IJCSI

International Journal of Computer Science Issues

Volume 10, Issue 1, No 2, January 2013
ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

IJCSI proceedings are currently indexed by:

IJCSI Publicity Board 2013

IJCSI Editorial Board 2013

EDITORIAL

In this first edition of 2013, we bring forward issues from various dynamic computer science fields ranging from system performance, computer vision, artificial intelligence, software engineering, multimedia, pattern recognition, information retrieval, databases, security and networking among others.

Considering the growing interest of academics worldwide to publish in IJCSI, we invite universities and institutions to partner with us to further encourage open-access publications.

As always we thank all our reviewers for providing constructive comments on papers sent to them for review. This helps enormously in improving the quality of papers published in this issue.

Google Scholar reported a large amount of cited papers published in IJCSI. We will continue to encourage the readers, authors and reviewers and the computer science scientific community and interested authors to continue citing papers published by the journal.

It was with pleasure and a sense of satisfaction that we announced in mid March 2011 our 2-year Impact Factor which is evaluated at 0.242. For more information about this please see the 3rd question in the FAQ section of the journal.

Apart from availability of the full-texts from the journal website, all published papers are deposited in open-access repositories to make access easier and ensure continuous availability of its proceedings free of charge for all researchers.

We are pleased to present IJCSI Volume 10, Issue 1, No 2, January 2013 (IJCSI Vol. 10, Issue 1, No 2). The acceptance rate for this issue is 33.09%.

IJCSI Editorial Board
January 2013 Issue
ISSN (Online): 1694-0814

www.IJCSI.org

IJCSI Reviewers Committee 2013

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

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. M. Azath, Anna University, India

Dr. P. Suresh Varma, Adikavi Nannaya University, India

Prof. V. N. Kamalesh, JSS Academy of Technical Education, India

Dr. D Gunaseelan, Ibri College of Technology, Oman

Mr. Sanjay Kumar Anand, CDAC, India

Mr. Akshat Verma, CDAC, India

Mrs. Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia

Mr. Hasan Asil, Islamic Azad University Tabriz Branch (Azarshahr), Iran

Prof. Dr Sajal Kabiraj, Fr. C Rodrigues Institute of Management Studies (Affiliated to University of Mumbai, India), India

Mr. Syed Fawad Mustafa, GAC Center, Shandong University, China

Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA

Prof. Selvakani Kandeeban, Francis Xavier Engineering College, India

Mr. Tohid Sedghi, Urmia University, Iran

Dr. S. Sasikumar, PSNA College of Engg and Tech, Dindigul, India

Dr. Anupam Shukla, Indian Institute of Information Technology and Management Gwalior, India

Mr. Rahul Kala, Indian Institute of Inforamtion Technology and Management Gwalior, India

Dr. A V Nikolov, National University of Lesotho, Lesotho

Mr. Kamal Sarkar, Department of Computer Science and Engineering, Jadavpur University, India

Prof. Sattar J Aboud, Iraqi Council of Representatives, Iraq-Baghdad

Dr. Prasant Kumar Pattnaik, Department of CSE, KIST, India

Dr. Mohammed Amoon, King Saud University, Saudi Arabia

Dr. Tsvetanka Georgieva, Department of Information Technologies, St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria

Mr. Ujjal Marjit, University of Kalyani, West-Bengal, India

Dr. Prasant Kumar Pattnaik, KIST,Bhubaneswar,India, India

Dr. Guezouri Mustapha, Department of Electronics, Faculty of Electrical Engineering, University of Science and Technology (USTO), Oran, Algeria

Mr. Maniyar Shiraz Ahmed, Najran University, Najran, Saudi Arabia

Dr. Sreedhar Reddy, JNTU, SSIETW, Hyderabad, India

Mr. Bala Dhandayuthapani Veerasamy, Mekelle University, Ethiopa

Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia

Mr. Rajesh Prasad, LDC Institute of Technical Studies, Allahabad, India

Dr. Mokhled AlTarawneh, PhD computer engineering/ Faulty of engineerin/ mutah university, jordan

Prof. Santhosh.P.Mathew, Saintgits College of Engineering, Kottayam, India

Ms. Suman Lata, Rayat Bahara institue of engg. & Nanotechnology,Hoshiarpur, India

Dr. Shaikh Abdul Hannan, Vivekanand College, Aurangabad, India

Prof. PN Kumar, Amrita Vishwa Vidyapeetham, India

Dr. P. K. Suri, Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, India

Dr. Syed Akhter Hossain, Daffodil International University, Bangladesh

Mr. Sunil, Vignan College, India

Mr. Ajit Singh, TIT&S Bhiwani, Haryana, India

Mr. Nasim Qaisar, Federal Urdu Univetrsity of Arts , Science and Technology, Pakistan

Ms. Rshma, Maharishi Markandeshwar University, India

Mr. Gaurav Kumar Leekha, M.M.University, Solan (Himachal Pardesh), India

Mr. Ordinor Tucker, Ministry of Finance Jamaica, Jamaica

Mr. Mohit Jain, Maharaja Surajmal Institute of Technology (Affiliated to Guru Gobind Singh Indraprastha University, New Delhi), India

Dr. Shaveta Rani, GZS College of Engineering & Technology, India

Dr. Paramjeet Singh, GZS College of Engineering & Technology, India

Dr. G R Sinha, SSCET, India

Mr. Chetan Sharma, TechMahindra India Ltd., India

Dr. Nabil Mohammed Ali Munassar, University of Science and Technology, Yemen

Prof. T Venkat Narayana Rao, Department of CSE, Hyderabad Institute of Technology and Management , India

Prof. Vasavi Bande, HITAM, Engineering College, India

Prof. S.P.Setty, Andhra University, India

Dr. C. Kiran Mai, J.N.T.University,Hyderabad/VNR Vignana Jyothi Institute of Engineering & Technology/, India

Ms. Bindiya Ahuja, Manav Rachna International University, India

Mrs. Deepa Bura, Manav Rachna International University, India

Mr. Vikas Gupta, CDLM Government Engineering College, Panniwala Mota, India

Dr Juan José Martínez Castillo, University of Yacambu, Venezuela

Mr Kunwar S. Vaisla, Department of Computer Science & Engineering, BCT Kumaon Engineering College, India

Mr. Abhishek Shukla, RKGIT, India

Prof. Manpreet Singh, M. M. Engg. College, M. M. University, Haryana, India

Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran

TABLE OF CONTENTS

An Efficient Distributed Data Extraction Method for Mining Sensor Network's Data

Azhar Mahmood[1], Ke Shi[1] and Shaheen Khatoon[1]

[1] **School of Computer and Applied Technology**
Huazhong University of Science & Technology (HUST) Wuhan, China

Abstract
A wide range of Sensor Networks (SNs) are deployed in real world applications which generate large amount of raw sensory data. Data mining technique to extract useful knowledge from these applications is an emerging research area due to its crucial importance but still it's a challenge to discover knowledge efficiently from the sensor network data. In this paper we proposed a Distributed Data Extraction (DDE) method to extract data from sensor networks by applying rules based clustering and association rule mining techniques. A significant amount of sensor readings sent from the sensors to the data processing point(s) may be lost or corrupted. DDE is also estimating these missing values from available sensor reading instead of requesting the sensor node to resend lost reading. DDE also apply data reduction which is able to reduce the data size while transmitting to sink. Results show our proposed approach exhibits the maximum data accuracy and efficient data extraction in term of the entire network's energy consumption.

Keywords: *Sensor Network, Data Mining, Data Extraction, Association Rules, Clustering, Frequent Pattern, Data Reduction.*

1. Introduction

Advances in wireless communication and microelectronic devices led to the development of low power sensors and the deployment of large scale sensor networks. With the capabilities of pervasive surveillance sensor networks has attracted significant attention in many applications domains, such as habitat monitoring [1, 2], object tracking [3, 4], environment monitoring [5-7], military [8, 9], disaster management [10], just to mention a few example[11]. These applications yield huge volume of dynamic, geographically distributed and heterogeneous data. The raw data if analyzed in an appropriate way might help to automatically and intelligently solve a variety of tasks thus making the human life more safe and comfortable. Recently, extracting knowledge from sensor data has been received a great deal of attention by the data mining community. However, the extremely constrained nature of sensors and the potentially dynamic behavior of SNs hinder the use of traditional mining approaches commonly applied on other domains. Traditional approaches are meant for multi-step methodologies and multi-scan algorithms, which cannot be straightforwardly applied to sensor network. Development of algorithms that consider the characteristics of sensor networks, such as energy and computation constraints, network dynamics, faults, constitute an active area of current research.

Several techniques have been proposed in the literature for knowledge extraction from sensor data e.g. association rules [12-14] frequent patterns mining, knowledge discovery over data streams[15, 16], and clustering [17] to enhance the performance of SNs. In these applications large numbers of sensors are distributed in the physical world and generate streams of data that need to be combined, monitored, and analyzed on central side. However, collecting all data in a central computing node with a high computational power does not optimize the use of energy-costly transmissions. Indeed in most cases all raw data are not needed, we are only interested in an estimate of a small number of parameters. Instead of computing such parameters on the sink node, a better approach suggests that each node contributes to the computation. Since accessing the data, processing data, and transmitting data are all tasks that consume energy which is a limited resource in sensor node. So, what should be the solution for theoretical and applicative research in SNs for efficient data extraction? This question motivates us to develop a distributed data extraction (DDE) method which pre-processes the raw data directly at sensor node. Hence, instead of sending the raw data to the central site, sensor nodes use their processing abilities to locally carry out simple computations and transmit only the required and pre-processed data. The processing performs at each sensor node is helpful for taking real time decision as well as can serve as prerequisite for development of scalable data mining technique on central side. In DDE method the major contributions are following:

1. Rule based clustering technique for efficiently extracting data from sensors nodes to optimize network lifetime in term of energy and data size. These rules are identified by applying association rule mining on cluster head (CH) node.
2. A significant amount of sensor readings sent from the sensors to the data processing point(s) may be lost or corrupted. In DDE this problem is addressed by estimating missing values from available sensor reading instead of requesting the

sensor node to resend lost reading. The key advantage of our missing value estimation is that it is done directly at sensor node and can be used to identify the behavior of the sensor nodes. Data Reduction is applied which is able to reduce the data size received from sensor nodes. The extracted data is more compact than raw sensor data and can therefore be more efficiently transmitted to sink from the sensor network.

The rest of this paper is organized as follow: after introducing basic concept of SNs data mining in Section 1; we provided an overview of related work of data extraction methods either centralized or distributed in Section 2. Proposed method, algorithms and its details are presented in section 3; Simulation results are presented in section 4 and finally sections 5 concludes the paper and suggest directions for the future work.

2. Related Work

Several techniques have been proposed in the literature to enhance the performance of SNs, such as frequent pattern mining, clustering, classification, prediction, just to mention a few examples. In this section we review past studies in term of three categories related to this research: Association Rule Mining, Missing value identification and Clustering methods.

Tanbeer et al. [18] and Boukerche and Samarah [12] proposed centralized data mining models to find association among the sensors nodes. They proposed tree-based data structure that used FP-growth approach to obtain the frequency of all events detecting sensor. Tanbeer et al. used Sensor Pattern Tree (SP-Tree) to construct a prefix-tree and reorganize the tree in a frequency descending order. Through the reorganization the SP-tree can maintain the frequently event-detecting sensors' nodes at the upper part of the tree, which provides high compactness in the tree structure. Once the SP-tree is constructed FP-growth mining technique is applied to find the frequent event-detecting sensor sets. Boukerche, and S. Samarah [19] used Positional Lexicographic Tree (PLT) structure for mining association rules in which the event-detecting sensors are the main objects of the rules regardless of their values. The mining begins with the sensor having the maximum rank by generating the frequent patterns from its PLT in a recursive way. The computation required at each recursion to update the PLT involved in the prefix part of a pattern. Therefore, the two database scans requirement and the additional PLT update operations during mining limit the efficient use of this approach in handling SNs data. K Romer, [20] and Chong et al. [21] link the problem of mining sensor data to the association rules' mining problem by proposing in-network models. Romer's approach takes into consideration the distributed nature of wireless sensor networks to discover frequent patterns of events with certain spatial and temporal properties. Whereas, Chong et al. finds strong rules from sensor readings and use these learnt rules as a triggers to control sensor network operations or supplement sensor operations. For example, triggers activated from the rules could be used to sleep sensors or reduce data transmissions to conserve sensor energy. Our proposed in-network technique is different from Romer's and Chong et al. approach in a way that extracted rules are used to cluster the sensor node and estimating missing sensor's values.

For missing values identification Halatchev and Gruenwald [22] proposed a centralized methodology called Data Stream Association Rule Mining (DSARM) to identify the missing sensor's readings. It uses Association Rule Mining algorithm to identify sensors that report the same data for a number of times in a sliding window called related sensors and then estimates the missing data from a sensor by using the data reported by its related sensors.

For the clustering issues in sensor networks, several methods have been proposed. Clustering protocol for node clustering such as LEACH [23], ACE [24], HEED [25], DEEH [26] and Energy Aware Protocol (EAP) [27] are proposed to solve energy consumption problems in SNs. These protocols probabilistically selects several nodes as cluster heads according to their residual energy, and then remainder nodes are joined into clusters to minimize the communication cost between them and corresponding cluster heads. Yoon and Shahabi [28], Beyens et al. [29], Yeo et al. [30] proposed data correlation clustering architecture for WSNs in which cluster-heads spatiotemporally correlate. In Beyens et al. approach cluster head maintains a local prediction model that is used to select a suitable node of the cluster to be activated. The idea is to put a sensor node to sleep when there are no objects in its sensing region. In Yoon and Shahabi approach nodes are groups based on similar values and only one reading per group is transmitted. Whereas, in Yeo et al. approach the size of data size is reduced at each cluster head by applying data suppression technique.

All the above techniques have focused on extracting data regarding the phenomenon monitored by the sensor nodes, in which the mining techniques are applied to the sensed data received from the sensor nodes and accumulated at a central database. In our work, we have proposed an in-network data extraction approach to extract the pre-processed sensor data required for mining by applying rule based clustering to save energy and in-network missing value estimation to increase the accuracy of extracted data. Furthermore a data reduction method is used to reduce the transmission energy and data size.

3. Proposed Distributed Data Extraction (DDE) Method

In this section we proposed distributed data extraction methodology for efficiently extracting data from SNs. The main goal is to overcome the challenges for mining continuous stream of data arrived from SNs. We adopted distributed solution where sensor nodes are using their processing capabilities to perform computation and instead of sending the raw data, preprocessed data should be transmitted from nodes to sink. The system workflow consists of three main phases: (1) Clustering of sensor nodes (2) identification of missing sensor and estimation of value (3) data reduction. Our clustering and missing value identification methods are based on association rule mining. To apply association rule mining in SNs we first define association rule mining problem for sensor network.

3.1 Association Rules Mining Problem in Sensors

The association rule mining problem define for transactional database are develop to work on static data and cannot be applied directly on SNs data, where the data is continuous and come with high speed. Static data base algorithms require multiple scans of the original database, which leads to high CPU and I/O costs. Therefore, they are not suitable for a SNs data, in which data can be scanned only once. In view of these challenges we aim to define sensor association rule mining problem. The definition of mining sensor association rules use in our DDE approach following the definition provided by Boukerche and Samarah[19] inspired by the definition of frequent patterns proposed in domain of transactional database by Agrawal et al. [31].

Let $S = \{s_1, s_2, \ldots s_n\}$ a set of sensors in a particular sensor network. We assume that the time is divided into equal-sized slots $(t_1, t_2, \ldots t_w)$ such that $t_{w+1} - t_w = \lambda$ for all $1 < w < n$, where λ is the size of each timeslot, and $T_{his} = t_n - t_1$ represents the historical period of data during data extraction process. The main step in the formation of association rules is to find the patterns of sensors that co-occur together and exceed a certain frequency (these patterns are called frequent patterns). After finding the frequent patterns association rules are generated. For instance, the rule ($s_1 s_2 \rightarrow s_3$) is generated from the pattern ($s_1 s_2 s_3$).

Definition1.1. Suppose sensors data is stored in epoch, where each epoch contains time slot, sensor id and sensor value which sense in given time slot. Let $P = \{s_1, s_2, \ldots s_k\}$ is a set of sensors that detect event in same time slot (Dts) and node value $N_V = \{v(s_1, s_2, \ldots s_k)\}$ then an epoch D is defined as following: $D(Dts, P, N_V)$.

Given a database of epochs (DS) generated after a particular historical period, the problem of mining sensors' association rules is to generate all the rules present in the DS.

Definition1.2. The frequency freq of the pattern P in DS is defined to be the number of epochs in DS that supports it.

Definition1.3. Let min sup represent the minimum number of epoch that P should satisfy. The P is said to frequent if its freq is greater than the min sup i.e.

$Freq\ (P, DS) = \{D\ (D_{ts}, P\} \geq min\ sup$.

Definition1.4. Sensor association rules between two sensor s_1 and s_2 in P are implication of form $s_1 \rightarrow s_2$ where $s_1, s_2 \subset S$ and $s_1 \cap s_2 = \phi$

Definition1.5. Support and confidence of the rule $s_1 \rightarrow s_2$ is defined as follow:

$Freq\ (s_1 \rightarrow s_2) = (s_1 \cup s_2, DS)$

$Conf\ (s_1 \rightarrow s_2) = freq\ (s_1 \cup s_2, DS) / freq\ (s_1)$

The rule $(s_1 \rightarrow s_2\text{: } 90\%, \lambda)$ means if we receive events from sensors s_1 then there is a 90 % chance of receiving an event from sensor s_2 within λ units of time. Note that frequency and support are used interchangeably and min *sup* represents the minimum number of epochs that the frequency of the rules should satisfy. The main challenges of mining these rules can be as follow:

1. How data can be extracted efficiently from the sensor network required for mining process
2. How the patterns that meet the given minimum support can be generated efficiently

3.2 Data Extraction Methodology

The network architecture used for extracting the data is shown in Fig.1. It consists of sensor nodes deployed randomly and the network is divided into groups based on distance from the sink. Each group has its own cluster number and member nodes. The database is attached at sink to store the preprocessed data from each Cluster-Head (CH).

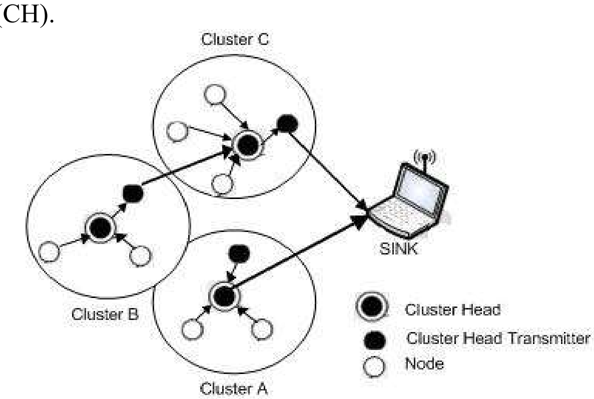


Fig.1. Network Model

1. N sensors are randomly deployed within circular field A. The sink is deployed far away from A

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

2. Every node and sink is at fixed position; the location of sink and distance is known to each node and can communicate directly to sink
3. CH nodes uses clustered based multiple-hop mode of transmission to route the data towards sink
4. All nodes are homogenous means each have same capacities

Table.1 Notations used in algorithms

Notation	Meaning	Notation	Meaning
H_P	Historical Period	T_S	Timeslot
S	Support	N_L	Node Location
S_L	Sink Location	N_V	Node Value
R_L	Rules	N_{TE}	Node Total Energy
R_N	Range	N_{ID}	Node ID
C_D	Cluster Distance	CH_{ID}	Cluster Head ID
C_I	Confidence	CHT_{ID}	Cluster Head Transmitter ID
C_H	Cluster Head	CHT	Cluster Head Transmitter
N_F	Node Frequency		

The data extraction process is shown in Algorithm 1. The notations used in algorithms are shown in Table 1. Algorithm 1 shows the data extraction process starts with the application that provides the mining parameters to the sink which includes *Timeslot size* T_S, *Historical Period* H_P, S*upport S*, *Range* R_N, *Cluster Distance* C_D, Rules R_L and Confidence C_I. The Sink broadcast these parameters to the network nodes. The nodes collect data and transfer it from *node→CH→Sink* or *node→CH→CHT→Sink* in multi-hop fashion. In this way computation load is distributed on sensor nodes especially on CH nodes within network.

3.2.1 Cluster Formation

Algorithm 2 shows the cluster formation process. At the end of each T_S network nodes checks its sensed data and broadcast messages to nodes within given cluster distance C_D for cluster formation. Cluster formation uses the R_N and C_D to group the sensor in same cluster. Upon receiving the broadcasted message each node checks the value of R_N. If its value is within R_N it saves in its buffer and compares C_D with each node's distance. If the distance between nodes is less than or equal to C_D and sensed value is within given R_N then those group of nodes forms a cluster.

In the second round association rules are scanned first for cluster formation. The nodes N_{ID} which are associated they will not broadcast message for cluster formation. These nodes join same cluster within C_D. Nodes which are not will join cluster formation process based on R_N and C_D. e.g. if rules says $S_1S_2 \rightarrow S_3$, $S_1S_3 \rightarrow S_2$, $S_2S_3 \rightarrow S_1$, in this case S_1, S_2, S_3 nodes are in same cluster and only participate in cluster head selection step. These nodes will not participate in cluster formation process in upcoming rounds which save sensor's energy and reduced number of messages broadcast.

Algorithm1. DDE
Input: *Raw Data Stream (DS)*
Output: *Pre-Processed Data (PS)*
SINK:
Broadcast parameters(H_P, T_S, S, R_N, R_L, C_D, C_I)
Upon Receiving all messages
For Slot Number=1to(H_P/T_S)
P=The set of the sensors identifies with in the same timeslot
D=(Slot Number, P)
Insert(D,DS)
NODE:
SET CHFound=False
TimeSlot=1
    For (i=1; to H_P/T_S; i ++)
    Sense Data(N_{ID}, T_S);
    Broadcast (N_{ID}, T_S, T_E, N_L, N_V)
    For (Network Nodes i to n)
    ScanRules (R_L)
    If (ScanRules (R_L)==False)
        {
    ClusterFormation (N_{ID}, T_S, T_E, N_L)
        {
        Range Datagroup(1 to n) within C_D and R_N
        MatchRulesID(R_L)
        CalculateDiatance(N_{ID},N_L)
        Return CH_{ID},CHT_{ID}
        }}
    Else
    Join(N_{ID} Clsuter) //If ScanRules() return True then join cluster within given C_D
    SET CHFOUND=True
    CHBroadcast (CH_{ID})
    CHEpoch=TransferData(N_{ID},N_v)
    TimeSlot= TimeSlot++
}
MissingValues(SensorAssociationEpoch);
DataEstimation(CHEpoch)
ApplyReducation(PEpoch)
//@TansfeEnergy=Amount of Energy required to transmit Epoch to Sink
If(CH_{ID} T_E<@TransferEnergy)
{
Transmit data CHT_{ID} (PreprocessedEpoc)
Send to SINK (PreprocessedEpoch DS, Rules R_L)
}
Else
Send to SINK (PreprocessedEpoch DS, Rules R_L)

3.2.2 Cluster Head Selection

Upon completion of cluster formation process each node has its cluster members N_{ID}, node location N_L, N_{TE} and sink location S_L in its buffer. The node having maximum energy will calculate minimum distance of each node within cluster called Cluster Head (CH) and broadcast CH_{ID} to network nodes. It also calculates the node having minimum distance from the sink called Cluster Head Transmission (CHT) and broadcast CHT_{ID} to CH. This node (CHT) will be used to transmit data toward sink if remaining CH energy is not sufficient for data transformation after computation. Once each node knows it's CH it transmits data to CH. Upon receiving data from each member node CH start computing associated sensors, missing values identification and data reduction. After this processing if CH residual energy is sufficient to transfer the data to sink it directly transfer to sink. If CH energy has not sufficient to transfer data then it handover data to CHT which will act as gateway and send the data towards sink.

Algorithm2. Cluster Formation
Input: *N_{ID}, N_L, T_S,N_{TE},N_V,R_N,C_D*
Output: *CH_{ID}, CHT_{ID}*

Node: *Total Received Request from Nodes R_N*
ClusterFormation (N_{ID}, T_S, T_E,N_L,C_D)
{
NodePower=0
Set NodeDistance=C_D
NodeBroadcast(N_{ID},N_L, T_S, T_E,N_V) //within given C_D
Receive (N_{ID}, N_L, N_V) //Receive with C_D sensor Ids
Compare data range R_N
For (Sensor S_i=1 to S_i=n, i++)
{
//Compare each S_i R_N and C_D to
If (R_{Ni}- R_{Ni+1} <=R_N && C_{Di}- C_{Di+1} <=C_D)
{
JoinCluster (R_{Ni+1})

For(i=1 to i_{Rn})
{
Calculate (NodePower<Maximum (T_{Ei}) && NodeDistance<Minimum(S_L-N_{Li}))
Set NodePowerID=Maximum (E_{Si})
Set NodeDistanceID=Minimum(S_L-N_{Li})
}
i ++

Return (NodePowerID CH_{ID}, NodeDistanceID CHT_{ID})

Following example along the test data explain the process of data extraction approach from a random cluster of network. Let S = {s_1, s_2, s_3 ...... s_n} be the sensors in a particular sensor network. Let the timeslot T_S size equals to 5 minutes and the historical period H_P is 35 minutes.

Assume the extraction process is initiated at time 08:00. At end of each time slot the nodes sensed in that timeslot will broadcast (N_{ID}, T_S, N_E, N_L, N_V) to neighboring nodes within given R_N and C_D provided by the sink to form clusters. Node having maximum energy will identify the two others nodes having minimum distance from sink called *Cluster Head Transmitter CHT* and maximum energy within cluster members called *Cluster Head* CH. Table.2 shows the detected events with in the sensor network within each timeslot. At the end of the first timeslot at (08:05), sensors *(s_2, s_3, s_4, s_5)* send the messages respectively. The same process is repeated periodically for each timeslot until the end of the historical period. Table.3 shows the extracted epochs after one historical period of 35 minutes from one network cluster as an example.

Table 2: Sensor readings each timeslot

T_S	N_{ID}	N_V	T_S	N_{ID}	N_V	T_S	N_{ID}	N_V
1	S_2	1	2	S_4	6	5	S_2	4
1	S_3	3	2	S_5	5	5	S_3	3
1	S_4	4	3	S_4	4	6	S_3	7
1	S_5	3	3	S_5	5	6	S_5	6
2	S_1	2	4	S_3	1	7	S_3	4
2	S_2	3	4	S_4	3	7	S_4	2
2	S_3	2	4	S_5	4			

After the historical period CH start processing to identify the frequent sensors to identify association rules among sensors and the estimation of missing values.

Table 3: Data stored at CH

T_S	(N_{ID}, N_V)
1	($S_2S_3S_4S_5$, 1343)
2	($S_1S_2S_3S_4S_5$, 23265)
3	(S_4S_5, 45)
4	($S_3S_4S_5$, 134)
5	(S_2S_3, 43)
6	(S_3S_5, 76)
7	(S_3S_4, 42)

Algorithm.3 is used to estimate missing sensor value which is identified by finding the frequent sensors where a sensor has its support *S* is higher than given *S* reported in set of epoch. If the numbers of sensors are denoted by *n* then the maximum number of possibly existing frequent sensors are:

$$\max\_num\_freq\_sensor = \sum_{i=1}^{n} \binom{n}{i}$$

Table 4.a. shows the individual frequency of each sensor reported data in a given historical period. Suppose given support S>3, Table 4.b. shows the frequent sensors at level-1 where support is higher than given support *S*. In the next step we generated the level-2 frequent sensor from level-1 by calculating the *S* in set of epoch where both

appear together in same timeslot as shown in Table.4.c. Table.4.d shows level-2 frequent sensors having $S>3$.

Table: 4.a Sensor NF

N_{ID}	N_F
S_1	1
S_2	3
S_3	6
S_4	5
S_5	5

Table: 4.b Sensor NF>3

N_{ID}	N_F
S_3	6
S_4	5
S_5	5

Table: 4.c Sensor level-2 NF

N_{ID}	N_F
S_3S_4	3
S_3S_5	4
S_4S_5	4

Table: 4.d Sensor level-2 NF>3

N_{ID}	N_F
S_3S_5	4
S_4S_5	4

After identification of frequent sensors at level-2 they are used to find association rules. Association rules are in the form ($S_1 \rightarrow S_2$). The frequency of rule ($S_1 \rightarrow S_2$) is the frequency of the ($S_1 \cup S_2$). The *confidence* of the rule is defined as*:*

$Conf\,(S_1 \rightarrow S_2) = Freq\,(S_1 \cup S_2, DS)/\,Freq\,(S_1,DS)$

Following rules are identified from frequent sensor (s_3s_5) and (s_4s_5). The confidence value is set to 60 %.

Conf (S3 → S5)= Freq(S3 ∪ S5, DS)/ Freq(S3,DS)=66%
Conf (S5 → S3)= Freq(S5 ∪ S3, DS)/ Freq(S5,DS)=80%
Conf (S4 → S5)= Freq(S4 ∪ S5, DS)/ Freq(S4,DS)=80%
Conf (S5 → S4) = Freq(S5 ∪ S4, DS)/ Freq(S5,DS)=80%

The rules are used to estimate the values of sensors in those timeslots where the associated sensors have not reported data. By analyzing the values of associated sensors we can identify the upper and lower bound value range reported in each historical period. Table.5 shows the value of associated sensors pairs identified from Table 3.

Table 5: Lower and upper bound sets

N_{ID}	Pairing Values
S_3S_5	(3-3), (2-5), (1-4), (7-6)
S_4S_5	(4-3), (6-5),(4-5), (3-4)

The missing pair of associated sensor is identified for each time slot. In Table.3 it can be observed that in timeslot 3, 5, 6 and 7 the associated sensor pair is missing. The value of missing pair is added by using the Average Round (AR) approach. For example in timeslot 3 value of s_3 is missing. It can be estimated from values available in Table.5. Now initial set of epoch contains values from missing pair of associated sensor shown in Table.6.

Table.6: Final DS along estimated values

T_S	(N_{ID}, N_V)
1	($S_2S_3S_4S_5$, 1343)
2	($S_1S_2S_3S_4S_5$, 23265)
3	($S_3S_4S_5$, 445)
4	($S_3S_4S_5$, 134)
5	($S_2S_3S_5$, 434)
6	($S_3S_4S_5$, 756)
7	($S_3S_4S_5$, 425)

Algorithm3. Missing Value Estimation
Input: *Epoch contains Missing Values*
Output: *Epoch contains Estimated Values*

CH Node:
```
//Traverse Readings from 1 to Max Reading Size for
frequent sensor readings
Set CountSensor =Count (Si)
        For(i=1 to 2, i++)
        {
         call Frequent(frequentepoch)
        }
        call Frequent()
        {
        For (i=1 to N= HP /TS; i ++)
        {
        Set CountSensor =Count (Si)
         If(CountSensor>=S) //where S is given
        Support value
        Set Frequency [SFi,]=Si[ CountSensor]
        Else
        Do Nothing
        Return frequentepoch
        }
\\Traverse epoch for frequent sensor's readings to
estimate missing value in each window slot
    For ([Si] to [Sn])
    {
        If( SFi>=S)
        {
        HighBound[Si]= Get (Max(Epoch))
        LowBound[Si]= Get (Min(Epoch))
        Estimated [Si]=Avg(HighBound[Si],
        LowBound[Si])
        }
    //Traverse Epoch for Find Missing Si
        For (i=1 to HT / WS; i ++)
        {
        If (Traverse Epoch=Found)
            Set PostionSi= Estimated [Si]
            Else
            Do nothing}}
 return ESTEpoch
```

Table.7: DS before reduction

T_S	(N_{ID}, N_V)
1	$(S_2S_3S_5S_4, 1334)$
2	$(S_1S_3S_2S_4S_5, 22365)$
3	$(S_3S_4S_5, 445)$
4	$(S_3S_4S_5, 134)$
5	$(S_3S_5S_2, 344)$
6	$(S_4S_5S_3, 567)$
7	$(S_4S_3S_5, 245)$

Table.8: DS after reduction

T_S	(N_{ID}, N_V)
1	$(S_2S_3S_5S_4, 134)$
2	$(S_1S_3S_2S_4S_5, 2365)$
3	$(S_3S_4S_5, 45)$
4	$(S_3S_4S_5, 134)$
5	$(S_3S_5S_2, 34)$
6	$(S_4S_5S_3, 567)$
7	$(S_4S_3S_5, 245)$

By applying estimation process the data size *DS* is increased to reduce *DS* we applied reduction process as shown in Algorithm 4. This process first sorts the reported values as shown in Table.7. It can be observed that it contains same reported value in same timeslot within cluster. Data reduction process identifies these values and removes it from the *DS* by using *Right Trim rule*. After data reduction process the *DS* as shown in Table.8 transmitted to sink.

Algorithm4. Data Reduction

Input: *ESTEpoch*
Output: *FinalEpoch*
CH Node:

```
//Traverse Epoch to find duplicate values from different sensor IDs in same WS
ApplyReduction(ESTEpoch)
{
For ( i=1 to HT/WS of ESTEpoch; i++)
 {
        While (NIDBuffer==Finish)

        {
                Traverse ([NID_i],[Value_i])
                If ([NID_i],[Value_i])== ([NID_i+1],[Value_i+1])
                Set Position=Position([NID_i],[]
                Else
                Next Match NID Value with Initial Value
        }

Return FinalEpoch }                        \\Send FinalEpoch to SINK
```

This same process is executed on each cluster within network and each CH computes these values before sending *DS* towards sink. After this computation it may be possible that CH energy will not remain enough to transmit, so it can transmit *DS* to CHT, because it has the minimum distance from sink or neighboring cluster head and having sufficient energy to transmit towards sink.

Sink will receive the *DS* along estimated values and identified association rules after each historical period. Before the start of next historical period sink will broadcast these rules along other parameters for clustering formation process. After each round new rules will be evaluated on sink from historical datasets for efficient network clustering.

4. Experimental Results

We evaluated the performance of DDE algorithm using NS2 simulator. All experiments are based on 2.2 GHz computer with 2GB RAM and Windows XP operating system. In the network of 300 nodes, all nodes are homogenous and deployed randomly. We compared the DDE with LEACH in term of network lifetime, number of cluster heads, messages delivered, data size and number of rounds.

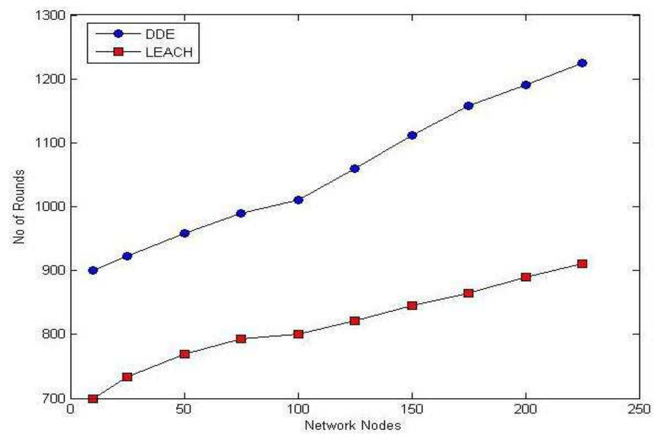


Fig.2 No of Rounds

Figure.2 shows the impact of number of rounds on network lifetime; DDE shows the good behavior if the networks size grows, whereas LEACH has less impact on network lifetime as compared to DDE.

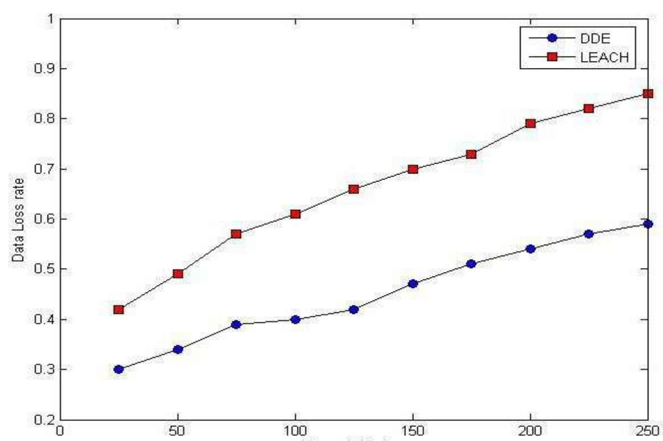


Fig.3 Data loss rate

Due to data estimation algorithm the data loss rate is also less in DDE as shown in Figure.3. The number of messages broadcast is higher in LEACH which results into more energy consumption. When the data loss is low it also consume energy but in DDE data loss is handled after data extraction step as compared to LEACH, DDE consume less energy and message broadcast during data extraction. The energy consumption and message broadcast during data extraction process is also improved as shown in Fig.4 and Fig.5.

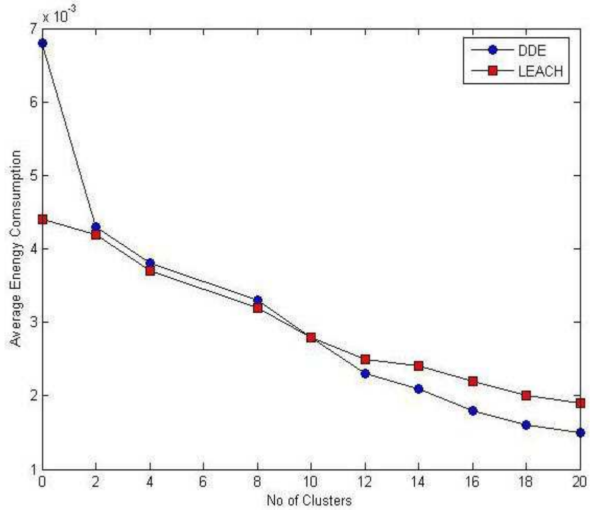


Fig.4 Avg. energy consumption

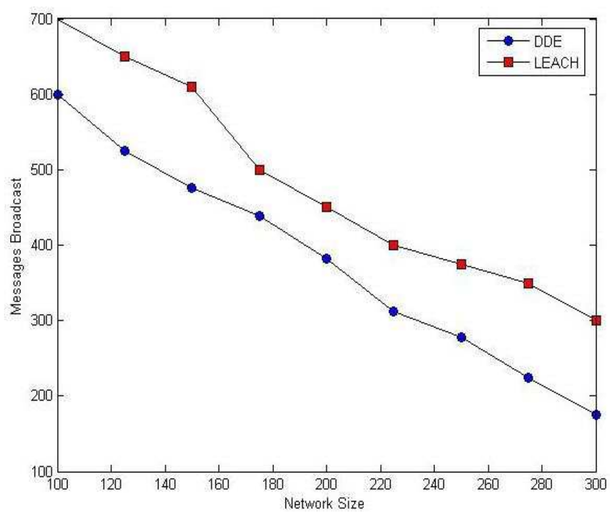


Fig.5 Messages broadcast

LEACH uses the random cluster head scheme in each network block so the numbers of force cluster heads are also increased whereas DDE uses data value range, sink distance and residual energy to create cluster and cluster heads. When number of rounds reaches more than 500 it nearly close to DDE because the numbers of still alive nodes and their residual energy remain less within network as shown in Fig.6 but during the initial rounds DDE has less no of force cluster heads.

