2) = -\frac{\partial p}{\partial y} + \frac{\partial}{\partial x}\left(\frac{1}{Re}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\right) + \frac{\partial}{\partial y}\left(\frac{2}{Re}\frac{\partial v}{\partial y}\right), \quad (3)$$

where  $u, v$  are the  $x$  and  $y$  components of velocity,  $p$  is the pressure. The bulk flow Reynolds number,  $Re = \rho U_0 D / \mu$ ,  $U_0$  being the inlet velocity,  $\rho$  the density,  $L$  the channel length, and  $\mu$  is the dynamic viscosity. Velocities are non-dimensionalized with respect to  $U_0$ , pressure with respect to  $\rho U_0^2$ .

The boundary conditions are prescribed as follows:

(1) Along the channel inlet:

$$u = 1; v = 0. \quad (4)$$

(2) Along the channel exit:

$$p = 0; \frac{\partial u}{\partial x} = 0; \frac{\partial v}{\partial x} = 0. \quad (5)$$

(3) Along the walls:

$$u = 0; v = 0. \quad (6)$$

The governing equations for laminar flow through a three-dimensional rectangular duct are presented below in the non-dimensional form. In three-dimensional

calculations, instead of the primitive  $u, v, w, p$  formulation, penalty approach is used to reduce the memory requirements.

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \quad (7)$$

$$\begin{aligned} \frac{\partial}{\partial x}(u^2) + \frac{\partial}{\partial y}(uv) + \frac{\partial}{\partial z}(uw) = & \lambda \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) + \frac{\partial}{\partial x}\left(\frac{2}{Re}\frac{\partial u}{\partial x}\right) \\ & + \frac{\partial}{\partial y}\left(\frac{1}{Re}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\right) + \frac{\partial}{\partial z}\left(\frac{1}{Re}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right)\right), \end{aligned} \quad (8)$$

$$\begin{aligned} \frac{\partial}{\partial x}(uv) + \frac{\partial}{\partial y}(v^2) + \frac{\partial}{\partial z}(vw) = & \lambda \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) + \frac{\partial}{\partial x}\left(\frac{1}{Re}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\right) \\ & + \frac{\partial}{\partial y}\left(\frac{2}{Re}\frac{\partial v}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{1}{Re}\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right)\right), \end{aligned} \quad (9)$$

and

$$\begin{aligned} \frac{\partial}{\partial x}(uw) + \frac{\partial}{\partial y}(vw) + \frac{\partial}{\partial z}(w^2) = & \lambda \frac{\partial}{\partial z}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) + \frac{\partial}{\partial x}\left(\frac{1}{Re}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right)\right) \\ & + \frac{\partial}{\partial y}\left(\frac{1}{Re}\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right)\right) + \frac{\partial}{\partial z}\left(\frac{2}{Re}\frac{\partial w}{\partial z}\right). \end{aligned} \quad (10)$$

where  $u, v, w$  are the  $x, y$  and  $z$  components of velocity,.

The bulk flow Reynolds number,  $Re = \rho U_0 D / \mu$ ,  $U_0$  being the inlet velocity,  $\rho$  the density,  $L$  the channel length,  $\mu$  is the dynamic viscosity and  $\lambda$  is the penalty parameter. Velocities are non-dimensionalized with respect to  $U_0$ .

The boundary conditions are prescribed as follows:

(1) Along the channel inlet:

$$u = 1; v = 0; w = 0. \quad (11)$$

(2) Along the channel exit:

$$\frac{\partial u}{\partial x} = 0; \frac{\partial v}{\partial x} = 0; \frac{\partial w}{\partial x} = 0. \quad (12)$$

(3) Along the walls:

$$u = 0; v = 0; w = 0. \quad (13)$$

## 4. Newton's Algorithm

The set of non-linear equations obtained by the discretization of Galerkin Finite element formulation is solved using Newton's iterative algorithm.

Let  $\underline{X}^{(k)}$  be the available vector of field unknowns for the  $k^{\text{th}}$  iteration. Then the update for the  $(k+1)^{\text{st}}$  iteration is obtained as

$$\underline{X}^{(k+1)} = \underline{X}^{(k)} + \alpha \delta \underline{X}^{(k)}, \quad (14)$$

where  $\alpha$  is an under-relaxation factor, and  $\delta \underline{X}^{(k)}$  is the correction vector obtained by solving the linearized system

$$[J] \{ \delta \underline{X}^{(k)} \} = - \{ \underline{R}_X \}^{(k)}. \quad (15)$$

Here,  $[J]$  is the Jacobian matrix,

$$[J] = \frac{\partial \underline{R}_X^{(k)}}{\partial \underline{X}^{(k)}}. \quad (16)$$

and  $\{ \underline{R}_X \}^{(k)}$  is the residual vector. Newton's iteration is continued till the infinity norm of the correction vector  $\delta \underline{X}^{(k)}$  converges to a prescribed tolerance of  $10^{-6}$ .

The modified Newton's algorithm employs the jacobian calculated during the first iteration repeatedly for all the successive iterations. The jacobian is not updated. This would reduce the rate of convergence of the Newton algorithm. Since the jacobian is not updated, factorization can be skipped during all the subsequent iterations.

## 5. MUMPS Solver

The most time consuming part is the solution of the set of linear equations. To solve these linear systems, we use a robust parallel solver MUMPS. It employs multifrontal techniques to solve the set of linear equations on parallel computers on a distributed environment using MPI. It relies on Level II and Level III optimized BLAS routines. It requires SCALAPACK and PBLACS routines. In this paper, the vendor optimized INTEL Math Kernel Library is used. The Software is written in Fortran 90 and has a C interface available.

The solution of the set of linear equations takes place in 3 essential steps

- (i) Analysis step: MUMPS offers various built in ordering algorithms and interface to external packages such as PORD [16] and METIS [17]. The Matrix is analyzed to determine an ordering and a mapping of the multifrontal computational graph is constructed. This symbolic information is passed on from the host to all the other processors.
- (ii) Factorization: Based on the symbolic information, the algorithm tries to construct several dense sub matrices that can be processed in parallel. The numerical factorization is carried out during this step.
- (iii) Solution step: Using the right hand side vector, the solution vector is computed using the distributed factors.

All these steps can be called separately or as a combination of each other. This can be exploited to save some computational effort during the solution of subsequent iterations in the solution of a set of nonlinear equations. For example if the structure of the matrix does not change during every iteration, the analysis step can be skipped after evaluating once. Similarly, if the left hand matrix does not change, both the analysis and the factorization steps can be skipped.

## 6. Parallel Implementation

The MUMPS solver is implemented using the MPI library, which makes the code very portable and usable on both, shared and distributed memory parallel computers. The parallelization is done internally in the code. The calling program should also be in a parallel environment to call the code. In the present formulation, each element is assigned to particular processor such the elements are equally (or almost equally) distributed amongst all the processors. The computation of the matrix coefficients and the right hand side vector are done in parallel corresponding to the set of local elements. Evaluation of the Jacobian matrix and the right hand side vector in a parallel environment is crucial for problems, which consume lot of time for the evaluation of matrix coefficients.

During the progress of overall iterations, the different set of linear equations obtained during every iteration is solved by successive calls to the MUMPS. For the modified Newton's algorithm, the left hand matrix remains the same (numerically). So both the analysis and the factorization steps are skipped during the subsequent iterations. Since the factorization is most costly step, it leads to a significant amount of savings in time for the subsequent iterations. The performance of Newton and Modified Newton's method is tested.

While implementing the finite element code in a parallel environment with the MUMPS code, the element matrix entries are calculated locally on each of the processors. Although the facility for element matrix input is available, only the option of centralized element entry is available in the current versions of MUMPS solver. To facilitate distributed matrix input (necessary for improving the parallel efficiency), the local element entries are converted into sparse matrix triplet entries in coordinate format and are input in a distributed fashion (using ICNTL(5) = 0 and ICNTL(18) = 3). There will be lot of duplicate entries due to contribution of all the neighboring elements at a given grid point. MUMPS solver automatically sums up all the duplicate entries. Different ordering can be chosen by using different values for ICNTL(7). The different ordering options that are available within MUMPS solver are (i) Approximate minimum degree (AMD), (ii)

Approximate minimum fill (AMF), (iii) PORD, (iv) METIS, (v) Approximate Minimum degree with automatic quasi-dense row detection (QAMD).

## 7. Performance of MUMPS Solver

Preliminary experiments have been done to study the effect of different ordering routines on the performance of MUMPS solver both in terms of memory and computational time. Table 1 shows the performance of different ordering routines for two-dimensional codes. Table 1 shows the comparison of different ordering routines for a 300x300 mesh using 12 processors. Results indicate the both PORD and METIS perform well in minimizing the computational time requirements but METIS performs well in terms of memory requirements. Based on this observation, METIS ordering is used for all subsequent calculations. Table 2 shows the performance of Newton's and modified Newton's method using MUMPS solver for a two-dimensional channel flow problem. Results indicate that modified Newton's method performs better than Newton's method. This is due to the fact that in modified Newton's method, factorization is done only during the first iteration. During the subsequent iterations, factorization is skipped and only the solution step is performed. This decreases the convergence rate, thereby increasing the number of iterations to obtain convergence. However, the solution step being computationally inexpensive compared to the factorization step, the overall solution time is less compared to the Newton's step. However it should be noted that the memory requirement for the Newton and modified Newton's method are the same. Table 2 shows that the memory requirement does not vary linearly with the number of processors. It behaves erratically. Table 2 also shows that the MUMPS solver does not scale so well. Using 20 processors, the computational time is approximately halved compared to the computational time using 2 processors. It does not scale much beyond 6 processors. The scalability of MUMPS solver for two-dimensional problems is observed to be poor.

Table 3 shows the performance of MUMPS solver using different ordering routines for a three dimensional channel flow problem. The table shows that METIS ordering is a better choice both in terms of computational speed and memory requirement. Hence METIS is used for all the subsequent computations of three dimensional problems. For a 100x20x20 mesh, memory was not sufficient to run on 2 processors. The table shows that the scalability of MUMPS solver for three-dimensional problems is better than that for the two-dimensional problems. When the number of processors increased from 4 to 20, the computational time of Newton's method has reduced to a factor of 3.6 approximately, while that of modified

Newton's method has reduced to a factor of 3 approximately. The maximum memory requirement for a single processor has reduced to a factor of 4. The use of MUMPS solver for three dimensional problems seems to be promising. However, memory requirement is a serious limitation for solving three dimensional problems using direct solvers.

## 8. Conclusions

Finite element based Navier-Stokes codes are parallelized using MUMPS solver. Both Newton and modified Newton's algorithms are used. It is observed that modified Newton's method leads to savings in computational time compared to the Newton's algorithm. It is also observed that METIS ordering enhances the performance of MUMPS solver both for two-dimensional and three-dimensional problems. MUMPS solver does not scale well for two-dimensional problems but it scales better for three dimensional problems.

Table 1: Comparison of the performance of different orderings in MUMPS solver for a 300x300 mesh on 12 processors

Ordering	CPU time/ iteration (sec)	Memory (MB)	
		avg	max
PORD	16.6	1525	1640
METIS	5.1	1560	1820
AMD	5.5	1586	1923
AMF	6.7	1592	2000
QAMD	5.3	1603	2056

Table 2: Comparison of the performance of Newton and Modified Newton's methods using MUMPS solver for a 200x200 mesh

# of processors	Time to solve (Seconds)		Memory Requirements (MB)		Ordering
	Newton	Modified Newton	max memory on one processor	Total memory	
2	35.7	27.4	412	806	Metis
4	23	18.4	259	977	Metis
6	21	16.8	227	1082	Metis
8	20	15.3	178	1209	Metis
10	19.2	14.8	159	1394	Metis
12	18.6	14.6	157	1700	Metis
14	19.1	15	171	2039	Metis
16	18.4	13	156	2352	Metis
18	18.3	13	145	2534	Metis
20	18	12	141	2675	Metis

Table 3: Comparison of the performance of different orderings in MUMPS solver for a 100x20x20 mesh on 12 processors

Ordering	CPU time/iteration (sec)	Memory (MB)	
		Avg	max
PORD	41	1385	1612
METIS	38	1286	1400
AMD	105	2296	2496
AMF	93	1246	1425
QAMD	102	2296	2593

Table 4: Comparison of the performance of Newton and Modified Newton's methods using MUMPS solver for a 100x20x20 mesh

# of processors	Time to solve (Seconds)		Memory requirements (GB)		Ordering
	Newton	Modified Newton	max memory on one processor	Total memory	
6	147	77.2	1.6	8.6	metis
8	112	61.6	1	7.5	metis
10	91	53.4	1.5	13.9	metis
12	99	42.5	1.4	15.4	metis
14	68	37.2	1.45	17	metis
16	58	37	0.7	9.6	metis
18	52	34.8	0.63	10.2	metis
20	50	31.1	0.56	9.9	metis

## Acknowledgments

Author would like to thank the Ohio Supercomputing Centre for providing the computing facility.

## References

- [1] B.M. Irons, "A frontal solution scheme for finite element analysis," **Numer. Meth. Engg**, Vol. 2, 1970, pp. 5-32.
- [2] T. A. Davis, and I. S. Duff, "A combined unifrontal/multifrontal method for unsymmetric sparse matrices," **ACM Trans. Math. Softw.**, Vol 25, No 1, 1997, pp. 1-19.
- [3] M. P. Raju, and J. S. T'ien, "Development of Direct Multifrontal Solvers for Combustion Problems," **Numerical Heat Transfer, Part B**, Vol. 53, 2008, pp. 1-17.
- [4] M. P. Raju, and J. S. T'ien, "Modelling of Candle Wick Burning with a Self-trimmed Wick," **Comb. Theory Modell.**, Vol. 12, No. 2, 2008, pp. 367-388.
- [5] M. P. Raju, and J. S. T'ien, "Two-phase flow inside an externally heated axisymmetric porous wick," **Journal of Porous Media**, Vol. 11, No. 8, 2008, pp. 701-718.
- [6] P. K. Gupta and K. V. Pagalthivarthi, "Application of Multifrontal and GMRES Solvers for Multisize Particulate

Flow in Rotating Channels," **Prog. Comput Fluid Dynam.**, Vol. 7, 2007, pp. 323-336.