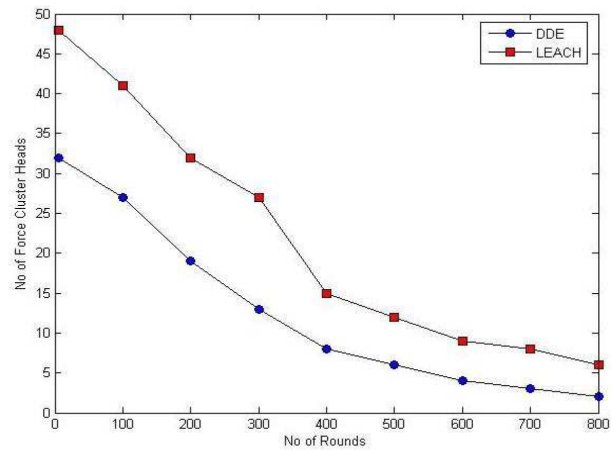


Fig.6 No of force cluster heads

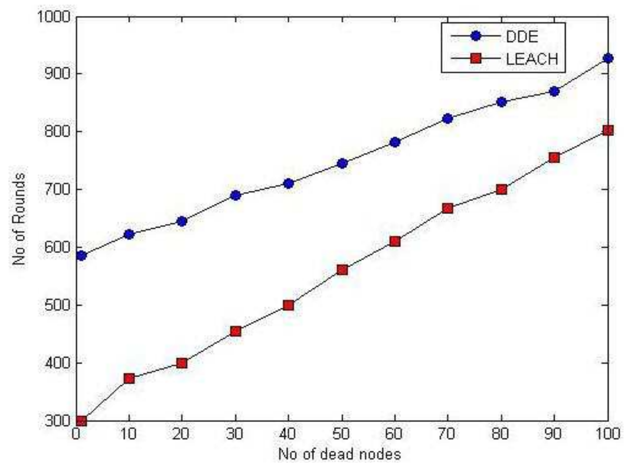


Fig.7 No of dead nodes

Sensor nodes are energy-constrained, so the network's lifetime is important for SNs application. When the number of dead node increases, the network cannot make more contributions. Thus, the network lifetime should be defined as the time when enough nodes are still alive to keep the network operational. As shown in Fig.7 LEACH has more no of dead nodes in initial rounds whereas DDE retains maximum number of nodes alive. If we compared for equal number of rounds in LEACH 100 nodes are dead in 802 rounds whereas in DDE same number of nodes are dead after 928 rounds.

5. Conclusion

In this paper, we have introduced a new Distributed Data Extraction (DDE) approach which consists of rule based cluster formation and identification of correlated sensor. DDE captures the temporal and data relation between the sensors by using association rule mining. The rules identified by DDE are also used to estimate the value of missing sensor within in cluster. In subsequent round these rules are used in cluster formation process where correlated sensors join the same cluster. Results show the DDE outperforms LEACH by significant margin particularly for network life time. DDE maximize the network lifetime by reducing the number of broadcast messages, energy consumption, number of dead nodes, forced cluster heads and data loss rate and maximize the number of rounds during data extraction process.

As future work, we are going to mine the extracted data on central side (SINK) to analyze the behavior of entire sensor network. By applying mining techniques at sink we can find global patterns that can be used for different purpose such as predicting the future sources of events and faulty node identification. The ongoing task of this research work is the building of adaptive data mining framework for sensor network applications

Acknowledgments

This work was supported in part by the Joint Funds of NSFC-Microsoft Research Asia under Grant No. 60933012, the Specialized Research Fund for the Doctoral Program of Higher Education under Grant No.20110142110062 and International S&T Cooperation Program of Hubei Province under Grant No. 2010BFA008.

References

[1] A. Rozyyev, H. Hasbullah, and F. Subhan, "Indoor Child Tracking in Wireless Sensor Network using Fuzzy Logic," *Research Journal of Information Technology,* vol. 3, no. 2, pp. 81-92.

[2] R. Szewczyk, E. Osterweil, J. Polastre *et al.*, "Habitat monitoring with sensor networks," *Communications of the ACM,* vol. 47, no. 6, pp. 34-40, 2004.

[3] S. Chauhdary, A. Bashir, S. Shah *et al.*, "EOATR: Energy Efficient Object Tracking by Auto Adjusting Transmission Range in Wireless Sensor Network," *Journal of Applied Sciences,* vol. 9, no. 24, pp. 4247-4252, 2009.

[4] P. K. Biswas, and S. Phoha, "Self-organizing sensor networks for integrated target surveillance," *Computers, IEEE Transactions on,* vol. 55, no. 8, pp. 1033-1047, 2006.

[5] L. Lee, and C. Chen, "Synchronizing Sensor Networks with Pulse Coupled and Cluster Based Approaches," *Information Technology Journal,* vol. 7, no. 5, pp. 737-745, 2008.

[6] N. Sabri, S. Aljunid, R. Ahmad *et al.*, "Wireless Sensor Actor Network Based on Fuzzy Inference System for Greenhouse Climate Control," *Journal of Applied Sciences,* vol. 11, pp. 3104-3116.

[7] D. Kumar, "Monitoring forest cover changes using remote sensing and GIS: A global prospective," *Res. J. Environ. Sci,* vol. 5, pp. 105-123, 2011.

[8] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer networks,* vol. 52, no. 12, pp. 2292-2330, 2008.

[9] T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks," in Proceedings of the 13th Mediterranean Conference on Control and Automation, 2005, pp. 719-724.

[10] Y. C. Tseng, M. S. Pan, and Y. Y. Tsai, "Wireless sensor networks for emergency navigation," *Computer,* vol. 39, no. 7, pp. 55-62, 2006.

[11] A. Mahmood, K. Shi, and S. Khatoon, "Mining Data Generated by Sensor Networks: A Survey," *Information Technology Journal,* vol. 11, pp. 1534-1543.

[12] A. Boukerche, and S. Samarah, "An Efficient Data Extraction Mechanism for Mining Association Rules from Wireless Sensor Networks," in *IEEE International Conference on Communications (ICC'07)* 2007, pp. 3936-3941.

[13] Y. Chi, H. Wang, P. S. Yu *et al.*, "Moment: Maintaining closed frequent itemsets over a stream sliding window," in Fourth IEEE International Conference on Data Mining ICDM'04, 2004, pp. 59-66.

[14] M. H. A. Awadalla, and S. G. El-Far, "Aggregate Function Based Enhanced Apriori Algorithm for Mining Association Rules," *International Journal of Computer Science Issues,* vol. 9, Issue 3, no. 3, pp. 277-287, May 2012.

[15] M. S. Gouider, and M. Zarrouk, "Frequent Patterns mining in time-sensitive Data Stream," *International Journal of Computer Science Issues,* vol. 9 Issue 4, no. 2, pp. 117-124, July 2012.

[16] K. P. Lakshmi, and C. Reddy, "Compact Tree for Associative Classification of Data Stream Mining," *International Journal of Computer Science Issues,* vol. 9, Issue 2, no. 1, pp. 624-628, March 2012.

[17] B. J. LAKSHMI, and M. NEELIMA, "Maximising Wireless sensor Network life time through cluster head selection using Hit sets," *International Journal of Computer Science Issues,* vol. 9, Issue 2 no. 3, Marh 2012.

[18] S. K. Tanbeer, C. F. Ahmed, B. S. Jeong *et al.*, "Efficient mining of association rules from wireless sensor networks," in *11th International Conference on Advanced Communication Technology (ICACT'09)* 2009, pp. 719-724.

[19] A. Boukerche, and S. Samarah, "A novel algorithm for mining association rules in wireless ad hoc sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, pp. 865-877, 2007.

[20] K. Romer, "Distributed mining of spatio-temporal event patterns in sensor networks," *Proc. of the 1st Euro-American Wkshp. on Middleware for Sensor Networks (EAWMS)*, 2006.

[21] S. K. Chong, S. Krishnaswamy, S. W. Loke *et al.*, "Using association rules for energy conservation in wireless sensor networks," in *Proceedings of the 2008 ACM symposium on Applied computing*, 2008, pp. 971-975.

[22] M. Halatchev, and L. Gruenwald, "Estimating missing values in related sensor data streams," in Proc. 11th Int'l Conf. Management of Data (COMAD '05), 2005.

[23] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 2000, pp. 10 pp. vol. 2.

[24] H. Chan, and A. Perrig, "ACE: An emergent algorithm for highly uniform cluster formation," *Wireless Sensor Networks*, pp. 154-171, 2004.

[25] O. Younis, and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, pp. 366-379, 2004.

[26] W. D. Wang, and Q. X. Zhu, "A hierarchical clustering algorithm and cooperation analysis for wireless sensor networks," *Journal of Software,* vol. 17, no. 5, pp. 1157-1167, 2006.

[27] M. Liu, J. Cao, G. Chen *et al.*, "An energy-aware routing protocol in wireless sensor networks," *Sensors,* vol. 9, no. 1, pp. 445-462, 2009.

[28] S. Yoon, and C. Shahabi, "The Clustered AGgregation (CAG) technique leveraging spatial and temporal correlations in wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN),* vol. 3, no. 1, pp. 3, 2007.

[29] P. Beyens, A. Nowe, and K. Steenhaut, "High-density wireless sensor networks: a new clustering approach for prediction-based monitoring," in Proceeedings of the Second European Workshop on Wireless Sensor Networks, 2005, pp. 188-196.

[30] M. H. Yeo, M. S. Lee, S. J. Lee *et al.*, "Data correlation-based clustering in sensor networks," in International Symposium on Computer Science and its Applications, CSA 08., 2008, pp. 332-337.

[31] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules," *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, Citeseer, 1994, pp. 487-499.

An Efficient Framework for Medical Image Retrieval System using Contribution Mechanism

Dr. B.P. Mallikarjunaswamy[1]

Mr. Shivamurthy R.C[2]

[1] *Prof.*: Dept of Computer Science & Engg
Sri Siddharatha Institute of Technology
Tumkur, Karnataka:572106, India

[2] *Research Scholar*: Dept of Computer Science & Engg.
Akshaya Institute of Technology
Tumkur, Karnataka: 572106, India

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. We apply the algorithm to content-based image retrieval and compare its performance with that of the k-means clustering algorithm. Unlike the k-means algorithm, our algorithm optimizes on both intra-cluster and inter-cluster similarity measures. It has three passes and each pass has the same time complexity as an iteration in the k-means algorithm. Our experiments on a bench mark image data set reveal that our algorithm improves on the recall at the cost of precision.**

Keywords-; Content based image Retreival System, clustering, k-means algorithm, medical image retrieval

I. 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 several years [14]. This finds application in image retrieval. Content-based image retrieval (CBIR) aims at 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.

We use the notion of 'contribution of a data point' for partitional clustering. The resultant algorithm requires only three passes and we show that the time complexity of each pass is same as that of a single iteration of the k-means clustering algorithm. While the k-means algorithm optimizes only on the intra-cluster similarity, our algorithm also optimizes on the inter-cluster similarity. Clustering has widespread applications in image processing. Color-based clustering techniques have proved useful in image segmentation [13]. The k-means algorithm is quite popular for this purpose. Clustering based on visual content of images is an area that has been extensively researched for

finding images of interest from a large image database using the visual content of the images. Traditional approaches to image retrieval were text-based, where individual images had to be annotated with textual descriptions [8]. Since this is a tedious manual task, image retrieval based on visual content is very essential.

Organizing the retrieved search results into clusters is an intuitive form of content representation [14] and facilitates user's browsing of images [15]. Image clustering can also be used to optimize the performance of a CBIR system [16]. While the performance of a number of clustering algorithms in image retrieval have been analyzed in previous works [17, 18, 19, 20], we apply our proposed algorithm to CBIR and compare its performance with that of the k-means clustering algorithm.

K. Kim and R. H. Park [3,4,5]" the image retrieval scheme for JPEG formatted image is presented. Content based image retrieval for JPEG images has attracted many people's attention and a series of algorithms directly based on the discrete cosine transform domain. And to take full advantage of DCT coefficients and consider the color and texture information for the retrieval of JPEG formatted images. Here decompressing the images and then performing in the spatial domain. The feature vectors are computed from several DCT coefficients. And this operation is performed in the partial decoded domain. It can greatly decrease the retrieval complexity. M. Flickner et.al [6,7,8] proposed Color histograms are computationally efficient, and generally insensitive to small changes in camera position.

However, a color histogram provides only a very coarse characterization of an image, An images with similar histograms can have dramatically different appearances. Here, to describe a method which imposes additional constraints on histogram based matching. In histogram refinement, the pixels within a given bucket are split into classes based upon some local property. Split histograms are compared on a bucket by bucket basis, similar to standard histogram matching. Within a given bucket, only pixels with the same property are compared. Two images with identical color histograms can have different split histograms, split histograms create a finer distinction than color histograms. This is particularly important for large image databases, in which many images can have similar color histograms.

To describe a split histogram called a color coherence vector (CCV), which partitions each histogram bucket based on spatial coherence. A database with 15,000 images can be queried using CCV's in under 2 seconds. And to demonstrate that histogram refinement can be used to distinguish images A. P. Berman et.al [9,10] found that technique fairly integrates a diverse and expandable set of image properties (color, texture, and location) in a retrieval framework, and allows end users substantial control over their use. We propose a novel set of evaluation methods in

The rest of the paper is organized as follows. In section 2 we give an overview of related work which identifies all the major research work being done in this area. Section 3 highlights about the medical image retrieval system. Proposed system is discussed in section 4 followed by result and performance analysis in Section 5 and finally in section 6 we make some concluding remarks.

II. Related Work

Jing Huang ed al [1, 2] discussed new feature called color correlogram for image indexing and comparison. These new features computed efficiently and show that performance very well. Sim, D. G., H.

addition to applying established tests for image retrieval; our technique proves competitive with state of art methods in these tests and does better on certain tasks. The Stairs algorithm can operate in a regional query mode with only a moderate increase in computational overhead. For certain queries this capability significantly increases the relevance of the images retrieved. Furthermore, it improves on many standard image retrieval algorithms by supporting queries based on subsections of images. The merits of drawing on different types of image features for Image retrieval are firmly established.

Our work capitalizes on this trend, providing a framework for fairly and consistently integrating diverse image properties into a description amenable to fast, reliable retrieval. J. Zhang et.al [11, 12, 13] suggest the image retrieval based on the textural information of an image, such as orientation, directionality, and regularity. Here, utilize texture orientation to construct the rotated Gabor transform for extraction of the rotation-invariant texture feature. The rotation-invariant texture feature, directionality, and regularity are the main features used in the proposed approach for similarity assessment. Using these features, we finally propose an efficient mechanism for CBIR and examine it through some applications. The system can now compare features of the query with features of images in the collection based on some matching criterions. Because three features are used in this work, three matching scores need to be computed. A weighted average of the matching scores is then calculated to get a final score for each image.

Finally, rank images based on these final scores and top-ranked images are displayed to the user as the result of retrieval. Haralick RM et.al [14,15]discussed the four image features are extracted by this system, which are color feature (HSV color histogram), texture feature (co-occurrence matrix), shape feature (moment invariant based-on threshold optimization), spatial relationship feature (based-on the Markov chains). According to the statistical analysis of the experiment results discover that the four visual features describe image characters variously. The retrieval precision based on color feature is better than based on texture feature. An image retrieval method combined color and texture features. According to image texture characteristic, a kind of image feature statistic is defined. By using feature weight assignment operators designed here, the method can assign weight to color and texture features according to image content adaptively and realize image retrieval based on combined image features. The retrieval results are more exact and efficient than other methods based on single feature and simple linear combined features of fixed weight, the retrieval results are more suitable to the human visual characteristic. The error matching is decreased and weight assignment is logical.

P.S.Hiremath ed al," [16,17,18,19] discussed four approaches such as multispectral Approach, HSV color space, YCbCr color space,and uses gray scale texture features for color texture analysis. the wavelet decomposed coefficient of image and its complements by using texture feature. Their experiments are carried out on Wang's dataset using JSEG for segmentation and compare the four different color space finally haar wavelet is more effective in texture feature compare with other wavelet so, the results are encouraging. P. S. Hiremath and Jagadeesh Pujari [20]discussed An integrated matching scheme based on higher priority of similar image and the adjacency matrix of a bipartite graph by using tiles of query shape information is computed by Gradient Vector Flow fields. This demonstration is efficiency compare with wavelet method. K.P.

Ajitha Gladis and K.Ramar [21,22,23,24] discussed mainly as the image can be represented on statistical properties, morphological features and fuzzy cluster features of the image in order to get more accurate results. He distance is measured through a back propagation network.

So, experimental results is quite effective in both performance and retrieval rate. Son Lam Phung and A. Bouzerdoum [25] proposed new feature called edge density. It differentiates objects from non-objects using image edge characteristics. This approach is based on a fast object detection method. The edge density, which measures the specific region of the object, that can be computed more efficiently.

Where each feature is the average edge magnitude in a specific subregion. The new feature capability compared to the Harr-like features[30]. Finally new feature show good discriminative capability. S.Nandagopalan ed al [26,27,28] discussed texture for texture co-occurrence matrix based entropy, energy, etc, and for edge density , Edge Histogram Descriptor (EHD).For retrieval of images, finally to reduce the computational complexity based on greedy stategy. So, its achieved better results for both local and global feature.

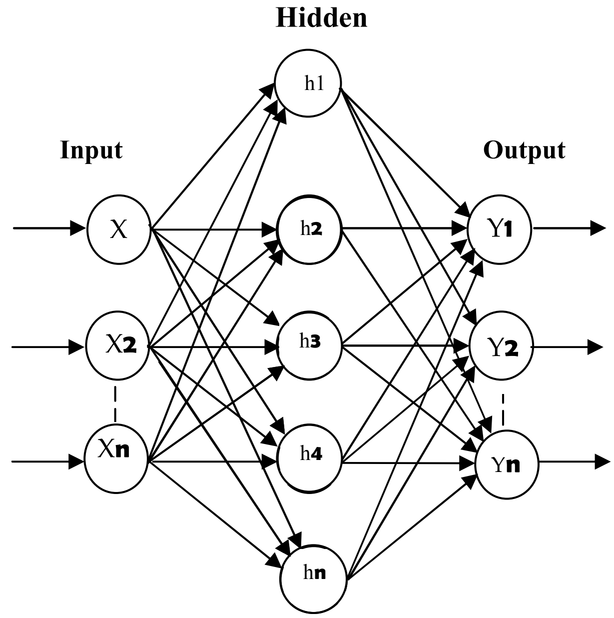


Fig 2: Back Propagation Network

Fig 3:Left: an image window. Middle: the edge magnitude.

Right: three edge density features

Mamta Juneja and Parvinder Singh Sandhu[29,30,31] proposed candy method for edge detection. Here to compare canny method with laplacian of Gaussian method .

So, the result shows canny's edge detection algorithms is performs better then laplacian method. Michele Saad [32.33] discussed to compare four color feature extraction algorithms such as

Operator	Canny	Lap of Gaussian
Canny	1	0.62386
Lap of Gaussian	1.602916	1
Prewitt	3.723412	2.322899

1) the conventional color histogram,

2) the fuzzy color histogram,

3) the color correlogram, and

4) a color/shape-based method and four texture feature extraction techniques such as

1) the steerable pyramid,

2) the contour let transform,

3) the Gabor wavelet transform, and

4) the complex directional filter bank.

Finally, the fuzzy color histogram and the Gabor wavelet transform were shows the highest color and texture retrieval results. In "J.Huang et.al" [34,35]proposed two stages for retrieving an image such as hierarchical clustering and then apply the clustered images to RBFN network.

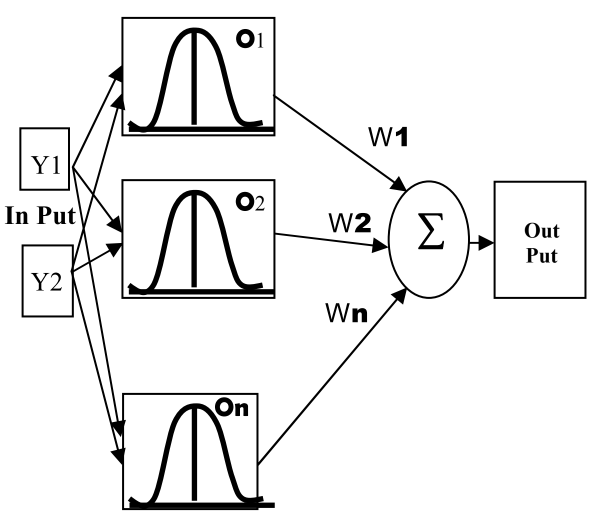


Figure 1 Give caption.

Hierarchical Algorithm is used to group similar images into clusters and RBFN Network which uses K-Means clustering and Gaussian function to retrieve the similar images. So that its get the better favored image results.N Gnaneswara Rao ed al [36,37,38] discussed the texture of an image is computed by using wavelet transformation, because its quite efficiently and also using clustering algorithm, to construct indexed image database based on the texture feature. Finally,clustering is to give the good matching and reduce the undesirable noice. P.AnandhaKumar and V.Balamurugan proposed two indexing technique such as Spatial assess method (SAM) and metric access method (MAM).

SAM providing good result on low dimensional feature. MAM-based balanced and dynamic indexing technique called feature based adaptive tolerance tree. Feature Based Adaptive Tolerance Tree (FATT), which brings effective solution and to increase efficiency of image retrieval. Rajshree S. Dubey et. al[39]discussed four techniques as Color Histogram, Color moment, Texture, edge histogram and it involves pattern recognition. Because it's most important tool for machine vision. Therefore, the combination of four techniques gives better result.

Kondekar V. H. et. al [40] discussed Image color quadratic distance for image histogram, Image Euclidian distance for image wavelet transform; image Hamming Distance for retrieval of an image. From these distance formulae is to increase the retrieval efficiency of an image. Ritu Shrivastava et. al [41,42] discussed to compare two clustering techniques such as K- mean and C-mean clustering for distance metric concept. Finally, K- mean algorithm is easy and fast to compute. C- Mean algorithm takes long computational time. Both converges but suffers from the problem of local minimum. In this survey paper candy method is easy and fast to compute the process. Image splitting and image compaction is to reduce the computation complexity by reducing feature vector size and Haar wavelets are used, since they are more effective compared to other wavelets. In each of the paper they provide several methods in that each method fulfills their works. The results are quite good for most of the query images and it is possible to further improve, to use genetic algorithm, cluster algorithm such as hierarchical clustering, Cure data Clustering, fusion algorithm and any other technique will including in CBIR, it will give the better and effective retrieval of an image.

III. MEDICAL IMAGE RETRIEVAL SYSTEM

Content-based image retrieval (CBIR) aims at finding images of interest from a large image database using the visual content of the images. Traditional approaches to image retrieval were text-based, where individual images had to be annotated with textual descriptions. Since this is a tedious manual task, image retrieval based on visual content is very essential. In this paper, we focus on the application of clustering to content-based image retrieval. A large collection of images is partitioned into a number of image clusters. Given a query image, the system retrieves all images from the cluster that is closest in content to the query image. We apply the proposed contribution-based clustering algorithm to image retrieval and compare its performance with that of the k-means algorithm.

Content Based Image Retrieval (CBIR) is a technique which uses visual contents, normally called as features, such as shape, color, texture, edge. etc…to search images from large scale image databases according to users' requests in the form of a query image. Content based retrieval of visual data requires a paradigm that differs significantly from both traditional databases and text based image understanding systems. The challenge in CBIR is to develop the methods that will increase the retrieval accuracy and reduce the retrieval time. Among them, Color feature is often broadly used to describe the images which are difficult to be segmented and needn't to consider space information. Texture is one of the most important ones, due to its presence in most real and synthetic world images, which makes it under high attention not only for CBIR but also for many other applications in computer vision, medical imaging, remote sensing, and so on .Finally the edge features that include five categories vertical, horizontal, 45 degree diagonal, 135 degree diagonal, and isotropic are added.

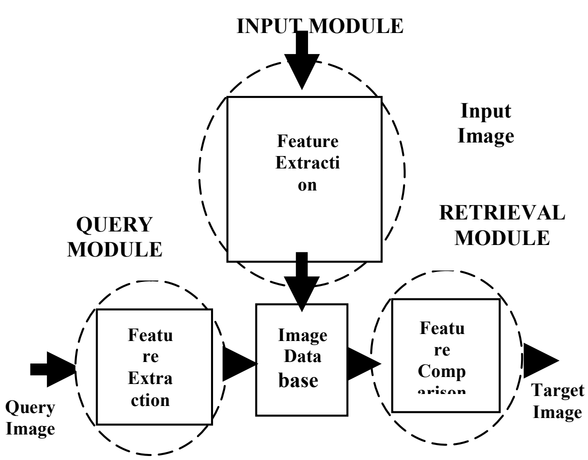


Fig 1: Block diagram of the content-based image retrieval system

In this paper, we focus on the application of clustering to content-based image retrieval. A large collection of images is partitioned into a number of image clusters. Given a query image, the system retrieves all images from the cluster that is closest in content to the query image. The overall system is shown in Fig. 1. We apply the proposed contribution-based clustering algorithm to image retrieval and compare its performance with that of the k-means algorithm.

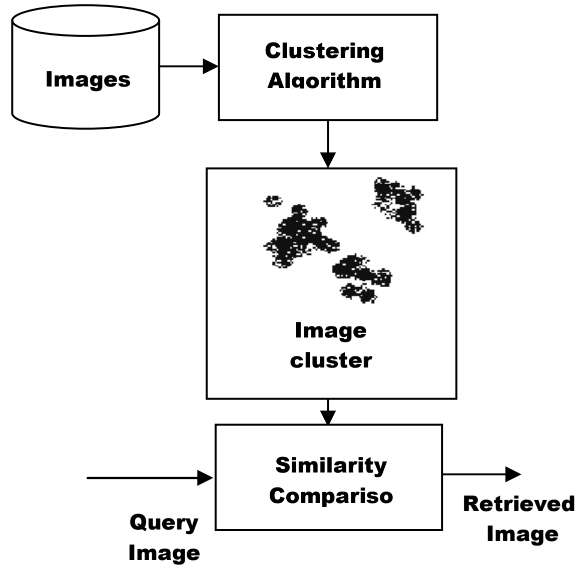


Fig. 2. Content-based Image Retrieval (CBIR) System.

Each image in the database is represented by a visual content descriptor consisting of a set of visual features [8]. A similarity/dissimilarity measure is then used to retrieve images whose features are closest to that of the query image. A common distance/dissimilarity metric is the Euclidean distance, which is used in our work. To represent the visual content of an image, we use a RGB color histogram. The color coordinates of the RGB color space are uniformly quantized into a number of bins. In our work, we use 8 bins each for the Red, Green and Blue coordinates, resulting in 512 bins/features.

IV. PROPOSED SYSTEM

The proposed system highlights a novel method to perform contribution based medical image retrieval (MIR) where the term 'contribution' is empirically designed considering clustering of image pixels. The proposed system is based on partitional clustering that aims at partitioning a group of data points into disjoint clusters optimizing a specific criterion [2]. When the number of data points is large, a brute force enumeration of all possible combinations would be computationally expensive. Instead, heuristic methods are applied to find the optimal partitioning. The most popular criterion function used for partitional clustering is the sum of squared error function given by,

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} (x - m_i)^2$$

where k is the number of clusters, C_i is the ith cluster, x is a data point and m_i is the centroid of the ith cluster. A widely used squared-error based algorithm is the k-means clustering algorithm [2]. In this paper, the model is design using clustering algorithm similar to the k-means algorithm. The contribution is empirically defined as a data point belonging to a cluster as the impact that it has on the quality of the cluster. This metric is then used to obtain an optimal set of '*k*' cluster from the given set of data points. The notion of contribution has its origin in game theory [9]. A recent work by Garg [10] focuses on the merger of game theory and data clustering. Garg mapped cluster formation to coalition formation in cooperative games and used the solution concept of Shapely value to find the optimal number of clusters for a given set of data points. While his work uses the concept of contribution to find the optimal cluster number, we use it in a different manner for optimal partitioning of the data points into a fixed number of clusters. Given a cluster C_i with n points and centroid mi, the average intra-cluster dispersion is given by,

$$dispersion(C_i) = \frac{1}{n}\sum_{x \in C_i}(x - m_i)^2$$

The contribution of a point x , C_i is measured as

Contribution(x, C_i)=Dispersion(Ci-{x})-Dispersion(C_i)

Clearly, if the contribution of a data point is negative, it has an adverse impact on its cluster. On the other hand, a positive contribution indicates that the removal of the point from the cluster would degrade its quality. In our work, we propose a clustering algorithm that treats points with negative contribution different from those with positive contribution. In the case of a negative contribution point, the point is shifted to a cluster, where its contribution is the highest, possibly positive. On the other hand, for a positive contribution point, a multi-objective optimization criterion is taken, where we try to optimize both the intra-cluster and inter-cluster dispersion measures.

The proposed outline presents contribution-based clustering algorithm. It optimizes on two measures, namely the intra-cluster dispersion given by

$$a = \frac{1}{n}\sum_{x \in C_i}(x - m_i)^2$$

and the inter-cluster dispersion given by

$$\beta = \frac{1}{k}\sum_{i=1}^{k}(m_i - \overline{m})^2$$

where k is the number of clusters and $\overline{m}$ is the mean of all centroids. The algorithm tries to minimize α and maximize β. The three steps (passes) in the algorithm are described below.

Algorithm: Contribution based MIR

Input: Query Image

Output: Retrieved Similar Images

Start

1. Initialization
2. Randomly select k centroids ($m_1, m_2, ..., m_k$)
3. **For** each point x
4. Find $1 \leq l \leq k$ such that distance($_{x, ml}$) is minimum
5. Add x to cluster C_l and update centroid m_l.
6. **End For**
7. Negative Contribution Points
8. **For** each cluster C_l
9. **For** each point x C_l
10. **If** contribution$(x, C_l) < 0$
11. Move x to a cluster C_p such that contribution
12. (x, C_p) is maximum
13. Update centroid m_p
14. **End If**
15. **End For**
16. End For
17. Positive Contribution Points
18. **For** each cluster C_l
19. **For** each point x □ C_l
20. **If** contribution(x, Cl) ≥ 0
21. Move x to a cluster C_p such that $\frac{a - a_{new}}{a} + \frac{\beta_{new} - \beta}{\beta_{new}} +$ is maximum
22. Update centroid m_p
23. **End if**
24. **End for**
25. **End for**

Stop

Note that α_{new} and β_{new} are values of α and β after the point x is moved to cluster C_p.

V. Result and Performance Analysis

The proposed framework for performing contribution based medical image retrieval system is evaluated on 777 medical images belonging to 22 categories of medical conditions pertaining to human body (e.g. spine, palm, MRI, skull, ankle etc) [AR]. Each category contained varying number of images. All the images contained a textual description mentioning the salient foreground objects. The images were clustered using our algorithm with the initial centroids chosen at random. The cluster whose centroid was closest in distance to the given test image was determined and the images belonging to the cluster were retrieved. The results were then compared with images retrieved using the k-means clustering algorithm with the same set of initial centroids. Some of the retrieved images for sample test images are given in Table I. The following performance measures were used to evaluate the performance of the algorithm,

Precision=(Total relevant images Retrieved)/(Total number of retrieved images)

Recall=(Total Number of Retrieved relevant images)/(Total number of relevant images in database)

For the purpose of performance comparison, the contribution based proposed MIR system is also compared with BTC Scheme [AR] and K-Means Algorithm [AR].

Scheme-1: Block Truncation Coding

Color Moment and Block Truncation Coding (BTC) are used to extract features for image dataset when the implementation was performed using BTC scheme. 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.

Scheme-2: K-means using DWT

It basically consists of 3 steps: feature extraction, image segmentation, and deciding similar images. The image has been partitioned into blocks of 4×4 pixels and a feature vector for each block consisting of 6 elements has been extracted. The LUV colour space has been used where L stands for luminance, U for hue and V for saturation, U and V contains colour information (chrominance). The first three of them are average of the values of the Luminance, Hue and Saturation, respectively of the 16 pixels present in the 4×4 blocks. For the other three features Haar (wavelet) transform [AR] has been used to L component of image. After a one-level wavelet transform, a 4×4 block is decomposed into 4 frequency bands of 2×2 block. The other three components of each feature vector are square root of second order moment of wavelet coefficients of the HL, LH and HH band, respectively because each of these bands provide the variations present in different directions. The k-means algorithm has been used to cluster the feature vectors into several classes with every class corresponding to one region in the segmented image. Then the K-means algorithm will do the three steps below until convergence – Iterate until stable (= set of centroids from the previous iteration equals the present set of centroids):

a) Determine the centroid coordinate.

b) Determine the distance of each object to the centroids.

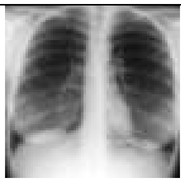
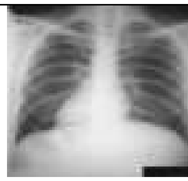
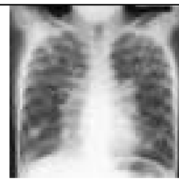
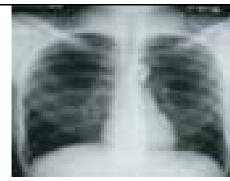
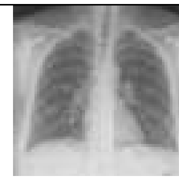
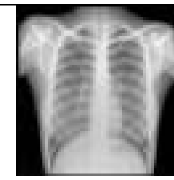
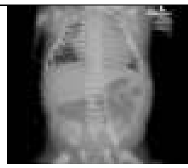
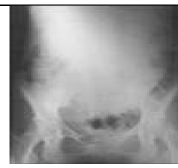
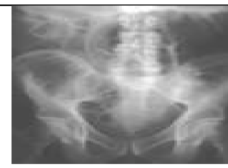
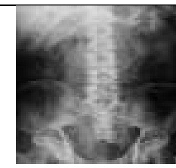
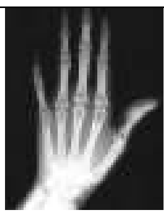
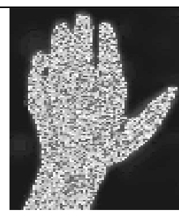
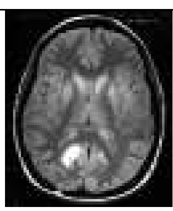
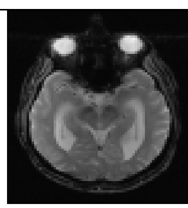
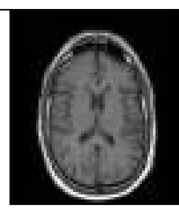
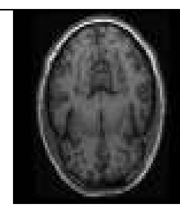
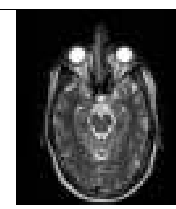
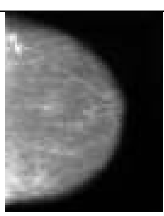
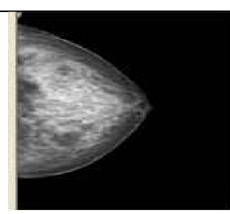
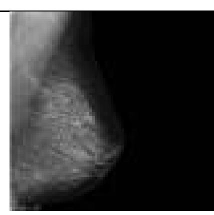
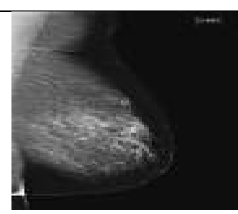
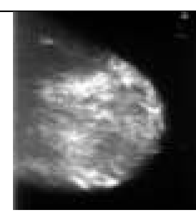
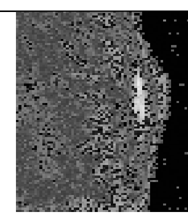
c) Group the object based on minimum distance.

An algorithm has been developed that determines the value of K (number of regions) required for the segmentation of the image, based on the variety of content present in the image.
1) First we start segmentation of feature vectors with K=2.
2) Then we find out the sum of the distances of each feature vector within a cluster from its centroid and then compare them to a experimentally determined critical value.
3) If there is more variety present, which means within a cluster the feature vectors differ considerably, then the value of this sum will be larger than the critical value.
4) If the value of the sum for each cluster present in the image is relatively higher than the critical value then we segment the feature vectors again by incrementing K by 1.
5) Steps 2 to 4 iterate until the sum becomes less than or equal to the critical value.

If two images have region sets $R_1 = \{r_1, r_2\}$ and $R_2 = \{r'_1, r'_2, r'_3\}$, first the significance of each region is calculated using the area percentage scheme. Significance of a region depends on the fraction of image occupied by the cluster. We then apply the Most Significant Highest priority [1] algorithm to get the priority of the significance of all the combinations of components from R_1 to R_2. Then we get the final distance between the images by calculating the weighted sum of the components of the distance matrix where the weights are the components of the matrix obtained from Most Significant Highest priority algorithm.

Thus, using these techniques we can match the query image with each image in the database and sort the images present in the database according to the distance in increasing order of distance and hence decreasing order of similarity

Table 1 Results Accomplished

TEST IMAGE	RESULTS OF RETRIEVED IMAGES				
					
					
					
					
					
					

The cumulative results accomplished are shown in Figure 1. After performing the comparative analysis of the proposed system with considered BTC scheme [AR] and K-Means Scheme [AR], it can be seen that proposed system has outperformed as shown in Table 5.1.

Table 5.1 Comparative Analysis of various techniques Adopted

COMPARATIVE ANALYSIS			
SL_No	Techniques	Recall(%)	Precision(%)
1	Contribution	92.86	94.05
2	BTC	92.86	92.82
3	K-means	91.63	91.15

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. The low recall in both cases is due to the fact that query output consists of images retrieved from a single cluster. Alternatively, the system could output a set of clusters ranked by relevance. This would improve the performance of the system and also, provide the user with a convenient interface for browsing through the retrieved images. It has to be noted that the motivation behind our work is not to improve the performance of existing CBIR systems, but instead to show that the proposed algorithm performs better at image clustering when compared to the popular k-means partitional clustering algorithm.

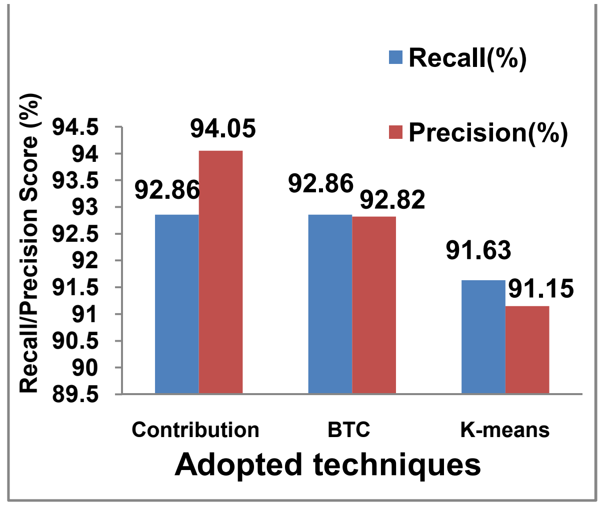


Figure 2. Results of performance Analysis of Proposed MIR System with BTC scheme [AR] and K-Means Scheme [AR]

VI. Conclusion

We have thus proposed a new partitional clustering algorithm based on the notion of 'contribution of a data point' Unlike the k-means algorithm, our algorithm optimized on both the intra-cluster and inter-cluster similarity measures and required fewer passes with each pass having the same time complexity as that of the k-means algorithm. We applied the clustering algorithm to content-based image retrieval and our experiments reveal that the algorithm improves on recall at the cost of precision. As with many other clustering algorithms, a limitation with our algorithm is that it requires the number of clusters to be known in prior. Various methods exist to determine the number of clusters for a given dataset [11] including the one based on game theory [10]. A problem with the k-means and k-medoids clustering algorithm is that they do not perform well when the clusters are non-spherical in shape. Whether the proposed algorithm suffers from such a limitation is yet to be investigated. Future lines of work would be to apply the concept of contribution to other clustering techniques such as hierarchical clustering. Our algorithm could also be extended to fuzzy partitioning of data points. The content-based image retrieval technique described in this paper uses only the RGB color histogram as the visual content descriptor of an image. The performance of the system with other visual features based on shape and texture and other distance metrics would have to be tested [8]. Also, learning through relevance feedback from the user could be incorporated in the proposed system [12].

REFERENCES

[1] J. Han and K. Micheline, “Data mining concepts and techniques,” Morgan Kauffman, 2006.

[2] R. Xu and D. Wunsch, “Survey of clustering algorithms,” IEEE Transactions on Neural Networks, Vol.16, Issue 3, pp. 645– 678, May

[3] 2005.

[4] E. Forgy, “Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications,” Biometrics, vol. 21, pp. 768–780, 1965.

[5] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in Proc. 5th Berkeley Symp., vol. 1, 1967, pp.281–297.

[6] L. Kaufman and P. Rousseeuw, “Finding groups in data: An introduction

[7] to cluster analysis,” Wiley, 1990.

[8] F. Höppner, F. Klawonn, and R. Kruse, “Fuzzy cluster analysis: Methods for classification, data analysis, and image recognition,” New York, Wiley, 1999.

[9] P. Bradley, U. Fayyad, and C. Reina, “Scaling clustering algorithms to large databases,” in Proc. 4th Int. Conf. Knowledge Discovery and Data Mining (KDD’98), 1998, pp. 9–15.

[10] F.H. Long, H.J. Zhang, and D.D. Feng, “Fundamentals of content-based image retrieval,” in D.D. Feng, W.C. Siu, and H.J. Zhang (Eds), ‘Multimedia information retrieval and management—technological fundamentals and applications’, Springer-Verlag, New York, 2003, pp. 1–26

[11] M.J. Osborne, “An introduction to game theory,” Oxford University Press, USA, 2007.

[12] V.K. Garg, “Pragmatic data mining: Novel paradigms for tackling key challenges,” Project Report, Computer Science & Automation (CSA), Indian Institute of Science, 2009.

[13] D. Pelleg and A. Moore, “X-means: Extending k-means with efficient estimation of the number of clusters,” in Proc. 17th Int. Conf. Machine Learning (ICML’00), 2000, pp. 727–734.

[14] Y. Liu, D. Zhang, G. Lu, and W. Ma, “A survey of content-based image retrieval with high-level semantics," Pattern Recogn., vol. 40(1), 2007, pp. 262-282.

[15] R. Datta, D. Joshi, J. Li, and J.Z. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” ACM Comput. Surv., 40(2), Apr. 2008, pp. 1-60.

[16] H. Zhang, J.E. Fritts, and S.A. Goldman, “Image segmentation evaluation: A survey of unsupervised methods,” Comput. Vis. Image Underst., 110(2), May. 2008, pp. 260-280.

[17] D. Cai, X. He, Z. Li, W. Ma, and J. Wen, "Hierarchical clustering of WWW image search results using visual, textual and link information," in Proc. 12th Annual ACM International Conference on Multimedia (MULTIMEDIA '04), Oct. 2004, New York, NY, pp. 952-959.

[18] D. Kinoshenko, V. Mashtalir and E. Yegorova, "Clustering method for fast content-based image retrieval," Computer Vision and Graphics, 32, Mar. 2006, pp. 946-952.

[19] Y. Chen, J.Z. Wang, and R. Krovetz, “Content-based image retrieval by clustering,” in Proc. 5th ACM SIGMM international Workshop on Multimedia information Retrieval, Berkeley, California, Nov. 2003, MIR '03, ACM, New York, NY, pp. 193-200.

[20] P.J. Dutta, D.K. Bhattacharyya, J.K. Kalita and M. Dutta, "Clustering approach to content based image retrieval," in Proc. Conference on Geometric Modeling and Imaging: New Trends (GMAI), 2006, IEEE Computer Society, Washington, DC, pp. 183-188.

[21] G. Liu and B. Lee, "A color-based clustering approach for web image search results," in Proc. 2009 International Conference on Hybrid information Technology (ICHIT '09), Daejeon, Korea, Aug. 2009, vol. 321, ACM, New York, NY, pp. 481-484.

[22] K. Jarrah, S. Krishnan, L. Guan , "Automatic content-based image retrieval using hierarchical clustering algorithms," in Proc. International Joint Conference on Neural Networks (IJCNN '06), Oct. 2006, Vancouver, BC pp. 3532 - 3537.

[23] “Object and concept recognition for content-based image retrieval,”[Online].Available:http://www.cs.washington.edu/research/ imagedatabase/groundtruth/.

Dr.B.P Mallikarjunaswamy: working as a professor in the Department of Computer Science & Engineering, Sri Siddhartha Institute of Technology, affiliated to Sri Siddhartha University. He has more than 20 years of Experience in teaching and 5 years of R & D. He is guiding many Ph.D scholars. He has published more than 30 technical papers in national and International Journals and conferences. His current research interests are in Pattern Recognition and Image Processing.

Mr.Shivamurthy R C received the BE degree from college of Engineering, Gulbarga University and received the M.Tech degree in Computer Science & Engineering from Malnad College of Engineering, Visvesvaraya Technological University, Belgaum. He served as a Bio Medical Engineer in AIMS Hospital & Cancer Research Center. He served as Assistant Professor and currently working as professor in the department of Computer Science at A.I.T, Tumkur, Karnataka and is also a Ph.D scholar in CMJ University, India.

Patterns Antennas Arrays Synthesis Based on Adaptive Particle Swarm Optimization and Genetic Algorithms

B. Kadri[1], M. Brahimi[1], I.K. Bousserhane[1], M. Bousahla[2], F.T. Bendimerad[2]

[1] Bechar University, Faculty of Sciences and Technologies, Departement of Electronic
P.O.Box 417, 08000, Bechar, Algeria

[2]Abou-Bakr Belkaid University, Engineering Sciences Faculty, Telecommunications Laboratory
P.O.Box 230, Tlemcen, Algeria

Abstract

In recent years, evolutionary optimization (EO) techniques have attracted considerable attention in the design of electromagnetic systems of increasing complexity. This paper presents a comparison between two optimization algorithms for the synthesis of uniform linear and planar antennas arrays, the first one is an adaptive particle swarm optimization (APSO) where the inertia weight and acceleration coefficient are adjusted dynamically according to feedback taken from particle's best memories to overcome the limitations of the standard PSO which are: premature convergence, low searching accuracy and iterative inefficiency. The second method is the genetic algorithms (GA) inspired from the processes of the evolution of the species and the natural genetics.

The results show that the design of uniform linear and planar antennas arrays using APSO method provides a low side lobe level and achieve faster convergence speed to the optimum solution than those obtained by a GA.

Keywords: *antennas arrays, planar arrays, synthesis, optimization methods; adaptive particle swarm algorithm, genetic algorithm.*

1. Introduction

Planar antenna arrays have been widely studied due to their importance in communications industry such us mobile, wireless communication, and other domains [1], in order to seek for an optimal planar antenna arrays feed laws so that the array complies with the requirements of the user and according to precise specifications, such us lower side lobes of planar antenna array pattern, controllable beamwidth, and the pattern symmetry in azimuth angles. The traditional optimization methods cannot bear the demand of such complex optimization problem. Particles Swarm Optimization (PSO) [2] is an evolutionary algorithm based on the swarm intelligence. Eberhart and Kennedy first introduced such algorithms in 1995. The original conception comes from the research of food hunting by birds. PSO algorithm can be used to solve the complex global optimization problems. Currently, the algorithm and its variations are applied to solve many practical problems. For the optimization of the antenna array, the parameters affecting antenna pattern are chosen as the design variables [3]. A desired pattern is presented according to the radiate requirement.

The simulation result show that the calculated pattern approaches the desired pattern and the SLL is very low. This kind of optimization improves the efficiency of antennas array.

2. Standard Particle Swarm Optimization

Recently, the PSO technique has been successfully applied to the design of antennas and microwave components [4-5]. The results proved that this method is powerful and effective for optimization problems. PSO is similar in some ways to Genetic Algorithms (GA) and other evolutionary algorithms, but requires less computational bookkeeping and generally fewer lines of code, including the fact that the basic algorithm is very easy to understand and implement. In the PSO mechanism, each potential solution of optimization problem is a bird in the solution space, which called "particle". Each particle has a value of fitness determined by objective functions. They also have a directional velocity to control its move tracks. The particles chase the optimal solution by searching the solution space. All particles have initial positions and velocities [6], where the positions and velocities are iterated. In each iteration, two "best position" are chased to update the particle. The first is the optimal solution found by particle, which called personal best position. The other is the optimal solution in the entire group, which called global best position. In PSO, the i -th particle in the solution space is determined by a fitness function's value. The fitness function is the optimal target, the position of ith particle can be presented by $x_i = (x_{i1}, x_{i2}, ..., x_{id})$, $v_i = (v_{i1}, v_{i2}, ..., v_{id})$ stand for the velocity of the ith particle, the optimal solution come into being through iterative searching, the positions and velocities of particles update by personal and global best positions in each

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

iteration. Let $p_i = (p_{i1}, p_{i2},..., p_{id})$ be the position vector for an individual particle's best fitness, which is personal best position, and $g_i = (g_{i1}, g_{i2},..., g_{id})$ be the global best position among all the agents. The positions and velocities of particles are updated according to the following equations (1) and (2) [7]:

$$v_{id} = \omega \times v_{id} + c_1 \times r_1 \times (p_{id} - x_{id}) + c_2 \times r_2 \times (g_{id} - x_{id}) \quad (1)$$

$$x_{id} = x_{id} + v_{id} \quad (2)$$

Where $\omega = 0.7$ is the inertia weight, c_1 and c_2 are the acceleration coefficients set to 1.7, r_1 and r_2 are random numbers in the range [0,1], The first part of (1) is the initial velocities of particles, the second part is "cognition", which expresses the cogitation of particles; the third part is "social", which expresses the registration of message and cooperation among particles.

The steps involved in standard PSO are shown by the flowchart drawn in figure 1.

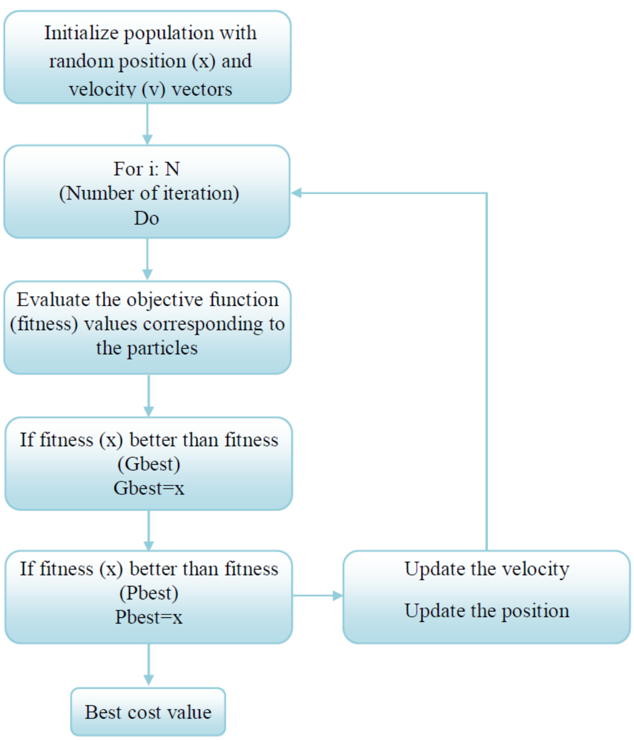


Fig. 1.Flowchart of PSO algorithm

3. Adaptive Particle Swarm Optimization

In this paper, the inertia weight and the acceleration coefficient are neither set to a constant value nor set as a linearly decreasing time varying function [8]. Instead they are defined as a function of local best (pbest) and global best (gbest) values of the fitness function of a minimization problem as given in Eqs. (3) and (4). The average of all the personal best values in that particular generation is termed as ((pbesti)average).

$$\text{Inertia weight,} \quad \omega_i = \left(1.1 - \frac{gbest}{(pbest_i)_{average}}\right) \quad (3)$$

Acceleration coefficient;

$$ac = ac_1 = ac_2 = \left(1 + \frac{gbest}{(pbest_i)}\right) \quad (4)$$

The inertia weight in (3) is termed global-average local best IW (GLbestIW) and the acceleration coefficient in (4) is called global-local best AC (GLbestAC).

4. Genetic Algorithm

By analogy with natural selection and evolution, in classical GA the set of parameters to be optimized (genes) defines an individual or potential solution X (chromosome) and a set of individuals makes up the population, which is evolved by means of the selection, crossover, and mutation genetic operators. The optimization process used by the GA follows the next steps [9].
The genetic algorithm generates individuals (amplitude excitations and phase perturbations of the antenna elements). The individuals are encoded in a vector of real numbers, that represents the amplitudes, and a vector of real numbers restrained on the range (0, 2π), that represents the phase perturbations of the antenna elements. Each individual generates an array factor of certain characteristics of the side lobe level and the directivity. Then, the genetic mechanisms of crossover, survival and mutation are used to obtain better and better solutions. The genetic algorithm evolves the individuals to a global solution that generates an array factor with minimum side lobe level and maximum directivity in the steering direction [10-11].
The steps involved in GA are shown by the flowchart drawn in figure 2.

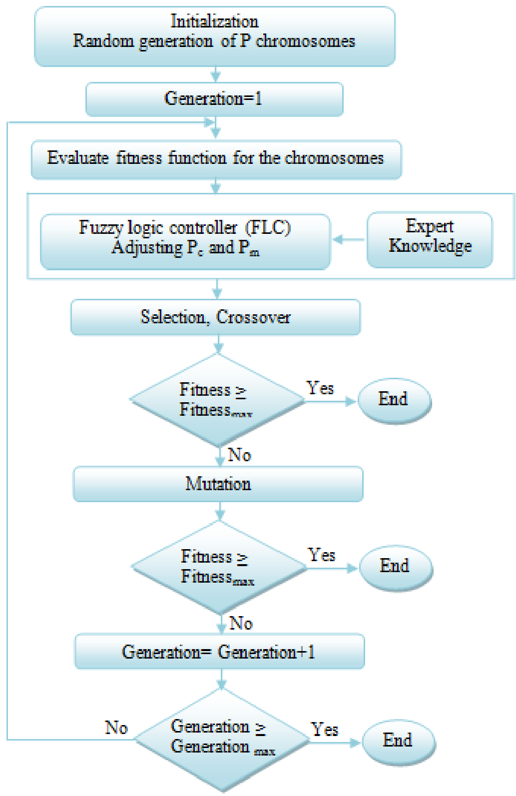


Fig. 2 Flowchart of GA algorithm

5. Linear Antenna Arrays Synthesis

In this section, the APSO and GA algorithms were implemented for the synthesis of uniformly spaced linear array constituted with 16 rectangular microstrip antennas (figure 3). Two examples of linear antenna array synthesis have been considered, the first one by optimizing only excitation weights for a desired radiation pattern specified by a symmetrical narrow beam pattern with a beam width of 8 degrees and maximum side lobe levels of -20dB. The second example for the same desired radiation pattern but pointed at 10°, the synthesis was carried out by optimizing both amplitude and phase weights. In our simulation, we have used a population size of 40 for GA.

For APSO, it set with adapting inertial weight and acceleration coefficients which is proposed by Ratnaweera and Halgamuge [12] and a population size equal to 30 individuals.

In figure 4 we present the result of the first example of linear antenna array synthesis by the optimization of amplitude excitation coefficients using both APSO and GA. It is clearly seen that the radiation pattern obtained by APSO meet better the desired pattern than the obtained by the GA. The side lobe level obtained by APSO optimization (-40dB) are much better than in the case of GA (-23dB).

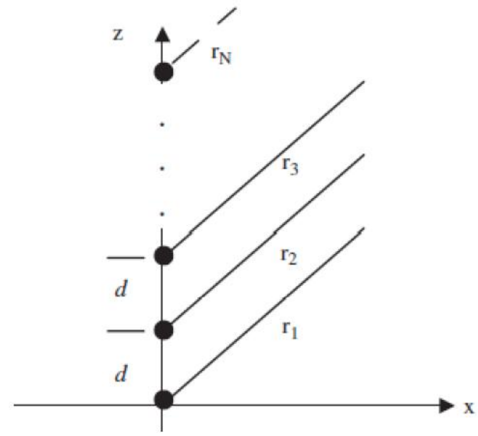


Fig. 3 Linear antennas array.

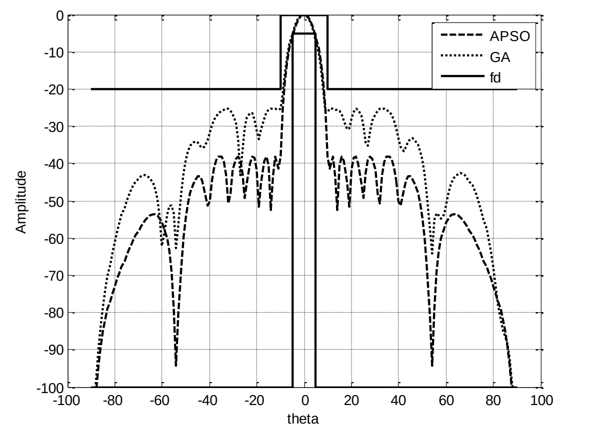


Fig. 4 Result of a linear array synthesis with 16 elements applying both APSO and GA.

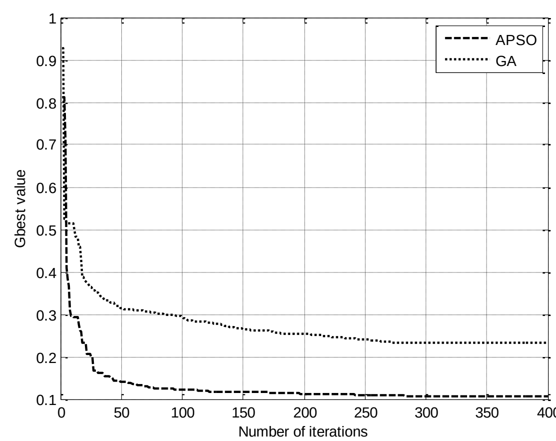


Fig. 5 Fitness evolution of APSO and GA algorithms

From figure 5, the speed approaching the global optimal of APSO is much quickly than that of GA, and the fitness values of the best individuals of APSO are almost higher than that of GA in every population.
In the second example, the synthesis result of a linear array with 16 uniformly spaced antennas for a desired radiation pattern, similar to the previous one but pointed at 10 degrees are shown in figures 6 and 7.

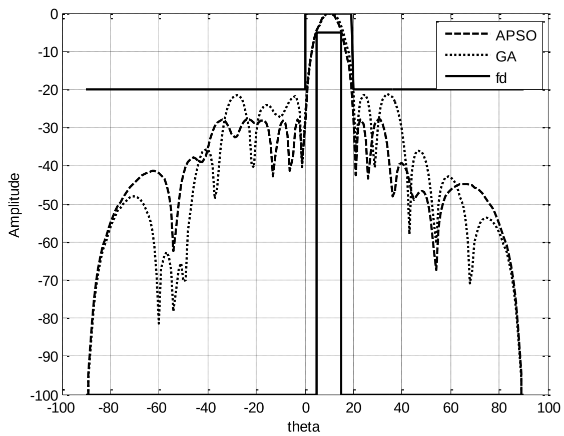


Fig. 6 Result of a linear array synthesis with 16 elements applying both APSO and GA.

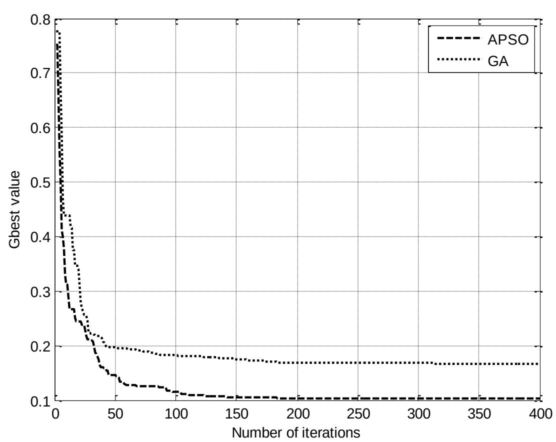


Fig. 7 Fitness evolution of APSO and GA algorithms

6. Planar Antenna Arrays Synthesis

A microstrip antenna have limited radiation diagram however, when we have an aggregate the performance of radiation diagram will be remarkable [13]. Let us consider a planar antenna array constituted of MxN equally spaced rectangular antenna arranged in a regular rectangular array in the x-y plane, with an inter-element spacing of $d = dx = dy = \lambda/2$ (figure 8), and whose outputs are added together to provided a single output. Mathematically, the normalized array far-field pattern is given by:

$$F_s(\theta,\varphi) = \frac{f(\theta,\varphi)}{F_{s\max}} \sum_{n=1}^{M} I_{mn} e^{\left(j(m-1)k_0 \sin\theta\cos\varphi dx + j\psi_{mn}\right)} \cdot \sum_{n=1}^{N} e^{\left(j(n-1)k_0 \sin\theta\sin\varphi dy + j\psi_{mn}\right)} \quad (5)$$

Where
$f(\theta,\varphi)$: Represents the radiation pattern of an element.
I_{mn} : Amplitude coefficient at element *(m, n).*
ψ_{mn} : Phase coefficient at element *(m, n).*
k_0 : Wave number.

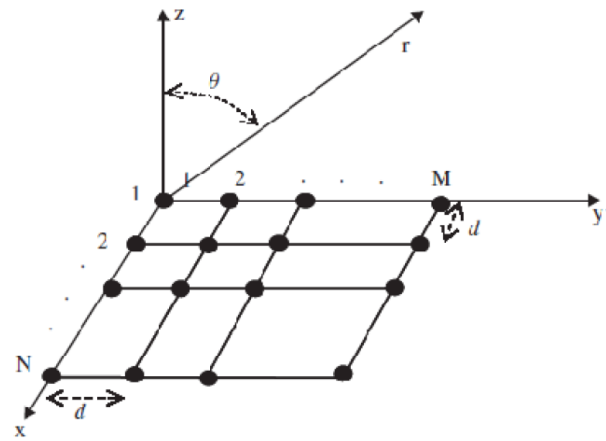


Fig. 8 Planar antennas array.

We use the APSO algorithm to find the appropriate excitation coefficients (amplitude and phase), which shall satisfy the desired radiation pattern.
We have chosen a suitable fitness functions that can guide the APSO optimization toward a solution that meets the desired radiation pattern. The fitness function to be minimized is selected from the work of Chuan Lin [14] which is described by the equation below

$$f(\bar{\rho}) = \mathrm{Max}_{\theta\in S} \left| \frac{A_F^{\bar{\rho}}(\theta)}{A_F^{\bar{\rho}}(\theta_0)} \right| \quad (6)$$

Where S is the space spanned by the angle θ excluding the mainlobe and ρ represents the unknown parameter vector, such as element positions and phases. This objective function minimizes all the sidelobe levels and maximizes the power in the main lobe located at $\theta=\theta_0$.
We implemented the two algorithms APSO and GA for the synthesis of uniformly spaced planar array of 16

rectangular patch antennas. Figures 9 and 11 represent respectively the synthesis result of our array constituted of 16 elements. It is a question respectively of the amplitude and phase optimization and the amplitude and phases pointed at 10 degree in order to as well as possible approach the radiation pattern resulting from a desired template specified by a symmetrical narrow beam pattern with a beam width of 8 degrees and maximum side lobe levels of -20dB. During the simulation we have used a population size of 40 for FGAs. Roulette strategy for "selection" one–point crossover and mutation to flip bits, the value of crossover and mutation probabilities (p_c and p_m) are determined according to FLC.

The figures represent the results of plane array synthesis consisted of 16 aerial elements.

It is noticed that the radiation pattern are contained within the limits imposed by the template and the maximum of side lobes level is lower than -20 dB in such way that the APSO is better than GA and reaches them respectively -35dB and -25 dB (figure 7), -30dB and -22dB (figure 9)

With each diagram, on associates the evolution of the quadratic error during the generations (figure 10 and 12). From this figures the best fitness obtained by the APSO is better than the obtained by the GA.

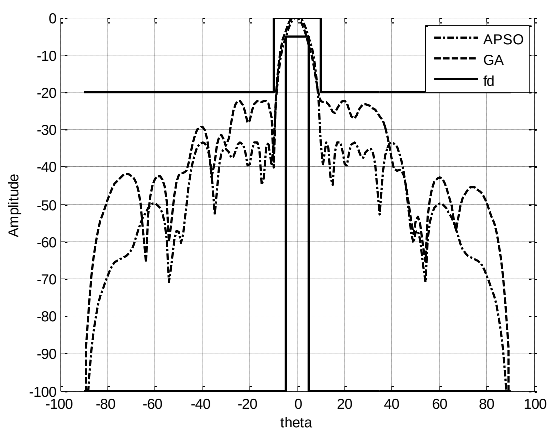


Fig. 9 Result of a linear array synthesis with 16 elements applying both APSO and GA (only amplitude synthesis).

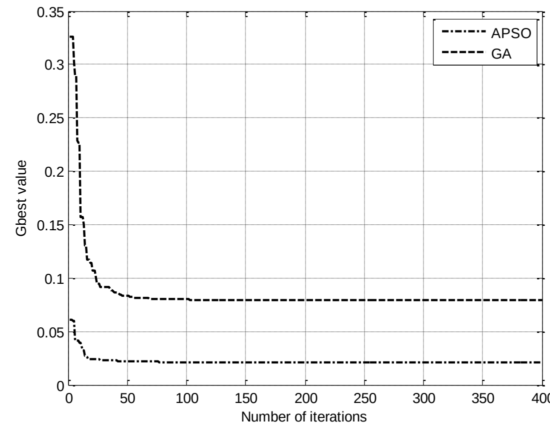


Fig. 10 Fitness evolution of APSO and GA algorithms

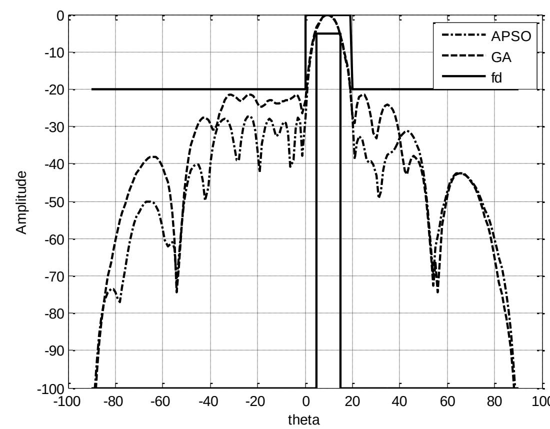


Fig. 11 Result of a linear array synthesis with 16 elements applying both APSO and GA (amplitude and phase synthesis).

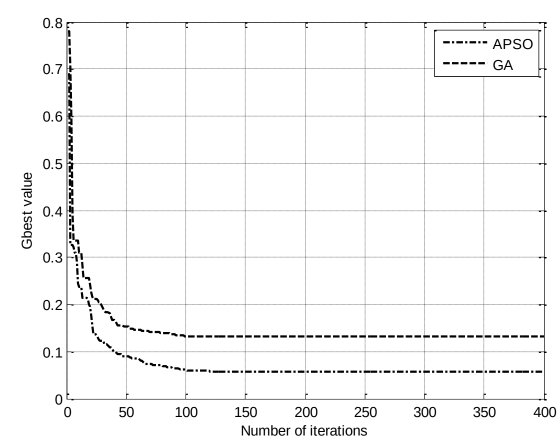


Fig. 12 Fitness evolution of APSO and GA algorithms

7. Conclusion

The optimization techniques seemed APSO and GA for the goal to obtain the global minimum and to avoid remaining to trap in a local minimum like in the case of the deterministic methods. However they present a major disadvantage which lies in their calculative cost and which believes according to the dimension of the problem considered and its difficulty.

The advantage of PSO on GA of is marked as much than the optimization variables number is important. Indeed for a synthesis of antennas array, GA requires an enormous computing time, because this one needs a great iteration number to converge towards an optimal solution.

Included examples on linear and planar antennas array synthesis demonstrate that PSO with adaptive scheme shows better performance than GA because of its simplicity in implementation and minor computing time.

References

[1] W.Y. Wei, D.M. Gong, and B.S. Chen, “Antenna Theory”, Xi’an: The publishing house of Xidian University, 1985.

[2] J. Robinson, Y. Rahmat-Samii, “Particle Swarm Optimization (PSO) in Electromagnetics”, IEEE Trans. Antennas Propagation, vol.52, no.2, February 2004, pp. 397-407.

[3] Cheng D K, “Optimization techniques for antenna arrays”, Proceedings for IEEE, vol.59 (12), 1988, pp. 1664-1674.

[4] Khodier, M.|Christodoulou,. “Linear array geometry synthesis with minimum sidelobe level and null control using particle swarm optimization”, IEEE Transactions on Antennas and Propagation 53 No. 8 (August 2005), 2674-2679.

[5] Ababneh, J.|Khodier, M.|Dib, N, “Synthesis of inter- digital capacitors based on particle swarm optimization and artificial neural networks”, International Journal of RF and Mi- crowave Computer-Aided Engineering 16 (July 2006), 322{330.

[6] J. Kennedy, R. Eberhat, “Particle Swarm Optimization”, IEEE Int. Conf. Neural Networks (Perth, Australia), vol. IV, 1995, pp.1942-1948.

[7] T.B. Chen, Y.B. Chen, Y.C. Jiao, and F.S. Zhang “Synthesis of Antenna Array Using Particle Swarm Optimization” National Laboratory of Antennas and Microwave Technology, Xidian University, Xi’an 710071, P. R. China 0-7803-9433 2005

[8] R. Eberhart, Y. Shi, “Comparison between genetic algorithms and particle swarm optimization,” Lect. Notes Comput. Sci. 611–618, 1998.

[9] R. L. Haupt,. “Ail Introduction to Genetic Algorithms for Electromagnetics”, IEEE Antenna and propagation Magazine. Vol. 37. pp. 7-15. 1995.

[10] S. A. Mitilineos, C A, Papagianni. G.. I. Verikaki. C. N. Capsalis, “Design of Switched Beam Planar Arrays Using the Method of Genetic Algorithms”, Progress In Electromagnetics Research. PIER 46. 105-126. 2004.

[11] D. Marcano, F. Duran, “Synthesis of Antenna Arrays Using Genetic Algorithms”. IEEE Antenna and propagation Magazine. Vol. 42. NO. 3. June 2000.

[12] A. Ratnaweera, S.K. Halgamuge, Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficient, IEEE Trans. Evol. Comput. vol. 8 (June (3)) (2004) 240–255.

[13] Boufeldja Kadri, Miloud Boussalila, Fetlii Tarik Beudimerad, “Phase-Only Planar Antenna Array Synthesis with Fuzzy Genetic Algorithms” IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 2, January 2010.

[14] Chuan Lin, Anyong Qing and Quanyuan Feng1, “Synthesis of unequally Spaced Antenna Arrays by a new Differential Evolutionary Algorithm”, International Journal of Communication Networks and Information Security (IJCNIS), Vol. 1, No. 1. April 2009

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

A Binary Quantum-behaved Particle Swarm Optimization Algorithm with Cooperative Approach

Jing Zhao[1,2], Jun Sun[2] and Wenbo Xu[3]

[1] School of Digital Media, Jiangnan University
Wuxi, 214122, Jiangsu, China

[2] School of Information, Shandong Polytechnic University,
Jinan, 250353, Shandong, China

[3] School of Internet of Things Engineering, Jiangnan University
Wuxi, 214122, Jiangsu, China

Abstract

A novel binary Quantum-behaved Particle Swarm Optimization algorithm with cooperative approach (CBQPSO) is introduced. In the proposed algorithm, the updating method of particle's previous best position and swarm's global best position are performed in each dimension of solution vector to avoid loss some components that have moved closer to the global optimal solution in the vector. Five test functions are used to test the performance of CBQPSO. The results of experiments show that the proposed technique can increase diversity of population and converge more rapidly than other binary algorithms.

Keywords: *Quantum-behaved Particle Swarm Optimization, Binary, Cooperative Approach, Test Functions.*

1. Introduction

Particle Swarm Optimization (PSO) is an evolutionary computation technique developed by Dr. Eberhart and Dr. Kennedy in 1995[1], inspired by social behavior of bird flocking or fish schooling. The optimal solution is obtained by exchanging information between individuals. However, the algorithm cannot converges to the global minimum point with probability one under suitable condition[2]. Jun Sun *et al* have proposed a global convergence-guaranteed PSO algorithm[3], Quantum-behaved Particle Swarm Optimization (QPSO) algorithm, which is inspired by quantum mechanics. It has been shown that QPSO outperforms PSO on several aspects, such as simple evolution equations, more few control parameters, fast convergence speed, simple operation and so on[4,5].

In 1997, Kennedy proposed the binary version of PSO (BPSO)[6], and Jun Sun *et al* proposed the binary version of QPSO (BQPSO) in 2007[7]. This paper will focus on developing the binary version of QPSO with cooperative method (CBQPSO). In the proposed algorithm, each dimension of particle's new solution vector replaces in turn the corresponding dimension of particle's previous best position and swarm's global best position to calculate the fitness value.

The rest structure of this paper is as follows. In section 2, a brief introduction of the BPSO is presented. The BQPSO is described in section 3. Next, the novel CBQPSO is depicted in section 4. Then the experiment results are given in section 5. Finally, the conclusion is put forward in section 6.

2. Binary Particle Swarm Optimization

In PSO, the population with M individuals, which is treated as a particle, is called a swarm X in the D-dimensional space. The position vector and velocity vector of particle i at the generation t represented as $x_i(t) = (x_{i1}(t), x_{i2}(t), \cdots, x_{iD}(t))$ and $v_i(t) = (v_{i1}(t), v_{i2}(t), \cdots, v_{iD}(t))$ .The particle moves according to the equations:

$$v_{id}(t+1) = wv_{id}(t) + c_1 r_1 (pbest_{id} - x_{id}(t)) + c_2 r_2 (gbest_d - x_{id}(t)) \quad (1)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (2)$$

Where $i = 1,2,\cdots,M; d = 1,2,\cdots,D$, w is the inertia weight, whose value is typically setup to vary linearly from 0.9 to 0.4. c_1 and c_2 are called the acceleration coefficients which usually are set as $c_1 = c_2$. r_1 and r_2 are random number uniformly distributed in $(0,1)$ .Vector $pbest_i = (pbest_{i1}, pbest_{2i}, \cdots, pbest_{iD})$ is the best previous position of particle i with the name personal best position($pbest$), while the global best position($gbest$), $gbest = (gbest_1, gbest_2, \cdots, gbest_D)$, is the best particle position among all the particles in the population.

In BPSO[6,8], Eq. (3) replaces Eq. (2).

if $(rand() < S(V_{id}))$ then $X_{id} = 1$ else $X_{id} = 0$ (3)

Where $S(v)$ is a sigmoid limiting transformation function($s(v) = \frac{1}{(1+e^{-v})}$), and $rand()$ is a random number selected from a uniform distribution in $(0,1)$.

3. Binary Quantum-behaved Particle Swarm Optimization

3.1 Quantum-behaved Particle Swarm Optimization

In PSO algorithm, the state of particle is depicted by its position vector and velocity vector, which determine the trajectory of the particle. The particle moves along a determined trajectory in Newtonian mechanics, but this is not the case in quantum mechanics. In quantum world, the term *trajectory* is meaningless, because position and velocity of a particle cannot be determined simultaneously according to *uncertainty principle*. Therefore, if individual particles in a PSO system have quantum behavior, the PSO algorithm is bound to work in a different fashion.

In quantum time-space framework, Jun Sun *et al.* introduce QPSO algorithm. The equations are as follows:

$$\begin{aligned} mbest &= \frac{1}{M}\sum_{i=1}^{M} pbest_i \\ &= (\frac{1}{M}\sum_{i=1}^{M} pbest_{i1}, \frac{1}{M}\sum_{i=1}^{M} pbest_{i2}, \cdots, \frac{1}{M}\sum_{i=1}^{M} pbest_{iD}) \end{aligned} \tag{4}$$

$$p_{id} = \varphi \times pbest_{id} + (1-\varphi) \times gbest_d \tag{5}$$

$$x_{id}(t+1) = p_{id} \pm \beta |mbest_d - x_{id}(t)| * \ln(\frac{1}{u}) \tag{6}$$

where φ is a random number uniformly distributed in $(0,1)$. $mbest$ is mean best position of the population. Parameter β is called the Contraction-Expansion coefficient, which can be tuned to control the convergence speed of the algorithm. From the results of stochastic simulation[9], it can be concluded that in QPSO, when $\beta < 1.782$, the particles will converge. In the process of iteration, $\pm$ is decided by the random number, when it is bigger than 0.5, minus sign (－) is proposed, others plus sign (＋) is proposed.

3.2 Binary Quantum-behaved Particle Swarm Optimization

In this section, a discrete binary version of QPSO (BQPSO) is proposed. Because the iteration equations of QPSO are far different from those of PSO, the methodology of BPSO does not apply to QPSO. In QPSO, there are no velocities and trajectories concepts but position and distance. In BQPSO, the position of the particle is represented as a binary string. The distance is defined as the Hamming distance between two binary strings. That is

$$|X - Y| = d_H(X,Y) \tag{7}$$

Where X and Y are two binary strings and represent two positions. The function $d_H()$ is to get the Hamming distance between X and Y. The Hamming distance is the count of bits different in the two strings.

The *j*th bit of the $mbest$ is determined by the states of the *j*th bits of all particles' $pbest$ in BQPSO. If more particles take on 1 at the *j*th bit of their own $pbest$, the *j*th bits of $mbest$ will be 1; otherwise the bit will be 0. However, if half of the particles take on 1 at the *j*th bit of their $pbest$, the *j*th bit of $mbest$ will be set randomly to be 1 or 0, with probability 0.5 for either state.

In BQPSO, the point p_i is obtained by crossover operation on $pbest_i$ and $gbest$. Firstly make one-point or multi-point crossover operation on $pbest_i$ and $gbest$ to generate two offspring. Then randomly select one of the offspring and output it as the point p_i.