- [7] S. Khaitan, J. McCalley, Q. Chen, "Multifrontal solver for online power system time-domain simulation," **IEEE Transactions on Power Systems**, Vol. 23, No. 4, 2008, pp. 1727-1737.
- [8] S. Khaitan, C. Fu, J. D. McCalley, "Fast parallelized algorithms for online extended-term dynamic cascading analysis," **PSCE**, 2009, pp. 1-7.
- [9] J. McCalley, S. Khaitan, "Risk of Cascading outages", Final Report, PSrec Report, S-26, August 2007. [http://www.pserc.org/docsa/Executive\\_Summary\\_Dobson\\_McCalley\\_Cascading\\_Outage\\_S-2626\\_PSERC\\_Final\\_Report.pdf](http://www.pserc.org/docsa/Executive_Summary_Dobson_McCalley_Cascading_Outage_S-2626_PSERC_Final_Report.pdf).
- [10] J. A. Scott, Numerical Analysis Group Progress Report, RAL-TR-2008-001.
- [11] P. R. Amestoy, I. S. Duff and J.-Y. L'Excellent, "Multifrontal parallel distributed symmetric and unsymmetric solvers," **Comput. Methods in Appl. Mech. Eng.**, Vol. 184, 2000, pp. 501-520.
- [12] P. R. Amestoy, I. S. Duff, J. Koster and J.-Y. L'Excellent, "A fully asynchronous multifrontal solver using distributed dynamic scheduling," **SIAM Journal of Matrix Analysis and Applications**, Vol. 23, No. 1, 2001, pp 15-41.
- [13] P. R. Amestoy, A. Guermouche, J.-Y. L'Excellent and S. Pralet, "Hybrid scheduling for the parallel solution of linear systems," **Parallel Computing**, Vol. 32, No. 2, 2006, pp. 136-156.
- [14] S. L. Xiaoye and W. D. James, "A Scalable Distributed-Memory Sparse Direct Solver for Unsymmetric Linear Systems," **ACM Trans. Mathematical Software**, Vol. 29, No. 2, 2003, pp. 110-140.
- [15] A. Gupta, "Recent advances in direct methods for solving unsymmetric sparse systems of linear equations," **ACM transaction in Mathematical Software**, Vol. 28, No. 3, 2002, pp. 301-324.
- [16] J. Schulze, "Towards a tighter coupling of bottom-up and top-down sparse Matrix ordering methods," **BIT Numerical Mathematics**, Vol. 41, No. 4, 2001, pp. 800-841.
- [17] G. Karypis, and V. Kumar, "A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs," **SIAM J Scientific Computing**, Vol. 20, 1999, pp. 359-392.

## Mandhapati P. Raju

Mandhapati P. Raju completed his MS (2002-2004) and Ph.D (2004-2006) in Mechanical Engineering department at Case Western Reserve University, Cleveland, OH. Later he worked as Postdoctoral fellow in Case Western Reserve University during 2006-2008. Later he worked at Caterpillar Champaign simulation centre as a CFD analyst. Currently he is working as a Post Doctoral fellow in General Motors Inc. His research interests are combustion, porous media flows, multifrontal solvers, fuel cell and hydrogen storage. This work was done during his presence in Case Western Reserve University. He has published 5 journal papers in reputed international journals.

# The Folklore of Sorting Algorithms

Dr. Santosh Khamitkar<sup>1</sup>, Parag Bhalchandra<sup>2</sup>, Sakharam Lokhande<sup>2</sup>, Nilesh Deshmukh<sup>2</sup>

School of Computational Sciences,  
Swami Ramanand Teerth Marathwada University,  
Nanded (MS ) 431605, India  
Email : <sup>1</sup>s\_khamitkar@yahoo.com , <sup>2</sup>pub1976@rediffmail.com

## Abstract

The objective of this paper is to review the folklore knowledge seen in research work devoted on synthesis, optimization, and effectiveness of various sorting algorithms. We will examine sorting algorithms in the folklore lines and try to discover the tradeoffs between folklore and theorems. Finally, the folklore knowledge on complexity values of the sorting algorithms will be considered, verified and subsequently converged in to theorems.

**Key words:** *Folklore, Algorithm analysis, Sorting algorithm, Computational Complexity notations.*

## 1. Introduction

Folklore is the traditional beliefs, legend and customs, current among people. Where as a theorem is a general conclusion which has been proved [1]. In view of these definitions one might be tempted to conclude simply that the folk theorem is a general conclusion which is traditional and can be proved. The objective of this paper is to narrate the folklore on complexity issues of sorting algorithms and prove them. Accordingly, we shall attempt to provide a reasonable definition for complexity or criteria for folklore, followed by a detailed example illustrating ideas. The latter endeavor might take one of the two possible forms. We could take a piece of folklore and show that it is a theorem or take a theorem and show that it is folklore. Literature review in present context highlights that the unified theory regarding their folklore knowledge and in terms of theorems is found missing.

Since the dawn of computing, the sorting problem has attracted a great deal of research. Till date, Sorting algorithms are open problems and many researchers in past have attempted to optimize them with optimal space and time scale. Here Optimization was thought as the process to reduce the time complexity so as to cross at least  $O(n \log n)$  milestone. Many new techniques were proposed and till today fine refinement of them is in

progress. Every researcher attempted optimization in past has found stuck to his / her observations regarding sorting experiments and produced his/ her own results. Since these results were specific to software and hardware environment therein, we found abundance in the complexity analysis which possibly could be thought as the folklore knowledge. We tried to review this folklore knowledge in past research work and came to a conclusion that it was mainly devoted on synthesis, optimization, effectiveness in working styles of various sorting algorithms.

Our work is not mere commenting earlier work rather we are putting the knowledge embedded in sorting folklore with some different and interesting view so that students, teachers and researchers can easily understand them. Synthesis, in terms of folklore and theorems, hereupon aims to check what past researchers have thought is really the same they were saying and to analyze research on sorting algorithms in such a way to get united understanding of and support for the research results so that they can be used directly and defended in public. Since we have resolved to introduce no new technical material in this paper except having technical elaboration of folklore, and as researchers seem to be less familiar with folklore than with theorems, we prefer to deal with both approaches stated in above Para. We will present strong evidence to the effect that a particular folk can be converged in to theorem and vice versa. Folklore knowledge on complexity values of the sorting algorithms will be considered, verified and subsequently converged in to theorems.

This paper is important as Sorting algorithms are often prevalent in introductory computer science classes, where the abundance of algorithms for the problem provides a gentle introduction to a variety of core algorithm concepts. Sorting algorithms illustrate a number of important principles of algorithm design; some of them are also counterintuitive [2]. Efficient sorting is important to optimizing the use of other algorithms such as Binary search and merge algorithms that require sorted lists to

work correctly [2,3]. We can not deploy binary search if data is not pre sorted otherwise the search process may get trapped into a blind alley thereby exhibiting worst case complexity. Literature review has highlighted mainly interesting things which could be the parts of folklore and or theorems. Usually these folklore and theorem parts are always omitted by the faculties teaching these topics and students find them extremely difficult to understand. We will examine sorting algorithms in these lines and try and discover the tradeoffs between folklore and theorems. This paper covers different complexity aspects of basic sorting models, such as big O notation, best, worst and average case analysis, time-space tradeoffs, lower bounds, etc.

## 2. Origin of Folklore

When we look to develop or use a sorting algorithm on large problems, it is important to understand how long the algorithm might take to run. The time for most sorting algorithms depends on the amount of data or size of the problem. In order to analyze an algorithm, we try to find a relationship showing how the time needed for the algorithm depends on the amount of data. This is called the "complexity" of the algorithm. [4]

A simple sorting algorithm like bubble sort may have a high complexity, whereas an algorithm which is very complex in its organization such as shell sort, sometimes pays off by having a lower complexity in the sense of time needed for computation[5,6]. Besides running time analysis, it is seen that, factors other than the sorting algorithm selected to solve a problem, affect the time needed for a program. It is just because different people carrying out a solution to a problem may work at different speeds, even when they use the same sorting method, different computers work at different speeds. The different speeds of computers on the same program can be due to different "clock speeds", the rates at which steps in the program are executed by the machine and different "architectures," the way in which the internal instructions and circuitry of the computer are organized. Literature review shows that, the researchers have attempted for rigorous analysis of sorting algorithms and produced prolific comments [7, 8, 9] discovering such complexity dependent factors.

Consequently, analysis of sorting algorithm can not predict exactly how long it will take on a particular computer [5]. What analysis can do is tell us how the time needed depends on the amount of data. For example, we always come across folklore like, for an algorithm, when we double the amount of data, the time needed is also doubled, and in other words the time needed is proportional to the amount of data. The analysis of

another algorithm might tell us that when we double the amount of data, the time is increased by a factor of four, which is the time needed is proportional to the amount of data squared. The latter algorithm would have the time needed increase much more rapidly than the first.

When analyzing sorting algorithms, it often happens that the analysis of efficiency also depends considerably on the nature of the data. For example, if the original data set already is almost ordered, a sorting algorithm may behave rather differently than if the data set originally contains random data or is ordered in the reverse direction. Similarly, for many sorting algorithms, it is difficult to analyze the average-case complexity. Generally speaking, the difficulty comes from the fact that one has to analyze the time complexity for all inputs of a given length and then compute the average. This is a difficult task. Using the incompressibility method, we can choose just one input as a representative input. Via Kolmogorov complexity, we can show that the time complexity of this input is in fact the average-case complexity of all inputs of this length. Constructing such a "representative input" is impossible, but we know it exists and this is sufficient [10]

For these reasons, the analysis for sorting algorithms often considers separate cases, depending on the original nature of the data. The price of this generality is exponential complexity; with the result that many problems of practical interest are solvable better than folklore of sorting, but the limitations of computational capacity prevent them from being solved in practice. The increasing diversity in computing platforms motivates consideration of multi-processor environment. Literature review suggests that no folklore is found mentioned regarding complexity in multiprocessor environment.

Recently, many results on the computational complexity of sorting algorithms were obtained using Kolmogorov complexity (the incompressibility method). Especially, the usually hard average-case analysis is amenable to this method. A survey [10] shows such results about Bubble sort, Heap sort, Shell sort, Dobosiewicz-sort, Shaker sort, and sorting with stacks and queues in sequential or parallel mode. It is also found that the trade-off between memory and sorting is enhanced by the increase in availability of computer memory and the increase in processor speed. Currently, the prices of computer memory are decreasing. Therefore, acquiring larger memory configurations is no longer an obstacle, making it easier to equip a computer with more memory.

Similarly, every year there is an increase in computer speed. The Moore's Law, a computer-industry standard

stated that computer chips should double in power approximately every 18 months. The speed of most of today's computers is really quite remarkable. Even the "slower" ones are quite fast, and the actual processing doesn't take much time. Delays, usually encountered in "I/O", either to/from disk, the network or peripheral devices have been eliminated because of innovative interfacing standards, thereby boosting speed. If a computer has an 800 MHz processor, sounds like it was made in the early to mid 90's. If we look at most computers coming out today we will notice that the dual core processor - which is in fact two processors on the computer, each processor is clocking in at around 1.85 Ghz, which alone is more than two times as fast as that 800 Mhz processor. Then we have to multiply that by two since there are two cores. It is something like today's computers are working about 4-5 times faster than the earlier ones. Such increase in computer speed causes acceleration of sorting algorithms. Thus knowledge proved by some researcher for a sorting algorithm on its complexity can not be considered absolute as it may be increased or decreased analogously. Therefore piece of knowledge seen can be thought as folklore knowledge.

Above discussion is potentially the source for folklore. Many people, same work on different environment leading to different folklores. Some examples of folklore knowledge are 1) A quick sort is an effective, but more complex, sorting algorithm, 2) Shell Sort is much easier to code than Quicksort, and it is nearly as fast as Quicksort.

In next part of paper, we try to express sorting algorithm's behavior using the folklore and then try to prove it by formulating theorems.

### 3. Background Knowledge

Sorting is always carried out technically. In computer science and mathematics; we can formulate a procedure for sorting unordered array or a file. Such procedure is always governed by an algorithm; called Sorting Algorithm. It is highly expected that we must converge to the definition of sorting concept. Basically a sorting problem is to arrange a sequence of objects so that the values of their "key" fields are in "non-decreasing" sequence.

That is, given objects

$$O_1, O_2, \dots, O_n$$

With key values

$$K_1, K_2, \dots, K_n \text{ respectively,}$$

Arrange the objects in an order

$$O_{i1}, O_{i2}, \dots, O_{in},$$

Such that

$$K_{i1} \leq K_{i2} \leq \dots \leq K_{in}. \quad (1)$$

Generally, Computational complexity (worst, average and best behavior) of element comparisons in terms of the size of the unsorted list is considered for the analysis of the efficiency of sorting algorithm. The complexity notational terminology is covered in [2]. If the size of unsorted list is  $(n)$ , then for typical sorting algorithms, good behavior is  $O(n \log n)$  and bad behavior is  $\Omega(n^2)$ . The Ideal behavior is  $O(n)$ . Sort algorithms which only use an abstract key comparison operation always need  $\Omega(n \log n)$  comparisons in the worst case.

An important criterion used to rate sorting algorithms is their running time. Running time is measured by the number of computational steps it takes the sorting algorithm to terminate when sorting  $n$  records. We say that an algorithm is  $O(n^2)$  ("of order  $n$ -squared") if the number of computational steps needed to terminate as  $n$  tends to infinity increases in proportion to  $n^2$ . Literature review carried out in [11, 5] indicate the man's longing efforts to improve running time of sorting algorithm with respect to above core algorithm concepts.

Research on efficiency analysis [3, 2] use big O, omega and theta notation to give asymptotic upper, lower, and tight bounds on time and space complexity of sorting algorithms. They determine the time complexity of sorting algorithms, use a list of functions and order them according to asymptotic growth.

### 4. Some Folklore and Convergence in to Theorems

4.1 Folklore 1: The running time of a sorting algorithm grows if the length of the input sequence grows.

Proof 1: Proof can be obtained from the following reformulation of the sorting problem. For  $n$  records there are  $n!$  possible linear arrangements, only one of which is the correct, sorted set of records. ( $n! = 1 \times 2 \times 3 \times \dots \times [n - 2] \times [n - 1] \times n$ .) One may imagine these  $n!$  arrangements as the leaves of a binary decision tree. The process of sorting then involves starting at the root of the tree (the unordered initial list) and traveling down the nodes, choosing either the left or right node based on results of comparisons. (In this terminology, the "root" of a tree is at the top, and all the branches expand downward.) The maximum number of steps needed would correspond to the height of the tree, which is  $\log n!$ . Using Stirling's approximation of  $\log n!$ . For large  $n$ , this process would be  $O(n \log n)$ .

Thus the folklore 1 can be converged in to a theorem as, **Theorem 1: For n records, linear sorting will take a complexity  $O(n \log n)$ .** If the input sequence is presorted, compared to an unsorted sequence possibly less steps are sufficient.

Seldom, the theorem does not hold true. It might be the case that the keys of the objects fall within a known range. Let there be a controlled environment, where we produce the keys. For example, we might issue employee numbers in a given order and in a given sequence, or we might issue part numbers to items that we manufacture. Therefore, if the keys are unique and known to be in a given range, we might allow the key to be the location of the object (imagine placing the objects in a vector such that object with key K is in (vector-ref v k)). However the algorithms presented only rely on the fact that we can compare keys, to determine if one is less than another

4.2 Folklore 2: In general, we can't sort faster than  $O(n \log n)$ .

Proof 2: Consider Heap Sort or Bubble Sort or Quick Sort or Merge Sort or anything else like those, whether already invented or still to be invented. They all will take at least  $O(n \log n)$  steps [5] to sort  $n$  items, because of the reason that, they all work by comparing data items and then moving data items around based on the outcome of the comparisons [5]. To illustrate this, here is a piece of Heap Sort:

*If ( $a[c/2] < a[c]$ ) // if parent is smaller than child*

*Swap ( $c/2, c$ ); // trade their values*

Imagine running such a sort on these numbers:

3, 2, 4, 5, 1

The sort will compare the numbers and based on the outcomes of that comparison move the values around, with the net effect of all the moves being summarized by these arrows:

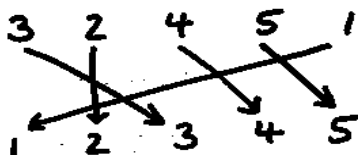


Fig.1. Suggested moves during swapping process

However it cannot run the same way on same but differently arranged inputs like 2, 3, 4, 5, 1, as output will

be misleading. Thus the execution must take a different path for every permutation of inputs.

We can visualize all these executions as an execution tree. Each node is some action, either an assignment or a comparison; the tree branches after each comparison; each path from the root to a leaf is one execution of the program; and the height of the tree is the (worst-case) running time of the program. If this is a sorting program that sorts the numbers 1 through  $n$  then execution must reach a different leaf for each of the  $n!$  permutations of  $1...n$ .

We know that a binary tree with  $L$  leaves has height at least  $\log L$  where the  $\log$  is base 2. Thus our execution tree has height at least  $\log(n!)$  which is  $O(n \log n)$ . To conclude, these sorting algorithms must take a different path of execution for every different permutation of the input values. There are lots of different permutations ( $n!$ ). To be able to run in that many different ways (i.e., for the execution tree to have that many leaves), the programs cannot run too quickly (i.e., the execution tree cannot be too shallow). We may want to conclude that sorting cannot be done in less than  $O(n \log n)$  time. This would be an overstatement because there is a subtle assumption in the above argument: that the program determines what actions to take (i.e., how to move values around) solely based on the outcome of comparisons between the input values. The lesson learned is, programs that *base their actions only on comparisons of the data* cannot beat the  $O(n \log n)$  barrier. To have a chance at sorting faster we have to avoid comparing data! , which is highly impossible. However, there are situation where we can say that we can sort faster than  $O(n \log n)$ . For example, Counting Sort and Radix Sort work faster when the range of values is limited .Thus the folklore 2 can be converged in to a theorem as, **Theorem 2: For n records, sorting will take minimum  $O(n \log n)$  time.**

4.3 Folklore 3: The exact number of steps required by insertion sort is given by the number of inversions of the sequence

Proof 3: Let  $a = a_0, \dots, a_{n-1}$  be a finite sequence. An inversion is a pair of index positions, where the elements of the sequence are out of order. Formally, an inversion is a pair  $(i, j)$ , where  $i < j$  and  $a_i > a_j$ . Example: Let  $a = 5, 7, 4, 9, 7$ . Then,  $(0, 2)$  is an inversion, since  $a_0 > a_2$ , namely  $5 > 4$ . Also,  $(1, 2)$  is an inversion, since  $7 > 4$ , and  $(3, 4)$  is an inversion, since  $9 > 7$ . There are no other inversions in this sequence. The inversion sequence  $v = v_0, \dots, v_{n-1}$  of a sequence  $a = a_0, \dots, a_{n-1}$  is defined by  $v_j = |\{(i, j) \mid i < j \wedge a_i > a_j\}|$  for  $j = 0, \dots, n-1$ .

The above sequence  $a = 5, 7, 4, 9, 7$  has the inversion sequence  $v = 0, 0, 2, 0, 1$ .

We now count the inversions  $(i, j)$  of a sequence  $a$  separately for each position  $j$ . The resulting value of  $v_j$  gives the number of elements  $a_i$  to the left of  $a_j$  which are greater than  $a_j$ . Obviously, we have  $v_i \leq i$  for all  $i = 0, \dots, n-1$ . If all  $v_i$  are 0, then and only then the sequence  $a$  is sorted. If the sequence  $a$  is a permutation, then it is uniquely determined by its inversion sequence  $v$ . The permutation  $n-1, \dots, 0$  has inversion sequence  $0, \dots, n-1$ .

Let  $a = a_0, \dots, a_{n-1}$  be a sequence and  $v = v_0, \dots, v_{n-1}$  be its inversion sequence. Then the number of steps  $T(a)$  required by insertion sort to sort sequence  $a$  is

$$T(a) = \sum_{i=0, \dots, n-1} v_i \tag{2}$$

Obviously, insertion sort takes  $v_i$  steps in each iteration  $i$ . Thus, the overall number of steps is equal to the sum of all  $v_i$ . This can be proved by an example: The following table shows the sequence  $a$  of the first example and its corresponding inversion sequence. For instance,  $v_5 = 4$  since 4 elements to the left of  $a_5 = 2$  are greater than 2 (namely 5, 7, 3 and 4). The total number of inversions is 17.

Table 1: Inversion Table

i	0	1	2	3	4	5	6	7
$a_i$	5	7	0	3	4	2	6	1
$v_i$	0	0	2	2	2	4	1	6

It is seen that insertion sort requires 17 steps to sort this sequence. Thus the folklore 3 can be converged in to a theorem as, **Theorem 3 : The sorting algorithm insertion sort has a worst case time complexity of  $T(n) = n \cdot (n-1) / 2$  comparison-exchange steps to sort a sequence of  $n$  data elements.**

4.4 Folklore 4: Any comparison sort type of algorithm requires lower bound comparisons in the worst case.

Proof 4: From the folklore 2, it suffices to determine the height of a decision tree in which each permutation appears as a leaf. Consider a decision tree of height  $h$  and  $l$  leaves corresponding to a comparison sort on  $n$  elements. Because each of the  $n!$  permutations of the input appear as some leaf, we have  $n! \leq l$ .

Since a binary tree of height  $h$  has no more than  $2^h$  leaves, we have to conclude that  $n! \leq 2^h$ , which by taking logarithms, implies  $h \geq \log_2(n!)$  (Since the log function is monotonically increasing).

Therefore,

$$\begin{aligned} h &\geq \log_2(n!) \\ &= \log_2(n \cdot n-1 \cdot n-2 \dots 2 \cdot 1) \\ &= \log n + \log(n-1) + \log(n-2) + \dots + \log 2 + \log 1 \\ &> \log n + \log(n-1) + \log(n-2) + \dots + \log(n/2) \\ &> \log(n/2) + \log(n/2) + \dots + \log(n/2) \\ &= n/2 \log(n/2) \\ &= n/2 \log n - n/2 \\ &= \Omega(n \log n) \end{aligned} \tag{3}$$

This folklore can be converged in to new theorem as, **Theorem 4: any comparison sort algorithm requires  $\Omega(n \log n)$  comparisons in the worst case.**

5. Some Convergence of Theorems into Folklore

Now we take a theorem and show that it is a piece of folklore. We know that time needed for these algorithms like selection sort varies like  $N^2$ . [12]. Let's deduce folklore from this.

5.1 Deduction 1:

Algorithms like selection sort, takes each unsorted list and goes through each item in order to identify the smallest. The first unsorted list consists of all  $N$  items in the original, and each successive unsorted list is one smaller, as the smallest item each time is placed in its correct position. So the first search for the smallest will take  $N$  operations, the next search for the smallest will take  $N-1$  operations, the next  $N-2$  operations, and so on until the last search for the smallest take 2 operations (we don't need to search when the list has only one item). Thus the time needed will be a constant times the following sum:

$$N + (N-1) + (N-2) + \dots + 3 + 2 + 1 \tag{4}$$

(We include the 1 to create a sum for which a standard formula applies; it adds only a small constant which does not affect the overall result.)

This sum, the sum of the first  $N$  integers, is known to be  $N(N+1) / 2$ , which is approximately  $N^2/2$ . Since we know that this is multiplied by some constant representing the time the operations take on a specific machine, we can see that the time needed for this algorithm varies like  $N^2$ , the square of the amount of data. Thus the time needed for the selection sort algorithm behaves like; each time the

size of the list is doubled, the amount of time needed is increased by a factor of four. **The folklore deduced is “Algorithms like Selection sort are quadratic in nature.”**

## 6. Conclusion

In this paper we have shown that the research on sorting algorithm has produced abundant results and can be thought as Folklore knowledge - momentarily belief among people. We proved that these results are not absolute as they are specific to some factors like person working, configuration of hardware, programming styles, whether parallelism is opted or not? Etc. Even with these uncertainties, we have analyzed sorting algorithms and presented strong evidence to the effect that a particular folk can be converged in to theorem and vice versa so as to show that folklore and converged theorems are interchangeable.

## References

- [1] Oxford Dictionary
- [2] Horowitz, E., Sahni, S, **Fundamentals of Computer Algorithms**, Computer Science Press, Rockville. Md.
- [3] Dr. D. E. Knuth, *The Art of Computer Programming*, 3rd volume, "**Sorting and Searching**", second edition.
- [4] Liu C. L., "Analysis of sorting algorithms", *Proceedings of Switching and Automata Theory*, **12th Annual Symposium, 1971**, East Lansing, MI, USA, pp 207-215.
- [5] Darlington, J, A synthesis of several sorting algorithms, *Acta Inf. II*, 1978, pp 1-30.
- [6] John Darlington, Remarks on "A Synthesis of Several Sorting Algorithms", **Springer Berlin / Heidelberg**, pp 225-227, Volume 13, Number 3 / March, 1980.
- [7] Talk presented by Dr. Sedgewick, "Open problems in the analysis of sorting and searching algorithms", at the **Workshop on the probabilistic analysis of algorithms**, Princeton, May, 1997.
- [8] A. Andersson & T. Hagerup, S. Nilsson, and R. Raman, "Sorting In Linear Time?" **Proceedings of the 27th Annual ACM Symposium on the Theory of Computing**, 1995.
- [9] Parag Bhalchandra, "Proliferation of Analysis of Algorithms with application to Sorting Algorithms", **M.Phil Dissertation**, VMRF, India, July 2009
- [10] Paul Vitanyi, "Analysis of Sorting Algorithms by Kolmogorov Complexity (A Survey)", appeared in **Entropy, Search, Complexity, Bolyai Society Mathematical Studies**, pp 209—232, Springer-Verlag, 2007
- [11] Richard Harter, Inefficient sort algorithms – "A Computer Environment for Beginners' Learning of Sorting Algorithms: Design and Pilot Evaluation", ERIC, **Journal Number 795978, Computers & Education**, v51 n2, pp708-723, Sep 2008
- [12] S Nilsson, "the Fastest Sorting Algorithm?" **Doctor Dobbs Journal**, 2000.

### Dr. S. D. Khamitkar, M.Sc. PhD

He is Associate Professor and has more than 14 years of teaching and research experience. He has published near about 6+ papers in international journals. He is research guide and currently ten students are working with him His interest area includes Network Security, Scientific computing, etc. He is life member of Indian Science Congress Association, ISTE.

### Parag Bhalchandra, Sakharam Lokhande, Nilesh Deshmukh, M.Sc., SET-NET, M.Phil

All of them are Assistant Professors and have registered for PhD research work. They have 8+ years teaching experience and have 2+ papers in international conferences. They are life members of ISCA, ISTE. Their interest lies in Algorithm Analysis, soft computing and web based developments.

# Color Image Clustering using Block Truncation Algorithm

Dr. Sanjay Silakari <sup>1</sup>, Dr. Mahesh Motwani <sup>2</sup> and Manish Maheshwari <sup>3</sup>

<sup>1</sup> Department of Computer Science and Engineering  
Rajiv Gandhi Technical University  
Bhopal, M.P., India  
*ssilakari@yahoo.com*

<sup>2</sup> Department of Computer Science and Engineering  
Jabalpur Engineering College  
Jabalpur, M.P., India  
*mahesh\_9@hotmail.com*

<sup>3</sup> Department of Computer Science & Application  
Makhanlal C. National University of Journalism and Communication  
Bhopal, M.P., India  
*manishbhom@yahoo.com*

## Abstract

With the advancement in image capturing device, the image data been generated at high volume. If images are analyzed properly, they can reveal useful information to the human users. Content based image retrieval address the problem of retrieving images relevant to the user needs from image databases on the basis of low-level visual features that can be derived from the images. Grouping images into meaningful categories to reveal useful information is a challenging and important problem. Clustering is a data mining technique to group a set of unsupervised data based on the conceptual clustering principal: maximizing the intraclass similarity and minimizing the interclass similarity. Proposed framework focuses on color as feature. Color Moment and Block Truncation Coding (BTC) are used to extract features for image dataset. Experimental study using K-Means clustering algorithm is conducted to group the image dataset into various clusters.