Consider iterative Eq. (6) and transform it as

$$b = d_H(x_i, p_i) = \beta \times d_H(x_i, mbest) \times \ln(\frac{1}{u}) \tag{8}$$

We can obtain the new string x_i by the transformation in which each bit in p_i is mutated with the probability computed by

$$c_d = \begin{cases} \frac{b}{l} \\ 1 \quad if \ \frac{b}{l} > 1 \end{cases} \tag{9}$$

Where l is the length of the *d*th dimension of particle i. In the process of iteration, if $rand() < c_d$, the corresponding bit in the position of particle i will be reversed, otherwise remains it.

With the above definition and modifications of iterative equations, the BQPSO algorithm is described as the following procedure:

Step 1 Initialize an array of binary bits for all particles, particle's personal best positions *pbest* and swarm's global best position *gbest* ;

Step 2 For each particle, determine the *mbest* and get a stochastic position p_i by exerting crossover operation on $pbest_i$ and *gbest* ;

Step 3 For each dimension, compute the mutation probability c_d and then update the particle's new position x_i by c_d ;

Step 4 Evaluate the objective function value of the particle, and compare it with the objective function value of *pbest* and *gbest* . If the current objective function value is better than that of *pbest* and *gbest* , then update *pbest* and *gbest* ;

Step 5 Repeat step 2~4 until the stopping criterion is satisfied or reaches the given maximal iteration.

4. Binary Quantum-behaved Particle Swarm Optimization with Cooperative Approach

As BPSO and BQPSO described, each particle represents a complete solution vector for the objective function $f(X) = f[(X_1, X_2, \cdots, X_N)]$. Each update step is also performed on a full *D*-dimensional vector. Then it may be appear the possibility that some dimension in the solution vector have moved closer to the global optimum, while others moved away from the global optimum. Whereas the objective function value of the solution vector is worse than the former value. BPSO and BQPSO take the new solution vector for a complete vector and neglect the deteriorated components during the iterations. As long as the current objective function value is better than the former value, then update *pbest* and *gbest* . Therefore, the current solution vector can be give up in next iteration and the valuable information of the solution vector is lost unknowingly. In order to make full use of the beneficial information, the cooperative method[10,11] is introduced to BQPSO. In the proposed method, we expect that the operation can avoid the undesirable behavior, which is a case of taking two steps forward (some dimension improved), and one step back (some dimension deteriorated).

4.1 Cooperative Approach

We expect that once for every time a component in the vector has been updated, resulting in much quicker feedback. Thus, a cooperative method for doing just this is presented. In the new method each dimension of the new solution vector replaces in turn the corresponding dimension of *pbest* and *gbest* , and then compare the new objective function value to decide whether to update *pbest* and *gbest* .

The process is as follows:

Step 1 For each particle i , initialize $cgbest = gbest$, $cpbest_i = pbest_i$;

Step 2 For each dimension of particle i , replace the dimension of *cpbest* and *cgbest* by the corresponding dimension of the particle;

Step 3 Evaluate the new objective function value of *cpbest* and *cgbest* , and compare them with the objective function value of *pbest* and *gbest* . If the current objective function value is better than that of *pbest* and *gbest* , then update *pbest* and *gbest* ;

Step 4 Repeat step 2~3 until all the dimension of the particle is compared.

4.2 CBQPSO

With above modifications, the iteration process of CBQPSO is described step-by-step below.

Step 1 Initialize an array of binary bits for all particles, particle's personal best positions *pbest* and swarm's global best position *gbest* ;

Step 2 Update the particle's new position x_i by BQPSO;

Step 3 Evaluate the objective function value of the particle, and compare them with the objective function value of *pbest* and *gbest* . If the current objective function value is better than that of *pbest* and *gbest* , then update *pbest* and *gbest* ;

Step 4 Use cooperative strategy to update *pbest* and *gbest* ;

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Step 5 Repeat step 2~4 until the stopping criterion is satisfied or reaches the given maximal iteration.

The proposed algorithm tries to improve convergence precision by comparing each dimension of solution vector. It must extend the search space and then increase the time consumption. Two adaptive control methods are proposed. Firstly, the cooperative strategy is adopted in a certain interval. In our method, it set to 5. Then the cooperative strategy is performed when the bit of the particle is different from the corresponding bit of $pbest$ and $gbest$.

5. Experiments

In this section, the performance of CBQPSO algorithm is tested on the following five different standard functions[7] to be maximized. Then the results are compared with BPSO and BQPSO.

$$f_1(X) = 78.6 - \sum_{i=1}^{3} x_i^2 \qquad (-5.12 \le x_i \le 5.12)$$

$$f_2(X) = 3905.93 - (100(x_1^2 - x_2)^2 - (1 - x_1)^2)$$
$$(-2.048 \le x_i \le 2.048)$$

$$f_3(X) = 25 - (x_1 + x_2 + x_3 + x_4 + x_5)$$
$$x_i \in Z, \quad (-5.12 \le x_i \le 5.12)$$

$$f_4(X) = 1248.2 - \sum_{i=1}^{30} x_i^4 \quad (-1.28 \le x_i \le 1.28)$$

$$f_5(X) = 500 - 1\Big/\left(0.002 + \sum_{j=1}^{25} \frac{1}{j + 1 + \sum_{i=1}^{2} (x_i - a_{ij})^6}\right)$$

$$a = \begin{pmatrix} 32.0 & 16.0 & 0 & 16.0 & 32.0 \\ -32.0 & -16.0 & 0 & 16.0 & 32.0 \end{pmatrix}$$

$$(-65.536 \le x_i \le 65.536)$$

In the numerical experiments, the algorithms parameters settings are described as follow: for BPSO, the acceleration coefficients are set to $c_1 = c_2 = 2$ and the inertia weight w is decreasing linearly from 0.9 to 0.4. In experiments for BQPSO and CBQPSO, the value of β is 1.4[12] . All experiments are run 50 independent times respectively with a population of 20, 40 and 80 particles on an Intel(R) Xeon(R) E5504 @2.00GHz 2.00GHz, 1GB RAM computer with the software environment of MATLAB2009a. All the algorithms terminate when the number of iterations succeeds 200.

The best fitness value (BFV), maximum value and minimum value are recorded after the algorithm terminates at each run. The performance of all the algorithms is evaluated by average BFV (Avg. BFV) and Standard Deviation (St. Dev.). All the measurements are listed on Table 1. Fig.1 illustrates the convergence process of average BFV of three algorithms over 50 runs with 40 particles on five test functions.

The optima of function $f1$, whose fitness value is 78.6, can be find out by BPSO, BQPSO and CBQPSO. As can be seen from Table 1, the average BFV and St. Dev. of CBQPSO is best. And BQPSO outperforms BPSO. As of solution quality, CBQPSO and BQPSO with 20 particles make 12 successful searches out of 50 trial runs, whereas BPSO find out the optima for 7 times. And the corresponding times is 14, 13 and 2 respectively with 40 particles. When the population number is 80, the optima are found out for 29, 20 and 4 times corresponding with CBQPSO, BQPSO and BPSO.

On the function $f2$, all the algorithms can be found the optimum fitness value 3905.93. However CBQPSO generates best average BFV and St. Dev.. And BQPSO takes second place. As can be seen from Table 1, BQPSO has the worst performance than other two algorithms with 40 particles. Note that the St. Dev. of BQPSO with 40 particles is better than that of BPSO.

The third function $f3$ is a simple integer function with an optimum of 55. CBQPSO, BQPSO and BPSO with 80 particles hit the optima for 50 times out of 50 runs. CBQPSO and BQPSO have better quality of solution than BPSO with 20 and 40 particles.

In order to measure the average fitness value over the entire population, Gaussian noise is introduced into f_4 function. In this function, the average BFV of BQPSO is inferior to CBQPSO but superior to BPSO. However the St. Dev. of BQPSO is the best results.

The last function $f5$ has an optimum 500. All the algorithms can be found out the best value 499.26991. CBQPSO with 40 and 80 partilces is able to hit the optimum beyond 47 times out of 50 runs. The number of successful searches of BPSO is better than BQPSO. However the average BFV and St. Dev. of BPSO is inferior to BQPSO.

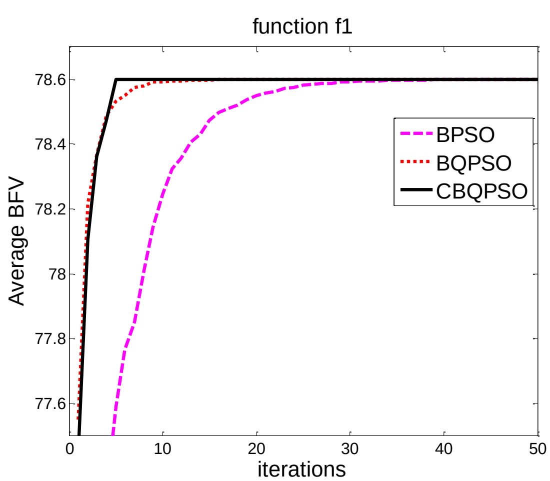
As is illustrated in Fig.1, we can see that the effectiveness of the proposed CBQPSO. CBQPSO can converge to the optimum more rapidly than BQPSO and BPSO on three functions except $f2$ and $f5$. On $f2$, BPSO converges more quickly but generates worse solution than CBQPSO. On $f5$, BPSO converges rapidly than other two

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

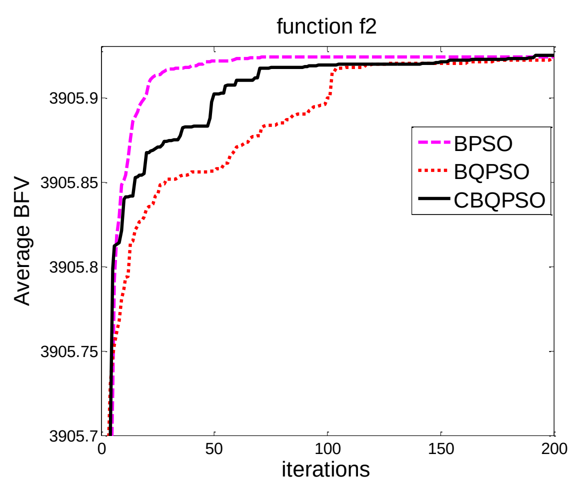
algorithms at the early stage of running, but CBQPSO exceeds BPSO soon and generates a slightly better solution. Compared with BPSO and BQPSO, experimental results show the effectiveness of the proposed CBQPSO.

Table 1: Results of BPSO, BQPSO and CBQPSO on five testing functions

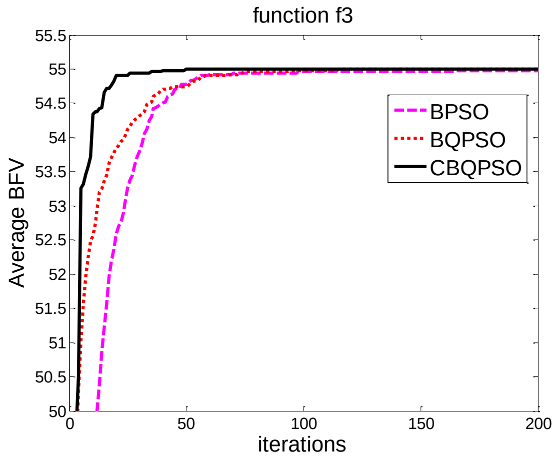
Function	***Particles***	***BPSO***		***BQPSO***		***CBQPSO***	
		Mean (St.Dev.)	***MAX (MIN)***	***Mean (St.Dev.)***	***MAX (MIN)***	***Mean (St.Dev.)***	***MAX (MIN)***
$f1$	20	78.59986 0.000086	78.6 78.5997	78.59987 0.000099	78.6 78.5997	78.59988 0.000087	78.6 78.5997
	40	78.59985 0.000076	78.6 78.5997	78.59989 0.000086	78.6 78.5997	78.59991 0.000070	78.6 78.5998
	80	78.59984 0.000086	78.6 78.5997	78.59992 0.000082	78.6 78.5997	78.59995 0.000058	78.6 78.5998
$f2$	20	3905.9002 0.110331	3905.93 3905.1536	3905.9102 0.037457	3905.93 3905.7815	3905.9144 0.038555	3905.93 3905.6873
	40	3905.9242 0.016954	3905.93 3905.8418	3905.9235 0.016072	3905.93 3905.8312	3905.9252 0.013599	3905.93 3905.8728
	80	3905.9292 0.001747	3905.93 3905.9188	3905.9292 0.001608	3905.93 3905.9214	3905.9296 0.000863	3905.93 3905.9246
$f3$	20	54.86 0.350510	55 54	54.96 0.197949	55 54	55 0.141421	55 54
	40	54.98 0.141421	55 54	55 0	55 55	55 0	55 55
	80	55 0	55 55	55 0	55 55	55 0	55 55
$f4$	20	1250.7889 3.918132	1258.2255 1240.5389	1253.5857 3.305561	1261.7592 1247.7543	1259.3092 3.602731	1266.8519 1252.1388
	40	1251.7949 4.256897	1263.1885 1241.2841	1252.9749 3.470581	1262.1260 1245.6837	1260.2212 4.843062	1274.1392 1250.1568
	80	1251.8510 3.902359	1264.5341 1243.9959	1254.2206 3.011134	1260.8333 1245.5395	1262.8852 4.664300	1270.5571 1252.4547
$f5$	20	498.71163 0.498272	499.2699 497.76306	498.75278 0.483852	499.2699 497.81977	499.25442 0.038667	499.2699 499.15955
	40	498.95986 0.366094	499.2699 498.10809	498.97292 0.422411	499.2699 497.62203	499.2699 0.000038	499.2699 499.26975
	80	499.03857 0.352775	499.2699 498.10906	499.13661 0.203900	499.2699 498.51791	499.2699 0.000022	499.2699 499.26975



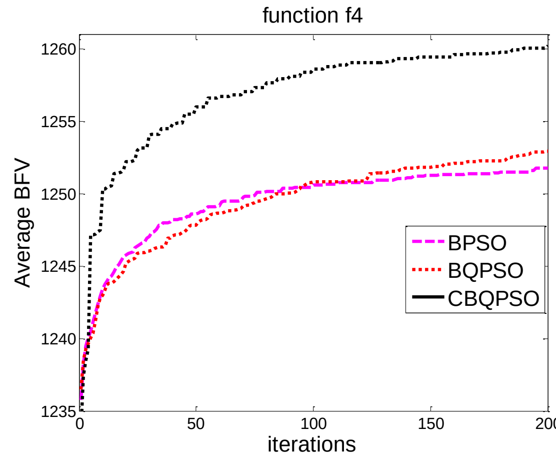
(a) $f1$



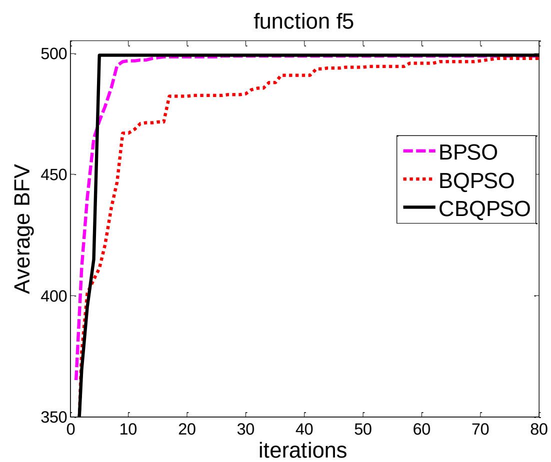
(b) $f2$



(c) $f3$



(d) $f4$



(e) $f5$

Fig. 1 The convergence process of three algorithms with 40 particles.

6. Conclusions

In BQPSO, an improvement in two components will overrule a potentially good value for a single component. In this paper, a discrete binary version of Quantum-behaved Particle Swarm Optimization algorithm with cooperative method (CBQPSO) is introduced to improve the undesirable behavior by decomposing the solution vector. In the proposed algorithm, each dimension update of particle can feed back to personal best positions and swarm best position. The results of experiment have showed that the CBQPSO algorithm performs better than other algorithm on global convergence and has stronger ability to escape from the local optimal solution during the search process. However it can be extend the search space with the increasing complexity of the problem, time consumption is the main deficiency of CBQPSO.

References

[1] Kennedy J., and Eberhart R., "Particle Swarm Optimization", in IEEE International Conference on Neural Networks, 1995, Vol.4, pp.1942-1948.

[2] Frans Van Den Bergh, "An Analysis of Particle Swarm Optimizers", Ph.D. thesis, University of Pretoria, South Africa, 2002.

[3] Jun Sun, Bin Feng, and Wenbo Xu, "Particle Swarm Optimization with Particles Having Quantum Behavior", in IEEE Congress on Evolutionary Computation, 2004, Vol.1, pp.325-331.

[4] Jun Sun, Wei Fang, Xiaojun Wu, Vasile Palade, and Wenbo Xu, "Quantum-behaved Particle Swarm Optimization: Analysis of Individual Particle Behavior and Parameter Selection", Evolutionary Computation, Vol.20, No. 3, 2012, pp. 349-393.

[5] Wei Fang, Jun Sun, Yanrui Ding, Xiaojun Wu, and Wenbo Xu, "A Review of Quantum-behaved Particle Swarm Optimization", IETE Technical Review, Vol.27, No.4, 2010, pp.336-348.

[6] Kennedy, J., and Eberhar, R. C, "A Discrete Binary Version of the Particles Swarm Algorithm". In IEEE International Conference on Systems, Man and Cybernetics, 1997, Vol.5, pp. 4104-4108.

[7] Jun Sun , Wenbo Xu , Wei Fang, and Zhilei Chai, "Quantum-behaved Particle Swarm Optimization with Binary Encoding", in International Conference on Adaptive and Natural Computing Algorithms, 2007, Vol.1, pp. 376-385.

[8] Seyyed Ali Hashemi,and Behrouz Nowrouzian, "A Novel Discrete Particle Swarm Optimization for FRM FIR Digital Filters", Journal of Computers, Vol.7, No.6, 2012, pp. 1289-1296.

[9] Jun Sun, Xiaojun Wu, Vasile Palade, Wei Fang, Choi-Hong Lai, and Wenbo Xu, "Convergence Analysis and Improvements of Quantum-behaved Particle Swarm Optimization", Journal of Information Science, Vol.193, 2012, pp.81-103.

[10] Van den Bergh F., and Engelbrecht A. P., "A Cooperative Approach to Particle Swarm Optimization", IEEE Transactions on Evolutionary Computation, Vol.8, No.3, 2004, pp.225-239.

[11] Mehdi Neshat, Shima F. Yazdi, Daneyal Yazdani, and Mehdi Sargolzaei, "A New Cooperative Algorithm Based on PSO and K-Means for Data Clustering", Journal of Computers Science, Vol.8, No.2, 2012, pp. 188-194.

[12] Jun Sun, "Particle Swarm Optimization with Particles Having Quantum", Ph.D. thesis, Jiangnan University, Wuxi, China, 2009.

Application of SketchUp in Coke Oven Three-Dimensional Digital Modeling

Shuguang Ouyang[1], Jie Xu[1], Xiaoli Su[1] and Zikan Wang[1]

[1] College of Chemical Engineering and Technology, Wuhan University of Science and Technology Wuhan, 430081, P. R. China

Abstract

Coke oven, which is a large industrial furnace, is complex in structure. A two-dimensional structure diagram can hardly help one observe the inner structure of a coke oven or master its working principle comprehensively. In order to solve this problem, a complete 3D digital model of a coke oven is generated by assembling the three-dimensional models of coke oven components created with SketchUp. It enables users to section the various components of the cove oven. The outer appearance and inner structure of the oven components also can be displayed visually from several different orientations. Moreover, it is convenient to storage and carry, operation easily and fast. It can be displayed on an ordinary computer and occupies no space at the laboratory. Meanwhile, a large sum of money that used for purchasing a physical coke oven model can be saved.

Keywords: *SketchUp, Coke Oven, 3D Digital Model, Animation.*

1. Introduction

Coke oven is the main equipment for a coking plant. It is mainly used to produce such chemical products as raw gas and coke which is indispensable for blast furnace ironmaking, cupola melting iron and nonferrous metal smelting by isolating coal from the air for dry distillation [1, 2]. Coke oven is a large industrial furnace and complex in structure. It is composed of four parts: regenerator, ramps, the coking chamber & combustion chamber and furnace roof [3]. The structures of all the four parts are complex, especially the ramp area in which thousands of ramps are laid intricately for gas, air and waste gas to pass respectively [4]. There are 56 combustion chambers in a 55 holes JN60 coke oven, each of which has 32 flues. As each flue is connected with two independent ramps, the whole ramp area has 3584 ramps. In addition, a coke oven with two kinds of heating methods includes another 1792 vertical brick gas roads [5].

Therefore, a two-dimensional structure diagram can hardly help one observe the inner structure of a coke oven and master its working principle comprehensively. In order to solve this problem, a set of 3D digital models of coke oven were created in this program with SketchUp software. The created 3D digital models of coke oven are based on a WH43K coke oven which is designed by the Design and Research Institute of Wuhan University of Science and Technology. The main structural features are: double flue, exhaust gas recirculation and two kinds of heating methods. The models of each part can be classified into two types: the bricklaying model of the coke oven components and the overall model of the coke oven components.

2. A Brief Introduction of SketchUp

SketchUp is a 3D design software program developed by @Last Software Company mainly used for 3D modeling. It also has an outstanding performance in modeling technology, material editor, animation creation and post-processing. The modeling process of SketchUp is so easy and fast compare with 3DMAX and PROE that it is widely used in architecture modeling [6, 7]. It can be used to create photo-realistic images and excellent animation works [6, 8]. In addition, it is convenient to generate sectional views and dissection animation that used for demonstrating from any orientation. Therefore, it has been widely applied in architecture, planning, landscape, indoor, and industrial design [9].

3. 3D Modeling of Coke Oven and Animation Creation Based on SketchUp

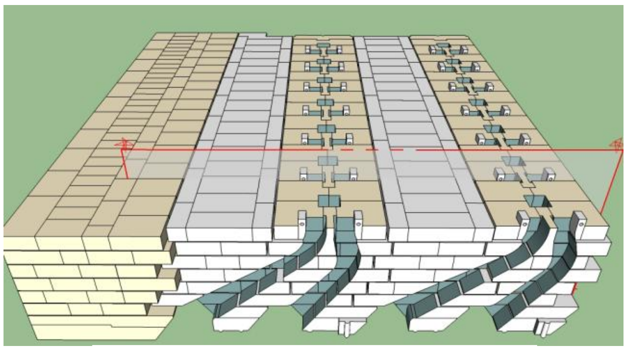
3.1 3D Modeling of Ramp Area of a Coke Oven

The ramp area is the most complex part in a coke oven whose inner structure is extremely irregular with many curved surfaces and notches. Therefore, it is also the most difficult and important part in 3D modeling. The process of generating a 3D model of a coke oven is exemplified by its ramp area in this paper.

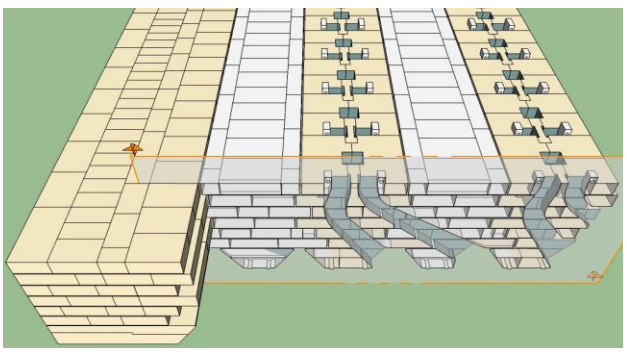
The bricklaying model of coke oven components. The modeling of the ramp area is done layer by layer from the bottom to the top. As the ramp area consists of many brick layers, and the bricks in every layer are quiet different in type and size, a large amount of 3D model monomers should be built in accordance with different brick types during the process of building each brick layer, which is shown by a large number of brick components in SketchUp. The process is as follows.

A. Some rectangle bricks are pulled out for later use according to the size of each brick in the CAD drawing.
B. The 3D model of the whole ramp area are generated strictly in accordance with the contours of the ramps in CAD Bricklaying Figure.
C. The above prepared bricks and models of ramps are precisely positioned according to their relative positions. The real brick types are created by using some SketchUp commands such as Intersect with Model. As there are a large number of expansion joints and sliding slits in the ramp area, it is necessary that a large number of guides should be drawn in SketchUp for precise positioning so as to ensure that each brick type is accurate.
D. Material, texture and color are given to each of the bricks above, which are precisely assembled according to the order of modeling process. Some important parts such as ramps and tube bricks can also be cut as required.

The bricklaying model and cutaway view of the Ramps are shown in Figure 1.



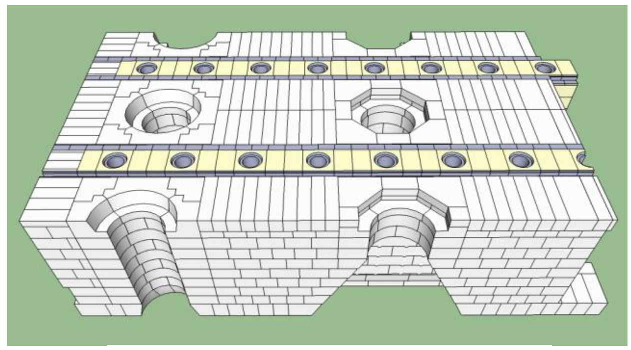
(a) The view of the Ramps.



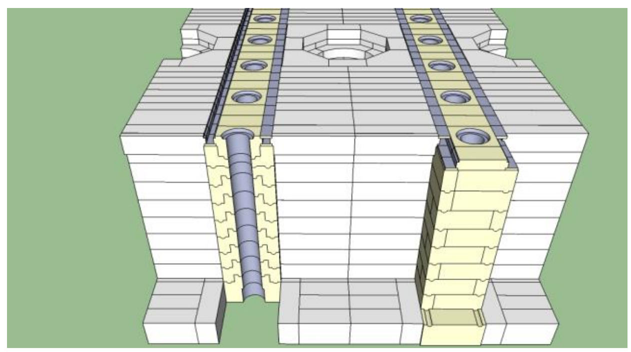
(b) The cutaway view of the Ramps.

(c) The bottom view of the ramps.

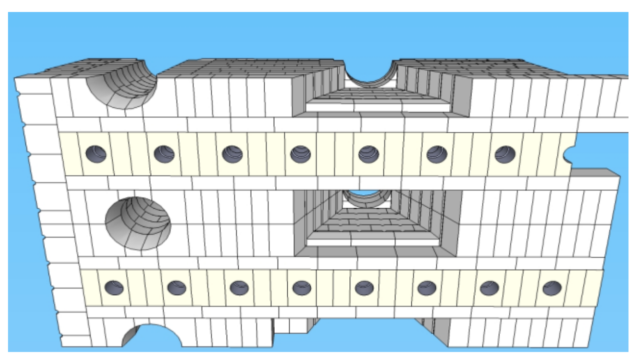
Fig. 1. The bricklaying model of the Ramps.



(a) The view of the furnace roof.



(b) The fire hole of the furnace roof.



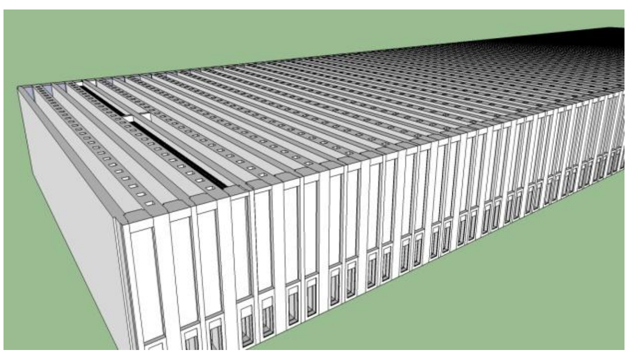
(c) The bottom view of the furnace roof.

Fig. 2. The bricklaying model of the furnace roof.

The overall model of coke oven components. Compared with bricklaying model, modeling of the overall structure of coke oven is easier. First, a rough 3D model of the ramp area is drawn according to the overall size of the ramp area in SketchUp. The 3D model of the ramps is built in accordance to the contour and size of ramps. Last, the ramps are generated by using Intersect with Model command after precise positioning, and the overall 3D model of the ramp area is obtained. In the 3D model of overall structure, expansion joints and sliding joint are not presented.

3.2 3D Modeling of Other Parts of the Coke Oven

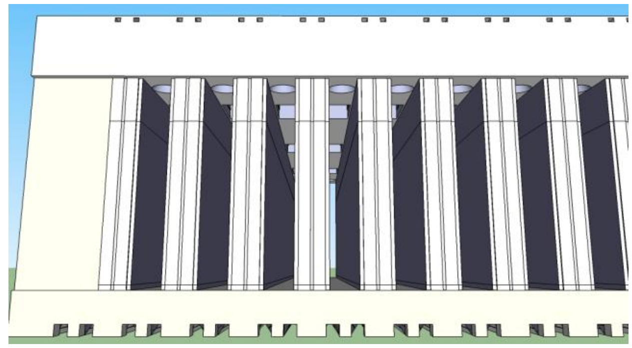
The other parts of a coke oven such as regenerator, the coking chamber & combustion chamber and furnace roof are simpler in structure when compared with ramp area because they do not have irregular curved surfaces. Consequently, modeling of these parts are much easier, and the needed models can be created only by some basic commands in SketchUp such as Pull/Push, Intersect With Model. The modeling process is similar to the ramp area, which is not repeated here as to the paper length. Models of the various parts of the coke oven are shown in Figures 2, 3, 4, and 5.



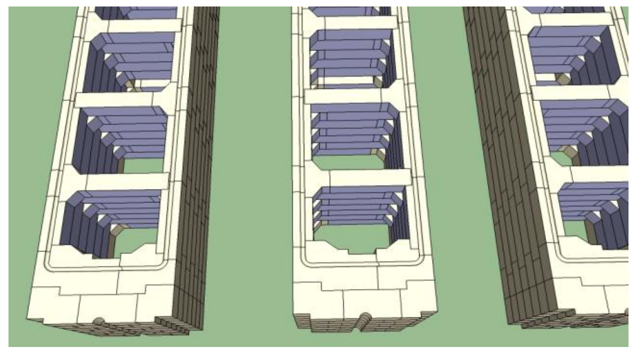
(a) The regenerative chamber's overall model.

(b) Partial bricklaying model of the regenerative chamber.

Fig. 3. The 3D model of the regenerative chamber.



(a) The coking chamber & combustion chamber's overall model.

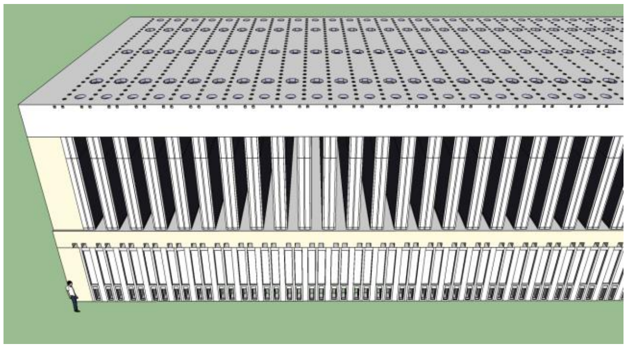


(b) Partial bricklaying model of the coking chamber & combustion chamber.

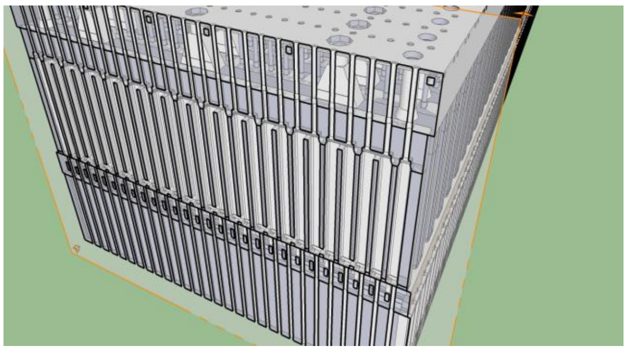
Fig. 4. The 3D model of the coking chamber&combustion chamber.

4. Animation Creation Based on 3D Digital Model of Coke Oven

SketchUp, which owns the feature of animation, can automatically generate smooth animation according to a variety of perspective views which can be set after the scene is created. Its animation is easy to achieve in which the obscure concept of keyframe is avoided [8]. In addition, SketchUp can also conveniently create sectional views from any orientation and generate dissection animation for display [10]. Moreover, the animation can be exported in AVI, Quicktime and MOV files [9].
In this program, there are 48 AVI animations in total. The content of the animations involve the movement and rotation animation of various parts of the coke oven, dissecting display, the growth animation of the coke oven, as well as the display of the structures of such coke oven parts as checker bricks, grate brick and furnace column etc. The animation vividly displays the relative positions of various parts of the coke oven as well as its complex inner structure.



(a) The regenerative chamber's overall model.



(b) The cutaway view of the coke oven.

Fig. 5. The 3D model of the whole coke oven.

In addition, the relation between the smoothness of the animation and size of AVI file should be balanced. So the size of AVI files should be decreased as far as possible on the premises of aesthetic appearance and clear display of the structure. If the size of the AVI file is too large, it will not only occupy a large disk space, but more importantly consumes a large amount of system resources which may cause inconvenience in animation exportation [9, 11]. Accordingly, two measures can be taken. First, the material which is sensitive to light and occupies a large space should be avoided when the 3D model is given the materials. Second, the time frame in the animation output should be strictly restricted. This design adopts 20 frames per second.

5. Conclusions

The created 3D digital model has many merits. First, the various parts of the coke oven could be sectioned and visually displayed. It enables users to observe the outer appearance and inner structure of the coke oven from many different perspectives. Second, it is convenient to storage and carry, and it can be displayed on any ordinary computer. Last but not the least, it occupies no space in the laboratory and saves the money that used for purchasing a physical coke oven model.

References

[1] Zhaozhang Yao and Mingdon Zheng, Coking Textbook, 3rd edn., Beijing, CHINA: Metallurgical Industry Press, 2008.
[2] Thomas Henry Byrom and John Edward Christopher, Modern Coking Practice, London: C. Lockwood and son, 1910.
[3] Rainer Worberg and Martin Reinke, "Progress in Coke Oven Technologies", Coke Oven Managers Association, 2009, pp. 1-15.
[4] Zhendon Yu and Wenhua Zheng, Modern coke oven production of technical manuals, Beijing, CHINA: Metallurgical Industry Press, 2010.
[5] Zhendon Yu and Chengyou Cai, Coke oven production technology, Shenyang, CHINA: Liaoning Science and Technology Press, 2003.
[6] Daniel Tal, Google SketchUp for site design; a guide to modeling site plans, terrain, and architecture, Hoboken, New Jersey: John Wiley & Sons, Inc., 2009.
[7] J. Rodr ǵuez, N. Omtzigt, E. Koomen, and F. S. de Blois, "3D Visualisations in Simulations of Future Land Use: Exploring the Possibilities of New, Standard Visualisation Tools", International Journal of Digital Earth, Vol. 1, No. 1, 2008, pp. 148-154.
[8] Tao Wei and Song Wang, SketchUp architectural design, Beijing, CHINA: China Electric Power Press, 2007.
[9] Chris Grover, Google SketchUp: The Missing Manual, Sebastopol, California: O'Reilly Media / Pogue Press, 2009
[10] Sandeep Singh, Beginning Google SketchUp for 3D Printing, New York: Apress, Inc., 2010.
[11] Wei Li, Deren Sheng, and Wei Li, "Development of Fluid Machinery Multimedia Demonstration System Based on 3DS MAX and Authorware", Power System Engineering, Vol. 22, No. 5, 2006, pp. 50-52.

Shuguang Ouyang received the B.Sc. and M.Sc. degrees in Chemical Engineering and Technology from the Wuhan University of Science and Technology (WUST), Wuhan, China, in 1992 and 1995, respectively.
He is currently a Professor at the College of Chemical Engineering and Technology, Graduate University of the Wuhan University of Science and Technology, Wuhan. His research interests include Coal Conversion, New Carbon Material, Computer Science & Technology, and 3D Digital Models of Coke Oven. His current research focuses on Image Recognition, Activated Carbon and 3D Digital Models of Coke Oven. He has published more than 30 research articles in refereed journals.
Pro. Ouyang is a member of the Hubei Coal Conversion and New Carbon Material Key Laboratory and was a Co-Chair of Internal Workshop on Modern Science and Technology 2012.. And he is a recipient of 2009 National Science and Technology Progress Awards of China (the Second Award).

Jie Xu received the B.S. degree in Chemical Engineering and Technology from the Wuhan University of Science and Technology (WUST), Wuhan, China, in 2010. He is currently working towards the M.S. degree at the College of Chemical Engineering and Technology, WUST. His research interests mainly include Coke-making and 3D Digital Models of Coke Oven.

Xiaoli Su received the B.S. degree in Chemical Engineering and

Technology from the Wuhan University of Science and Technology (WUST), Wuhan, China, in 2011. He is currently working towards the M.S. degree at the College of Chemical Engineering and Technology, WUST. His research interests mainly include Activated Carbon and Coking Fly Ash.

Zikan Wang received the B.S. degree in Chemical Engineering and Technology from the Guangdong Ocean University, Guiyang, China, in 2011. He is currently working towards the M.S. degree at the College of Chemical Engineering and Technology, Wuhan University of Science and Technology (WUST), Wuhan, China. His research interests mainly include Coking Fly Ash and 3D Digital Models of Coke Oven.

Analysing Word Importance for Image Annotation

Payal Gulati, A. K. Sharma

Computer Engineering Department, YMCA UST
Faridabad, Haryana, India

Abstract

Image annotation provides several keywords automatically for a given image based on various tags to describe its contents which is useful in Image retrieval. Various researchers are working on text based and content based image annotations [7,9]. It is seen, in traditional Image annotation approaches, annotation words are treated equally without considering the importance of each word in real world. In context of this, in this work, images are annotated with keywords based on their frequency count and word correlation. Moreover this work proposes an approach to compute importance score of candidate keywords, having same frequency count.

Keywords: *Image Annotation, Keyword Importance, Frequency Count, Word Correlation, Indexing*

1. Introduction

The importance of image annotation has been increasing with the growth of the worldwide web. Finding relevant digital images from the web and other databases is a difficult task because many of these images do not have relevant annotations [4,5]. There are various methods to deal with image annotations but it is observed that these approaches [6,7,9] treat each candidate word equally. In the light of this, in this work, words are associated with importance score instead of being considered equally. Annotation words should be associated with importance measurement, instead of being considered equally. In this work, a method to compute an importance scores for every candidate keyword is proposed. In this approach, collaboration of word frequency and word correlation is used for annotating images. Moreover importance score of keywords with same frequency is determined, for instance, suppose $word_1$ and $word_2$ have the same frequency count, but $word_1$ is correlated with $word_3$ which has higher importance as compared with $word_2$ which is correlated with $word_4$ having low importance; thus $word_1$ will be considered important as compared to $word_2$.