**Key words:** *Image features, Clustering, Color moments, BTC*

## 1. Introduction

The rapid progress in computer technology for multimedia system has led to a rapid increase in the use of digital images. Rich information is hidden in this data collection that is potentially useful in a wide range of applications like Crime Prevention, Military, Home Entertainment, Education, Cultural Heritage, Geographical Information System (GIS), Remote sensing, Medical diagnosis, and World Wide Web [1, 2]. Rich information is hidden in these data collection that is potentially useful. A major challenge with these fields is how to make use of this useful information effectively and efficiently. Exploring

and analyzing the vast volume of image data is becoming increasingly difficult.

The image database containing raw image data cannot be directly used for retrieval. Raw image data need to be processed and descriptions based on the properties that are inherent in the images themselves are generated. These inherited properties of the images stored in feature database which is used for retrieval and grouping. The strategy for earlier image retrieval system focused on “search-by-query”. The user provides an example image for the query, for which the database is searched exhaustively for images that are most similar.

Clustering is a method of grouping data objects into different groups, such that similar data objects belong to the same group and dissimilar data objects to different clusters [3,4]. Image clustering consists of two steps the former is feature extraction and second part is grouping. For each image in a database, a feature vector capturing certain essential properties of the image is computed and stored in a feature base. Clustering algorithm is applied over this extracted feature to form the group.

In this paper we propose a data mining approach to cluster the images based on color feature. Concept of color moment is extended to obtain the features and k\_means algorithm is applied to cluster the images. The rest of paper is organized as follows: In section two we provide overview of the previous work related to image retrieval and mining. Section three introduces the concept of 3.1 Color moments, 3.2 Block Truncation Coding Algorithm, 3.3 K-means clustering algorithm. In section four we

present the results of our experiments and finally section five concludes the paper.

## 2. Related Work

Feature extraction is the process of interacting with images and performs extraction of meaningful information of images with descriptions based on properties that are inherent in the images themselves. Color information is the most intensively used feature for image retrieval because of its strong correlation with the underlying image objects. Color Histogram [2] [5] [6] is the commonly and very popular color feature used in many image retrieval system. The mathematical foundation and color distribution of images can be characterized by color moments [7]. Color Coherence Vectors (CCV) have been proposed to incorporate spatial information into color histogram representation [8]. Texture refers to the presence of a spatial pattern that has some properties of homogeneity. Textures are replications, symmetries and combinations of various basic patterns or local functions, usually with some random variation. There are a number of texture features which have been used frequently liked Tamura Texture feature [5,6], Simultaneous Auto-Regressive (SAR) models[9], Gabor texture features[10] and Wavelet transform features[11,12].

Intelligently classifying image by content is an important way to mine valuable information from large image collection. [13] Explore the challenges in image grouping into semantically meaningful categories based on low-level visual features. The concept of fuzzy ID3 decision tree for image retrieval was discussed in [14]. ID3 is a decision tree method based on Shannon's information theory. Given a sample data set described by a set of attributes and an outcome, ID3 produces a decision tree, which can classify the outcome value based on the values of the given attributes like Color, Texture and Spatial Location. Image dataset were defined in 10 classes (concepts): grass, forest, sky, sea, sand, firework, sunset, flower, tiger and fur. At each level of the ID3 decision tree, the attribute with smallest entropy is selected from those attributes not yet used as the most significant for decision-making.

The SemQuery [15] approach proposes a general framework to support content-based image retrieval based on the combination of clustering and querying of the heterogeneous features. Given a query image, the SemQuery compares the features of query image to those of database images, resulting in a group of retrieved image sets based on individual feature classes. Database images were categorized into five categories of cloud, floral, leaves, mountain and water. Hierarchical clustering approach was performed on texture and color feature. Wavelet transforms extract the texture features while color

histograms method used for color feature. [16] Describe data mining and statistical analysis of the collections of remotely sensed image. Large images are partitioned into a number of smaller more manageable image tiles. Then those individual image tiles are processed to extract the feature vectors.

## 3. Proposed Work

An image is a spatial representation of an object and represented by a matrix of intensity value. It is sampled at points known as pixels and represented by color intensity in RGB color model. A basic color image could be described as three layered image with each layer as Red, Green and Blue as shown in fig 1.

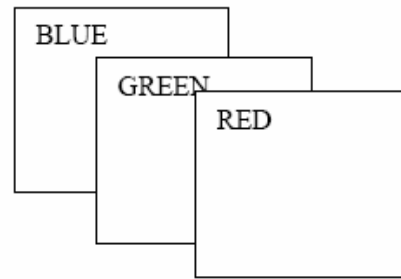


Fig. 1 Image Components

### 3.1 Color Feature Extraction

Color moments are measures that can be used differentiate images based on their features of color. The basis of color moments lays in the assumption that the distribution of color in an image can be interpreted as a probability distribution. Probability distributions are characterized by a number of unique moments (e.g. Normal distributions are differentiated by their mean and variance). It therefore follows that if the color in an image follows a certain probability distribution, the moments of that distribution can then be used as features to identify that image based on color. Stricker and Orengo [7] use three central moments of an image's color distribution in which  $p_{ij}^k$  is the value of the  $k$ -th color component of the  $ij$ -image pixel and  $P$  is the height of the image, and  $Q$  is the width of the image. They are Mean, Standard deviation and Skewness.

MOMENT 1 – Mean:

$$E_k = \frac{1}{PQ} \sum_{i=1}^P \sum_{j=1}^Q p_{ij}^k \quad (1)$$

Mean can be understood as the average color value in the image.

MOMENT 2 - Standard Deviation:

The standard deviation is the square root of the variance of the distribution.

$$SD_k = \text{SQRT} \left( \frac{1}{PQ} \sum_{i=1}^P \sum_{j=1}^Q (p_{ij}^k - E_k)^2 \right) \quad (2)$$

MOMENT 3 – Skewness:

$$S_k = \left( \frac{1}{PQ} \sum_{i=1}^P \sum_{j=1}^Q (p_{ij}^k - E_k)^3 \right)^{1/3} \quad (3)$$

Skewness can be understood as a measure of the degree of asymmetry in the distribution.

### 3.2 Block Truncation Algorithm

Steps in Block Truncation Coding Algorithm:

1. Split the image into Red, Green, Blue Components
2. Find the average of each component
  - Average of Red component
  - Average of Green component
  - Average of Blue component
3. Split every component image to obtain RH, RL, GH, GL, BH and BL images
 

RH is obtained by taking only red component of all pixels in the image which are above red average and RL is obtained by taking only red component of all pixels in the image which are below red average. Similarly GH, GL, BH and BL can be obtained.
4. Apply color moments to each splitted component i.e. RH, RL, GH, GL, BH and BL.
5. Apply clustering algorithm to find the clusters.

### 3.3 K Means Clustering Algorithm

K-means is one of the simplest unsupervised learning algorithms in which each point is assigned to only one particular cluster. The procedure follows a simple, easy and iterative way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The procedure consists of the following steps,

Step 1: Set the number of cluster k

Step 2: Determine the centroid coordinate

Step 3: Determine the distance of each object to the centroids

Step 4: Group the object based on minimum distance

Step 5: Continue from step 2, until convergence that is no object move from one group to another.

## 4. Experiments

The proposed scheme has been performed using a image database of 1000 images including 10 classes, which can be downloading from the website <http://wang.ist.psu.edu/iwang/test1.tar>. Each class has 100 images. Each image is of size 384\*256 pixels. The system is developed in Matlab. We define unichrome feature as values that are extracted from a single color layer of Red, Green and Blue.

Original Image (Bus Image)

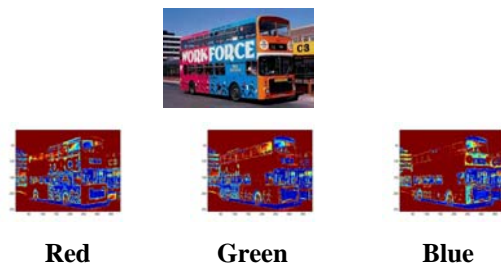


Fig. 2 RGB Components of Sample Image

The first part of evaluation computes color moments for each of the three color components. Each color component yields a feature vector of three elements as discussed in section 3.1 i.e. mean, standard deviation and skewness. Thus total nine feature vectors are calculated for one image.

In the second part apply Block Truncation Coding Algorithm discussed in section 3.2 over RH, RL, GH, GL, BH and BL. Thus total 18 feature vectors are calculated for one image.

Table 1: Recall and Precision using Color Moments

Classes	Recall	Precision
African People and villages	30	25.43
Beaches	47	40.17
Buildings	33	23.57
Buses	37	35.24
Dinosaurs	100	92.59
Elephants	32	35.56

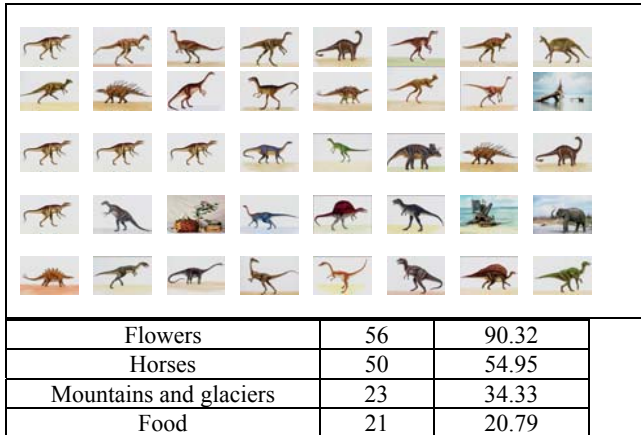


Fig. 3 Sample Dinosaurs Clusters using color moments

Based on commonly used performance measures in information retrieval, two statistical measures were computed to assess system performance namely Recall and Precision.

Recall consists of the proportion of target images that have been retrieved among all the relevant images in the database.

$$\text{Recall} = \frac{\text{Number of Relevant Images Retrieved}}{\text{Total Number of Relevant Images}} \quad (4)$$

Precision consists of the proportion of relevant images that are retrieved.

$$\text{Precision} = \frac{\text{Number of Relevant Images Retrieved}}{\text{Total Retrieved Images}} \quad (5)$$

Table 2: Recall and Precision using BTC

Classes	Recall	Precision
African People and villages	44	33.58
Beaches	42	42.42
Buildings	8	7.92
Buses	52	44.83
Dinosaurs	99	97.06
Elephants	39	44.83
Flowers	80	94.12
Horses	50	58.82
Mountains and glaciers	35	43.21
Food	25	22.32

Figure 3 showing the sample images in Dinosaurs cluster using color moments and Figure 4 showing the sample images in Flowers cluster using BTC algorithm.

Table 1 and Table 2 shows the values of recall and precision of each classes using color moments and BTC algorithms. Precision is maximized for dinosaur’s images in both the clustering algorithm. For maximum classes recall and precision of BTC is better than Color moment feature extraction.



Fig. 4 Sample Flowers Clusters using BTC

### 5. Conclusion

In image retrieval system, the content of an image can be expressed in terms of different features such as color, texture and shape. These low-level features are extracted directly from digital representations of the image and do not necessarily match the human perception of visual semantics. We proposed a framework of unsupervised clustering of images based on the color feature of image. Test has been performed on the feature database of color moments and BTC. K-means clustering algorithm is applied over the extracted dataset. Results are quite acceptable and showing that performance of BTC algorithm is better than color moments.

### References

- [1] H.J.Zhang et al., “Video Parsing, Retrieval and Browsing: an Integrated and Content-Based Solution”, Proc. ACM Multimedia 95, San Francisco, Nov 95
- [2] B.Furht, S.w.Smoliar, and H.J.Zhang, “Image and Video Processing in Multimedia systems, kluwer Academic Publishers, Norwell MA, 1995
- [3] Han and M.Kamber, “Data Mining concepts and Techniques”, Morgan Kaufmann Publishers, 2002
- [4] A.K.Pujari, “Data Mining Techniques”, University Press, 2001
- [5] Wayne Niblack, Ron Barber, William Equitz, Myron Flickner, Eduardo H. Glasman, Dragutin Petkovic, Peter Yanker, Christos Faloutsos, Gabriel Taubin: “The QBIC Project: Querying Images by Content, Using Color, Texture, and Shape”, Storage and Retrieval for Image and Video Databases (SPIE) 1993: 173-187

- [6] Alex Pentland, Rosalind W. Picard, Stan Sclaroff, "Photobook: Tools for Content-Based Manipulation of Image Databases", *Storage and Retrieval for Image and Video Databases (SPIE)* 1994: 34-47
- [7] M.Stricker and M.Orengo, "Similarity of color images", *Storage and Retrieval for Image and Video Databases III (SPIE)* 1995: 381-392
- [8] Greg Pass, Ramin Zabih, Justin Miller: *Comparing Images Using Color Coherence Vectors*. *ACM Multimedia* 1996: 65-73
- [9] Jianchang Mao, Anil K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models", *Pattern Recognition* 25(2): 173-188 (1992)
- [10] B. S. Manjunath, Wei-Ying Ma, "Texture Features for Browsing and Retrieval of Image Data", *IEEE Trans. Pattern Anal. Mach. Intell.* 18(8): 837-842 (1996)
- [11] Ingrid Daubechies, "The wavelet transform, time-frequency localization and signal analysis", *IEEE Transactions on Information Theory* 36(5): 961-1005 (1990)
- [12] Tianhorng Chang, C. C. Jay Kuo, "Texture analysis and classification with tree-structured wavelet transform", *IEEE Transactions on Image Processing* 2(4): 429-441 (1993)
- [13] Y. Uehara, S. Endo, S. Shiitani, D. Masumoto, and S. Nagata, "A computer-aided Visual Exploration System for Knowledge Discovery from Images", In *Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001)*, San Francisco, CA, USA, August, 2001.
- [14] Ying Liu<sup>1</sup>, Dengsheng Zhang<sup>1</sup>, Guojun Lu<sup>1</sup>, Wei-Ying Ma<sup>2</sup>, "Deriving High-Level Concepts Using Fuzzy-Id3 Decision Tree for Image Retrieval", *IEEE* 2005, pp 501-504
- [15] Gholamhosein Sheikholeslami, Wendy Chang, Aidong Zhang, "SemQuery: Semantic Clustering and Querying on Heterogeneous Features for Visual Data", *IEEE Trans. Knowl. Data Eng.* 14(5): 988-1002 (2002)
- [16] Krzysztof Koperski, Giovanni Marchisio, Selim Aksoy, and Carsten Tusk, "Applications of Terrain and Sensor Data Fusion in Image Mining", *IEEE* 2002, pp 1026-1028
- [17] H.B.Kekre, Tasneem Mirza, "Content Based Image Retrieval using BTC with Local Average Thresholding". In *Proceeding of ICCBIR 2008*, pp 5-9
- [18] <http://wang.ist.psu.edu/iwang/test1.tar>

# DAMQ-Based Schemes for Efficiently Using the Buffer Spaces of a NoC Router

Mohammad Ali Jabraeil Jamali, Ahmad Khademzadeh

Department of Computer Science, Islamic Azad University, Science & Research Branch, Tehran, Iran  
*m\_jamali@itrc.ac.ir* , *zadeh@itrc.ac.ir*

## Abstract

In this paper we present high performance dynamically allocated multi-queue (DAMQ) buffer schemes for fault tolerance systems on chip applications that require an interconnection network. Two or four virtual channels shared the same buffer space. On the message switching layer, we make improvement to boost system performance when there are faults involved in the components communication. The proposed schemes are when a node or a physical channel is deemed as faulty, the previous hop node will terminate the buffer occupancy of messages destined to the failed link. The buffer usage decisions are made at switching layer without interactions with higher abstract layer, thus buffer space will be released to messages destined to other healthy nodes quickly. Therefore, the buffer space will be efficiently used in case fault occurs at some nodes.

Key words: *Network on chip, Fault tolerance, DAMQS, DAMQAS, Buffer space, Odd-even routing algorithm.*

## 1. Introduction

Future system-on-chip (SoC) designs require predictable, scalable and reusable on-chip interconnect architecture to increase reliability and productivity. Current bus-based interconnect architectures are inherently non-scalable, less adaptable for reuse and their reliability decreases with system size. To overcome these problems, it has been proposed to build a message passing network for on-chip communication - network-on-chip (NoC).

Due to the constraints of being in a single chip, using an interconnection network on chip needs be restricted in terms of area. Thus, it is extremely important to design the schemes that require less hardware resources and still provide a good performance. Virtual channel multiplexing across a physical channel is extensively used to boost performance and avoid deadlock.

As virtual channels are not equally used in many applications, if they share a common buffer, the whole buffer space will be better utilized. In this paper we present schemes that are based on a dynamically allocated multi-queue (DAMQ) buffer. These schemes provide

similar performance as other statically allocated multiple-queue (SAMQ) buffers using less hardware and, therefore, requiring less hardware.

In order to improve the reliability of SoCs, their interconnect infrastructures must be designed such that fabrication and life-time faults can be tolerated. These irrecoverable faults influence the behavior of NoC fabrics and consequently degrade the system performance. Therefore, achieving on-chip fault tolerant communication is becoming increasingly important in presence of such permanent faults. A fault tolerant algorithm distinguishes from deterministic one according to the fact that it can provide an alternative path so that the message wouldn't be blocked by a faulty component [1], [2]. Consequently we investigate the applicability of partially adaptive algorithms to achieve a certain degree of fault tolerance in NoC communication fabrics.

The paper is organized as follows. In section 2, we review the fault tolerant odd-even algorithm in mesh-based NoC while in section 3, the DAMQ shared (DAMQS) and DAMQ all shared (DAMQAS) buffer schemes are discussed. Experimental results, that show the performance of the proposed approaches, are presented in Section 4 followed by conclusions in Section 5.

## 2. Fault Tolerant Odd-even Algorithm in Mesh-Based NoC

Fault tolerance is the ability of the network to function in the presence of component failure. The turn model is a well known partial adaptive routing algorithm, widely investigated for multi-processor environments [3]. The odd-even turn model facilitates deadlock-free routing in mesh network of a NoC.

The detail of odd-even turn model is explained with the help of Fig. 1. In the figure, the odd-even routing algorithm has been illustrated.

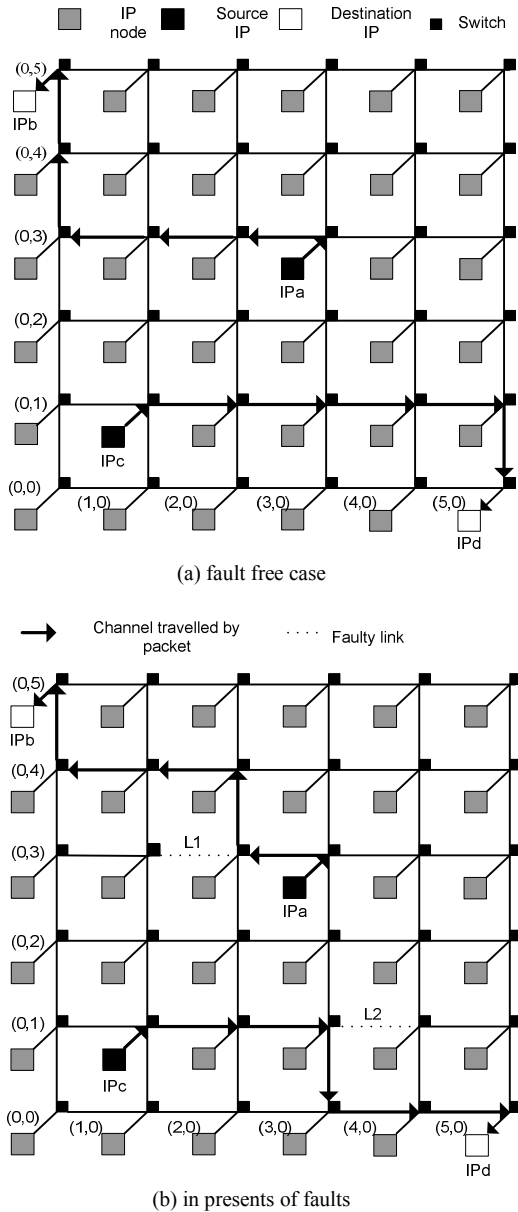


Fig. 1 A sample for Odd-even turn model algorithm.

In Fig. 1. (a), source IPa is trying to send a packet to destination IPb. Two situations are presented in Fig. 1. (a) and (b) respectively. When the network is fault-free, the packet is first routed in x direction before being routed to y direction. When one link L1 is faulty in the path, if normal deterministic x-y routing was adopted, there is no path for the packet to move towards the destination. For odd-even turn algorithm, the packet is routed one hop perpendicular to the top, then toward the destination. As shown in Fig. 1. (b), suppose source IPc is trying to communicate to the destination IPd, and the link L2 is faulty. In this case the packet is routed, one hop toward the down, then toward the destination.

### 3. DAMQS and DAMQAS Buffer Schemes

DAMQ dynamically allocates buffer blocks according to the packet received. Compared with SAMQ, the advantage of DAMQ is that it uses efficiently the buffer space by applying free space to any incoming packet regardless its destination output port.

DAMQ All (DAMQA) is based on self-compacting buffer scheme. Two buffer slots are reserved for each virtual channel before the buffer accepts any incoming flit and during buffers operation [4] [5] [6].

DAMQS buffer combines the buffer for virtual channels from two different physical channels. We combine the buffer space for east X and south Y virtual channels to build one physical buffer, and west X and north Y forms another buffer group. There are two buffers for four physical channels; each buffer is shared by eight virtual channels, and has two read ports and write ports respectively.

As shown in Fig. 2, in this way, the shared space is placed in the middle of the two buffer regions of two virtual link groups then the two buffer regions expand towards center of the free buffer space. This way, there will be less data shifts when a flit is saved into buffer because the movement of one region doesn't depend on shift of another group.

When a node or a physical channel is deemed as faulty, the previous hop node will terminate the buffer occupancy of messages destined to the failed link. Therefore, the buffer space will be efficiently used in case fault occurs at some nodes. For example, if node X is connecting to node Y and Z and there are message flows from X to Y and X to Z. When Z fails, there are probably still a number of message flits left in X's buffer. It will improve the system performance if the buffer space occupied by Z's message can be allocated to Y's message quickly.

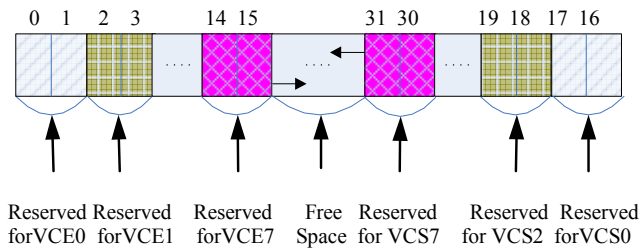
DAMQAS buffer combines the buffer for virtual channels from four different physical channels. There is one buffer for four physical channels; the buffer is shared by sixteen virtual channels, and has four read ports and write ports respectively.

As shown in Fig. 2. (a,c), two buffer slots are reserved for each virtual channel before any flit comes in the buffer. Virtual channels on X dimension start to occupy the buffer from lower end of the whole buffer space while virtual channels on Y dimension start using buffer on another end. The buffer space of groups expands and compresses in opposite direction. When a buffer group accepts a flit it expands, when it dispatches a flit it compresses.

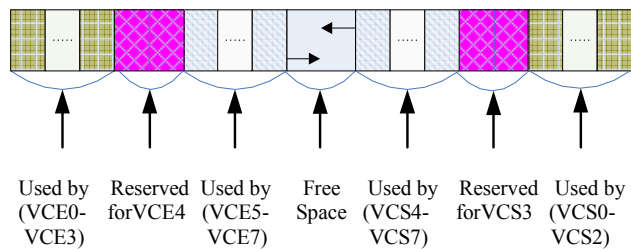
The reserve space (RS) is always kept if there is no flit or only one flit in the buffer region for a specific virtual channel [4].

As shown in Fig. 2. (b,d), when the buffer performs shift up or shift down operations, the RSs are also shifted. When a virtual channel accepts a flit, it first uses its RS. If RS is used up, buffer space of the whole group expands toward another group's buffer space to produce a slot. At any time, the number of current flits in buffer plus the number of reserved slots equals to the total amount of buffer slots. Therefore, one or more virtual channels which have the flits come into the buffer at earlier time can never deprive the chance for other virtual channels which get flits later than them to get buffer.

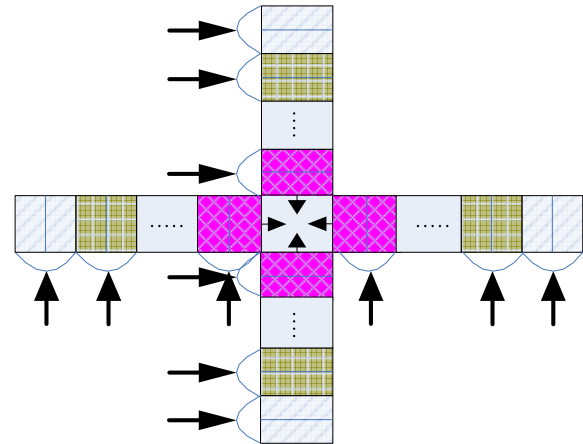
Also, in case the earlier coming packets are blocked in the buffer, since there is still reserved space for other virtual channels, the network traffic will keep flowing; therefore the performance of the switch is enhanced. Moreover, as virtual channels from two or four physical channels are sharing the buffer, the buffer space is more efficiently used by the incoming flits. The results will be shown in next section.



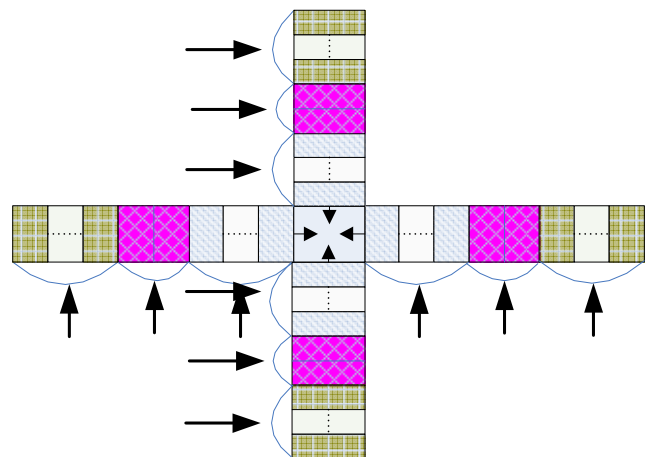
(a) DAMQS at initial state (# VC=8).



(b) DAMQS at operation state (# VC=8).



(c) DAMQAS at initial state (# VC=8).



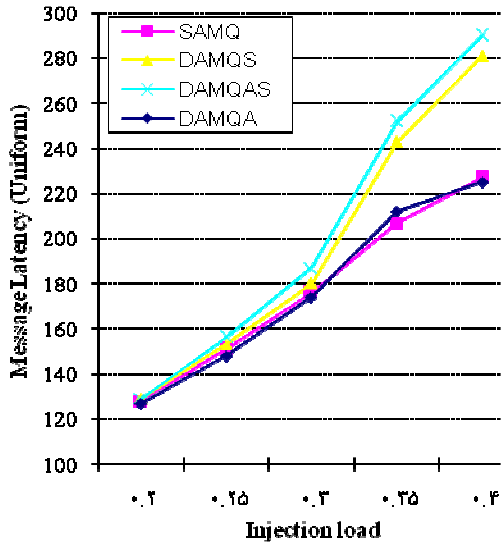
(d) DAMQAS at operation state (# VC=8).