This paper is organised in the following way. Section 2 discusses the relevant work done in this domain. Section 3 presents the proposed architecture. Algorithm and implementation results are discussed in section 4 and 5 respectively. Finally Section 6 comprises of the conclusion.

2. Related Work

Keywords are most important Search Engine Optimizing [1,8] element as they are the search strings that are matched against. Therefore importance of keyword needs to be analyzed before annotating and indexing them. Some researchers have worked on analysis of keyword on the basis of keyword density [2,3]. The general idea is higher the keyword density, more relevant to the search string the page is. Some work is done on tracking the keyword on special places which proves that keywords are not only selected on the basis of their quantity i.e. frequency but on quality as well. For instance keywords in page title, headings are of more importance compared to the other text. In the light of this, this work focuses on calculating the importance of keyword on the basis of word correlation irrespective of their same frequency/occurrence count.

3. Proposed Architecture

The proposed architecture shown in Fig. 1 is discussed in this section.

3.1 WWW and Web Page

WWW is a vast repository of web pages. Crawl manager takes the seed URL from the URL list and fetches the page from www.

3.2 Parser

Parser parses the web page fetched by crawl manager from www. It then parses for tags such as image tags, alt text, metadata, page title tags etc. This phase returns all the keywords fetched during parsing to the frequency calculator.

3.3 Frequency Calculator

In this phase, frequency count of all the keywords returned by the parser is determined. Candidate keywords are further determined based on the threshold value.

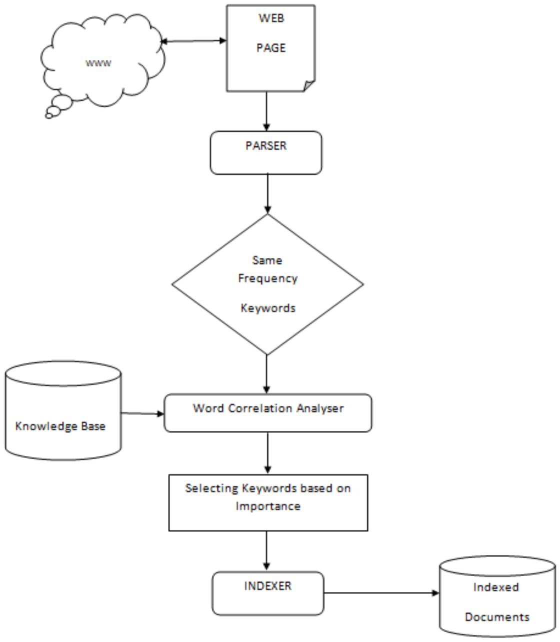


Fig. 1 Proposed Architecture

3.4 Same Frequency Keyword Detector

Same frequency keyword detector detects the candidate keywords having same frequency count.

3.5 Knowledge base

The knowledge base is a repository of extracted rules which have been derived using the association rule mining. A knowledge base containing rules is shown in Fig.2.

Table1

ID	TermsX	TermsY
3	water	mountain
4	himalaya	mountain
5	nature	mountain
6	tree	nature
7	leaves	tree
8	greenary	leaves
9	bird	nature
10	greenary	grass
11	nature	sky
12	sun	sky

Fig. 2 Knowledge Base

3.6 Word Correlation Analyser and Important Keyword Selector

In this module correlation between words is determined based on association knowledge base. In this module, keyword importance is analysed based on word correlation. Suppose word w1 and w2 have the same frequency, but w1 was correlated to word w3 which has higher importance, while w2 was correlated with w4 which has lower importance. Thus w1 will be considered important compared to w2.

3.7 Indexer

This modules indexes www images along with their keywords in indexed repository. Without an index, the search engine would scan every document in the corpus for the queried image, which would require considerable time and computing power.

3.8 Indexed Repository

Image Search engine component crawler collects, parses, and stores images in an indexed repository to facilitate fast and accurate image retrieval.

4. Algorithm

ImportanceScoreCalculation
Input : Same Frequency Keywords
Output: Important Keyword

Algorithm ImportanceScoreCalculation
Begin
for every keyword with same frequency count
calculate the Rank of keyword using
$R(b) = \sum_{b \in B}(T(a,b) + R(a)$ *where a is the backword of b i.e.* $a \rightarrow b$, *R(b) is the rank of word b that belong to set B that contains all the forward words in the knowledge base, R(a) is the rank of a.*
If $a \rightarrow b$ *(a is a back word of b)*
$T(a,b) = 1/Ntotal$ *where Ntotal is the total no. of keywords in the knowledge base*
else
$T(a,b) = 0$
end

Let A,B,C,D,E,F,G,H,I be keywords and their corresponding frequencies are shown in Table 1.

Table 1. Keywords and their Frequency

Keywords	Frequency Count
A	4
B	6
C	3
D	6
E	5
F	2
G	2
H	1
I	3
J	1

Considering 40% threshold value, Candidate keywords obtained from Table 1 is shown in Table 2.

Table 2. Candidate keywords

Candidate Keywords	Frequency Count
A	4
E	5
B	6
D	6

Importance score for same frequency keywords B and D can be calculated using the proposed algorithm ImportanceScoreCalculation.

Table 3. below show the word correlations

Table 3. Word Correlation

Word Correlations
a → b
c → b
d → b
e → d
f → e
g → f
h → d
g → j

According to the algorithm, Rank of B and D is calculated

$N_{Total} = 1/10 = 0.1$

$$
\begin{aligned}
R(b) &= [0.1 + R(a)] + [0.1 + R(c)] + [0.1 + R(d)] \\
&= 0.1 + 0.1 + [0.1 + [[0.1 + R(e)] + [0.1 + R(h)]] \\
&= 0.1 + 0.1 + [0.1 + 0.1 + [0.1 + R(f)] + 0.1 \\
&= 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + [0.1 + R(g)] + 0.1 \\
&= 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 \\
&= 0.7 \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
R(d) &= [0.1 + R(e)] + [0.1 + R(h)] \\
&= [0.1 + [0.1 + R(f)] + 0.1 \\
&= 0.1 + 0.1 + [0.1 + R(g)] + 0.1 \\
&= 0.1 + 0.1 + 0.1 + 0.1 \\
&= 0.4 \qquad (2)
\end{aligned}
$$

From (1) and (2) it is seen that keyword b is more important compared to keyword d as per their correlation score.

5. Implementation

To compute the importance score of candidate keywords, the proposed algorithm has been implemented in java. The implemented result for the same is shown in Fig. 3.

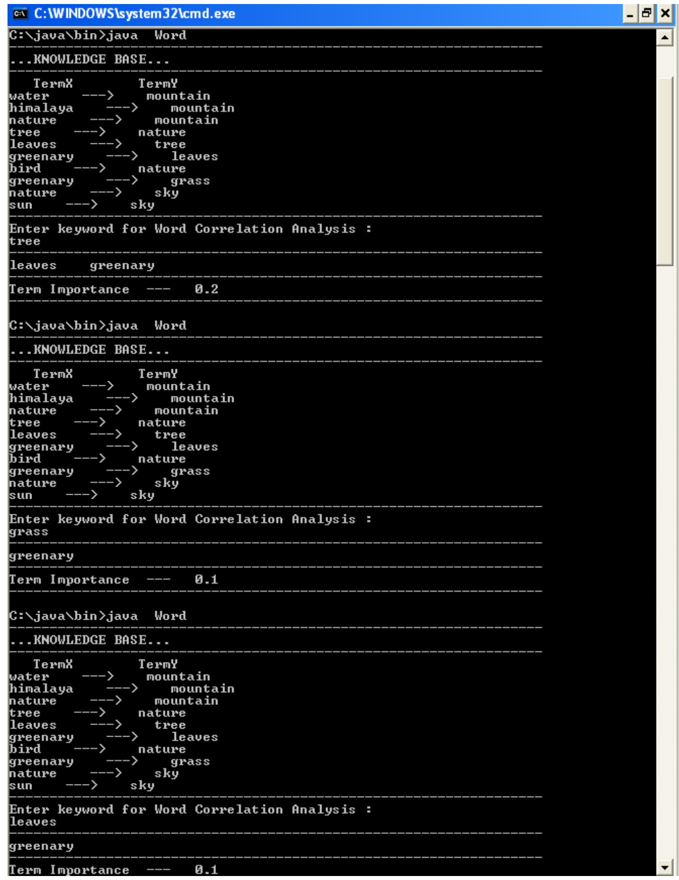


Fig. 3 Implementation of the proposed approach

6. Conclusion

This work proposes an approach for analysing word importance for image annotation based on frequency count and word correlation information. An approach proves to be efficient, as important keywords are analysed for annotating images.

References

[1]Wayne Hulbert, "Keyword Density: SEO Considerations", May.2005
[2] Lee Underwood, "A Brief History of Search Engines".
[3] Yogesh Yadav et al, Site Content Analyzer in Context of Keyword density and Key Phrase, Int. J. Comp. Tech. Appl., Vol 2 (4), 860-872, 2011.
[4] Chai, C. and C. Hung. 2008. Automatically annotating images with keywords: A review of image annotation systems. Recent Patents on Computer Science, 55–68.
[5] S. Gao, D.-H. Wang and C.-H. Lee, "Automatic Image Annotation through Multi-Topic Text Categorization", Proc. ICASSP, Toulouse, France, May 2006.
[6] M. Srikanth, J. Varner, M. Bowden and D. Moldovan, "Exploiting Ontologies for Automatic Image Annotation", Proceedings of the 28th ACM SIGIR, 2005.
[7] T. Sumathi1, C.Lakshmi Devasena2, and M.Hemalatha, "An Overview of Automated Image Annotation Approaches", International Journal of Research and Reviews in Information

Sciences Vol. 1, No. 1, March 2011 Copyright © Science Academy Publisher, United Kingdom.

[8] Ville Viiyaniemi and Jorma Laaksonen: "Keyword detection approach to automatic image annotation": In proceedings of 2nd European workshop on the integration of knowledge, semantic and Digital media technologies (EWIMT 2005) London, UK, November 2005.

[9] C. Wang, F. Jing, L. Zhang, H. Zhang, Content-Base Image Annotation Refinement, Proc. of CVPR, 2007.

[10] Hemlata Sahu, Hitesh Gupta, "An Approach to review on Automatic Image Annotation, A Survey", International Conference on Computing and Control Engineering (ICCCE 2012), 12 & 13 April, 2012.

A Review on: Issues Related to Security on Tree Structure Data

Vivek Waghamre, Dr. Ravindra Thool

**Ph. D. Research Scholar, S. G. G. S. I. E. & T, NANDED, SRTMUN, NANDED
NANDED, MAHARASHATRA -431605, INDIA**

**Professor & HOD of Information Technology , S. G. G. S. I. E. & T, NANDED
NANDED, MAHARASHATRA -431605, INDIA**

Abstract

The problem of tree structured data is that it requires different Integrity and confidentiality for different portion of same content. Integrity assurance technique not only applies integrity to received data by user, but also any compromise to data that must be precisely determined. In tree structures each node contains some content and structural relationships between the nodes. Therefore, it considers structural integrity and content integrity. Confidentiality means a user receives only those nodes as well as organized information to user has privilege according to access control policies, wherever one should not infer others information.

In this paper we reviewed different techniques related to dissemination of tree structured data that exploits structural properties of tree based data model (such as XML document). The approach is based on notion of encrypted post-order numbers which is based on post-order number properties. It facilitates efficient identification, extraction, distribution of selected content portions.

Keywords: *Confidentiality, Integrity, Tree Structured Data.*

1. Introduction

An XML (Extensible Markup Language) has become the standard document interchange language for the web. XML document contain information of different sensitivity degrees that must be shared by possible large user communities. Data sharing among multiple parties require both data integrity & data confidentiality [1]. Data that a consumer is not authorized to access, but belongs to the complete data set is called extraneous data. Flow of extraneous data to a consumer may leak information, even when this data is encrypted.

In particular, extraneous data is prone to offline dictionary attacks even by a legitimate consumer that can exploit contextual knowledge from the data elements it has access to. Therefore, it is important that extraneous data, even if encrypted with keys that the consumer does not have, be removed from the content before its delivery. Efficiency and scalability must however be provided by assuring at the same time security of contents and privacy of the parties acquiring and disseminating contents. It is useless to provide high-bandwidth content distribution systems if integrity of the disseminated contents is not as sure or the property of the contents not protected. Such problems are further complicated when dealing with contents encoded in XML, in that, because of the hierarchical organization of the content, different confidentiality and integrity requirements may exist for different portions of the same content. Thus need a dissemination approach specifically tailored to XML that addresses the issues of security, privacy and scalability in a holistic manner [2].

The structural properties also contribute towards the efficiency and scalability of the dissemination framework. This solution is based on the simple notion of encrypted post-order numbering [1] and its properties. By using such notion, they develop a novel content routing scheme called Structure Based Routing of XML-data. Such routing scheme prevents information leaks and at the same time improves efficiency and scalability of the structure based dissemination model. A key feature of this approach is that it directly takes into account access control policies, that is, policies specifying which entity can access which portion of the contents, so that contents is disseminated according to these policies. The resulting dissemination model is a multicast model for XML dissemination that based, on the content structure and access control policies, builds an overlay topology. Moreover, we exploit the properties of post-order numbers towards integrity assurance. This technique allows consumers to verify the integrity of data they receive, and in the case in which data have been tampered with, allows the consumers to determine the affected portions of the data.

2. Related Work

Many of different design approaches are related to this topic. The models proposed Fernandez [9] and Rabitti [10] are specifically tailored to an object-oriented DBMS storing conventional, structured data. As such, great attention has been devoted to concepts such as versions and composite objects, which are typical of an object-oriented context. Those models support concepts such as positive and negative authorizations, and authorization propagation. This model also supports such concepts, even though it has a larger variety of authorization propagation options.

Three different options are supported by which the SA can specify,

- That an authorization defined at a given level in the hierarchy Propagates to all lower levels.
- That the propagation stops at a specified level down in the hierarchy.
- That no propagation has to be enforced.

By contrast ORION, Rabitti [10] has only one propagation policy, which is equivalent to option (1). Moreover, none of the above mentioned models provide support for secure information push mechanisms. This is the most innovative feature of access control model, which is not found in any access control model previously proposed for object oriented DBMSs. An access control model for WWW documents has been proposed by Samarati [11]. In this model, HTML documents are considered, organized as unstructured pages connected by links. Authorizations can be given either to the whole document or to selected portions within the document. Although [11], the idea of selectively granting access to a document (by authorizing a subject to see only some portions and/or links in the document), this work substantially differs from the proposal. Differences are due to the richer structure of XML documents with respect to HTML documents and to the possibility of attaching a DTD to an XML document, describing its structure. Such aspects require the definition and enforcement of more sophisticated access control policies, than the ones devised for HTML documents. The access control model proposed [11] has great limitations deriving from the fact that it is not based on a language able to semantically structuring the data, as in this model for XML. As such, administering authorizations is very difficult. In particular, if one wants to give access to portions of a document, he/she has to manually split the page into different slots on which different authorizations are given.

This problem is completely overcome by providing semantic information for various document components. Authorizations can thus be based on this semantic information. An access control model for XML documents has been recently proposed by Damiani [11]. This model is very similar to previous models for object-oriented databases and does not actually take into account some peculiarities of XML.

In particular, this model has two main shortcomings,

- The first one is that it does not consider the problem of a secure massive distribution of XML documents and thus considers only the information pull mode.
- Second, the model proposed in Damiani [11] does not provide access control modes specific to XML documents, it only provides the read access mode.

By contrast, it provides a number of specialized access modes for browsing and authoring, which allow the SA to authorize a user to read the information in an element and/or to navigate through its links, or to modify/delete the content of an element/attribute. Because of the widespread use of XML and due the relevance of XML security, the Worldwide Web Consortium (W3C) has set up several working groups to address the various security aspects related to XML. For example, The XML Working Groups of the W3C are working on standards for both an XML representation of digital signatures (W3C XML Signature Working Group) and encrypted contents W3CXML Encryption Working Group). The goal of the OASIS Consortium [OASIS Consortium] is the design and development of industry standard specifications for XML-based interoperability. In this framework, the XACML Technical Committee is studying the definition of a standard model for XML based security policies. However, the draft proposal for XACML is based on very simple access control policies, in that notion such as credentials, positive/negative policies, convict management, and dissemination strategies are not taken into account.

Other related work deals with approaches proposing more flexible methods to qualify subjects with respect to traditional identity-based schemes for access control. One of the most relevant research efforts in this area are role-based access control (RBAC) models. In particular, note that the concept of credential has some similarity with that of role [12]. Roles can be seen as a set of actions or responsibilities associated with a particular working activity. Under role-based models, all authorizations needed to perform a certain activity are granted to the role associated with that activity, rather than being granted directly to users. Users are then made members of roles, thereby acquiring the roles authorizations. User access to data is mediated by roles, each user is authorized to play certain roles and, on the basis of the role, he/she can perform accesses on the data. Whenever a user needs to perform a certain activity, the user only needs to be granted the authorization of playing the proper role, rather than being directly assigned the required authorizations. A basic distinction between roles and credentials is that credentials are characterized by a set of attributes, and this allows us to grant access authorizations only to users whose credentials satisfy certain conditions (e.g., access to a document can be granted to all the users with a given age or with a given nationality). This can of course be done also through roles but it requires the creation of a distinct role for each condition to enforce (e.g., enforcing the access control policy of the previous example requires the creation of two distinct roles, one corresponding to the users with the specified age, and the other corresponding to the users with the specified nationality).

This makes the specification and management of authorizations very difficult, given also the large variety of users that typically access XML documents. The concept of subject credential was first presented by Winslett [13], whereas the access control model proposed by Adam [15] provides a formalization of the concept of subject credential by proposing the credential logic-based specification language that we use in this article. Other related work on credential specification for stranger parties. In particular, the work by IBM on Trust Policy Language (TPL), Herzberg and Mass [3] is devoted to the enforcement of an XML-based framework for specifying and managing role based access control in a distributed context. This framework has been extended for mapping subject certificates to a role, based on policies defined by the owner of the resource and on the roles of the issuers of the certificates. In another area, Bertino and Ferrari approach supporting access control in both pull and push based distribution of data [3]. This approach is depending on encrypting different portion of the data with different keys & then distributing the keys to data consumers according to access control policies. Information pull is based on authorization. Consumer sends request to source for XML document. When consumer submits an access request then access control system checks authorization of consumer. Based on this authorization, consumer is returned a view of the requested document that contains all and only those portions. When no authorizations are found, the access is denied. Information push approach is used for distributing documents to users which based on broadcast data to clients. Also in this case, different users may have privileges to see different, selected portions of the same document. Thus, different views of same document are sent to different consumer [5]. Example, the case of a newsletter sent once a week to all users. Different users have different privilege to see different, selected portion of same document, supporting an information push approach for generating different physical views of the same document and sending them to proper users. The main problem with Information pull and Information push approach is number of views becomes large and such approach cannot be practically applied. Bertino [3] have also investigated the problem of integrity of XML data by using notion of Merkle Hash Tree. Merkle proposed a digital signature scheme [1] [2] [3] based on a secure conventional encryption function over a hierarchy (tree) of document fragments. Merkle trees are binding (integrity-preserving) but not hiding (confidentiality-preserving). The use of commutative hash operations to compute the Merkle hash signature prevents leakage related to the ordering among the siblings. However it cannot prevent the leakage of signatures of a node and the structural relationships with its descendants or ancestors. Moreover, one-way accumulation is very expensive in comparison to the one-way hash operation. Merkle hash trees are a well-known mechanism used in several computer areas for certified query processing. For instance, it has been exploited by Naor and Nissim in [16] to deal with the task of introducing and maintaining efficient authenticated data structures that holds the information of certificates about its validity. More precisely, the paper proposes as data structure a sorted hash tree scheme, such that tree leaves correspond to revoked certificates. Thus, verifying that a certificate is revoked or not implies verifying the existence of certain leaves in the tree.

A similar approach has been proposed by Devanbu et al. [17] to prove the completeness and authenticity of queries on relational data. Similar schemes have also been used for micropayments [18], where Merkle hash trees are used to reduce the number of public key signatures that are required in to provide or authenticating a sequence of certificates. By contrast, the use of such trees for handling XML documents is still a novel aspect, which, to the best of our knowledge, has been so far investigated only by Devanbu et al. in 179. In this work, the authors have developed a scheme, based on Merkle hash trees, allowing clients to validate the answers to certain type of queries against XML sources managed by untrusted publishers. The method developed in [17] is based on the definition of a data structure, called Xtrie, which stores the set of possible paths that can be specified on a given DTD. However, the work presented in [5] has many differences with regard to this proposal. A first difference is the type of XML documents supported by the two approaches. In this approach, they have no limitation on the structure of XML, documents, whereas the approach presented in does not consider attributes and it imposes that data content be only present in leaf nodes. Another important difference is the kinds of queries for which the subject is able to verify authenticity. In our approach, we can certify the authenticity for each possible kind of XPath queries, whereas the approach presented in considers only queries returning whole sub trees. A further distinction is that it also considers completeness with regard to access control rights, besides data authentication, and provides a comprehensive architecture and related mechanisms to support data authentication and completeness services. Bertino proposed a technique based on the Merkle hash technique for selective dissemination of XML data in a third party distribution framework [4]. Gladney and Lopspiech proposed solution for above problem which is based on Multilevel Encryption [3]. In multilevel encryption, different portions of same document are encrypted with different keys and same encrypted copy is broadcasted to all subjects.

Issues related to multilevel encryption are,

- Which and how many keys should be distributed to which subjects?
- How to securely and efficiently distribute keys?
- How to encrypt document?

Solution for these issues:

- Encryption of document according to specified access control policies.
- According to policies apply key.
- Therefore number of policies equal to number of keys.

The Data Dissemination Problem has been studied by a number of projects [19, 2]. However, none of them attempt to reduce costs by automatically merging similar queries. The Query Merging Problem is also related to Client-side caching in client/server configurations [21]. In this approach, data is loaded into each client cache as answers to other queries are broadcast by the server. When a client is ready to make a query, it first checks in its own cache to see if the cache already contains the answer. The difference with this work is that in the client-caching approach, queries are not known, so the server cannot optimize the global cost. This research is also related with the Semantic Query Optimization Problem [23]. The goal of the Semantic Query Optimization problem is to use semantic knowledge (such as integrity constraints) for transforming a query into a form that may be answered more efficiently than the original version. The main difference with this work is that in semantic query optimization only one query at a time is optimized. This limits the opportunities for improvement versus work, where they considered a set of queries.

There are a number of data dissemination products and services in the market [24, 27]. However, as far as know they do not attempt to do any real query merging. Most of these products are very simple, requiring clients to maintain their subscriptions and to "pull" from the server any new information. Servers normally unicast the results to each client, making this approach non-callable and resulting in a very high cost. The Cellular Telephony and Telecommunication research community has also consider the problem of improving the bandwidth use on broadcast channels [25][24]. The difference between this effort and work is the level of abstraction. While the telephony community focus on random memory page requests (and therefore, there is little information available to the optimizer), specially this work focus on queries and query answers which allows to have more sophisticated schemes. The BADD problem [26, 28] has generated a wealth of research in the data dissemination arena. References [33] and [34] have proposed multicast protocol, which can be used as a low level support to this algorithm. Deployment of Internet services through a satellite broadcast channel has been studied in [26] and smart information “push” by [33]. Reference [30] extends the client-side caching by considering caches not only at the client, but also at intermediate locations "close" to the clients. Finally, in [31] the data staging problem is described and heuristics to solve it are presented. The query merging problem in a geographical database is closely related to the polygon covering problem, and to the set covering problem [32]. However, the special characteristics of the Query Merging Problem make it difficult to directly use the well known solutions to those problems.

3. Conclusions

RBAC model supports the formulation of high-level access control policies. Such policies take into account both user characteristics, document contents and structure. In Merkle hash tree, values of sub-trees that are not accessible to consumer also to be forwarded. Also it, avoids sending the hash values of sub-trees that are not accessible to consumer. Query Merging Problem, presented a general framework and cost model for evaluating merging, and also a variety of merging algorithms. It shows that dissemination costs can be significantly decreased by applying a merging algorithm, and that heuristic algorithms work well.

References

[1] A. Kundu and E. Bertino, “A New Model for Secure Dissemination of XML Content,” IEEE Transaction on, May 2008.

[2] A. Kundu and E. Bertino, “Secure Dissemination of XML Content Using Structure Based Routing,” in Proc. 10th IEEE Int. Enterprise Distrib.Object Computing. Conf. (EDOC06), 2006, pp. 153164.

[3] E. Bertino and E. Ferrari, "Secure and Selective Dissemination of XML Documents," ACM Trans. Inf. Syst. Secure, vol. 5, no. 3, pp. 290331,2002.

[4] E. Bertino, B. Carminati, E. Ferrari, B. M. Thuraisingham, and A. Gupta, "Selective and Authentic Third-party Distribution of XML documents," IEEE Trans. Knowl. Data Eng., vol. 16, no. 10, pp. 12631278,Oct. 2004.

[5] Su Cheng Haw, and G. S. V. Radha Krishna Rao,"Query Optimization Techniques for XML Databases," 2005.

[6] C. Wang, A. Carzaniga, D. Evans,"Security Issues and Requirements for Internet-scale Publish subscribe Systems," 2002.

[7] A. K. Datta, M. Gradinariu, M. Raynal, and G. Simon, "Anonymous Publishsubscribe in p2p Networks," presented at the Int. Parallel Distrib.Process. Symp., Nice, France, 2003.

[8] F. Cao and J. Singh, "Efficient Event Routing in Contentbased Publish Subscribe Service Networks," in Proc. of IEEE INFOCOM 2004, pp. 929-940.

[9] Fernandez, E., Gudes, E., & Song, H. "A Model for Evaluation and Administration of Security in Object-oriented Databases," IEEE Trans. Knowl. Data Eng,1994, 275292.

[10] Rabitti, F., Bertino, E.,Kim, W., Ndwoelk, D. "A Model of Authorization for Next-Generation Database Systems," ACM rans. Datab. Syst. 1991, 16, 1, 88131.

[11] Damiani, E., De Captitani Di Vimercati, S. Paraboschi, S., & Samarati, P. "Securing XML Documents," In Proceedings of the 6th International Conference on Extending Database Technology (Konstanz, Germany), 2000 pp. 121135.

[12] Osborn, S. Ed. Proceedings of the 5th ACM Workshop on "Role-Based Access Control," (Berlin, Germany) ACM, New York, 2000.

[13] Winslett, M., Ching, N., Jones, V., & Slepchin, I. "Using Digital Credentials on the World Wide Web," J. Comput. Secu.197, 7.

[14] Adam, N., Atluri, V., Bertino, E., & Ferrari, E. "A Content-Based Authorization Model for Digital Libraries," IEEE Trans. Knowl. Data Eng.2002, 14, 2,296315.

[15] Herzberg, A., Mass, Y., & Mihaeli, J. "Access Control Meets Public key Infrastructure or Assigning Roles to strangers," In Proceedings of the IEEE Symposium on Security and Privacy (Oakland, Calif.). IEEE Computer Society Press, Los Alamitos, Calif. 2000.

[16] M. Naor and K. Nissim, "Certificate Revocation and Certificate Update," Proc. Seventh USENIX Security Symp., 1998.

[17] P. Devanbu, M. Gertz, C. Martel, and S. Stubblebine, "Authentic Third-Party Data Publication," Proc. 14th Ann. IFIP WG 11.3 Working Conf. Database Security, Aug.2000.

[18] S. Charanjit and M. Yung, Paytree: "Amortized Signature for Flexible Micropayments," Proc. Second Usenix Workshop Electronic Commerce, 1996.

[19] D. Giord. Polychannel, "Systems for Mass Digital Communication," CACM, February 1990.

[20] T. Imielinski and B. Badrinath, "Mobile Wireless Computing," CACM, 37(10):18-28, Oct 1994.

[21] A. M. Keller and J. Basu, "A Predicate-Based Caching Scheme for Client-Server Database Architectures," The VLDB Journal, 5(1), January 1996.

[22] T. Bowen, G. Gopal, G. Herman, T. Hickey, K. Lee, J. Raitz, and A. Weinrib. "The Data cycle Architecture," CACM, 35(12), December 1992.

[23] J. J. King. Quist: "A System for Semantic Query Optimization in Relational Databases," In Very Large Data Bases, 7th International Conference, September 9-11, 1981,Cannes, France, Proceedings, pages 510- 517. IEEE Computer Society Press, 1981.

[24] A. Chan. Transactional publish/subscribe: "The Proactive Multicast of Database Changes," In ACM SIGMOD, June 1998.

[25] Q. Hu, D. L. Lee, and W. C. Lee, "Optimal Channel Allocation for Data Dissemination in Mobile Computing Environments," In 18th International Conference on Distributed Computing Systems, May 1998.

[26] M. Lazaro and P. Sage, "Any Information, Anywhere, Anytime for the Warghter," In Proceedings of the SPIE, volume 3080, pages 35-42, 1997.

[27] M. Tan, M. D. Theys, H. J. Siegel, N. B. Beck, and M. Jurczyk. "A Mathematical Model, Heuristic, and Simulation Study for a Basic Data Staging Problem in a Heterogeneous Networking Environment," In Proceedings of the 7th International Computing Workshop (HCW'98). IEEE, 1998.

[28] R. Douglass, J. Mork, and B. Suresh. Battleeld "Awareness and Data Dissemination (badd) for the Warghter," In Proceedings of the SPIE, volume 3080, pages 18-24, 1997.

[29] R. Lindell, J. Bannister, C. DeMatteis, M. O'Brien, J. Stepanek, M. Campbell, and F. Bauer. "Deploying Internet Services Over a Direct Broadcast Satellite Network: Challenges and Opportunities in the Global Broadcast Service," In MILCOM. IEEE, 1997.

[30] T. Stephenson, B. DeCleene, G. Speckert, and H. Voorhees. Badd phase ii. Dds "Information Management Architecture," In Proceedings of the SPIE, volume 3080, pages 49-58, 1997.

[31] H. Salkin & J. Saha. "Set Covering: Algorithms, Results and Codes," In Bulletin of the Operations Research Society of America, volume 20, suppl.2, Nov 1972.

[32]. J. Dukes-Schlossberg, Y. Lee, and N. Lehrer. "Iids: Intelligent Information Dissemination Server," In MILCOM 97 Proceedings, volume 2, pages 635-639. IEEE, 1997.

Vivek N. Waghmare completed his B. Tech., M. Tech. from SGGS Institute of Engineering & Technology, Nanded, and Walchand College of Engineering, Sangli, India in 2008 and 2010 respectively. He is currently pursuing his Ph. D. at SGGS Institute of Engineering & Technology, Nanded under SRTMUN, Nanded, India. His research area includes HPC and Network Security.

Dr. Ravindra C. Thool received his BE, ME and Ph.D. in Electronics from SGGS Institute of Engineering & Technology, Nanded, India, in 1986, 1991 and 2003 respectively. He is currently working as professor and head with Information Technology department in the same organization. His research area includes Computer Vision, Image processing and multimedia information systems. He has published several research papers in refereed journals and professional conference proceedings. He is member of IEEE, Life member of ISTE.

Accurate Image Search using Local Descriptors into a Compact Image Representation

Soumia Benkrama[1], Lynda Zaoui[2] and Christophe Charrier[3]

[1] University of Science and Techenology Mohamed Boudiaf, Department of Computer Science, Laboratory Systems Signals Data Oran, Algeria

[2] University of Science and Techenology Mohamed Boudiaf, Department of Computer Science, Laboratory Systems Signals Data Oran, Algeria

[3] University of Caen-Basse Normandie, GREYC, UMR CNRS 6072 Caen, France

Abstract

Progress in image retrieval by using low-level features, such as colors, textures and shapes, the performance is still unsatisfied as there are existing gaps between low-level features and high-level semantic concepts.

In this work, we present an improved implementation for the bag of visual words approach. We propose a image retrieval system based on bag-of-features (BoF) model by using scale invariant feature transform (SIFT) and speeded up robust features (SURF). In literature SIFT and SURF give of good results. Based on this observation, we decide to use a bag-of-features approach over quaternion zernike moments (QZM). We compare the results of SIFT and SURF with those of QZM.

We propose an indexing method for content based search task that aims to retrieve collection of images and returns a ranked list of objects in response to a query image. Experimental results with the Coil-100 and corel-1000 image database, demonstrate that QZM produces a better performance than known representations (SIFT and SURF).

Keywords: *Content-Based Image Retrieval Systems, feature detection, Bag of visual words.*

1. Introduction

Content-based image retrieval (CBIR) is a long standing challenging problem in computer vision and multimedia, seek to represent the content of images automatically, using visual descriptors representing the multimedia data.

CBIR system views the query image and the images in the database as a collection of features, and ranks the relevance between the query and any matching image in proportion to a similarity measure calculated from the features. These features, or signatures of images, characterize the content of images; the key idea in the image search is an approximate search by using concept of proximity, similarity, and distance between objects [1]. The similarity measure is often based on the calculation of a distance in the feature space, one then seek the nearest neighbors of the query [1].

Recent works on Content Based Image Retrieval rely on Bag of Visual Words (BoVW) to index images. In BoVW, local features are extracted from the whole image dataset and quantized (termed as visual words). For compact representation, a visual vocabulary is usually constructed to describe BoF through the clustering of keypoint features. Each keypoint cluster is treated as a "visual word" in the visual vocabulary. This approach employs histogram based features for image representation.

In this paper, we address the problem of searching the most similar images in image database. We put an emphasis on the joint optimization of three constraints: storage, computational cost, and recognition performance.

2. State of the art

Many different approaches for CBIR have been proposed in the literature. Swain and Ballard [2] were the first to use color histograms features to describe images. Since, many other introduced other features like texture or colorimetric moments. These descriptors allow a quite efficient retrieval in many cases, but fail in precision, because global features lose most of local information expressed in the image. Recent approaches propose to use local features to describe interest regions in the image.

The idea is to detect interesting local patches, represent the patches as numerical vectors and consider image which allows comparing images by measuring the similarity between signatures.

Many approaches have been proposed to extract local features from images. In [4] and [5] the authors extract local patches using a regular grid. Other authors use also random sampling [6], [7] and segmentation methods. A more interesting approach is to extract keypoints.

The most popular approach today, initially proposed in [3], relies on a bag-of-features (BOF) representation of the image. The idea is to quantize local invariant descriptors.

Bag of features

This model allows describing an image as bag of elementary local features called visual words. As a result, an image is represented by a vector of weights, where weight corresponds to the importance of visual word in the image. The choice of local features and the weighting schema are very important to perform image retrieval.
Recent approaches that address the problem of indexing techniques and image search, trying to find areas in images that contain visual information for robust visual variations. In particular, the extraction and description of the regions of interest are successfully applied to detect these areas.

Because of the success of BoF approaches for image retrieval, several authors propose extensions of BoVF representations. [16] propose the random locality sensitive vocabulary (RLSV) scheme towards visual vocabulary construction in such scenarios, this method is not advantageous in terms of query complexity. A several range of methods has been proposed to incorporate spatial information to improve the BoVW model [17, 18]. [19] [19] propose a image retrieval system based on bag-of-features (BoF) model by integrating scale invariant feature transform (SIFT) and local binary pattern (LBP). Using a weighted Kmeans clustering algorithm, the image-based SIFT-LBP integration achieves the superior performance on a given benchmark problem. Most of them aimed at human actions recognition.
Current techniques are based on the extraction of many local information like SIFT [8] or SURF [9] in that produce average over 1000 points per image. Each of these points is then characterized by a dictionary of visual (visual pattern characteristic of a portion of the image).
Each image is thus represented by a "bag of visual words". Indexing techniques involve comparing each visual word in an image with all the words that exist in database. However, the computational cost of such techniques makes them difficult to use in an environment where many images are available. The method bag of words is to quantify the local descriptors calculated on regions of interest, which are previously extracted by a detector affine invariant. The quantization indices of these descriptors are called visual words, by analogy with the search of textual records. The image is represented by a histogram of the frequency of occurrence of visual words, this representation allows efficient computation of similarity between images.

The steps for the construction of the bags of words are as follows: (i) the identification or extraction of information in the image, (ii) quantifying with matching characteristics with local visual words, (iii) creating a histogram of frequency used to evaluate a global representation of the image. So using the method of word bag, it is possible to reduce the size of the vectors descriptors of images, while providing a compact representation of the images.
The BOF representation groups local descriptors. It requires the definition of a codebook of k “visual words” usually obtained by *k*-means clustering [10]. Each local descriptor of dimension *d* from an image is assigned to the closest centroid. The BOF representation is obtained as the histogram of the assignment of all image descriptors to visual words. Therefore, it produces a *k* dimensional vector, which is subsequently normalized.
There are several variations on how to normalize the histogram. When seen as an empirical distribution, the BOF vector is normalized using the Manhattan distance. Another common choice consists in using Euclidean normalization.
Several variations have been proposed to improve the quality of this representation. One of the most popular [11, 12] consists in using soft quantization techniques instead of a k-means.
The main advantages of the BOF representation are 1) its compactness, i.e., reduced storage requirements and 2) the rapidity of search. [13] propose to use more visual words in the BoW algorithm, and showed that using multiple independent dictionaries built from different subsets of the features increases significantly the recognition performance of BoW systems. [14] propose the structural features for object recognition are nested multi-layered local graphs built upon sets of SURF feature points with Delaunay triangulation. This representation conserves the invariance to affine transformations of image plane which the initial SIFT/SURF features have. A Bag-of-Visual-Words framework is applied on these graphs, giving birth to a Bag-of-Graph-Words representation. For each layer of graphs its own visual dictionary is built.

Our contribution consists in proposing a representation that provides excellent search accuracy with a reasonable vector dimensionality. We propose three descriptors, derived from both BOF, to produce a compact representation.

3. Proposed approach

This article addresses precise image search based on local descriptors. A variety of feature detection algorithms have been proposed in the literature to compute reliable descriptors for image matching [8], [9]. SIFT and SURF descriptors are the most promising due to good performance and have now been used in many applications. [20] summarize the performance of two robust feature detection algorithms namely Scale Invariant Feature Transform (SIFT) and Speeded up Robust Features (SURF) on several classification datasets. Based on this observation, we decide to use a bag-of-features approach over quaternion zernike moments (QZM). We compare the results of SIFT and SURF with those of QZM.

SIFT, SURF and QZM features are used because reasonably invariant to changes in illumination, image noise, rotation, scaling, and small changes in viewpoint. We propose an indexing method for content based search task that aims to retrieve a large collection of images and returns a ranked list of objects in response to a query image. The visual words quantize the space of descriptors. Here, we use the k-means algorithm to obtain the visual vocabulary. For each query image, the signature is calculated from these points can be compared over the signature of the query image with all signatures from the base.

Algorithm

1. Extract local features $\{d_n\}$ from every image i.
2. Constructing codebooks: randomly selecting centres from among the sampled training patches, and online k-means initialized using this.
3. The dictionaries are built using Approximate K-Means. Histograms are normalized to have unit l_1 norm, then the l_1 distance is used to measure similarity between histograms.