Fig. 2. DAMQS and DAMQAS buffer spaces status. VCE (VC East), VCS (VC South)

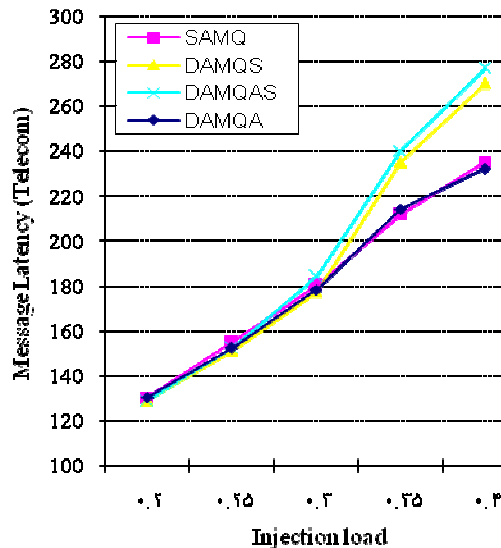
### 4. Experimental Results

In this section, we consider a system consisting of 64 IP blocks mapped onto mesh-based NoC architecture as shown in Fig. 1. Packets size is set to 32 flits. Virtual channels number for each physical channel is 4. Faults are generated randomly in the network. We set the buffer size (BS) for each virtual channel to 4 flits when DAMQA and SAMQ are used. Since four virtual channels are multiplexing cross one physical channel, the buffer size for each direction of a duplex physical channel is 16 flits when these two buffer schemes are evaluated. To examine the performance of DAMQS and DAMQAS with regard to the relationship between buffer size and network performance, we use flits buffer with different sizes. Since the network performance is greatly influenced by the traffic pattern, we applied two different traffic patterns, including synthetic traffic pattern (uniform) and a real-life

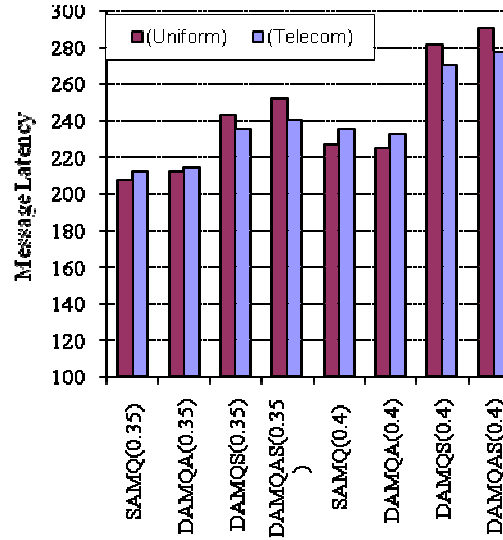
traffic pattern (telecom) is retrieved from "E3S" benchmark suite [7], which contain 30 tasks. Fig. 3 (a, b) shows the message latency as function of injection load with 2% fault rate (fr) for two uniform and telecom traffic patterns, respectively. When we further increase the traffic load and the fault rate after the network starts getting saturated, DAMQAS shows higher latency than other schemes. The reason is DAMQAS holds much more flits in the buffer than other schemes.



(a) Uniform traffic



(b) Telecom traffic



(c) Latency Comparison

Fig. 3. Message latency (2% fr)

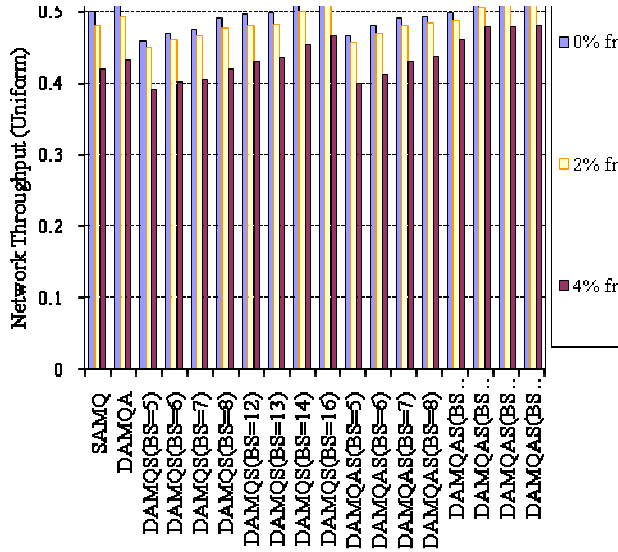
In Fig. 3 (c), we compare the message latency of network for two different traffic patterns under different injection rates.

Fig. 4 (a, b) shows the throughput characteristics of the NoC for each buffer scheme in presence of permanent faults with different fault rates for two uniform and telecom traffic patterns, respectively.

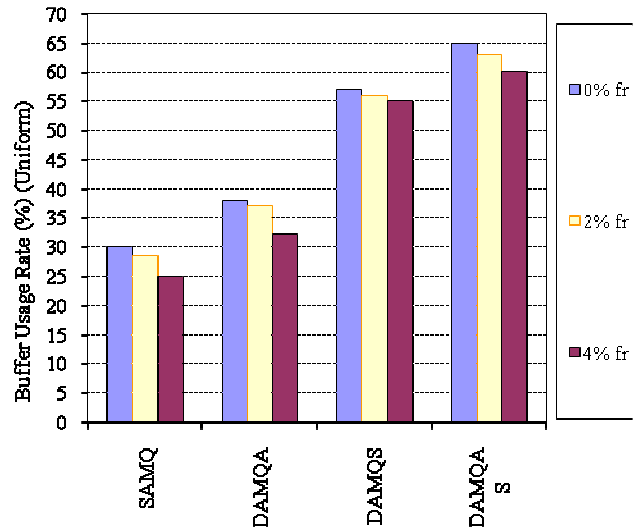
As shown in Fig. 4, DAMQAS and DAMQS can provide a better performance in presence of faults than other schemes. For example in uniform traffic, 14-flit DAMQS with 0% fr and 8-flit DAMQS with 4% fr achieves approximately the same maximum throughput as a 16-flit DAMQA as shown in Fig. 4 (b). Also, a 13-flit DAMQAS with 0% fr and 6-flit DAMQAS with 4% fr achieves approximately the same maximum throughput as 16-flit DAMQA. This is to say, to provide a similar network performance on very limited buffer resource, DAMQA with 0% fr achieves similar throughput with 12.5% and 18.75% more buffer space than DAMQS and DAMQAS with 0% fr, respectively and DAMQA with 4% fr achieves similar throughput with 50% and 62.5% more buffer space than DAMQS and DAMQAS with 4% fr, respectively. For telecom traffic, 13-flit DAMQS with 0% fr and 7-flit DAMQS with 4% fr achieves approximately the same maximum throughput as a 16-flit DAMQA as shown in Fig. 4 (b). Also, a 12-flit DAMQAS with 0% fr and 5-flit DAMQAS with 4% fr achieves approximately the same maximum throughput as 16-flit DAMQA. This result shows that, DAMQA with 0% fr achieves similar throughput with 18.75% and 25% more buffer space than

DAMQS and DAMQAS with 0% fr, respectively and DAMQA with 4% fr achieves similar throughput with 56.25% and 68.75% more buffer space than DAMQS and DAMQAS with 4% fr, respectively.

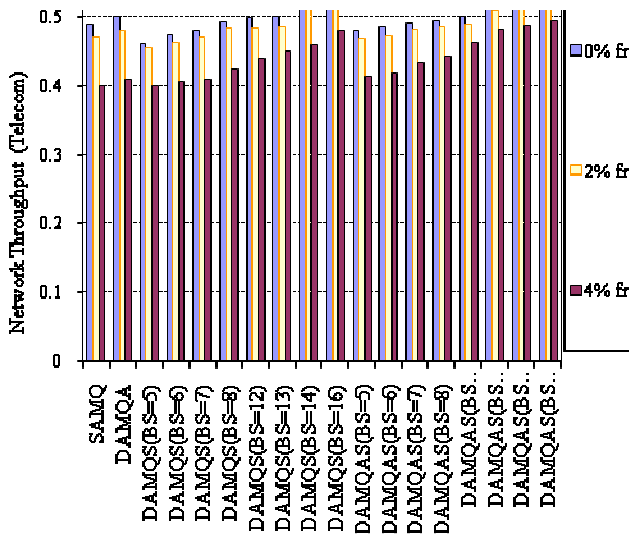
where PHY is the physical channel corresponding to a node port, N is the count of nodes, VC is the count of virtual channel multiplexing a physical channel and VB is the buffer size of a virtual channel. The total available buffer space is 3584 flits in our simulations. We considered high injection load of 0.35. As shown in Fig. 5, DAMQS and DAMQAS can provide a better performance than DAMQA and SAMQ.



(a) Uniform traffic



(a) Uniform traffic

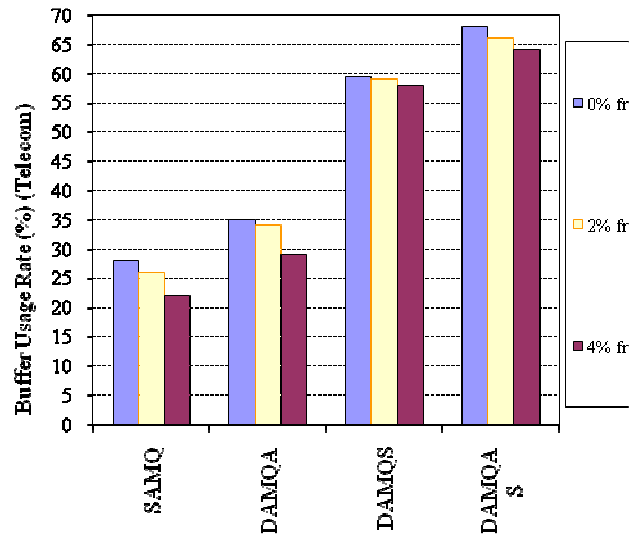


(b) Telecom traffic

Fig. 4. Network throughput

The total buffer space ( $BUF_{total}$ ) can be obtained by the following formula:

$$BUF_{total} = (N \times PHY - 4\sqrt{N}) \times VC \times VB$$



(b) Telecom traffic

Fig. 4. Buffer usage rate

Comparing uniform traffic results with telecom traffic results, it is clear that the DAMQS and DAMQAS are even more beneficial for real benchmarks. Indeed,

compared to the uniform traffic pattern in Fig. 3 (a), Fig. 4 (a) and Fig. 5 (a), the traffic patterns for real benchmarks are much more unbalanced (for instance, some of the channels have even a load of zero), which makes the idea of share buffer more attractive.

## 5. Conclusions

For deep sub-micron VLSI processes, the life-time reliability of devices is likely to be compromised by effects such as electromigration and material ageing. Consequently, the performance of NoC interconnect architectures will be severely affected due to presence of permanent faults. Though deterministic routing is very easy to implement, it fails to sustain the desired level of performance in presence of permanent faults. When an adaptive routing protocol such as odd-even turn algorithm is used for the NoC, DAMQAS and DAMQS are excellent schemes to optimize buffer management providing a good throughput when the network has a larger load in presence of faults. They can utilize significantly less buffer space with further increase of the fault rate without sacrificing the network performance. As shown in our simulation results, these buffer schemes can provide marginally higher throughput than traditional SAMQ when same amount of resource is used, this is due to the fact that buffer cannot play a major role in determining the network performance in terms of throughput or latency. However, the results show that these schemes can use significantly less hardware to provide a same performance as traditional SAMQ buffer. The simulation results show that the proposed DAMQ-Based schemes are indeed beneficial for real-world applications characterized by real-life traffic patterns.

## References

- [1] T. Schonwald, J. Zimmermann, O. Bringmann, W. Rosenstiel "Fully Adaptive Fault-tolerant Routing Algorithm for

- Network-on-Chip Architectures", in **10th Euromicro Conference on Digital System Design Architectures, (DSD 2007)**.
- [2] J. Zhou, F. C. M. Lau, "Adaptive Fault-tolerant Wormhole Routing with Two Virtual Channels in 2D Meshes", in **7th International Symposium on Parallel Architectures, Algorithms and Networks**, May 2004, pp. 142 - 148.
- [3] C.J. Glass and L.M. Ni, "Adaptive Routing in Mesh-connected Networks", in **12th International Conference on Distributed Computing Systems**, June 1992, pp. 12-19.
- [4] J. Liu, J. G. Delgado-Frias, "DAMQ Self-Compacting Buffer Schemes for Systems with Network-On-Chip," **Int. Conf. Computer Design, Las Vegas**, June 2005, pp. 97-103.
- [5] Y. Tamir and G. L. Frazier, "Dynamically-allocated multi queue buffers for VLSI communication switches," **IEEE Transactions on Computers**, vol. 41, no. 2, June 1992, pp. 725-737.
- [6] J. Liu, J. G. Delgado-Frias, "A Shared Self-Compacting Buffer for Network-on-Chip Systems," **49th IEEE Int. Midwest Symposium on Circuits and Systems**. August 2006.
- [7] R. Dick. Embedded system synthesis benchmarks suites (E3S), <http://www.ece.northwestern.edu/~dickrp/e3s>

**Mohammad Ali Jabraeil Jamali**. received his B.Sc. in Electrical Engineering and M.Sc. in Computer Hardware Engineering from Oromie University in 1994 and Tehran Science and Research Branch of Islamic Azad University in 2003, respectively. Currently, he is a Ph.D. student and Faculty Member at the Tehran Science and Research Branch of Islamic Azad University under supervision of Dr. Ahmad khademzadeh and Islamic Azad University, Shabestar Branch, respectively. His research interests include VLSI Design, Interconnection Network, Fault Tolerant, Ad hoc Network, Sensor Network and Computer Architectures.

**Ahmad Khademzadeh**. is currently associate professor in Iran Telecommunication Research Center. His research interests include VLSI Design, Interconnection Network, Fault Tolerant and Computer Architectures.

# Distributed Object Medical Imaging Model

Ahmad Shukri Mohd Noor<sup>1</sup> and Md Yazid Md Saman<sup>2</sup>  
Department of Computer Science ,Faculty of Science and Technology  
University Malaysia Terengganu  
21030 Kuala Terengganu, Malaysia  
096683159 / 09-6694660(FAX)  
E-mail address: ashukri@umt.edu.my<sup>1</sup> / yazid@umt.edu.my<sup>2</sup>

## ABSTRACT

Abstract- Digital medical informatics and images are commonly used in hospitals today,. Because of the interrelatedness of the radiology department and other departments, especially the intensive care unit and emergency department, the transmission and sharing of medical images has become a critical issue.