In the conventional bag-of-visual-words model, at first each image I is represented in terms of image descriptors:
$I = \{d_1, d_2, \ldots, d_n\}$, where d_i is the description of an image patch and n is the total number of patches in the image. By this way, we get numerous descriptors from all the local patches of all the images for a given dataset. Typically, K-means unsupervised clustering is applied on these descriptors to find clusters $W = \{w_1, w_2, \ldots, w_N\}$, that constitutes the visual vocabulary, where N is the predefined number of clusters.
Figure 1 illustrates how each of these operations impacts our representation.

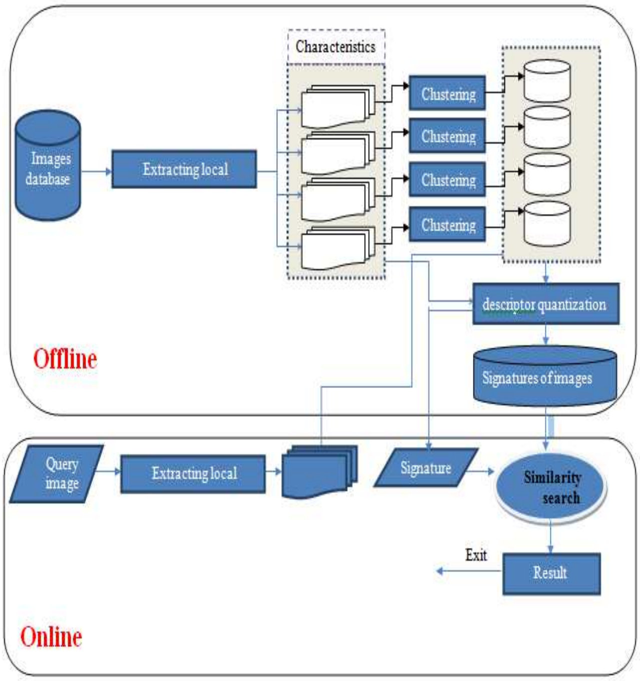


Fig. 1 The architecture of the approach suggested

3.1 Experiments

In this section, we first evaluate our descriptors and the joint dimensionality reduction.

3.1.1 Datasets and Evaluation

Two datasets have been used to perform the k-means clustering: the COIL-100 set, as well as the Corel dataset.
Columbia Object Image Library (COIL-100) is a database of color images of 100 objects. The objects were placed on a motorized turntable against a black background. The turntable was rotated through 360 degrees to vary object pose with respect to a fixed color camera. Images of the objects were taken at pose intervals of 5 degrees. Figure 2 shows The COIL-100.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Fig. 2 Samples of Coil-100 dataset. The dataset includes 128x128.

COREL database composed of 1000 images distributed in 10 classes. The tags of the classes are: Africans, Beach, Architecture, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains and Food. Figure 3 shows one sample per each class.

Fig. 3 Samples of Corel 1000 dataset. The dataset includes 256x384 or 384x256 images.

The objective of this work is focused on finding images, and not detection. SIFT and SURF are two recent and competitive alternatives to image local featuring that we compare through QZM.
To do this, a set of points of interest is extracted from each descriptor, SIFT (128 dimensions), SURF (64 dimensions), and QZM (34 dimensions), it is mandatory for the construction of the visual dictionary, using the *K*-means algorithm applied to the associated descriptors. Each corresponds to a visual word class center.
For each query image, the signature is calculated from these points than we can compare the signature of the query image with all signatures from the database, then the user can send the images most similar to his query. In a retrieval system images, the user is interested in answers relevant system. So the image search systems require the evaluation of the accuracy of the response. This type of evaluation is considered performance evaluation research. We describe the two most common measures: recall and precision. This relationship is often described by a curve of recall and precision.

3.1.2 Results and analysis

In this section, we present some results, giving details of the curves - reminder for each descriptor.
Indeed it has been demonstrated in the literature, these methods currently give the best results.
The approach is based on an unsupervised classification (K-means) that we use to build the dictionary. The comparison result by varying the number of classes (K) to obtain the visual vocabulary. We note that for large dictionaries, the approach provides better results.

Regarding the descriptors according to their performance varis bases and therefore the application.
The mosr common evaluation measures used in CBIR are precision and recall, usually presented as a precision vs. recall.

$$\text{Precision} = \frac{\text{Number.relevant images retrieved}}{\text{Total number.images retrieved}} \quad (1)$$

$$\text{Recall} = \frac{\text{Number relevant images retrieved}}{\text{Total relevant images in the collection}} \quad (2)$$

of images of the class that the target image belongs to. In literature SIFT and SURF give of good results, we have to compare this with QZM.
From these results, we can calculate the recall/ precision curve:

Table: Impact of the dataset used for k-means clustering (uncorrelated Corel dataset or the Coil-100) and of the vocabulary size.

Descriptors	Image database	Vocabulary size	Recall	Precision
SIFT	Coil	10	30%	60%
	Coil	50	40%	80%
	Corel	30	25%	48%
	Corel	50	35%	70%
SURF	Coil	10	20%	40%
	Coil	50	30%	60%
	Corel	30	30%	45%
	Corel	50	40%	60%

QZM	Coil Coil Corel Corel	10 50 30 50	50% 65% 35% 40%	100% 100% 70% 80%

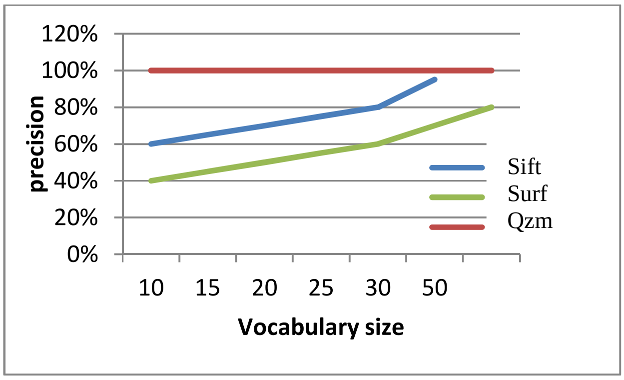


Fig. 4 Comparing results for SURF-based, SIFT-based and QZM- based approaches at various vocabulary sizes for coil-100

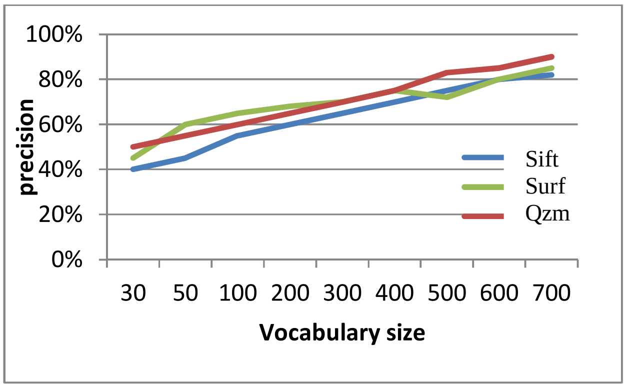


Fig. 5 Comparing results for SURF-based, SIFT-based and QZM- based approaches at various vocabulary sizes for corel-1000

3.1.3 Discussion

In this paper we have reported a analysis. We explored ways to boost the performance of BoW image search methods by using more visual words. We presented algorithm, QZM. Experimental results with Coil-100 and Corel image databases, demonstrate that the method QZM produces better performance.

4. Conclusion

The method allowed bags of words obtaining a compact representation of images, so that from a set of descriptors for each image, we obtain a single feature vector for each image, then we could therefore reduce the size of the search space using clustering, and the size of the descriptors (signatures) with an index structure that will limit the number of images to check to accelerate research (early treatment the query).

Furthermore, we develop descriptor QZM based on the bag-of-features approach and use the benchmark to demonstrate that they significantly outperform other descriptors in the literature such as SIFT and SURF.

The proposed approach is based on a reduction in the size of the feature vectors. The originality consists in an adaptation of a bag of words representation of the use of descriptors point of interest SIFT, SURF and QZM.

References

[1] I. Daoudi, K. Idrissi, et S. E. Ouatik, "A multi-class metric learning for contentbased image retrieval," IEEE international Conference on Image processing, ICIP 2009.
[2] M. J. Swain and D. H. Ballard, "Color Indexing," International Journal of Computer Vision, pp. 11-32, 1991.
[3] J. Sivic and A. Zisserman. "Video Google: A text retrieval approach to object matching in videos," In ICCV, pages 1470– 1477, 2003.
[4] L. Fei-Fei and P.Perona. "A Bayesian hierarchical model for learning natural scene categories," Proceedings of IEEE International Conference Computer Vision and Pattern Recognition, pp. 524- 531, 2005.
[5] J. Vogel and B. Schiele, "On Performance Characterization and Optimization for Image Retrieval," Proceedings of European Conference on Computer Vision, pp. 51-55, 2002.
[6] M. Vidal-Naquet and S.Ullman. "Object recognition with informative features and linear classification," Proceedings of IEEE International Conference Computer Vision, pp. 281-288, 2003.
[7] R. Maree, P. Geurts, J. Piater, J. And L. Wehenkel, "Random subwindows for robust image classification," Proceedings of IEEE International Conference Computer Vision and Pattern Recognition, pp. 34-40, 2005.
[8] David G. Lowe. "Object recognition from local scale invariant features," In ICCV, pages 1150–1157, 1999.
[9] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc J. Van Gool. "Speeded-up robust features (surf),". Computer Vision and Image Understanding, 110(3) :346– 359, 2008.
[10] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," CVPR, 2007.
[11] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. "Lost in quantization: Improving particular object retrieval in large scale image databases," In CVPR, June 2008.
[12] J. van Gemert, C. Veenman, A. Smeulders, and J. Geusebroek. "Visual word ambiguity,". PAMI. 2009
[13] Mohamed Aly 1 Mario Munich 2 Pietro Perona 1 "MULTIPLE DICTIONARIES FOR BAG OF WORDS LARGE SCALE IMAGE SEARCH," IEEE International Conference on Image Processing (ICIP), Brussels, Belgium, September 2011.
[14] Svebor Karaman, Jenny Benois-Pineau, Rémi Mégret and Aurélie Bugeau. " Mots visuels issus de graphes locaux

multi-niveaux pour la reconnaissance d'objets," RFIA (Reconnaissance des Formes et Intelligence Artificielle). 2012.
[15] H. Mahi, N. Benkablia, " Moments de Zernike Quaternioniques pour la Classification des Bâtiments sur des données à Très Haute Résolution Spatiale," Proceedings of International Conference on image processing, Avril 2012.
[16] Y.Mu, J. Sun, Tony X. Han, Loong-Fah Cheong, Shuicheng Yan, "Randomized Locality Sensitive Vocabularies for Bag-of-Features Model," The 11th European Conference on Computer Vision (ECCV), 2010
[17] S.Kim, X. Jin, and J. Han. "Disiclass: discriminative frequent patternbased image classification," In Proceedings of the Tenth International Workshop on Multimedia Data Mining, MDMKDD'10, pages 7:1–7:10, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0220-3.
[18] D. Liu, G. Hua, Paul A. Viola, and Tsuhan Chen. "Integrated feature selection and higher-order spatial feature extraction for object categorization,". In CVPR, 2008.
[19] X. Yuan, Jing Yu, Z. Qiny , and T.Wan "A SIFT-LBP IMAGE RETRIEVAL MODEL BASED ON BAG-OF-FEATURES," 18th IEEE International Conference on Image Processing. 2011.
[20] N. Younus Khan, B. McCane *and* G. Wyvill. "SIFT and SURF Performance Evaluation Against Various Image Deformations on Benchmark Dataset," International Conference on Digital Image Computing: Techniques and Applications. 2011.

Soumia BENKRAMA
PhD Student at the University of Sciences and Technology of Oran

Lynda ZAOUI
HdR at the University of Sciences and Technology of Oran

Christophe CHARRIER
HdR at the University of Caen Basse-Normadie

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Exploration of Success Factors of Information System

Bikram Pal Kaur[1] and Himanshu Aggrawal[2]

[1] Ph.D Scholar, Department Of Computer Science Engineering,Punjabi University Patiala, 147002,Punjab, India

[2] Professor, Department Of Computer Science Engineering,Punjabi University Patiala, 147002,Punjab, India

Abstract

In current decades, the organization is highly effected by the successful implementation of Information system (IS). In today's business there are also no boundaries and hence the joint venture between customers and challengers are usual. IS provides increased competitiveness and flexibility to business. It also affects the trade relationships in the market. Whenever IS is properly implemented ,it facilitates business redesigning process, supports an organization to offer good products, and services. It also enables organizations to participate in new advertising programmes and initiate operational effectiveness. However the better performance of the organization depends upon the successful implementation of IS. Therefore the objective of this paper is to review the success factors of Information system on the basis of literature survey and proposing the new model for the critical success factors for the Information Systems in an organization.
.

Keywords: *Information System,Critical Success Factors,IS Model, ERP,IS Life Cycle.*

1. Introduction

This research paper highlights the previous research work and studies conducted by different researchers on Critical Success Factors (CSFs) of Information System. This paper presents the holistic observation of all the researches done earlier. The literature review has been done to collect the overall observations of the subject under study. In this paper the development related to IS systems relevant for the study of the critical success factors have been highlighted.

The literature review is an analytical and in-depth evaluation of the researches done earlier. The information has been collected from various sources which are further documented. It also helps in recognizing the gap that exists in the area of research. For justification of the literature review, the literature has been classified into different sections.

2. Information System: Basic Concept

The studies of Yeo[1] suggests that IS are user-interfaced systems which are required for supplying the information and also for processing the information to support the strategies ,different functions of the various departments as well as ,decision making processing of the management in the business dealings.

The studies done by Delone & McLean's[2] fig 2. represents the three most important dimensions: Service quality, Information quality, System quality which must be monitored for the user satisfaction. Earlier also in Delone & McLeon[3] proposed this model but it was lacking in showing the benefits of services and support for the e-commerce success in business(see Figure 1).

Kumar et al.[4] considered certain management related issues regarding implementation of Information System like vendor selection of IS, collaborators in implementation process, project team structure and their training.

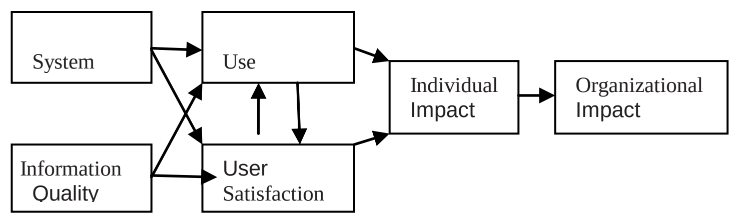


Figure 1 DeLone and McLean IS success model (1992).

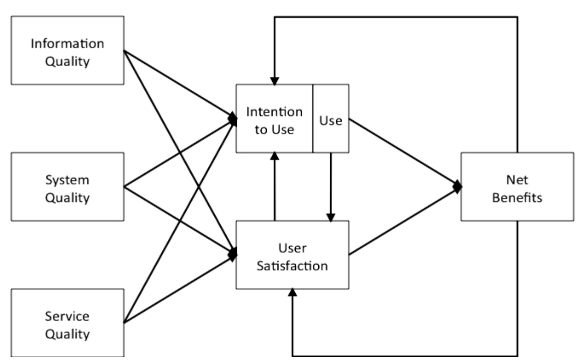


Figure 2: Depiction of the Updated ISs Success Model (DeLone & McLean 2003)

Alter[5] elaborates that the organization in which the information systems are implemented are multifaceted, synthetic, and reason fully designed. The information system is composed of people, structure, technologies, and work systems; The IS developers are concerned with the design which is according to the goal of the organization. Figure3 shows the alignments between business and information technology strategies. It also depicts the alignment between organizational and information systems infrastructures Henderson and Venkatraman[6] . The business strategy conversion into successful Information System Infrastructure requires a lot much design processes like organizational design for the formation of organization strategy and information system design for the successful information system infrastructure.

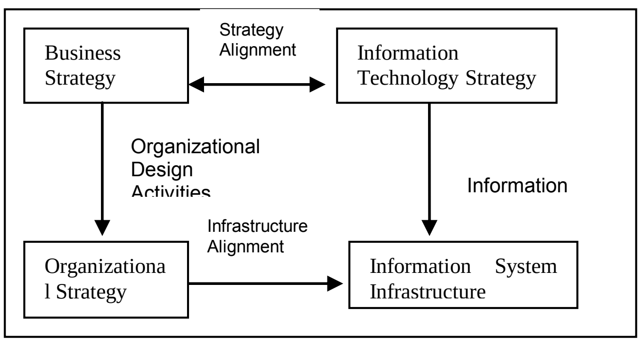


Figure 3. Organizational Design and Information Systems Design Activities (Adapted from J. Henderson and N. Venkatraman), 1993.)

Allen S. Lee[7] elaborates the framework for the evaluation of IS in which the combination of behavioral sciences and design science has been shown in Fig 4.

Shehab et al[8] illustrates that to maintain the competitive edge in the market and also for increasing the efficiency and effectiveness of business process the IS (ERP) is the best enable.

Turban et al.[9] reviews yearly Datamation (a leading practitioner journal of information systems) and then suggests why IS are important for a business organization.
The information systems are required for the following reasons:

• For the business process reengineering.
• To meet the company's goals and objectives.
• Better decision making.
• For the development of the productivity.
• Enhancing the quality of the product.
• Building the competitive edge.
• Retention of change management environment.
• Creation of Research and Innovation environment.

According to Turban et al. the IS projects can be classified into four categories:
a) Commercial e.g. customer relationship management (CRM),e-commerce, knowledge management
b)Strategic e.g. re-engineering, information architecture
c)Organizational e.g. centralization vs. decentralization, outsourcing, resource management;
d) Technological e.g. database, internet and intranet.

Diniz [10],proposed a three dimensional model for the evaluation of virtual business environments from the user's perspective by doing the case study of three banks in a Brazil. The studies includes the services offered, functionality, reliability, security of transactions on the sites and also the users transaction quality. This evaluation approach is useful to known the quality of the sites used for Internet banking .

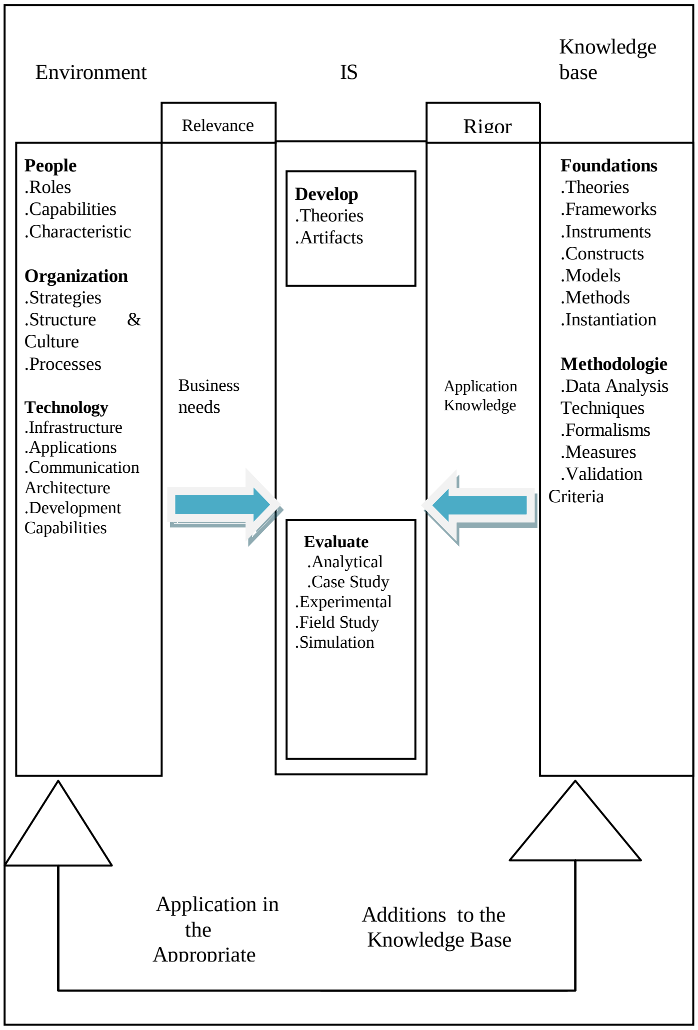


Figure4 IS research framework Hevner et al.

Khaddaj [11] suggests the software quality factor have a great impact on the information system performance and should be taken care of large information systems. Khaddaj exposed, many elements which leads to the failure of system's performance. Usability and portability are major problems that need to be considered which affects software quality.

Almutairi & Subramanian [12], supports through their pragmatic application of Delone & McLean model in private sector organization of Kuwait, find the direct linkage between the variables of Delone & McLean modal. They proved that information quality and system quality influences user satisfaction significantly. Similarly system usage has a great control on in individual impact

Wixon & Todd [13] formed an integrated model which differentiate beliefs and attitude about the system from belief and attitudes using system (behavior based).He built a theoretical logic which links the user satisfaction and technology acceptance. This model bridges the gap between the system characteristics and its usage as shown in Figure.5.

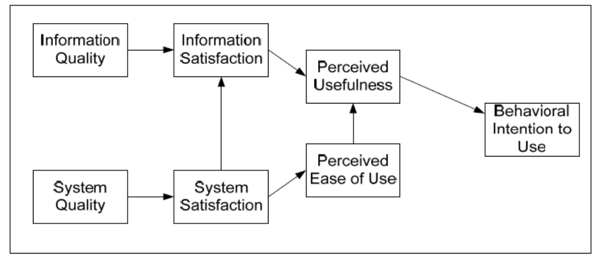


Figure 5: Integrating System Characteristics to TAM, (Wixon & Todd 2005).

Laudon & Laudon [13], describes factors which are highly enhancing the importance of ISs ;

- the extreme use of internet and communication technologies,
- Technology forces renovates of business enterprise
- Globalization of business
- Big prospects on the international level
- Growth of information and knowledge market,
- Appearance of the digital firm leading to business relationships digitally facilitated

Al-adaileh, [14] examines the use of IS not only meant for gaining competitive advantage but also attain valuable benefits for an organization including,

- Raising productivity,
- Small product cycle,
- Automation of operational decision
- Supporting of strategic and tactical decisions

Sørensen et al.,[15] indicates through the Figure 6 ,the decomposition of the management system of an organization. To enhance the profits and low operating cost to the management can only be achieved through a well-constructed and well-organized IS . IS provides the information about operations for the rational decision making which helps in increasing effectiveness and efficiency.

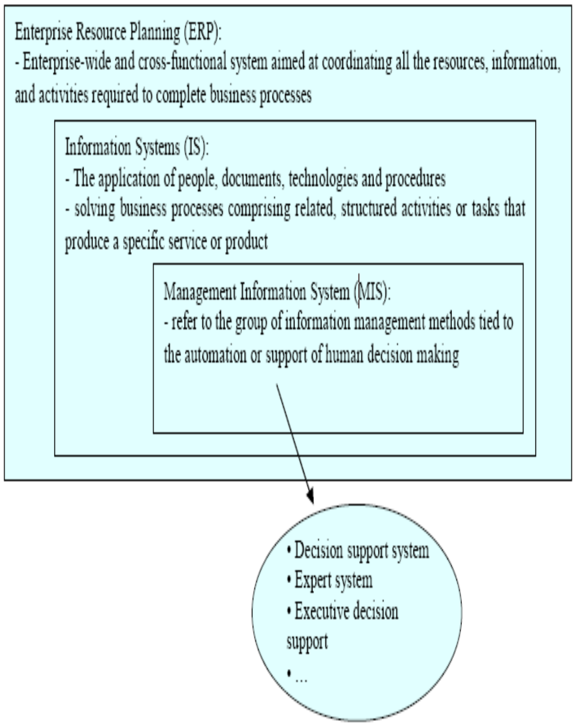


Figure 6. Concept of management information systems. [Sørensen, C., Bildsøe, P., Fountas, S.]

Bechor et al [16], suggests that the strategically planning of IS is must. Strategically planning is the process of strategic thinking that identifies the most desirable IS on which the organization can implement its long-lasting IS activities and policies.

Naser Khani, Khalil Md Nor[17] explains that IS capability is related to maintaining competitive advantage in IS field. Information Technology's(IT) competence is in delivering the best IS system, manage IT costs, and shaping goals of the business .In this study the model showing relationship between IS capabilities and strategic planning for IS for the effective IS has been developed. Depending on RBV(Resource Based View) perspective the variation between two variables were studied i.e. IS capabilities and SISP(Strategic Information System Planning) success.

Analysis

IS is the business element required for the future and success of the business. The successful managers are those who use IS so that their business communicates rapidly and effectively over a global network. IS is the foundation stone of this capability. The key reasons why IS are required for the current business processes:

- IS is a competitive weapon
- The line manager take charge of IS
- Information literacy is must
- Pragmatic planning is must.
- Recognition of the Human Resource Factor
- Third Party alliances
- Executive support systems that aid decision making process
- Business reengineering is required

The beneficiaries of these IS system are the top management, IT industry and the operational management. Therefore IS is a thriving force of the current business scenario.

3. IS :Life Cycle

Providing effective information technology (IT) support or business processes has become crucial for enterprises to stay competitive in their market. Thus, process life cycle support W. M. P. Vander Aalst [18] and continuous process improvement adopt a key role in contemporary and future enterprise computing. The process life cycle (Figure 7) starts with the (re)design of a business process. Process modeling and process analysis tools can be used during this phase. Thereafter, the business process has to be implemented resulting in a process-oriented IS. Organizations constantly search for ways to grow and maintain their competitive edge and today's business activities depend greatly on information technology (IT) enablement. This demands that IT maintenance is regarded as a critical process that needs to be performed with the highest possible quality. It requires organization specific knowledge about internal business operations and explicit technical knowledge to achieve the desired IT quality.

Hanlie Smuts,Alta van der Merwe[19] The survey consisted of 12 constructs

• Construct 1 : requirement definition -It includes business and functional requirements, operational processes and procedures and performance requirements.

Construct 2: design made up of solution architecture, reporting solution design, IS operations process modelling and IS infrastructure and environments.

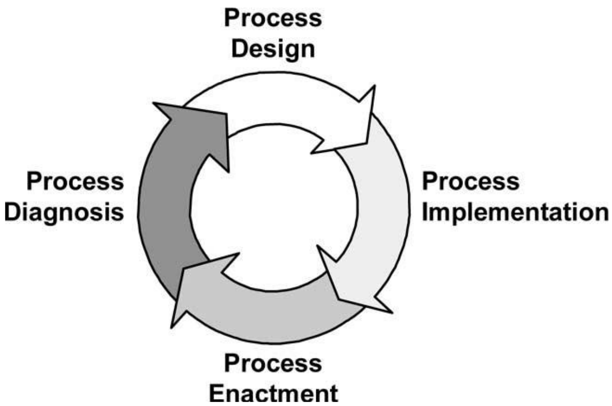


Figure7.Process Life Cycle

• Construct 3 Programming and testing refers to solution and reporting solution development, legacy application development, release management and solution end to end integration.

• Construct 4 is integration and system test and includes business readiness testing, parallel run, user acceptance testing and systems integration testing.

• Construct 5 is Data migration is included data migration planning, data migration testing and data migration execution.

• Construct 6 is cutover and transition and cover transition planning for project to operations, cut over planning, cut over testing and cut over execution.

• Construct 7 is business integration and focuses on end user training, behavioral change management and communication.

• Construct 8 is project control and management and includes scheduling / planning, progress measurement and reporting, roles and responsibilities definition, knowledge management, leadership and project management.

• Construct 9 Governance and sponsorship is and comprises of client organization decision-making, client organization sponsorship, outsources vendor's sponsorship, governance structures and forums.
• Construct 10 is commercial management and points to commercial agreements and administration thereof, as well as value add to delivery milestones.

• Construct 11 is auditing and risk management and refers to macro project risk management, data transformation risk management and internal controls.

• The last construct, 12, is relationships between project team and outsource vendors, between project team and client organization and among outsource vendors.

Analysis

From the above discussion on the life cycle stages of IS, it is concluded that for successful and effective IS implementation, the IS of organization must go through the following phases as shown in diagram:

- Planning process
- Analysis & Design process
- Implementation process
- Testing process
- Stabilization process
- Continuous Updation process
- Evaluation process

4. Information System : Integrated IS Model

It can be concluded from literature discussed above that the information life cycle are important for of managing the corporate application assortment. The approach presented here does not correspond just to the extension of the traditional information system development life cycle. This approach is based in the generic lifecycle employed in other contexts like manufacturing or marketing. In this paper it is proposed a model of an information system life cycle,
supported in the assumption that a system has a limited life. But, this limited life may be extended. This model is also applied in several cases; being reported here two examples of the framework application in a construction enterprise, and in a manufacturing enterprise. this paper is presenting a framework based in the information system life cycle that could be effective to analyze information systems and information technology adopted by a specific enterprise. Although we admit that this framework could be used in other contexts, our purpose is analyzing its effectiveness in corporations whose business is in manufacturing and construction industries.

Information System Life Cycle -----------different phases

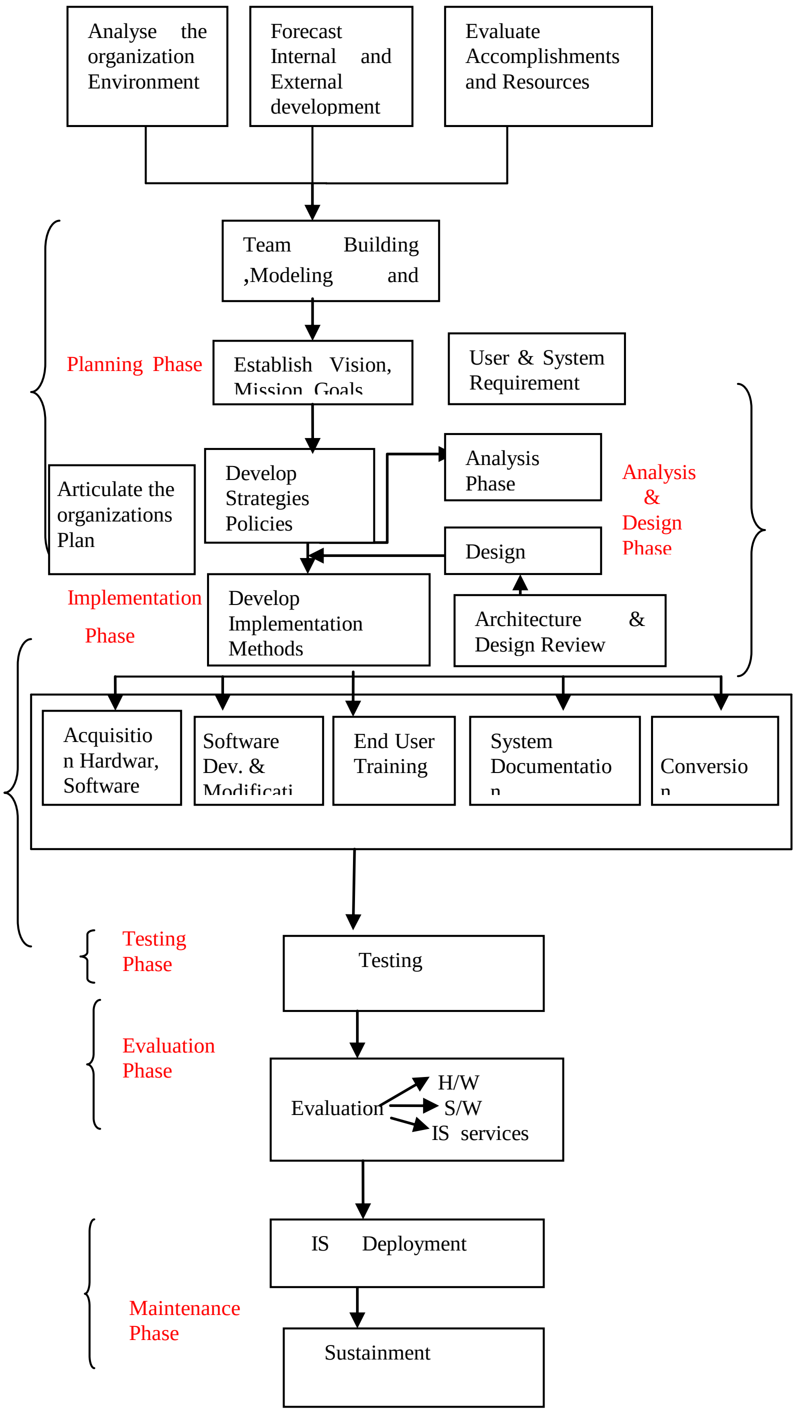


Figure8 Information System Life Cycle

Successful planning and implementation is the obvious goal of any organization that has chosen to go in for IS. Planning is very important as well as difficult. So, it is valuable to have a framework to act as guideline for planning process. A large number of models have been suggested and practiced for the Information Systems. An integrated model for successful IS is developed and shown in the Figure.

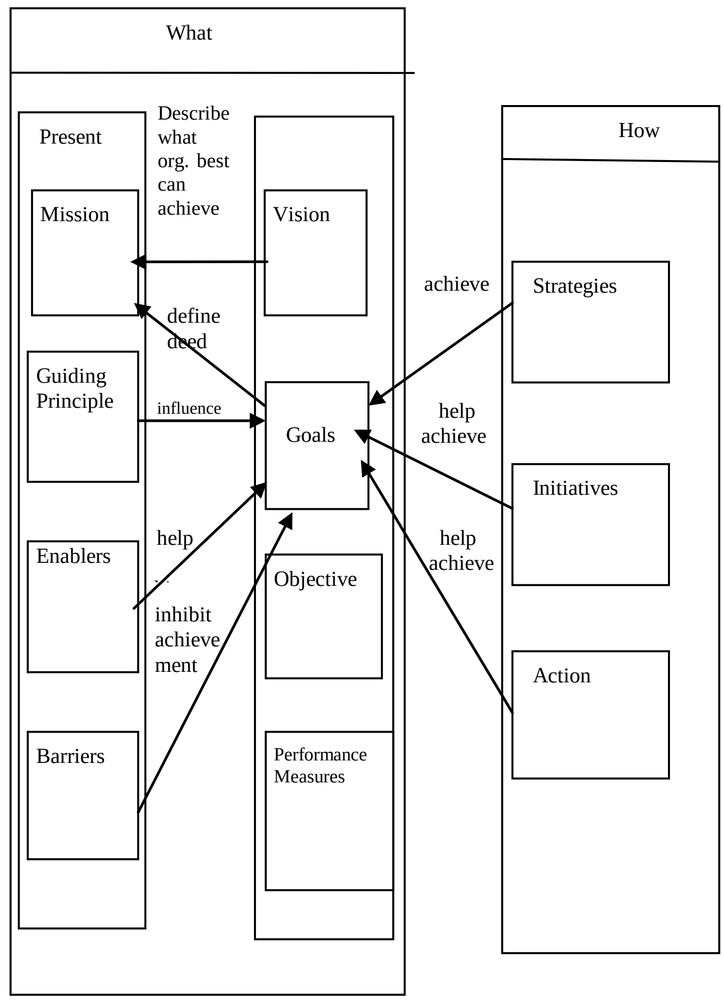


Figure9 An integrated model for successful IS

Typical strategic planning processes examine an organization's current environment and abilities (the present situation), considerations about how it would like to grow or evolve (the desired future), its aspirations as an organization (what it will strive to do), and its intentions for moving forward (how it will move forward) as shown in figure 9. These high-level elements are described below:

• The What: These are descriptions of what the organization does and what it aspires to achieve—its organizational targets—including its goals, objectives, and quantitative performance measures.

• The Present: The present situation, or current environment, is typically described in terms of the organization's mission, guiding principles (or values), organizational strengths (or enablers), and organizational barriers (weaknesses or challenges).

• The Future: The desired future is described by the organizational vision and targets.

• The How: The preferred route to achieving the organizational goals, objectives, and mission is communicated as a strategy or as strategic goals.6

The studies done by Finch[20] indicates the project implementation profile (PIP) framework can be utilized for supporting future project's success in terms of defining

- Project mission,
- Client consultation ,
- Schedule/plan ,
- Top management support ,
- Personnel Client acceptance,
- Technical tasks ,
- Monitoring and feedback,
- Communication,
- Trouble-shooting . -

Finch & Olson[21] indicates that the achievement of planning, budgetary, and business goals are the motives of the successful implementation of IS projects. The study has indicated that the successful IS projects has seven characteristics also known as 'critical success factors' (CSFs);within time limit, within the forecasted budget, alignment with the business performance, user's acceptance, minimum disordering of the work flow of organization, slightest effect on the business culture.

Belout and Gauvrea[22] proposed the seven items for the IS project implementation. They are

- Completion of project within time limit.
- Stick schedule of the projects followed.
- project must be within budget
- Users feel good about IS project implementation
- All the requirements must be considered
- Cost objectives must be met
- Intended clients use the project

Horine [23] from an optimistic viewpoint reviews a broad qualities and behavior common among those most successful projects. A thriving project must have:
IS alignment with the company objectives.

- Clear defined scope, deliverables and approach in the process of planning.
- Role and responsibility has to be clearly defined to each user and team members.
- Accuracy must be the priority.
- Realistic schedule need to be developed.
- Project team must be customer- focused and outcome concerned.
- Constant, efficient, and alerts must be present on understanding project communications.
- Progression of the project must be measured.
- Project problems and consequent action items must be followed.
- Cultivate a strong teamwork.
- Prospects and changes in surrounding scope, excellence, agenda, and outlay must be managed.
- Availability of experts as project assets as and when required.
- Identification of risk and its management...
- Predict and conquer barriers to guarantee project accomplish objectives.

P. Soja [24] ,study suggest that the success of IS implementation is only possible if the organization is capable of spending money, time and also provides the resources. The risk management is also required to done on the IS project. The user considers IS a failed project if implementation of projects is late and it cost a more.

D. Aloini, R. Dulmin, V. Mininno [25] reveals factors user satisfaction, computer functions, cost, quality of products, time, that establishes the accomplishment of a project.

Bartis and Mitev[26] defines the IS success is a multi-faceted phenomenon and is socially constructed which is affected by stakeholders' viewpoints.

Thomas and Fernández [27] investigates the organizations of Australia and do the measurment of IS project success. There are three categories of success criterion

- Project management success criteria,
- Technical success criteria,
- Business success criteria.

Therefore the finding of the research is that a clear definition of success and effective measurement of the defined criteria is a factor that causes success.

Kanaracus[28]reveals that even in the economic downera, the companies continue to pay outing on information technology (IT) and their budgets continue to climb. Almost every organization uses one criteria from every category. He states that if there is a project management criteria of 'on-time' and 'on budget', the other organization may have another type of category like business success criteria may have 'delivery of benefits', 'met business objectives' and 'business continuity.

De Lone and McLean (D&M Model)[29] predicts that due to the economic downtime and increasing competition ,cutting the cost is required, which further intricate companies to assess and observe the advantages of costs of technology for calculating the ROI(return on these investments). Human factors, organizational, environmental factors are indirectly affecting the impacts of IT; therefore, measurement of information systems (IS) success is both multifaceted and deceptive.

Ika[30] explains the objective of IS success may follow the criteria of time, cost, and quality where as the subjective perceptions are also important for measuring project success. The evaluation of subjectively perceptions may be done by the satisfaction of different stakeholder groups with regard to the project and its results.

Hisham, B. M. B. & Mohd[31] reveals that the strategic information system planning focuses on budget establishment , goal defining ,the selection of the best methodology and deciding internal and external contributors.