Our research group has developed a Java-based Distributed Object Medical Imaging Model(DOMIM) to facilitate the rapid development and deployment of medical imaging applications in a distributed environment that can be shared and used by related departments and mobile physicians.DOMIM is a unique suite of multimedia telemedicine applications developed for the use by medical related organizations. The applications support real-time patients' data, image files, audio and video diagnosis annotation exchanges. The DOMIM enables joint collaboration between radiologists and physicians while they are at distant geographical locations. The DOMIM environment consists of heterogeneous, autonomous, and legacy resources. The Common Object Request Broker Architecture (CORBA), Java Database Connectivity (JDBC), and Java language provide the capability to combine the DOMIM resources into an integrated, interoperable, and scalable system. The underneath technology, including IDL ORB, Event Service, IIOP JDBC/ODBC, legacy system wrapping and Java implementation are explored. This paper explores a distributed collaborative CORBA/JDBC based framework that will enhance medical information management requirements and development. It encompasses a new paradigm for the delivery of health services that requires process reengineering, cultural changes, as well as organizational changes

## KEY WORDS

Java, CORBA, DICOM , Medical Image ,Medical Informatics , Distributed Object Computing.

## 1. Introduction

Digital medical images are commonly used in hospitals today, even outside the radiology department. Because of the interrelatedness of the radiology department and other departments, especially the intensive care unit and emergency department, the transmission of medical images has become a critical issue. The use of World

Wide Web and network related technologies in radiology is not new. These technologies have been used in radiology teaching files to access information in multimedia integrated picture archiving and communication systems (PACS), for teleradiology

purposes . Web technology has also been used to access the images stored in a Digital Imaging and Communications in Medicine (DICOM)archive in PACS environments[1][2].

## 2. Project Background

This research develops a framework of distributed medical informatics that can be used for multimedia data exchange. framework can be expand it any distributed object oriented, collaborative applications, for example, distance learning modeling and simulation

A DOMIM system based on distributed object computing system. The system can be viewed as a set of object services and a set of client applications. Each client application has a defined, interactive user interface. The object services provide and manage the information for the DOMIM clients. The ultimate goal is to have a complete set of services with a single fine-grained framework. The DOMIM strategy is an approach towards a single architecture where hardware and software from multiple vendors coexist in harmony. This is achieved by categorizing information into components or services (object services) as they communicate, by passing the information via interface invocations of objects. These object services are manufactured by different vendors and can run on different computers on networks. The architecture must have certain key characteristics:

- (1) *distributed*: it must support a service object model that is distributed across a regional area over LAN and WAN networks.
- (2) *platform independent*: it must support multiple computing platforms, from mainframes to servers to desktop PCs.
- (3) *heterogeneous*: it must support all different types and classes of medical equipment and software tools from many different vendors.
- (4) *location insensitive*: it must allow components in the system to be replaced, repaired, upgraded and changed without compromising its ability.

Obviously, interoperability is a key technology that allows this exchange to scale. Interoperability is also the ability to leverage and reuse system content and functionality to an end user or to another system

### 3. Approach

The distributed and heterogeneous nature of today's computing systems requires a middleware infrastructure capable of supporting a three-tier computing architecture such as Common Object Request Broker Architecture (CORBA). Business logic can be built, or existing applications encapsulated, into middle-tier components that interact with end users via standard interfaces such as web browsers and standard GUI desktops, and back-end data repositories [4].

#### 3.1 Distributed Object .

Common Object Request Broker Architecture (CORBA) was introduced by OMG in 1991 to go a step beyond OMA to specify technologies for interoperable distributed OO systems. With the CORBA specification, a broad and consistent model for building distributed applications is defined :

- An object-oriented based model for developing applications
- A common application programming objects in the network to be shared by client and server applications
- A syntax to define and describe the interfaces of objects used in the environment
- Support for multiple programming languages and platforms

Therefore, CORBA model formally separates the client and server portions of the application and also logically separates an application into objects that can perform certain functions. It also provides data marshaling to send and receive data with remote or local machine applications without direct knowledge of the information source or its location. In the CORBA environment, client and server applications communicate using Object Request Broker (ORB).

#### 3.2 Java Based CORBA.

The Java programming language is a strongly typed, object-oriented language that borrows heavily most of its syntax from C and C++. Java is a simple, object-oriented, distributed, interpreted, robust, secure, architecture neutral, portable, high performance, multithreaded and dynamic language. This language was primarily used for developing applets-downloadable mini-applications that could be embedded inside Web pages and performed in browsers. However, since 1995, Java has emerged as a first-class programming language that is being used for everything from embedded devices to enterprise servers. Nowadays the Java language can be seen in use in a wider range of applications. When an application is written and compiled in one place it can run on any machine under any operating system. Sometimes the "Write Once, Run

Anywhere" slogan is called the synonym of Java. Anyway, platform independence is the ability of a program to move from one computer system to another. Java is platform independent at both the source and the binary level. The secret of the Java has been hidden into Java Virtual machine (JVM). Instead of creating machine dependent code, the Java compiler creates a bytecode format, which can be run on any Virtual Machine (VM). Somehow, Java makes programming easier because it is object-oriented and has automatic garbage collection.

Java offers tremendous flexibility for distributed application development. To do this, Java needs to be augmented with a distributed object infrastructure, which is where OMG's CORBA comes into the picture. Using CORBA requires more than just a knowledge of the CORBA architecture. CORBA should be part of a well designed system architecture.

CORBA technology as part of the Java 2 platform consists of an Object Request Broker (ORB) written in Java. Java IDL adds CORBA capability to the Java platform, providing standards-based interoperability and connectivity. Java IDL enables distributed Web-enabled Java applications to transparently invoke operations on remote network services using the industry standard OMG IDL (Interface Definition Language) and IIOP (Internet Inter-ORB Protocol) defined by the Object Management Group. CORBA is an distributed object-oriented middleware protocols, used for the DOMIM development. By using CORBA it gives us several benefits in the system distributed computing environment. For example, we are able to interface legacy database by developing CORBA wrapper that allows us to access the data structures in the database without disturbing the existing database. Interoperability and scalability are other benefits of using CORBA. The CORBA IDL for streaming medical image in our model as follow :-

```
struct Info {  
    string name; module Student_App {
```

```
        struct Info {  
            string name;  
            string matric;  
            string address;  
            string city;  
            string state;  
            string zip;  
            string country;  
            string email;  
            string phone;  
            string program;  
        };
```

```
typedef sequence<octet> Data;
```

```
interface project {
    string execute(in short operation, in Info
info_student);
    Info execute2(in short operation, in Info
info_student);
    Data downloadFile(in string fileName);
};

};
string matric;
string address;
string city;
string state;
string zip;
string country;
string email;
string phone;
string program;
};
```

```
typedef sequence<octet> Data;
```

```
interface project {
    string execute(in short operation, in Info
info_student);
    Info execute2(in short operation, in Info
info_student);
    Data downloadFile(in string fileName);
};
```

```
};
module Student_App {
```

```
struct Info {
    string name;
    string matric;
    string address;
    string city;
    string state;
    string zip;
    string country;
    string email;
    string phone;
    string program;
};
```

```
typedef sequence<octet> Data;
```

```
interface project {
    string execute(in short operation, in Info
info_student);
    Info execute2(in short operation, in Info
info_student);
    Data downloadFile(in string fileName);
```

```
};
```

```
typedef sequence<octet> Data;
interface project {
    string execute(in short operation, in Info info_patient);
    Info execute2(in short operation, in Info info_patient);
    Data downloadFile(in string fileName);
```

### 3.3 Digital Imaging and Communications in Medicine

The Digital Imaging and Communications in Medicine (DICOM) standard was created by the National Electrical Manufacturers Association (NEMA) to aid the distribution and viewing of medical images, such as CT scans and ultrasound. New technologies such as Java should always be used as complements of the de facto standard in medical imaging, DICOM. DICOM allows the interchange of images from different modalities, archives, and workstations from different vendors. Java technology can be used to build a storage system and to make this service accessible for different clients. However, this storage service should also incorporate DICOM services to store and access examination data from DICOM workstations and DICOM modalities. Since Java version 1.4, the Java standard includes a specification for working with images stored in files and accessed across the network. This specification is called Java Image I/O. It provides a pluggable framework for easily adding support for alternate image formats using third-party plug-ins. The DICOM Image I/O Plug-in connects the DICOM® standard to the Java™ standard. DICOM is the universal standard for sharing medical imaging resources between heterogeneous and multi-vendor equipments (acquisition device, workstation, storage server, patient management system, etc.).

### 3.4 Distributed Medical API Implementation

A toolkit, which is referred to as *NeatMed*, is intended to reduce development time by eliminating the need for the application developer to deal directly with medical image data. The medical imaging application developers interface (API), NeatMed interface (API), was developed using the Java programming language (Sun Microsystems, Mountain View, Calif). An extension API is a set of classes that can be instantiated by a programmer to create a particular type of application, thus facilitating software reuse. NeatMed is an example of an extension API that can be used for the development of applications that deal with off-line medical image data. NeatMed currently provides support for the Digital Imaging and Communications in Medicine and Analyze medical image file formats. The NeatMed API is a group of core and support classes that can be used to interpret, represent, and manipulate images and related data that are stored in DICOM-compliant files. The toolkit has been

created specifically for interpreting medical image data; it thus acts as a platform for development of medical imaging applications.. NeatMed was implemented by using Java[1]. NeatMed currently provides support for the DICOM in Medicine and Analyze medical image file formats.The Neat-Med API is distributed in accordance with the terms and conditions laid out in the GNU Lesser General Public License[2]. This license was selected to ensure that the NeatMed. NeatMed was developed using the Java programming languages. It was initially intended for the development of software for use in the consumer electronics industry (eg, set-top boxes). The core Java libraries maintained by Sun Microsystems are used as the foundation for the development of any Java application. These libraries can be used in conjunction with an extension API in order to develop specialized applications. An extension API is a set of classes that can be instantiated by a programmer to create a particular type of application, thus facilitating software reuse. NeatMed is an example of an extension API that can be used for the development of applications that deal with off-line medical image data. A large number of Java APIs exist; these deal with a broad range of applications ranging from communicating with the serial and parallel ports to advanced image processing. The set of classes representing the API is deployed in some type of library. Java provides a packaging tool that can be used to package a set of class files and associated resources into a Java archive or *JAR* file. In order to be useful, an API must be well documented. Java provides a documentation tool called *Javadoc* that allows an API developer to document software as it is being written. The resulting documentation provides detailed information about each class, method, and variable that is defined in the associated API. The structure of Javadoc documentation is more or less the same for every API. This makes it very easy for programmers to familiarize themselves with a new API once they are comfortable with the basic Javadoc documentation structure. The API documentation is generated in HTML and can be viewed using any standard Web browser. Java has a wide range of benefits associated with it; however, there are also some limitations. One example is performance: Java is a multiplatform programming language; the byte code (ie, binary form) that represents a Java program is interpreted and not executed directly. This reduces the performance of a Java program compared to a natively executed program. Overall, however, the benefits associated with Java (listed previously) far outweigh the drawbacks, hence its selection for the development of NeatMed.. In

Fig 2, Client Programming structure show flexibility, and ease of use of the NeatMed API in java Environment

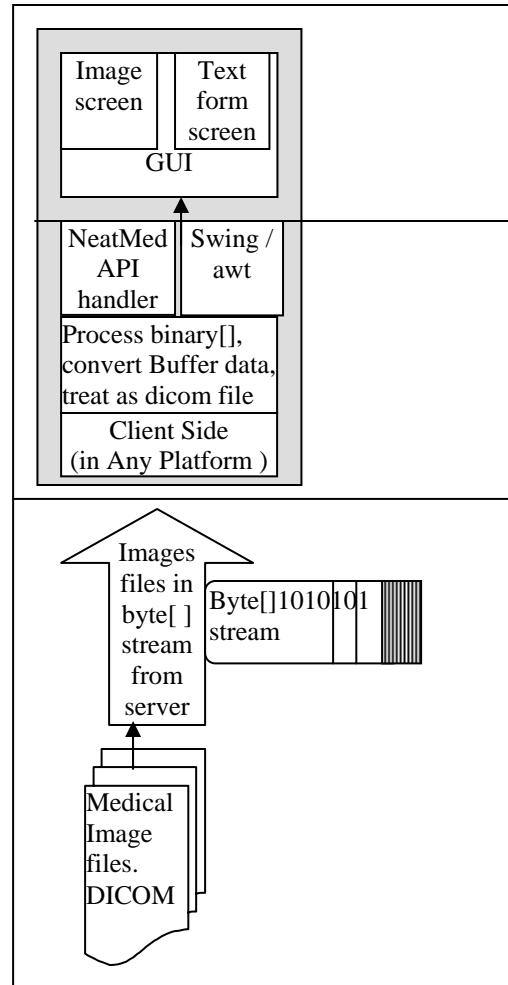


Figure 2: Client Programming structure.