Abdel Nasser H. Zaied[32] has created a new model for the evaluation of success of an information system .He applied the two models, one is Technology Acceptance Model (TAM) and other one is DeLone & McLean update IS success model (D&M).

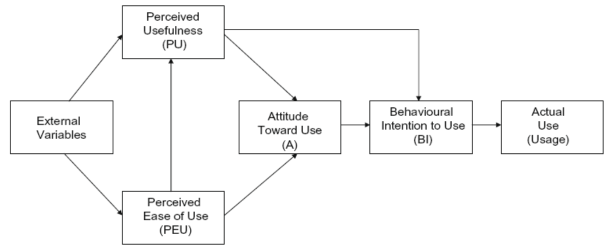


Fig 10: Original Technology Acceptance Model (TAM)

TAM Model & D & M updated IS model were used to generate the integrated evaluation model for IS success. This model is helpful in appraising the critical success factors affecting information systems in public sector in Egypt. This model is useful in supporting the decision makers in evaluating and developing the information systems.

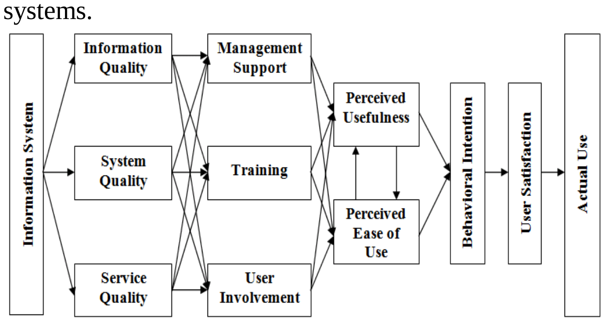


Fig 11: Proposed integrated success model (ISM)

Accordingly, for measuring the IS success ten dimensions were projected .They are

- Service quality;
- System quality;
- Information quality;
- Management support;
- Perceived ease of use;
- Perceived usefulness;
- Training;
- User satisfaction;
- User involvement.
- Behavior intention

The model assumes that information quality, system quality and service quality are linked to management support, training and user involvement, and these in turn; influence perceived usefulness and perceived ease of use which affect on behavior intention and user satisfaction as shown in Figure 11.

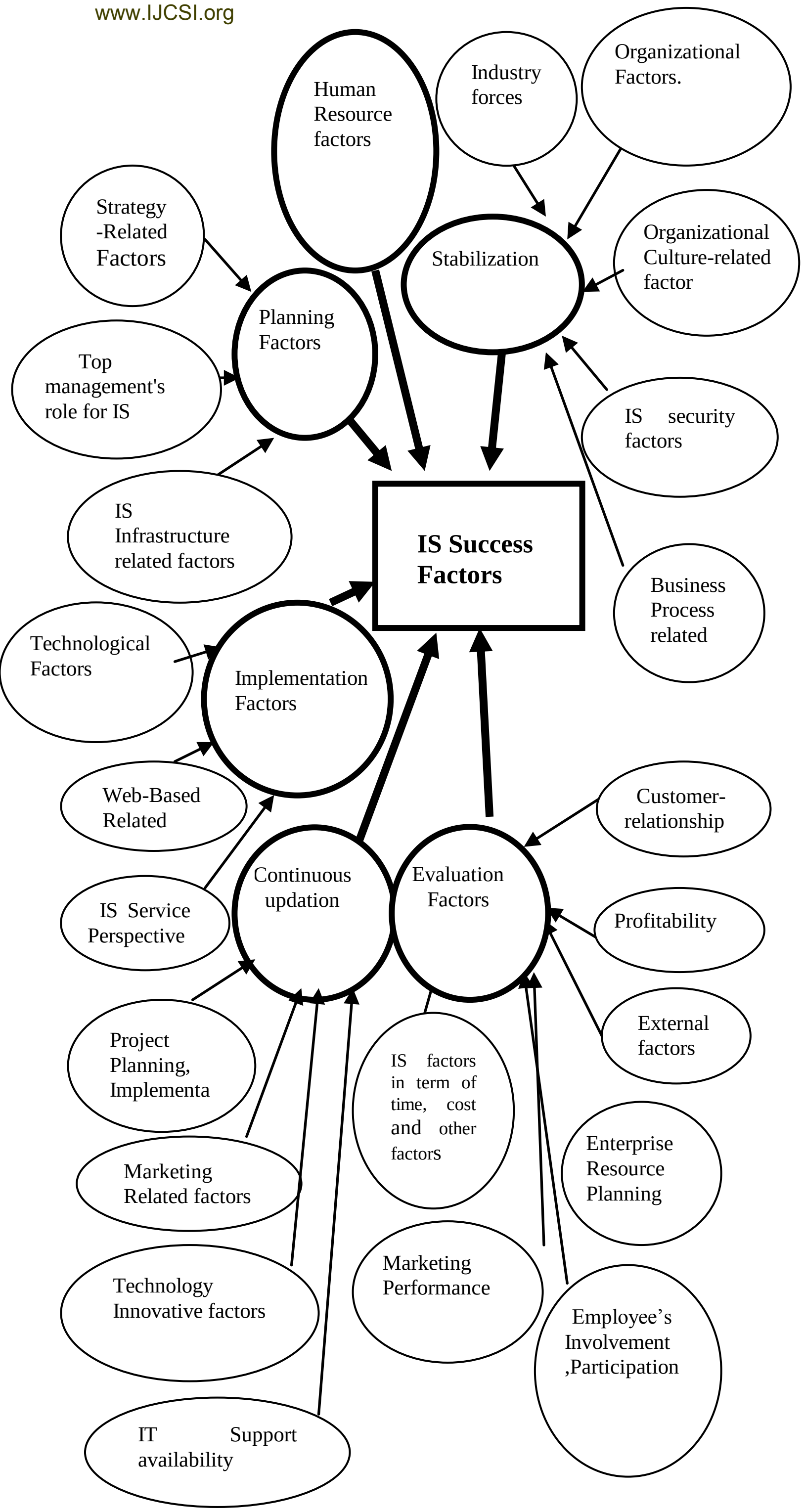


Figure12 Critical Success Factors For Information System

Analysis

The critical success factor within the organization is hierarchical in nature and can be evaluated by many different factors. The organizations are not operating in vacuum however they are operating in a some surroundings, society or in a business. Present business is becoming more competitive and global in a society and industry, therefore certain critical success factors are required to consider for the tomorrow's business. Every organization has critical success factors that must direct individuals in turn department and in turn the whole organization for the IS success. At this stage it is important to understand the scope of critical success factors for IS .The greatest success factors for IS are relating to most the human resource factors then the planning factors then implementation factors, stabilization continuous updation and performance factors . There are certain underlying sub factors as shown in fig 12

References

[1] K.T. Yeo ,Critical failure factors in information system projects,Division of Systems and Engineering Management, School of Mechanical & Production Engineering, Nanyang Technological University,Nanyang Avenue, 639798 Singapore

[2] DELONE WH And MCLEAN ER (2003) The DeLone and McLean model of information systems success: a ten-year update. Journal of Management Information Systems 19(4), 9–30.

[3] DELONE WH and MCLEAN ER (1992) Information systems success: the quest for the dependent variable. Information Systems Research 3(1), 60–95..

[4] V. Kumar, B. Maheshwari, and U. Kumar. "An investigation of critical management issues in ERP implementation: Empirical evidence from Canadian organization". Technovation, 2003.23(9), pp.793-807

[5] Alter, S. (2004): Desperately seeking systems thinking in the information systems discipline,Proceedings of the Twenty-Fifth International Conference on Information Systems, 757-769

[6] J. Henderson and N. Venkatraman, "Strategic Alignment Leveraging Information Technology for Transforming Organizations," IBM Systems Journal(32:1), 1993.

[7] 2004: Mårtensson, P. and Lee, A. S., "Dialogical Action Research at Omega Corporation," MISQuarterly (28:3), pp. 507-536.

[8] E.M. Shehab, M.W. Sharp, L. Supramaniam, T.A. Spedding, (2004) "Enterprise resource planning: An integrative review", Business Process Management Journal, Vol. 10 Iss: 4, pp.359 - 386

[9] Turban, E. & McLean, E. & Wetherbe, J. (2005). Information Technology forManagement: Transforming Organizations in the Digital Economy (5th Ed.). New York:John Wiley & Sons, Inc.

[10] Diniz E, Porto R & Adachi T, September-2005. "Internet Banking in Brazil: Evaluation ofFunctionality, Reliability

and Usability", Electronic Journal of Information SystemsEvaluation, Vol. 8, Issue 1, pp41-50

[11] .Khaddaj, S, 2005. "The Evaluation of Software Quality Factors in Very Large InformationSystems", Electronic Journal of Information Systems Evaluation Vol 8, Issue 1, pp 43-48

[12] Almutairi, H & Subramanian, G, April-2005. "An Empirical Application Of The Delone AndMclean Model In The Kuwaiti Private Sector", Journal Of Computer Information Systems

[13] Laudon, K. & Laudon, J. (2006). Management Information Systems: Managing the Digital Firm. Ninth edition. Pearson/Prentice Hall: Upper Saddle River, NJ .

[14] Raid Moh'd Al-adaileh,An Evaluation of Information Systems Success: A UserPerspective - the Case of Jordan Telecom Group, Assistant Professor of Management Information Systems, Department of Businessand Management Information Systems, Mutah University, B. O. Box (7)Mutah – Al-karak – JordanE-mail: radaileh@mutah.edu.jo

[15] Sørensen, C., Bildsøe, P., Fountas, S.,Pesonen, Pedersen, S., Basso, B., Nash, E.Integration of Farm Management Information Systems to support real-timemanagement decisions and compliance ofmanagement standards. Center for research & technology, Thessaly, Greece. 2009.Available online at: http://www.futurefarm.eu.

[16] Bechor, t., neumann, s., zuiran, m. & glezer, c. 2010. A contingency model forestimating success of strategic information systems planning. Information & Management, 47, 17-29

[17] Khani, N., Md Nor, K., and Bahrami, M. (2011), A Proposed Research Model for the Relationship between Organizational IS capabilities, Organizational Performance and Strategic Planning, 16th IBIMA Conference (Innovation and Knowledge Management: A Global Competitive Advantage) Malaysia 29-30 June 2011.

[18] W. M. P. van der Aalst, "Business process management—A personal view," BPM J., vol. 10, no. 2, pp. 248–253, 2004.

[19] Hevner, A.R., S.T. March, J. Park & S. Ram (2004): Design science in information systems research, *MIS Quarterly*, 28(1), 75-105.

[20] Finch,Peter (2003) Applying the stevin Pinto Project Implementation Profile to an Information Projects,Project MANAGEMENT Journal 34(3)

[21] Olson, D. (2004), Introduction to Information Systems Project Management, 2nd ed.,McGraw-Hill, New York, NY.

[22] Belout, A. and Gauvreau, C. (2004), "Factors influencing project success: the impact of human resource management", International Journal of Project Management, Vol. 22 No. 1,pp. 1-11.

[23] Horine, G. An Absolute Beginner's Guide to Project Management. (2005).].

[24] P.Soja, Success factors in ERP systemsimplementations, Journal of Enterprise Information Management, 19 (2006) 418-433.

[25] D. Aloini, R. Dulmin, V. Mininno, Riskmanagement in ERP project introduction:Review of the literature, Information &Management, 44 (2007) 547-567 .

[26] Bartis, E. and Mitev, N. (2007). The Effect of Strategic Alignment on Strategic Information System Planning (SISP) Success: An Exploratory Study in Public Universities in Malaysia. International Journal of Scientific Research in Education, 2, 76-87.

[27] Thomas, G. and Fernández, W. (2008) Success in IT projects: a matter of definition. International Journal of Project Management, 26 (7), 733-742.

[28] KANARACUS C (2008) Gartner: global IT spending growth stable. InfoWorldApril 3, 2008.

[29] DeLone, W.H., and McLean, E.R. Information systems success: The quest for the dependent variable. information Systems Research, 3, 1 (2008), 60–95

[30] Ika, L.A. (2009) Project Success as a Topic in Project Management Journals. Project Management Journal, 40 (4), 6-19. .

[31] Khani, N., Md Nor, K., and Bahrami, M. (2010), Factors Moderating the Relationship between IS Capabilities and Strategic Information System Planning (SISP) Success, 3 rd International GraduateConference on Engineering, Science, and Humanities 2010 (IGCESH2010), 2nd-4th November 2010,UTM, Johor

[32] Naser Khani, Iran Khalil Md Nor, Mojgan Bahrami Samani, Malaysia, MalaysiaHossein Hakimpoor, Islamic Azad University–Birjand,International Journal of Strategic Information Technology and Applications, 3(2), 1-17, April-June 2012 , Strategic Information, Strategic Information Systems Planning (SISP),An Empirical Investigation of Capability Factors AffectingStrategic Planning Success.

Bikram Pal Kaur is an Assistant Professor in the Deptt. Of Computer Science & Information Technology and is also heading the Deptt. Of Computer Application in Chandigarh Engineering College, Landran ,Mohali. She holds the degrees of B.tech. M.Tech, and M.Phil.. And is currently pursuing her Ph.D.in the field of Information Systems from Punjabi University, Patiala. She has more than 13 years of teaching experience and served many academic institutions. She is an Active Researcher who has supervised many B.Tech. Projects and MCA ,M.tech. dissertations and also contributed 12 research papers in various national & international conferences. Her areas of interest are Information System, ERP .

Dr. Himanshu Aggarwal, is Professor (Reader) in Computer Engineering at University College of Engineering, Punjabi University, Patiala. He had completed his Bachelor's degree in Computer Science from Punjabi University Patiala in 1993. He did his M.E. in Computer Science in 1999 from Thapar Institute of Engineering & Technology, Patiala. He had completed his Ph.D. in Computer Engineering from Punjabi University Patiala in 2007.He has more than 18 years of teaching experience. He is an active researcher who has supervised 15 M.Tech. Dissertations and guiding Ph.D. to seven scholars and has contributed more than 40 articles in International and National Conferences and 22 papers in research Journals. Dr. Aggarwal is member of many professional societies such as ICGST, IAENG. His areas of interest are Information Systems, ERP and Parallel Computing. He is on the review and editorial boards of several refereed Research Journals.

IT outsourcing service provider dynamic evaluation model and algorithms based on Rough Set

Li Shi[1,2]

[1] International School of Software, Wuhan University, Wuhan, China
[2] School of Information Management, Hubei University of economics, Wuhan, China

Abstract

The traditional supplier evaluation methods mostly belong to the static evaluation, and the actual information service provider selection is a dynamic process, need to use dynamic evaluation method to measure. In order to select the IT outsourcing service provider, the tool which uses complex scientific management system-thinking - Exploration diagram, established the index system for the selection of IT outsourcing. Based on it, the decision table of dynamic IT-Outsourcing service provider selection is made. The decision rule set of IT outsourcing service provider prediction is obtained by applying rough set theory of the decision table attribute reduction and value reduction. Finally, a calculation example of IT outsourcing service provider selection is illustrated, which shows that the mentioned evaluation method is feasible and efficient for dynamic IT outsourcing service provider selection and prediction. Thus, it supplies reasonable analysis and policy making of IT outsourcing service provider.

Keywords: *IT outsourcing; Service evaluation; Exploration diagram; reduction; rules gaining*

1. Introduction

With the development of Information Technology and thinking of outsourcing, enterprise and researcher pay more attention on the IT outsourcing. Domestic and foreign scholars from different angles on information technology outsourcing are discussed. The information technology outsourcing is a new management method and it will be an effective way for all middle and small scale enterprises to realize informationizion. Currently, 60% of U.S. companies with professional IT outsourcing services rapidly expand their own business [1-2]. However, IT outsourcing is a very complex business processes and accompanied by a variety of risks, results show that IT outsourcing success is still to be improved. Among them, the one of the greatest risks is the selection of service providers. [3].

The traditional supplier evaluation methods mostly belong to the static evaluation, and the actual information service provider selection is a dynamic process, need to use dynamic evaluation method to measure. There is no unified framework of IT outsourcing service provider evaluation index system in the present study. And evaluation is often focused on the current situation, unable to evaluate the future continuous development of IT outsourcing service provider.

So, exploration diagram tool is put forward in this paper, it use the complex scientific management system thinking mode to apply IT outsourcing service provider evaluation, absorb its systematic, modular and visual thinking ways of thinking [4-5]. Through this method, IT-outsourcing service provider evaluation of influence factors are found, and the evaluation index system is established. IT outsourcing service provider evaluation about an enterprise is taken as an example to set up the IT outsourcing service provider evaluation decision table, the decision table is conducted by applying rough set theory of the decision table attribute reduction and value reduction. And IT outsourcing service provider of the evaluation results decision rule set is obtained. According to it, the dynamic IT outsourcing service provider evaluation results and development trend are given. It provides the basis for enterprises to select the suitable IT outsourcing service provider.

2. Index System based on Exploration Diagram

Major Exploration diagram is the system thinking tool, which mainly helps to solve the problem of how to conduct visual thinking. Exploration diagram can help us to solve such kind of decision making problems: advance unknown, such as new product development decision-making, investment project evaluation and decision-making, through the creation of Exploration diagram to find all the factors that affects or may affect the research topic, helping to make decision.

Exploration diagram via the researchers to the whole environment of observation, according to own mastery of knowledge and information, plus full imagination, apply the point of bigger environment consideration to created a map. The picture presents all affects or may affect the subject factors, at the same time, it also reflects the relationship of factors, and its formation is a collective creation process.

Exploration diagram use ellipse to represent factors, elliptic bidirectional arrow indicates interactive relationship, ruleless circles represent the same factors, and ligature used to connect factors of the same class, a tail on the ellipse indicates deleting factors. The process of drawing Exploration diagram for Evaluation of IT outsourcing service providers is as follows:

Step 1: According to the needs of the research theme or decision-making problems, the relevant experts convened to discuss the research topic or decision-making problems.

Step 2: Moderator or the responsible raise questions in accordance with the research topic or the question of the decision-making, allowing you to think .This article mainly discusses how to evaluate and select IT outsourcing service providers.

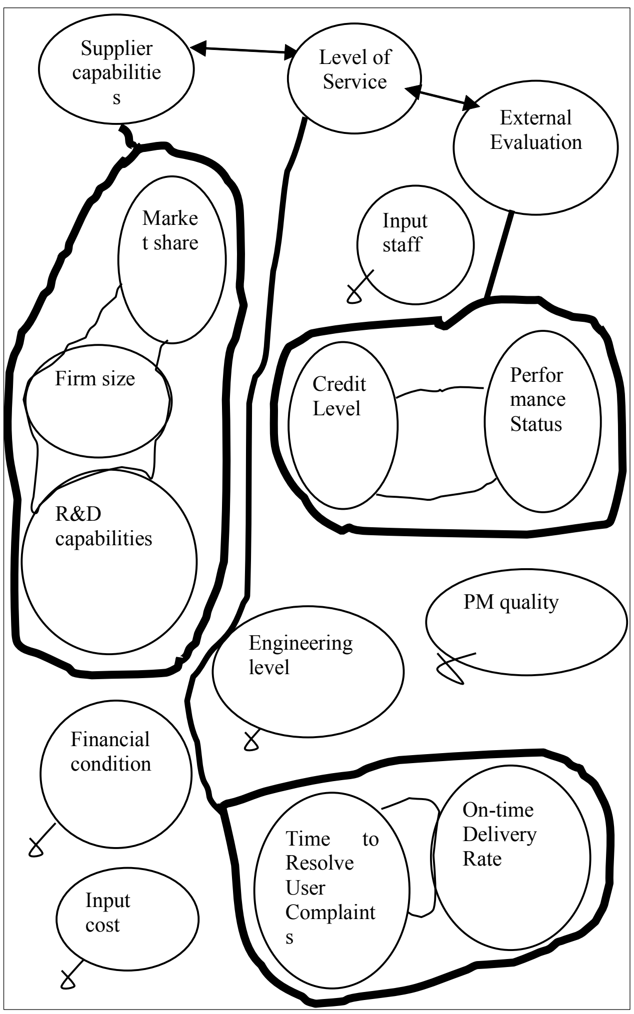


Fig. 1: classified exploration diagram

Step 3: Each member give full play to their imagination and express their views.

Step 4: After they have fully express their views, everybody are guided to conduct the visual think about all the proposed factors aimed at these oval, starting from the overall, integrated the same, Eliminate redundant, connect the factors of the same class with ligature, use the ruleless circles to enclose the same type factors, plus a tail in the oval for the deleted factors. the classified exploration diagram is shown in fig.1.

Step 5: To further the visual thinking, named each irregular circle, Such as commodity prices, types of goods, commodities performance, and Brand and commodity picture information can be classified as commodity information. The naming of this category is only an initial name, can eventually named after the analysis of the causal association.

Based on above, IT-outsourcing service provider selection index system is obtained and shown in table 1.

Table 1：IT-outsourcing supplier selection index system

Evaluation target	Factor indicators	Specific indicators	Character
IT-outsourcing supplier selection (A)	Supplier capabilities (B1)	Firm size (C1)	Quantitative
		R & D capabilities (C2)	Qualitative
		Market share (C3)	Quantitative
	Level of Service (B2)	On-time Delivery Rate (C4)	Quantitative
		Time to Resolve User Complaints (C5)	Quantitative
	External Evaluation (B3)	Credit Level (C6)	Qualitative
		Performance Status (C7)	Qualitative

3. Application of Rough Set Theory in IT Outsourcing Service Provider Evaluation

3.1 Rough Set Theory

Rough set theory [6-7] is a new mathematical tool to deal with imprecise, incomplete and inconsistent data. It can effectively analyze each kind of incomplete information such as imprecise, inconsistent, not integrity and so on, but also discovers the concealed knowledge and promulgates the latent rule according to analyzing and reasoning data.

In the rough set theory, computation of approximations and edge and attributes reduction of decision table is import part of them. Decision rules can be mined from given data using rough set theory. Comparing to others theories which processes indefinite and imprecise question, the most remarkable difference is it does not need outside the data acquisition which provides the question to examine the information, therefore it is quite objective to the indefinite description or processing questions. Because this theory has not been able to contain processing imprecise or indefinite primary data mechanism, so this theory and the theory of probability, the fuzzy mathematics and the evidence theory and so on other theory which processes indefinite or imprecise questions are complementary.

3.2 The Concepts of Rules Gaining

In this section we describe the concepts of rules gaining. Reducing the decision tables, searching the minimal attribute subsets and getting the succinct decision rules are not only the basic but also the toughest problems of Rough Set theory. The main idea of the theory of rough sets is to find out decision making and classification rules through knowledge reduction without changing the classification capacity of the information system.

The basic thought that discovers the classified rule in the decision support system based on rough set theory as follows:

Step1: the user proposes the duty of discover. The user takes some or many attributes as the classified policy-making attributes in the database, according to different values of these attributes, the data divides into the different category in the database, the duty of discover is produces these different determination rules.

Step2: using the algorithm based on rough set theory for gaining classification rules.

Some definitions are given follows:
Definition1: Discernibility Matrix [8-11].A mathematician named Skowron in Warsaw University proposes a discernibility matrix. There is a information system $S=(U,A,V,F)$, $U=\{x_1,x_2,...,x_n\}$ is the universe of discourse, A is the attribute set, $A=C\bigcup D$, C is the condition attribute, D is the policy-making attribute, $a(x)$ is a value with x on the attribute a , the resolution matrix is

$$c(ij)=\begin{cases} a\in A & :a(x_i)\neq a(x_j) \quad D(x_i)\neq D(x_j) \\ 0 & \phi \quad D(x_i)=D(x_j) \\ -1 & a(x_i)=a(x_j) \quad D(x_i)\neq D(x_j) \end{cases}$$

Definition 2: Equal Class. Regarding an attribute set $B\subset A$ in information system $S=(U,A,V,F)$, if it satisfies $IND(B)=\{(x,y)\in UXU \mid a(x)=a(y),\forall a\in B\}$, then equal relates $IND(B)$ called dual relates that can't distinguish each other. $B(x)$ express a equal class which object x in it.

Definition 3: CORE [9]. Opposite to attribute set D, core is an attribute set, which is intersect of reduction belongs to the attribute set C, records is CORE (C, D). The core is these attributes with its group divisible number is 1 in discernibility matrix.

Definition 4: Reduction [12]: U is the universe of discourse, R is an equal relational race, r $\in$ R, $IND(R)$ is the intersect ion of equal relation in R, if $U/IND(R)$ is equal to $U/IND(R-\{r\})$, then r is may be canceled in R. Otherwise r is not be canceled in R. If any element in p ($P=R-\{r\}$) is not being canceled, then called P is the reduction of R.

Definition 5: Equal set description: if an equal set named E_i , then we descript its character by using $Des(E_i)=(a=v)$. $a\in A$, $v\in V_a$

Definition 6: Rules gaining [13]. Assume that the division of A in U is E, the division of A' is Y. E Look upon as the classification condition; Y looks upon as the classification conclusion. We may get classified rule as follows:

(1) If $E_i\bigcap Y_j\neq\phi$, then we get $r_{ij}:Des(E_i)\rightarrow Des(Y_j)$

a. If $E_i\bigcap Y_j=E_i$, then, the rule r_{ij} is ascertained. Rule confidence level is one ($cf=1$).

b. If $E_i\bigcap Y_j\neq E_i$, then, the rule r_{ij} isn't ascertained. Rule confidence level is $cf=\frac{|E_i\bigcap Y_j|}{|E_i|}$

(2) if $E_i\bigcap Y_j=\phi$, there isn't establish rule.

When the rule confidence level is one, this kind of rule can simplify. The rule reduction is that some attributes are deleted from condition attributes; the rule confidence level was still one.

3.3 Model of IT Outsourcing Service Provider Evaluation based on Exploration Diagram and Rules Gaining

IT outsourcing service provider evaluation results can be divided into good, general, bad. According to the index system by applying explore diagram, selecting evaluation data of provider, using rough set theory to input space dimension reduction, completing input feature extraction work, it can achieve purpose: reduce the size of the data processing. The more sample data, the more the rules the higher reliability are. When the new supplier need to evaluation, using decision rule set, it can be evaluated. The whole procession is shown in Fig.2.

Now we describe the algorithm based on rough set for rules gaining.

Step 1: Foundation data processing. In this step, we need input $S=(U,A,V,F)$, then, according to given classification method, data are standardized.

Step 2: The smallest attributes set is obtained.
The algorithm as follows:

Begin
For i=1 to n
For j=i to n
M=$[c(ij)]$

$$/*\ \left(c(ij)=\begin{cases} a\in A & : a(x_i)\neq a(x_j) \quad D(x_i)\neq D(x_j) \\ 0 & \phi \quad D(x_i)=D(x_j) \\ -1 & a(x_i)=a(x_j) \quad D(x_i)\neq D(x_j) \end{cases}\right)\ */$$

End for
End for
P={union of single attribute in M.}
End

Step 3: Rules gaining and reducing. When attributes are reduced, we get a new policy-making table. First, equal sets of condition attributes and equal sets of policy attributes are obtained. Second, according to definition six, we can gain rules and reduce rules. At last, the same conclusions of rules are union.

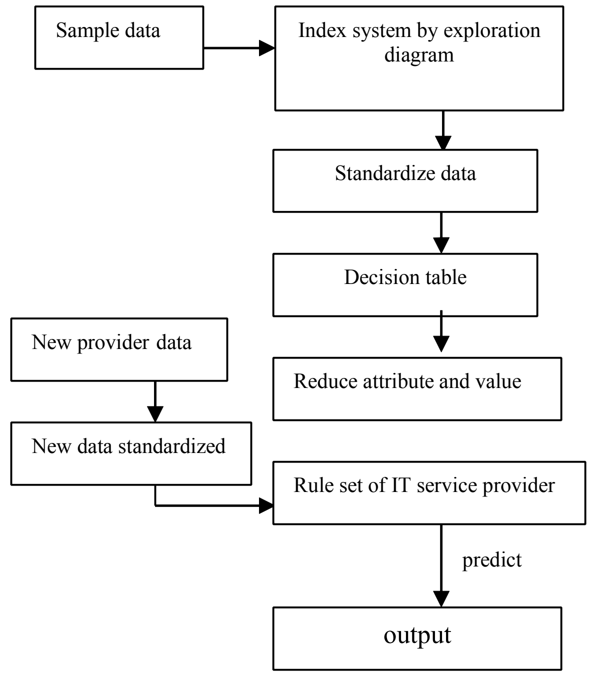


Fig. 2: The whole procession

4. Analysis of Example

In this section, an example in IT-outsourcing service provider selection is introduced to confirm the algorithm's validity. Suppose a company has ten IT-outsourcing service providers to be evaluated, we use model in section 3 to gain the rule and predict IT-outsourcing service providers.

Step 1: Foundation data processing. Each column of the attribute value is divided into three grades: 3-good, 2-general, 1-bad. The result is divided into three grades: 1-bad, 2-general, and 3-good. Table 2 is a dimension unitize policy-making table.

Table 2: Policy table

ID	C1	C2	C3	C4	C5	C6	C7	D
1	1	2	3	2	3	1	2	1
2	1	3	1	1	2	2	2	3
3	2	3	1	1	1	1	1	2
4	1	2	3	2	3	1	1	2
5	2	3	1	1	1	1	3	3
6	3	1	1	1	3	2	2	2
7	1	3	2	2	1	1	1	2

Step 2: Obtaining the smallest attributes set. According to definition three and step two from section 3, we use algorithm to calculate discernibility matrix M, and obtain the smallest attributes set P. In order to simply calculation,

we use the computer programming to solve it. Program interface is shown in Fig.3. We can get $P = \{c1, c2, c7\}$. Table 3 is Policy-making table of reduction.

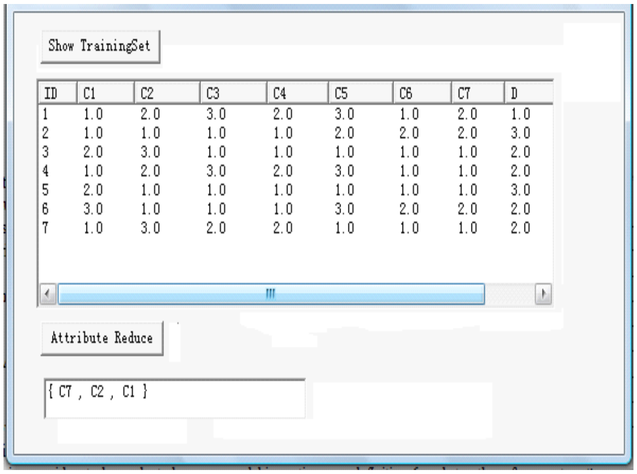


Fig.3: Program interface

Table 3: Policy-making table of reduction

ID	C1	C2	C7	D
1	1	2	2	1
2	1	3	2	3
3	2	3	1	2
4	1	2	1	2
5	2	3	3	3
6	3	1	2	2
7	1	3	1	2

Step 3: Rules gaining and reducing. According to definition 6 and step three from section 3, we can get the last policy.

When $cf = 1$, policy rules are below:

(1) $c_1 = 1 \wedge c_2 = 2 \wedge c_7 = 2 \rightarrow f_1$

(2) $c_7 = 1 \vee (c_1 = 3 \wedge c_2 = 1 \wedge c_7 = 2) \rightarrow f_2$

(3) $c_2 = 3 \vee (c_1 = 1 \wedge c_7 = 3) \rightarrow f_3$

The first rule shows that, if one service provider's firm size is bad, and R&D capability is general and performance status is general, we can deem this provider is bad, and rule confidence level is 1. The second rule shows that, if one service provider's performance status is bad, or firm size is good, and R&D capability is bad, and performance status is general, we can deem this provider is general, and rule confidence level is 1. The third rule shows that, if one service provider's firm size is bad, and R&D capability is good, or performance status is good, w can deem this provider is good, and rule confidence level is 1. When the new supplier need to evaluation, using decision rule set, it can be evaluated.

5. Conclusions

In the paper, according to the index system by applying explore diagram, selecting evaluation data of provider, using rough set theory to input space dimension reduction, completing input feature extraction work. An algorithm based on rough set for rules gaining is introduced to analyze and process data, minimal decision-making rules are proposed in IT outsourcing service provider selection. a calculation example of IT outsourcing service provider selection is illustrated， which shows that the mentioned evaluation method is feasible and efficient for dynamic IT outsourcing service provider selection and prediction.

Acknowledgments

The author wish to thank Science and Technology Research Funds of Educational Ministry of Hubei Province(Grant B20121905), under which the present work was possible.

References

[1] Lacity, M., Khan, S., Yan, A., and Willcocks, L. "A Review of the IT Outsourcing Empirical Literature and Future Research Directions," Journal of Information Technology, Vol. 25, no.4, 2010, pp. 395-433.

[2] Guofeng Tang, Bin Dan, Han Song,Xumei Zhang, "Research on Incentive Mechanism of Application Services Outsourcing Under Asymmetric Information", IJIPM: International Journal of Information Processing and Management, Vol. 3, No. 4, 2012, pp. 39 - 47.

[3] Lacity M C, Willcocks L P. "Inside Information Technology Outsourcing: A State-of-the-Art Report ", Oxford Institute of Information Management, 2000

[4] Xu Xusong, Zheng Xiaojing. "Qualitative Analysis Tools: Probe Graph, Cycle Graph and Structure Graph", Technology Economics, Vol. 29, no. 5, 2010, pp.1-6,

[5] Xu Xusong, Zheng Xiaojing. "Analysis Framework for Combination of Qualitative and Quantitative", Technology Economics, Vol. 29, no. 4, 2010, pp.1-6.

[6] Thanh Nguyen Chi, Yamada Koichi. Document representation and clustering with wordnet based similarity rough set model[J]. International Journal of Computer Science Issues, vol.8, no. 5, 2011, pp.1-8.

[7] Pawlak Z. Rough classification [J]. International Journal of Human-Computer Stdies, vol. 51, 1999,pp.369-383.

[8] Wang Jue, Miao Duo-qian, Zhou Yu-jian. Rough set theory and its application: a survey. Pattern Recognition and Aritificial Intelligence, vol.9, no. 4,1996, pp.337-344.

[9] Hu X,Cercone N. Learning in relational databases: a rough set approach. International Journal of Computational Intelligence, vol. 11, no. 2, 1995, pp.323-338.

[10] Miao duoqian,Wang jue.An Information Representation of the Concepts and Operations in Rough Set Theory[J]. Journal of software, vol. 2, no. 10, 1999, pp.113-116.
[11]Li Shi, Hong Wang, Improvement of the Intelligent Decision-Making Model based on AHP-Fuzzy Method [J]. Proceedings of 2006 International Conference on Artificial Intelligence, 2006, 8, pp.330-331.
[12]Liu Qing, Rough Set and Rough Reasoning[M].Beijing: Science Press,2005, pp.40-42.
[13] Acharjya, D.P, Ezhilarasi, L. A knowledge mining model for ranking institutions using rough computing with ordering rules and formal concept analysis[J]. International Journal of Computer Science Issues, vol.8, no. 2, 2011, pp.417-426.

Li Shi is a lecturer in School of Information Management of HuBei University of Economics. Currently, she is a PhD student in international school of Software of WuHan University. She has authored a number of different journal and paper. Her research interests include service science, data mining, and decision support. She is a member of CCF.

Review of Security Approaches in Routing Protocol in Mobile Adhoc Network

Sumati Ramakrishna Gowda
Research Scholar
Vinayaka Missions University, Salem, India

P.S Hiremath
Department of Computer Science.
Gulbarga University, Gulbarga, India

***Abstract*— In this paper the objective is to present a review of routing protocols in mobile ad hoc network (MANET) exclusively from security viewpoint. In MANET, the mobile nodes often move randomly for which reason the cumulative network experiences rapid and much unpredictable topology alterations. Due to presence of dynamic topology as well as limited range of transmission, very often some nodes cannot communicate directly with each other. Because of this phenomenon, all the QoS and security issues surface. Till now there is abundant literature work being formulated towards designing routing protocols. But security features designed till now are not able to provide optimal security towards secure routing. Routing protocols, data, bandwidth and battery power are the common target of the attackers. Therefore, in this paper the attempts are to throw light on the work that were focused exclusively for maintaining security in routing protocols in MANET**

Keywords-component: Routing, MANET, security

I. INTRODUCTION

A mobile ad hoc network (MANET) is a self-configuring infrastructureless network of mobile devices connected wirelessly [1]. People and vehicles can thus be internetworked in areas without a preexisting communication infrastructure or when the use of such infrastructure requires wireless extension [2]. In the mobile ad hoc network, nodes can directly communicate with all the other nodes within their radio ranges; whereas nodes that are not in the direct communication range use intermediate node(s) to communicate with each other. In these two situations, all the nodes that have participated in the communication automatically form a wireless network. Therefore, this kind of wireless network can be viewed as mobile ad hoc network. The mobile ad hoc network has the following typical features [3]:

- Unreliability of wireless links between nodes: Because of the limited energy supply for the wireless nodes and the mobility of the nodes, the wireless links between mobile nodes in the ad hoc network are not consistent for the communication participants.
- Constantly changing topology: Due to the continuous motion of nodes, the topology of the mobile ad hoc network changes constantly. The nodes can continuously move into and out of the radio range of the other nodes in the ad hoc network, and the routing information will be changing all the time because of the movement of the nodes.
- Lack of incorporation of security features that exist in statically configured wireless routing protocol and are not meant for ad hoc environments. Because the topology of the ad hoc networks is changing constantly, it is necessary for each pair of adjacent nodes to incorporate security in the routing issue so as to prevent some kind of potential attacks that try to make use of vulnerabilities that exist in the statically configured routing protocol.

Because of the features listed above, the mobile ad hoc networks are more prone to suffer from the malicious behaviors than the traditional wired networks. Therefore, we need to pay more attention to the security issues in the mobile ad hoc networks. In such an environment, there is no guarantee that a path between two nodes would be free of malicious nodes, which would not comply with the employed protocol and attempt to harm the network operation. The mechanisms currently incorporated in MANET routing protocols cannot cope with disruptions due to malicious behavior. For example, any node could claim that, it is one hop away from the sought destination, which is causing all routes to the destination to pass through itself. Alternatively, a malicious node could corrupt any in-transit route request (reply) packet and cause data to be misrouted. The presence of even a small number of adversarial nodes could result in repeatedly compromised routes, and, as a result, the network nodes would have to rely on cycles of time-out and new route discoveries to communicate. This would incur arbitrary delays before the establishment of a non-corrupted path, while successive broadcasts of route requests would impose excessive transmission overhead. In particular, intentionally falsified routing messages would result in a denial-of-service (DoS) experienced by the end nodes. The proposed scheme combats such types of misbehavior and safeguards the acquisition of topological information. Section 2 discusses about the preliminary study of the survey of literature on the domain followed by security issues in Section 3. Related work is discussed in Section 4, while MANET routing security is examined in Section 5 and briefing on other routing protocols in Section 6 and 7. Section 8 illustrates about scope of future research, with concluding remarks in Section 9.