The central class in the API is the *DICOMImage* class. A *DICOMImage* object can be instantiated by specifying a reference to a suitable data source in the constructor. The constructor will accept data from a number of sources (eg, local file, data stream, and remote uniform resource locator (URL)). Once constructed, a *DICOMImage* object provides direct access to all of the data elements stored within the specified DICOM source. Other classes in the API are used to represent individual components within a DICOM *RadioGraphics*.

```

DICOMImagePlus image = new DICOMImagePlus(ss);
JLabel label = new JLabel(new
ImageIcon(image.getAsBufferedImage(0)));
JFrame frame = new JFrame("Patient NO : "
+Patient_info2.id + " NAME : " +
Patient_info2.name + " Image File :"+
Patient_info2.file_id);
frame.getContentPane().setLayout(new
BorderLayout());
frame.getContentPane().add(label);
frame.setVisible(true);
frame.pack();}
    
```

The choice of Java for implementing the NeatMed API was also influenced by a number of its key features:

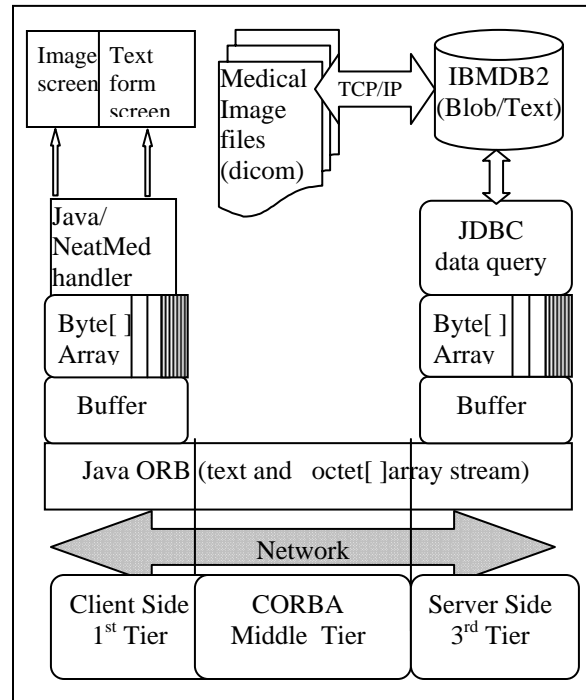
- Ease of use: Java is a modern programming language that was designed with simplicity in mind. Many of the complexities that are associated with other programming languages have been omitted, whereas much of the power and flexibility has been retained. This makes Java very easy to learn and use, particularly in the case of novice programmers.
- Level of support: Although Java is a relatively new programming language, there is a significant amount of support material available. Numerous texts have been written dealing with all aspects of the language. In addition, tutorials, sample source code, API documentation, and freely available integrated development environments (IDEs) can all be accessed via the Internet.
- Portability: Java is a multiplatform programming language. This means that a Java

### 5. Distributed Medical informatics Architecture.

The distributed medical informatics architecture design includes client applications at the 1<sup>st</sup> tier that access remote medical imaging data query server and database at 3<sup>rd</sup> tier via Java ORB as a middle agent at 2<sup>nd</sup> tier,

The client's implementations comprised of a Java application . All the user interfaces were created using Java APIs while NeatMed API for presenting and displaying the medical image files in DICOM format.

The server is comprised of the some logical algorithms that responsible for executing an input query statement from the client and returning the query results back to the client. The connection between the servers and the IBM DB2 Database is accomplished via its native JDBC driver. This server is placed on the local area network (LAN) with Java ORB acting as the middleware. The ORB using octet-streaming services is utilized to transfer multimedia data such as medical images, It also can be used for audio and video in a 3-tiers heterogeneous environment.



**Figure3: Distributed Object Medical Imaging Model (DOMIM) Architecture**

The back-end tier of the architecture involves the storage and retrieval of multimedia data on the database server.

In this paper, the Object Relational database management system (ORDBMS) is used for the development of medical imaging and multimedia database as it allows queries to be performed on complex data, e.g. images, video, audio, etc. The following components are utilised in the development of the server application:

- IBM DB2 v 8.1 PE database.
- Server application (ProjectServer.class) for receiving object and sending back the object from/to client application.
- Server application logic (ProjectImpl.class) for executing as input query

IBM DB2 has been chosen as the implementation database. Due to IBM DB2 ability to support the binary large object blocks (BLOBs) data. All the image data are stored in their native binary format in a particular column of the database table.

The following features are provided by the client's implementations(ProjectClient.class) via a Graphical User Interface (GUI):

- Binding to the servers' implementations.
- Invoking the servers' implementation with the appropriate commands.
- Displaying and presenting the query results(text and images) to the user.

The design depicts that ORB can be an integral part of deploying Java applets/applications, including those that access database. The diagram also shows that the client

application does not directly connect to the databases. Instead, Java ORB facilitates database connectivity by allowing client-side objects to communicate with server-side objects that assume responsibility for performing database access. That is, server-side objects is written to handle database access on behalf of a client object(s) that instantiated it. The client main function is to display multiple media data to the user in the specified format. Such architecture can provide adequate database support for medical applications demanding interactive medical imaging presentations .the modules involved are:-

- GUI Object Class

The Graphical User Interface (GUI) class represents the tools used for human-to-computer interaction . The GUI may include menus, buttons, graphics, textual and visual information, image annotation, full-duplex audio, and video information. The GUI class is responsible for collecting relevant information to be passed onto the user. The GUI class also distributes user input to objects whose states depend on the user. The objects in the class are linked with other objects in the system , e.g., management and control objects.

- Audio Object Class

The audio object class handles real-time and playback of audio sequences at the user workstation. Digital audio is encoded to sequences of 8-bit or 16-bit samples using standard telephone Coder/Decoder (CODEC) conversion. The audio sequences are two-way conversations between physicians or audio notes to be stored with a patient record.

- Video Object Class

The video object class handles real-time and playback video sequences. The video sequences are digital format using MPEG compression frames. Video sequences can be generated from digital camera or pre-recorded video on CD disk formats. Video sequences can be stored as video object classes in the multimedia database archive system.

- Buffer Object Class

The buffer object class controls jitter by queuing data until it is synchronized and ready for use. Buffering takes place on all data traversing the network stack to abstract the application from network timing idiosyncrasies and dependencies.

- File Stream Object Class

The file stream objects include files, e.g., images, which must be transferred between nodes over networks. These

nodes can be workstations or the Database Archive System. The objects represent a flow of packets, and the objects inherent the format of their types, i.e., audio, video, annotation commands, etc.

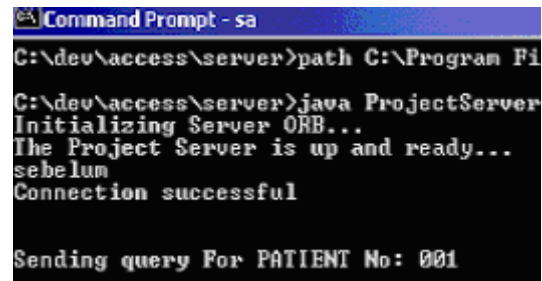
- Communication Object Class

The communication object class represents a group of connectivity mechanism. For example, objects conduct communication via middleware APIs or TCP/IP sockets. These objects provide the communication paradigm to the data objects described above, i.e., annotation commands, file streams, etc. and link the multimedia information exchanging over the distributed computing environment.

## 6. Implementantion

In this application, client patient's detail screen and medical image in DICOM format screen are displayed in separated windows.

- The client request patient data by providing patient case retrieving procedures such as patient id and notifies the server(services Perovider) via corba event service .
- The Server retrieve a patient demographic data and image(s) from client and send the DBA via JDBC Server using JDBC connectivity to the DBMS, e.g., IBM DB2 database. As in figure 5



```
Command Prompt - sa
C:\dev\access\server>path C:\Program Fi
C:\dev\access\server>java ProjectServer
Initializing Server ORB...
The Project Server is up and ready...
sebelum
Connection successful

Sending query For PATIENT No: 001
```

Fig 6:Server response screen shot

- The Server passes the patient demographic data and image file identification (where those files *resid ein remote storage*) to the client via ORB.
- The remote Client fetches the demographic data and display then details on client GUI screen as depict in figure 6 . Then it point the patient image(s) based on the given file identification. then, the system pop-up the medical image as illustrated figure 7 below

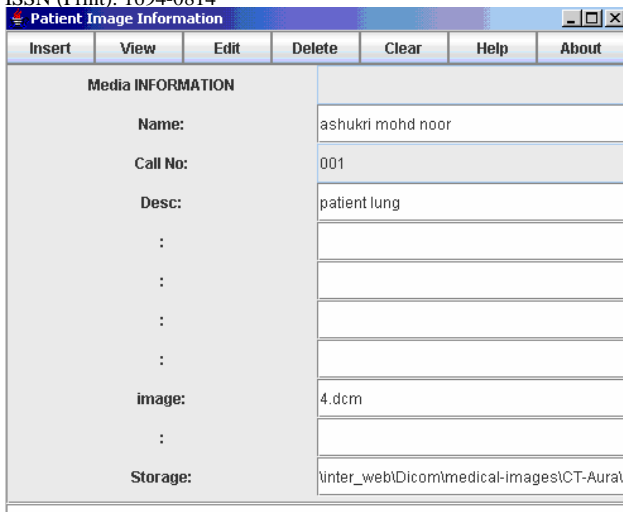


Fig 7. Screenshot of GUI patient's data



Fig 8. Screenshot of patient's Medical Image

## 7. Conclusion

The three-tier distributed medical image application allow the application interoperability and independence of platform, operating system, programming language and even of network and protocol. The application architecture or framework is an important and common stage in the development of any medical imaging application in distributed environment.

NeatMed removes the need to deal directly with encoded medical image data, thus increasing productivity and allowing the developer to concentrate on other aspects of application development. NeatMed is written in Java, a multiplatform programming language with a large amount of freely available support material that is straightforward to learn and use. These and other features of Java make NeatMed accessible to a large group of potential users. Most important, NeatMed is a freely available research tool whose ongoing development is driven by the needs and requirements of its users.

## References

- [1] Mildenberger P, Eichelberg M, and Martin E. (2002) Introduction to the DICOM standard. European Radiology, Publisher: Springer Berlin / Heidelberg ISSN: 0938-7994.
- [2] Rosslyn, V. (2003). Digital Imaging and Communications in Medicine (DICOM): National Electrical Manufacturers Association, 2003; PS 3.1-2003–3. 16-2003.
- [3] Robinson, J. and Wakeman I (2004) "Middleware for service composition in pervasive computing". The 2nd Workshop on Middleware for Pervasive and Ad-hoc Computing, Toronto, Canada
- [4] A. S. Mohd Noor and M. Y. Saman (2006). "Traditional CORBA Framework Re-engineering For Distributed Medical Informatics Model Development". IEEE International Conference on Computing & Informatics (ICOCI 2006) 6 – 8 June, 2006, Kuala Lumpur, MALAYSIA.
- [5] Togni D, Ribas P and Lisboa. M (2005) "Tool integration using the web-services approach" Proceedings of the 15th ACM Great Lakes symposium on VLSI, Chicago, Illinois, USA
- [6] Aaron W. Jr. (2004) "Flexible distributed programming in an extended Java". ACM Transactions on Programming Languages and Systems (TOPLAS), Volume 26 Issue 3,
- [7] Wan Zahari, W. N. I. (2003), Model Pengkomputeran Teragih Bagi Memaparkan Rupa Bumi Berdigit Bersaiz Besar. Master Thesis. Univerisiti Teknologi Malaysia.
- [8] Gokhale, A. and Schmidt, D. C. (1998), Measuring and Optimizing CORBA Latency and Scalability Over High-speed Networks. Transactions on Computing, 47(4).
- [9] Morgan G and Fengyun L. (2005) "Visibility & games: Interest management middleware for networked games". Symposium on Interactive 3D graphics and games, Washington, USA
- [10] Gokhale, A. and Schmidt, D. C. (1996), Measuring the Performance of Communication Middleware on High-Speed Networks. In Proceedings of SIGCOMM '96, pages 306–317, Stanford, CA. ACM.

# **IJCSI CALL FOR PAPERS MARCH 2010 ISSUE**

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas. See authors guide for manuscript preparation and submission guidelines.

**Accepted papers will be published online and authors will be provided with printed copies and indexed by Google Scholar, Cornell's University Library, ScientificCommons, CiteSeerX, Bielefeld Academic Search Engine (BASE), SCIRUS and more.**

**Deadline: 31<sup>st</sup> January 2010**

**Notification: 28<sup>th</sup> February 2010**

**Revision: 10<sup>th</sup> March 2010**

**Online Publication: 31<sup>st</sup> March 2010**

- Evolutionary computation
- Industrial systems
- Evolutionary computation
- Autonomic and autonomous systems
- Bio-technologies
- Knowledge data systems
- Mobile and distance education
- Intelligent techniques, logics, and systems
- Knowledge processing
- Information technologies
- Internet and web technologies
- Digital information processing
- Cognitive science and knowledge agent-based systems
- Mobility and multimedia systems
- Systems performance
- Networking and telecommunications
- Software development and deployment
- Knowledge virtualization
- Systems and networks on the chip
- Context-aware systems
- Networking technologies
- Security in network, systems, and applications
- Knowledge for global defense
- Information Systems [IS]
- IPv6 Today - Technology and deployment
- Modeling
- Optimization
- Complexity
- Natural Language Processing
- Speech Synthesis
- Data Mining

All submitted papers will be judged based on their quality by the technical committee and reviewers. Papers that describe research and experimentation are encouraged.

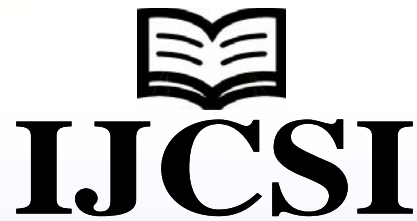
All paper submissions will be handled electronically and detailed instructions on submission procedure are available on IJCSI website (<http://www.ijcsi.org>).

For other information, please contact IJCSI Managing Editor, ([editor@ijcsi.org](mailto:editor@ijcsi.org))

Website: <http://www.ijcsi.org>

**© IJCSI PUBLICATION 2009**

**[www.IJCSI.org](http://www.IJCSI.org)**



The International Journal of Computer Science Issues (IJCSI) is a refereed journal for scientific papers dealing with any area of computer science research. The purpose of establishing the scientific journal is the assistance in development of science, fast operative publication and storage of materials and results of scientific researches and representation of the scientific conception of the society.

It also provides a venue for researchers, students and professionals to submit on-going research and developments in these areas. Authors are encouraged to contribute to the journal by submitting articles that illustrate new research results, projects, surveying works and industrial experiences that describe significant advances in field of computer science.

#### **Indexing of IJCSI:**

1. Google Scholar
2. Bielefeld Academic Search Engine (BASE)
3. CiteSeerX
4. SCIRUS
5. Docstoc
6. Scribd
7. Cornell's University Library
8. SciRate
9. ScientificCommons