II. PRELIMINARY STUDY

In MANETs, some form of routing protocol is required in order to dynamically detect the multi-hop paths through which packets can be sent from one node to another. Active research work on MANETs is carried out mainly in the fields of Medium Access Control (MAC), routing, resource management, power control, and security. Because of the importance of routing protocols in dynamic multi-hop networks, a lot of MANET routing protocols have been proposed in the last few years. Considering the special properties of MANET, many routing protocol, generally the following properties are expected, though all of these might not be possible to be incorporated in a single solution.

- A routing protocol for MANET should be distributed in a manner in order to increase its reliability.
- A routing protocol must be designed considering unidirectional links, because wireless medium may cause a wireless link to be opened in uni-direction only due to physical factors.
- The routing protocol should be power-efficient.
- The routing protocol should consider its security.
- A hybrid routing protocol should be much more reactive than proactive to avoid overhead.
- A routing protocol should be aware of Quality of Service (QoS).

There are basically two categories of routing protocols for MANETs:

1. Table Driven (Proactive): DSDV, GSR, WRP
2. Source Initiated On-Demand (Reactive): ABR, AODV, DSR, LAR

Based on the method of delivery of data packets from the source to destination, the classification of MANET routing protocols could be done as follows:

- Unicast Routing Protocols: The routing protocols that consider sending information packets to a single destination from a single source.
- Multicast Routing Protocols: Multicast routing is the delivery of information to a group of destinations simultaneously, using the most efficient strategy to deliver the messages over each link of the network only once, creating copies only when the links to the destinations split. Multicast routing protocols for MANET use both multicast and unicast for data transmission.

Multicast routing protocols for MANET can be classified again into two categories: Tree-based multicast protocol and Mesh-based multicast protocol. Mesh-based routing protocols use several routes to reach a destination, while the tree-based protocols maintain only one path.

Much of the research has been done focusing only on the efficiency of the MANETs. There are quite a number of routing protocols that are excellent in terms of efficiency. But the security requirement of these protocols has changed the situation and a more detailed research is currently underway to develop secure ad hoc routing protocols. MANETs are extremely vulnerable to attacks due to their dynamically changing topology, absence of conventional security infrastructures and open medium of communication, which, unlike their wired counterparts, cannot be secured. To address these concerns, several secure routing protocols have been proposed: Secure Efficient Distance Vector Routing (SEAD), Ariadne, Authenticated Routing for Ad hoc Networks (ARAN), Secure Ad hoc On-Demand Distance Vector Routing (SAODV), and Secure Routing Protocol (SRP). Although researchers have proposed several secure routing protocols, their resistance towards various types of security attacks and efficiency have been primary points of concern in implementing these protocols. Hence, there is a need for review. In a MANET, attacks can be classified into Passive Attacks and Active Attacks, which are discussed below.

A. Passive Attacks

In passive attacks, attackers don't disrupt the operation of routing protocol but only attempt to discover valuable information by listening to the routing traffic. The attacker only looks and watches the transmission and does not try to modify or change the data packets. Two types of passive attacks are possible:

- Traffic analysis: In this attack, attacker monitors packet transmission to infer important information such as a source, destination and source-destination pair.
- Eavesdropping: In Eavesdropping, attackers obtain some confidential information e.g. private key, public key, location or even password of the node that should be kept secret during transmission.

B. Active Attacks

In the active attacks, the malicious nodes introduce false information to confuse the network topology. They can either attract traffic to them and then drop or compromise the packets. They can also send false information and lead packets to the wrong node and cause congestion in one area. The attacks can either target at the routing procedure or try to flood the networks. Various types of active attacks are:

- Sinkhole Attack: A sinkhole node tries to attract the data toward itself from all neighboring nodes. In this attack, a malicious node generates fake routing information and show itself as legal nodes for the route. Sinkhole node attempts to regulate all network traffic according to itself, modifies the data packets, decrease the network life time, create complicated network topology and finally destroy the network.
- Flooding Attack: In this attack, a malicious node may also inject false packets to consume the available resources onto the network, so that valid user can not be able to use the network resources for valid communication. The flooding attack is possible in most of the on demand routing protocols such as SRP, SAODV, ARAN etc.
- Replay Attack: This attack usually targets the freshness of routes. In this attack, an attacker firstly records the message and then resend the old message to the other nodes to update their routing table with stale routes.
- Rushing Attack In rushing attack, attacker forwards routing packets as quick as possible to gain access to

multicast forwarding group before the legal node .By this way, rushing attack can slow down the performance of network .The rushing attack can act as an effective DoS attack against all currently proposed on demand MANET routing protocols.

C. Common attacks in MANETs

- Denial-of-service with modified source route: In the denial-of-service, a malicious node in between can successfully send an erroneous route message to the source route to disrupt the service.
- Tunneling Attack: The tunneling attack is where two or more nodes may collaborate to encapsulate and exchange messages between them along existing data routes.
- Wormhole Attack: In wormhole attack, an attacker records packet at one location in the network, tunnels them to another location, and retransmits them back into the network. This attack is possible even if the attacker has not compromised any hosts and even if all communication provides authenticity and confidentiality
- Black hole Attack: In Black hole attack, a malicious node uses the routing protocol to advertise itself as having the shortest path to the node whose packets it wants to intercept and in this way it can compromise the service.
- Spoofing Attack: In Spoofing, a single malicious node in the ad hoc network can spoof the nodes identity in order to forward packets through it. Later the information can be used to create DoS attacks.

D. Security Services

Security services include the functionality required to provide a secure networking environment. The main security services can be summarized as follows:

- Authentication: This service verifies a user's identity and assures the recipient that the message is from the source that it claims to be from. Firstly, at the time of communication initiation, the service assures that the two parties are authentic, that each is the entity it claims to be. Secondly, it must assure that a third party does not interfere by impersonating one of the two legitimate parties for the purpose of authorized transmission and reception. Authentication can be provided using encryption along with cryptographic hash functions, digital signatures and certificates. Details of the construction and operation of digital signatures can be found in RFC2560.
- Confidentiality: This service ensures that the data/information transmitted over the network is not disclosed to unauthorized users. Confidentiality can be achieved by using different encryption techniques such as only legitimate users can analyze and understand the transmission.
- Integrity: The function of integrity control is to assure that the data is received in verbatim as sent by authorized party. The data received contains no modification, insertion or deletion.
- Access Control: This service limits and controls the access of such a resource, which can be a host system or an application.
- Availability: This involves making the network services or resources available to the legitimate users. It ensures the survivability of the network despite malicious incidences.

As the currently available wireless networking and mobile computing hardware is now capable of fulfilling the promise of this technology, it is the need of the hour to design and develop routing protocols which should support the performance with endurance. The correct execution of these routing protocols is mandatory for smooth functioning of a MANET A variety of protocols have been proposed targeted at securing MANETs. The performance comparison of these protocols needs to be analyzed. In the present work, we have compared these protocols by highlighting their features, differences and characteristics. It can be summed up that each protocol has definite advantages and disadvantages, and can be appropriate for a particular application environment. The security in routing protocols addressed in the recent past decade has been discussed in the following sections.

The provision of security services in the MANET context faces a set of challenges specific to this new technology. The insecurity of the wireless links, energy constraints, relatively poor physical protection of nodes in a hostile environment, and the vulnerability of statically configured security schemes have been identified. Even if such services were assumed, their availability would not be guaranteed, either due to the dynamically changing topology that could easily result in a partitioned network, or due to congested links close to the node acting as a server. The absence of infrastructure and the consequent absence of authorization facilities impede the usual practice of establishing a line of defense, separating nodes into trusted and non-trusted. Such a distinction would have been based on a security policy, the possession of the necessary credentials and the ability for nodes to validate them.

Table 1: Exploring Research gap

Performance Parameters	ARAIDNE	ARAN	SEAD	SRP	SAODV	SAR	SLSP	SANEDNA
Type	Reactive	Reactive	Proactive	Reactive	Reactive	Reactive	Proactive	Reactive
MANET Protocol	DSR	AODV/ DSR	DSDV	DSR/ ZRP	AODV	AODV	ZHLS	DSR
Encryption	Sym	Asym	Sym	Sym	Asym	Sym/Asym	Asym	Sym
Synchronization	Yes	No	Yes	No	No	No	No	Yes
Trust Authority	KDC	CA	CA	CA	CA	CA/KDC	CA/KDC	No
Authentication	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Confidentiality	No	Yes	No	No	No	Yes	No	Yes
Integrity	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Non-Repudiation	No	Yes	No	No	Yes	Yes	Yes	Yes
Anti-Spoofing	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
DoS Attacks	Yes	No	Yes	Yes	No	No	Yes	Yes

Source: Tarun Dalal, Gopal Singh [4]

III. SECURITY ISSUES

The buildup of an ad hoc network can be envisaged where support of wireless access or wired backbone is not feasible. Ad hoc wireless network does not have any predefined infrastructure and all network services are configured and created on the fly. Thus, it is obvious that with lack of infrastructure support and susceptible wireless link attacks, security in ad hoc network becomes inherently weak. Achieving security within a MANET is challenging due to following reasons.

- Dynamic Topologies and Membership: A network topology of ad hoc network is very dynamic as mobility of nodes or membership of nodes is very random and rapid. This stipulates the need for secure solutions to be dynamic.
- Vulnerable wireless link: Passive/Active link attacks like eavesdropping, spoofing denial of service, masquerading, impersonation are possible.
- Roaming in dangerous environment: Any malicious node or misbehaving node can create hostile attack or deprive all other nodes from providing any service.

The main issues for providing security in MANET are briefly discussed below.

- Identification issue: Nodes having access to common radio link can easily participate to set up ad hoc infrastructure. But the secure communication among nodes requires the secure communication link to communicate.
- Before establishing secure communication link, the node should be capable enough to identify another node. As a result, node needs to provide his/her identity as well as associated credentials to another node.
- The delivered identity and credentials need to be authenticated and protected so that authenticity and integrity of delivered identity and credentials cannot be questioned by receiver node. Every node wants to be sure that delivered identity and credentials to recipient nodes are not compromised.
- Privacy Issue: The identification issue simultaneously leads to privacy issue for MANET Mobile node uses various types of identities and that varies form link level to user/application level. Also in mobile environment very frequent mobile node is not ready to reveal his/her identity or credentials to another mobile node from privacy point of view. Any compromised identity leads attacker to create privacy threat to user device. Unfortunately the current mobile standards do not provide any location privacy and in many cases revealing identity is inevitable to generate communication link. Hence a seamless privacy protection is required to harness the usage of ad hoc networking.

Therefore, due to the issues discussed above, it is essential to provide security architecture to secure ad hoc networking. In the literature, there are many works that address the security issues in MANETs.

IV. RELATED WORK

Panagiotis Papadimitratos et al. [5] have presented a route discovery protocol that is considered one of the standard works on secure routing in mobile ad hoc networks. John Marshall et al. [6] have proposed the SRP algorithm for routing in ad hoc networks. Oscar F. Gonzalez et al. [7] presented a mechanism that enables the detection of nodes that exhibit packet forwarding misbehavior. Stephan Eichler et al. [8] have introduced a novel secure routing protocol based on AODV for vehicular ad hoc networks. M. Rajesh Babu et al. [9] have proposed to develop an energy efficient secure authenticated routing protocol (EESARP). Steffen Reidt et al. [10] have introduced a trust metric in the cluster head selection process to securely determine constituting nodes in a distributed Trust Authority (TA) for MANETs. Muhammad Nawaz Khan et al. [11] have proposed distributed-ID, a smart agent in each mobile node analyzes the routing packets. Lu Jin et al. [12] introduced the secure the delivery of routing packets and the strategy to determine the most secure routes. Panagiotis Papadimitratos et al. [13] have proposed the securing the delivery of routing packets and the strategy of determine the most secure routes. Shivasharanappa Allur et al. [14] have proposed a cross-layer design to achieve an unswerving data transmission in ADHOC networks. Venkat Balakrishnan et al. [15] introduced Trust Enhanced security Architecture for MANET (TEAM), in which a trust model is overlaid on the following security models key management mechanism, secure routing protocol, and cooperation model.

Kimaya Sanzgiri et al. [16] have introduced solution to one, the managed-open scenario where no network infrastructure is pre-deployed, but a small amount of prior security coordination is expected. Poonam Yadav et al. [17] have introduced the on-demand routing protocols AODV, DSR and DYMO based on IEEE 802.11 and the characteristic summary of these routing protocols are presented. Parma Nand et al. [18] have introduced the on demand routing protocols AODV, DSR and DYMO. David B. Johnson et al. [19] have presented a protocol for routing in ad hoc networks that uses dynamic source routing. Xiaodong Lin et al. [20] have presented a novel anonymous secure routing protocol for mobile ad hoc networks (MANETs). Xu Su et al. [21] have proposed mechanisms to complement the existing secure routing protocols to resist the creation of in-band tunnels. Mohd Anuar Jaafar et al. [22] introduced some evaluation and performance comparisons of AODV, SAODV and A-SAODV routing protocols in MANETs. Umang Singh et al. [23] have introduced various existing routing protocols that were reviewed. Julien Francq et al. [24] have proposed countermeasure that provides a high level of fault detection. Karim El Defrawy et al. [25] have presented the PRISM protocol which supports anonymous reactive routing in MANETs. Satoshi Kurosawa et al. [26] have proposed an anomaly detection scheme using dynamic training method. Amit N. Thakare et al. [27] made an attempt to compare the performance of two prominent on demand reactive routing protocols for MANETs. Kimaya Sanzgiri et al. [28] have proposed a solution to one, the managed-open scenario where no network infrastructure is pre-deployed.

Claude Crrepeau et al. [29] have presented secure Robust Source Routing (RSR). Liana Khamis Qabajeh et al. [30] have proposed a new model of routing protocol called ARANz, which is an extension of the original Authenticated Routing for Ad-Hoc Networks .

Feng He et al. [31] have proposed a novel secure routing protocol S-MAODV which is based on MAODV. Arun Kumar Mondal et al. [32] have presented the analytical results for the probability of success of data transmission over the networks taking the probability of success or failure of individual paths. Pietro Michiardi et al. [33] have carried out a simulation study that identifies security issues which are specific to MANET R. Kalpana et al. [34] have addressed anonymity and trust issues for a wireless network containing selfish and malicious nodes. Mike Burmester et al. [35] have analyzed provable secure route discovery algorithm which is vulnerable to a hidden channel attack. Himani Yadav et al. [36] have carried out survey on different existing techniques for detection of black hole attacks in MANETs with their defects. Jiajia Liu et al. [37] have explored the capability of these networks to support multicast traffic. Stefaan Seys et al. [38] have studied anonymous routing protocol for mobile ad hoc networks (MANETs). Subash Chandra Mandhata et al. [39] have analyzed the black hole attack in MANET using AODV as its routing protocol. Saikat Chakrabarti et al. [40] have proposed an efficient, single round multi signature scheme, CLFSR-M, constructed using cubic (third-order) linear feedback shift register (LFSR) sequences. K.Seshadri Ramana et al. [41] have proposed a routing protocol that is based on securing the routing information from unauthorized users. Sridhar Subramanian et al. [42] have examined a trust based reliable protocol TBRAODV.

V. MANET ROUTING SECURITY

In an ad hoc network, all the nodes may not be within the transmission range of each other; hence, nodes are often required to forward network traffic on behalf of other nodes. Consider for example the scenario in Fig.1. If node S sends data to node D, which is three hops away, the data traffic will reach its destination only of A and B forward it. The process of forwarding network traffic from source to destination is termed routing.

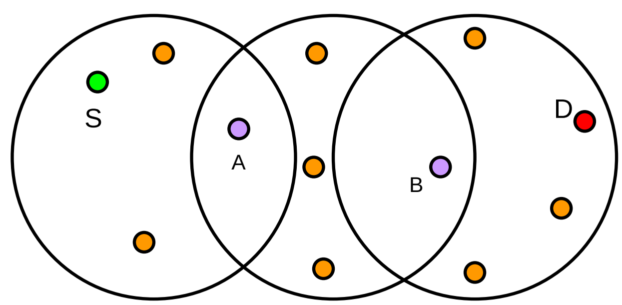


Figure 1 Multihop scenario

There are two general categories of MANET routing protocols: topology-based and position-based routing protocols. The list of some desirable qualitative properties of MANET routing protocols as adopted from an Internet Engineering Task Force (IETF) MANET Working Group memo [43] is as following:

- Loop-free: It is desirable that routing protocols prevent packets from circling around in a network for arbitrary time periods.
- Demand-based operation: In order to utilize network energy and bandwidth more efficiently, it is desirable that MANET routing algorithms adapt to the network traffic pattern on a demand or need basis rather than maintaining routing between all nodes at all time.
- Proactive operation: This is the IP-side of demand-based operation. In cases where the additional latency, which demand-based operations incur, may be unacceptable, if there are adequate bandwidth and energy resources, proactive operations may be desirable in these situations.
- "Sleep" period operation: It may be necessary, for reasons such as the need for energy conservation, for nodes to stop transmitting or receiving signals for arbitrary time periods. Routing protocols should be able to accommodate sleep periods without adverse consequences.
- Security: It is desirable that routing protocols provide security mechanisms to prohibit disruption or modification of the protocol operations.

VI. OTHER ROUTING PROTOCOLS

In addition to the above mentioned routing protocols for MANET, there are some other routing protocols that do not rely on any traditional routing mechanisms, instead rely on the location awareness of the participating nodes in the network. Generally, in traditional MANETs, the nodes are addressed only with their IP addresses. But, in case of location-aware routing mechanisms, the nodes are often aware of their exact physical locations in the three-dimensional world. This capability might be introduced in the nodes using Global Positioning System (GPS) or with any other geometric methods. Based on these concepts, several geo-cast and location-aware routing protocols have already been proposed. The major feature of these routing protocols is that, when a node knows about the location of a particular destination, it can direct the packets toward that particular direction from its current position, without using any route discovery mechanism. Recently, some of the researchers proposed some location-aware protocols that are based on such ideas. Some examples of these protocols are Geographic Distance Routing (GEDIR)[44], Location-Aided Routing (LAR)[45], Greedy Perimeter Stateless Routing (GPSR)[46], Geo-GRID[47], Geographical Routing Algorithm (GRA)[48], etc. Other than these, there are a number of multicast routing protocols for MANET Some examples of the multicast routing protocols are: Location-Based Multicast Protocol (LBM)[49], Multicast Core Extraction Distributed Ad hoc Routing (MCEDAR)[50], Ad hoc Multicast Routing protocol utilizing Increasing id-numbers (AMRIS)[51], Associativity- Based Ad hoc Multicast (ABAM)[52], Multicast Ad hoc On- Demand Distance-Vector (MAODV) routing [53],Differential Destination Multicast (DDM)[54],On-Demand Multicast Routing Protocol (ODMRP)[55], Adaptive Demand-driven Multicast Routing (ADMR) protocol [56], Ad hoc Multicast Routing protocol (AMRoute)[57], Dynamic Core-based Multicast routing Protocol (DCMP)[58], Preferred Link-Based
Multicast protocol (PLBM)[59],etc. Some of these multi cast protocols use location information and some are based on other routing protocols or developed just as the extension of another unicast routing protocol. For example, MAODV is the multicast-supporting version of AODV.

VII. OTHER RECENT WORKS ON MANET ROUTING

In this section, the recent works on routing in MANET that could be used as a reference by the practitioners, are considered. Some of these works have taken the major routing protocols as their bases and some of them have enhanced performances of the various previous routing protocols. Some recent works are: node-density-based routing [60], load-balanced routing [61], optimized priority based energy-efficient routing [62], reliable on-demand routing with mobility prediction [63], QoS routing [64], secure distributed anonymous routing protocol [65], robust position based routing [66], routing with group motion support [67], dense cluster gateway based routing protocol [68], dynamic backup routes routing protocol [69], gathering-based routing protocol [70], QoS-aware multicast routing protocol [71], recycled path routing [72], QoS multicast routing protocol for clustering in MANET [73], secure anonymous routing protocol with authenticated key exchange [74], self-healing on-demand geographic path routing protocol [75], stable weight-based on demand routing protocol [76], fisheye zone routing protocol [77], on-demand utility-based power control routing [78], secure position-based routing protocol [79], scalable multi-path on-demand routing [80], virtual coordinate-based routing [81], etc.

VIII. SCOPE FOR FUTURE RESEARCH

Many more efficient routing protocols for MANET might be developed in the coming future, which might take security and QoS (Quality of Service) as the major concerns. So far, the routing protocols mainly focused on the methods of routing, but in future a secured but QoS-aware routing protocol could be worked on. Ensuring both of these parameters at the same time might be difficult. A very secure routing protocol surely incurs more overhead for routing, which might degrade the QoS level. So an optimal trade-off between these two parameters could be searched. In the recent years some multicast routing protocols have been proposed. The reason for the growing importance of multicast is that this strategy could be used as a means to reduce bandwidth utilization for mass distribution of data. As there is a pressing need to conserve scarce bandwidth over wireless media, it is natural that multicast routing should receive some attention for ad hoc networks. So it is, in most of the cases, advantageous to use multicast rather than multiple unicast, especially in ad hoc
environment where bandwidth comes at a premium. Ad hoc wireless networks find applications in civilian operations (collaborative and distributed computing) emergency search-and-rescue, law enforcement, and warfare situations, where setting up and maintaining a communication infrastructure is

very difficult. In all these applications, communication and coordination among a given set of nodes are necessary. Thus, in future, the routing protocols might especially emphasize the support for multicasting in the network.

IX. Conclusion

This paper presents a number of routing protocols for MANET, which are broadly categorized as proactive and reactive. Proactive routing protocols tend to provide lower latency than that of the on-demand protocols, because they try to maintain routes to all the nodes in the network all the time. But the drawback for such protocols is the excessive routing overhead transmitted, which is periodic in nature without much consideration for the network mobility or load. On the other hand, though reactive protocols discover routes only when they are needed, they may still generate a huge amount of traffic when the network changes frequently. Depending on the amount of network traffic and number of flows, the routing protocols could be chosen. When there is congestion in the network due to heavy traffic, in general case, a reactive protocol is preferable. Sometimes the size of the network might be a major considerable point. For example, AODV, DSR, OLSR are some of the protocols suitable for relatively smaller networks, while the routing protocols like TORA, LANMAR, ZRP are suitable for larger networks. Network mobility is another factor that can degrade the performance of certain protocols. When the network is relatively static, proactive routing protocols can be used, as storing the topology information in such case is more efficient. On the other hand, as the mobility of nodes in the network increases, reactive protocols perform better. Overall, the answer to the debating point might be that the mobility and traffic pattern of the network must play the key role for choosing an appropriate routing strategy for a particular network. It is quite natural that one particular solution cannot be applied for all sorts of situations and, even if applied, might not be optimal in all cases. Often it is more appropriate to apply a hybrid protocol rather than a strictly proactive or reactive protocol as hybrid protocols often possess the advantages of both types of protocols.

References

[1] Azzedine Boukerche (Ed.), Algorithms and Protocols for Wireless and Mobile Ad Hoc Networks, John Wiley and Sons, New Jersey, 2009.
[2] M.S. Corson, J.P. Maker, and J.H. Cernicione, Internet-based Mobile Ad Hoc Networking, IEEE Internet Computing, July-August 1999, pp. 63-70.
[3] Mohammad Ilyas (Ed.), The Handbook of Ad Hoc Wireless Networks CRC Press LLC, Florida, 2003.
[4] Tarun Dalal, Gopal Singh, An Analysis of ASRP Secure Routing Protocol for MANET, IJCSMS International Journal of Computer Science & Management Studies, Vol. 12, Issue 02, Apr. 2012, pp. 132-137.
[5] Panagiotis Papadimitratos et al., Secure Routing for Mobile Ad hoc Networks, Proceedings of the SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS 2002), San Antonio, TX, Jan.2002.
[6] Marshall, j et al., Identifying flaws in the secure routing protocol. IEEE Conference on Performance, Computing, and Communications, 2003.
[7] Oscar F. Gonzalez, Michael Howarth, and George Pavlou, Detection of Packet Forwarding Misbehavior in Mobile Ad-Hoc Networks, Springer-Verlag Berlin Heidelberg 2007.
[8] Eichler, S. et al., Secure Routing in a Vehicular Ad Hoc Network. IEEE Vehicular Technology Conference, Sep. 2004.
[9] M. Rajesh Babu et al., An Energy Efficient Secure Authenticated Routing Protocol for Mobile Ad hoc Networks. International Journal of Reviews in Computing, Vol. 7, Sep. 2011.
[10] Steffen Reidt et al., Efficient, Reliable and Secure Distributed Protocols for MANETs. 2nd International Conference on Mobile Technology, Applications and Systems, Nov. 2005.
[11] Muhammad Nawaz Khan et al., Intrusion Detection System for Ad hoc Mobile Networks. Int. Conf. on Information Technology: Research and Education, Jun.2005.
[12] Lu Jin et al., Implementing and Evaluating An Adaptive Secure Routing Protocol for Mobile Ad Hoc Network. Wireless Telecommunications Symposium, Apr. 2006.
[13] Panagiotis Papadimitratos et al., How to Specify and How to Prove Correctness of Secure Routing Protocols for MANET, IEEE-CS Broad Nets, San Jose, CA, USA, Oct. 2006
[14] Shivasharanappa Allur et al., Efficient SNR/RP Attentive Routing Algorithm: Cross-Layer Design for Ad hoc Networks, IEEE International Conference on Mobile Ad hoc and Sensor Systems (MASS), Oct.2006.
[15] Venkat Balakrishnan, Vijay Varadharajan, Uday Tupakula, and Phillip Lucs, TEAM: Trust Enhanced Security Architecture for Mobile Ad-hoc Networks, IEEE International Conference on Networks, Nov. 2007, pp. 182-187.
[16] Kimaya Sanzgiri, Bridget Dahill, A Secure Routing Protocol for Ad Hoc Networks, IEEE Conference on Network Protocols, Nov. 2002, pp. 78-87.
[17] Poonam Yadav, Rakesh Kumar Gill, Naveen Kumar, A Fuzzy Based Approach to Detect Black Hole Attack, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, Jul. 2012.
[18] Parma Nand et al. (2007) Performance study of Broadcast based Mobile Adhoc Routing Protocols AODV, DSR and DYMO, International Symposium on Wireless Pervasive Computing, Feb.2007.
[19] David B. Johnson et al., Dynamic Source Routing in Ad Hoc Wireless Networks, International Conference on Wireless Communications, Networking and Mobile Computing Sep.2007.
[20] Xiaodong Lin et al., ASRPAKE: An Anonymous Secure Routing Protocol with Authenticated Key Exchange for Wireless Ad Hoc Networks, IEEE International Conference on Communications, Jun. 2007.

[21] Xu Su et al., On Mitigating In-band Wormhole Attacks in Mobile Ad Hoc Networks, IEEE International Conference on Communications, Jun.2007.
[22] Mohd. Anuar Jaafar et al., Performance Comparisons of AODV, Secure AODV and Adaptive Secure AODV Routing Protocols in Free Attack Simulation Environment. Communications Magazine, IEEE ,Feb. 2008.
[23] Umang singh et al., Secure Routing Protocols in Mobile Ad Hoc Networks-A Survey and Taxonomy, IEEE Conference on Wireless Communications and Networking, Mar.2008.
[24] Julien Francq et al., Error Detection for Borrow-Save Adders Dedicated to ECC Unit. , 5th Workshop on Fault Diagnosis and Tolerance in Cryptography, Aug.2008.
[25] Karim El Defrawy et al., Privacy-Preserving Location-Based On-Demand Routing in MANETs, Third International Conference on Risks and Security of Internet and Systems, Oct.2008.
[26] Satoshi Kurosawa et al., Detecting Black Hole Attack on AODV-based Mobile Ad Hoc Networks by Dynamic Learning Method, 16th IEEE International Conference on Networks, Dec. 2008.
[27] Amit N. Thakare et al., Performance Analysis of AODV & DSR Routing Protocol in Mobile Ad hoc Networks, International Conference on Advanced Information, Networking and Applications, May. 2009.
[28] Kimaya Sanzgiri et al., A Secure Routing Protocol for Ad Hoc Networks, 18th International Conference on Computer Communications and Networks, Aug.2009.
[29] Claude Crrepeau et al., A secure MANET routing protocol with resilience against byzantine behaviours of malicious or selfish nodes. IEEE Transactions on Vehicular Technology, Jan.2009
[30] Liana Khamis Qabajeh et al., A Scalable, Distributed and Secure Routing Protocol for MANETs, International Conference on Computer Technology and Development, Nov. 2009.
[31] Feng He et al., S-MAODV: A Trust Key Computing Based Secure Multicast Ad-hoc On Demand Vector Routing Protocol, 3rd IEEE International Conference on Computer Science and Information Technology, Jul. 2010.
[32] Arun Kumar Mondal et al., The Success of Data Transmission in Multipath Routing for MANET, 2nd International Asia Conference on Informatics in Control, Automation and Robotics, Mar. 2010.
[33] Pietro Michiardi et al., Simulation-based Analysis of Security Exposures in Mobile Ad Hoc Networks, 2nd International Conference on Education Technology and Computer (ICETC), Jun. 2010.
[34] R. Kalpana et al., Mobile Anonymous Trust Based Routing Using Ant Colony Optimization, IEEE Asia-Pacific Services Computing Conference (APSCC), 2010.
[35] Mike Burmester et al., Towards provable security for route discovery protocols in mobile ad hoc networks. Global Information Infrastructure Symposium(GIIS),2011, Aug. 2011.
[36] Himani Yadav et al., A Review on Black Hole Attack in MANETs., 13th International Conference on Advanced Communication Technology (ICACT), Feb.2011
[37] Jiajia Liu et al., Multicast Capacity, Delay and Delay Jitter in Intermittently Connected Mobile Networks. IEEE Global Telecommunications Conference (GLOBECOM 2011),Dec.2011.
[38] Stefaan Seys et al., ARM: Anonymous Routing Protocol for Mobile Ad hoc Networks, International Symposium on Computer Networks and Distributed Systems (CNDS), Feb. 2011.
[39] Subash Chandra Mandhata et al., A counter measure to Black hole attack on AODV based Mobile Ad-Hoc Networks, IEEE International Conference on Networking, Sensing and Control (ICNSC), Apr. 2011.
[40] Saikat Chakrabarti et al., Authenticating DSR Using a Novel Multi signature Scheme Based on Cubic LFSR Sequences. Date of Conference: 26-28 Sept. 2011
[41] K.Seshadri Ramana et al., A Trust-Based Secured Routing Protocol for Mobile Ad hoc Networks. , International Conference on Recent Trends In Information Technology (ICRTIT), Apr.2012.
[42] Sridhar Subramanian et al., Trust Based Scheme for QoS Assurance in Mobile Ad-Hoc Networks, Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP), May 2012.
[43] S. Corson and J. Macker. Mobile ad hoc networking (MANET): Routing protocol performance issues and evaluation considerations. Internet Request for Comments (RFC 2501), January 1999.
[44] Lin X, Stojmenovic I ,GEDIR: Loop-Free Location Based Routing in Wireless Networks, International Conference on Parallel and Distributed Computing and Systems, pp.1025–1028
[45] Ko Y-B, Vaidya NH, Location-Aided Routing (LAR) in Mobile Ad Hoc Networks. Wireless Networks, Volume 6, pp.307–321
[46] Karp B, Kung HT , GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. ACM MOBICOM 2000, pp. 243–254
[47] Liao W-H, Tseng Y-C, Lo K-L, Sheu J-P, GeoGRID: A Geocasting Protocol for Mobile Ad Hoc Networks based on GRID, Journal of Internet Technology, Volume 1, Issue 2,pp. 23–32.
[48] Jain R, Puri A, Sengupta R , Geographical Routing Using Partial Information for Wireless Ad Hoc Networks. IEEE Personal Communications, Volume 8, Issue 1, pp. 48–57.
[49] Ko Y-B, Vaidya NH , Location-based multicast in mobile ad hoc networks. Technical Report, Texas A&M University
[50] Sinha P, Sivakumar R, Bharghavan V, MCEDAR: Multicast Core-Extraction Distributed Ad Hoc Routing, IEEE WCNC, Volume 3, pp. 1313–1317.
[51] Wu CW, Tay TC, AMRIS: A Multicast Protocol for Ad Hoc Wireless Networks, IEEE MILCOM 1999, Volume 1, pp. 25–29
[52] Toh C-K, Guichal G, Bunchua S, ABAM: On-Demand Associativity-Based Multicast Routing for Ad Hoc Mobile

Networks. Proceedings of IEEE VTS-Fall VTC 2000, Volume 3, pp. 987–993

[53] Royer EM, Perkins CE, Multicast Ad Hoc On-Demand Distance Vector (MAODV) Routing. IETF Draft, draft-ietf-manet-maodv-00, 15 Jul. 2000,

[54] Ji L, Corson MS, Differential Destination Multicast-A MANET Multicast Routing Protocol for Small Groups, IEEE INFOCOM 2001, Volume 2, pp. 1192–1201.

[55] Lee S, Su W, Gerla M, On-Demand Multicast Routing Protocol in Multihop Wireless Mobile Networks, ACM/Kluwer Mobile Networks and Applications (MONET), volume 7, Issue 6,pp. 441–453.

[56] Jetcheva JG, Johnson DB, Adaptive Demand-Driven Multicast Routing in Multi-Hop Wireless Ad Hoc Networks. ACM MobiHoc 2001, pp. 33–44.

[57] Xie J, Talpade RR, Mcauley A, Liu M, AMRoute: Ad Hoc Multicast Routing Protocol. ACM Journal of Mobile Networks and Applications, Dec.2002, Volume 7, Issue 6, pp. 429–439.

[58] Das SK, Manoj BS, Murthy CSR, A Dynamic Core Based Multicast Routing Protocol for Ad Hoc Wireless Networks, ACM MobiHoc 2002,pp. 24–35.

[59] Sisodia RS, Karthigeyan I, Manoj BS, Murthy CSR,A Preferred Link Based Multicast Protocol for Wireless Mobile Ad Hoc Networks, IEEE ICC 2003, Volume 3, pp. 2213–2217.

[60] Quintero A, Pierre S, Macabe´o B, A routing protocol based on node density for ad hoc networks. Ad Hoc Networks, Volume 2, Issue 3,pp. 335–349.

[61] Saigal V, Nayak AK, Pradhan SK, MallR, Load balanced routing in mobile ad hoc networks, Computer Communications,1995, Volume 27, Issue 3, pp. 295–3054.

[62] Wei X, Chen G, Wan Y, Mtenzi F, Optimized priority based energy efficient routing algorithm for mobile ad hoc networks, Ad Hoc Networks, Volume 2, Issue 3, pp. 231–239.

[63] Wang N-C, Chang S-W, A reliable on-demand routing protocol for mobile ad hoc networks with mobility prediction. Computer Communications, 2005, Volume 29, Issue 1, pp. 123–135.

[64] Bu¨ r K, Ersoy C, Ad hoc quality of service multicast routing, Computer Communications, 2005, Volume 29, Issue 1, pp. 136–148.

[65] Boukerche A, El-Khatib K, Xu L, Korba L An efficient secure distributed anonymous routing protocol for mobile and wireless ad hoc networks. Computer Communications, 2005, Volume 28, Issue 10, pp. 1193–1203.

[66] Moaveninejad K, Song W-Z, Li, X-Y Robust position-based routing for wireless ad hoc networks, Ad Hoc Networks, 2005, Volume 3, Issue 5, pp. 546–559.

[67] Rango FD, Gerla M, Marano S, A scalable routing scheme with group motion support in large and dense wireless ad hoc networks. Computers & Electrical Engineering, 2006 Volume 32, Issues 1–3, pp. 224–240.

[68] Ghosh RK, Garg V, Meitei MS, Raman S, Kumar A, Tewari N, Dense cluster gateway based routing protocol for multi-hop mobile ad hoc networks. Ad Hoc Networks, 2006, Volume 4, Issue 2, pp. 168–185.

[69] Wang Y-H, Chao C-F, Dynamic backup routes routing protocol for mobile ad hoc networks. Information Sciences, 2006, Volume 176, Issue 2, pp. 161–185.

[70] Ahn CW, Gathering-based routing protocol in mobile ad hoc networks. Computer Communications, 2006, Volume 30, Issue 1, pp. 202–206.

[71] Sun B, Li L, QoS-aware multicast routing protocol for Ad hoc networks, Journal of Systems Engineering and Electronics, 2006, Volume 17, Issue 2, pp. 417–422.

[72] Eisbrener J, Murphy G, Eade D, Pinnow CK, Begum K, Park S, Yoo SM, Youn J-H, Recycled path routing in mobile ad hoc networks. Computer Communications, 2006, Volume 29, Issue 9, pp. 1552–1560.

[73] Layuan L, Chunlin L, A QoS multicast routing protocol for clustering mobile ad hoc networks, Computer Communications, 2007, Volume 30, Issue 7, pp. 1641–1654.

[74] Lu R, Cao Z, Wang L, Sun C, A secure anonymous routing protocol with authenticated key exchange for ad hoc networks, Computer Standards & Interfaces, 2007, Volume 29, Issue 5, pp. 521–527.

[75] Giruka VC, Singhal M, A self-healing On-demand Geographic Path Routing Protocol for mobile ad-hoc networks, Ad Hoc Networks, 2007, Volume 5, Issue 7, pp. 1113–1128.

[76] Wang N-C, Huang Y-F, Chen J-C, A stable weight-based on demand routing protocol for mobile ad hoc networks, An International Journal of Information Sciences, 2007, Volume 177, Issue 24, pp. 5522–5537.

[77] Yang C-C, Tseng L-P, Fisheye zone routing protocol: A multilevel zone routing protocol for mobile ad hoc networks, Computer Communications, 2007, Volume 30, Issue 2, pp. 261–268.

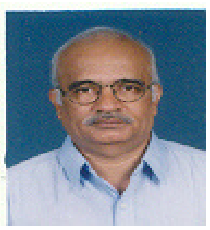
[78] Min C-H, Kim S, On-demand utility-based power control routing for energy-aware optimization in mobile ad hoc networks. Journal of Network and Computer Applications, 2007, Volume 30, Issue 2, pp. 706–727.

[79] Song J-H, Wong VWS, Leung VCM, Secure position-based routing protocol for mobile ad hoc networks, Ad Hoc Networks, 2007, Volume 5, Issue 1, pp. 76–86.

[80] Reddy LR, Raghavan SV, SMORT: Scalable multipath ondemand routing for mobile ad hoc networks, Ad Hoc Networks, Volume 5, Issue 2, pp. 162–188.

[81] Zhao Y, Chen Y, Li B, Zhang Q, Hop ID: A Virtual Coordinate-Based Routing for Sparse Mobile Ad Hoc Networks. IEEE Transactions on Mobile Computing, 2007, Volume 6, Issue 9, pp. 1075–1089.

Sumati Ramakrishna Gowda has obtained B.E Dégrée in Computer Science from Mysore, Master of Science in Information Technology from KSOU Mysore, and she also obtained Master of Philosophy in Computer Science from Madhurai Kamraj University, Madurai. She is currently working as a Lecturer, DOS in Computer Science, at the Karnataka State Open University, Mysore, India. KSOU was inspired by the concept of open learning and distance education, also providing IT education to a host of students at different levels. Currently pursuing her research work in the area of Security Approaches in Routing Protocol.



Dr. P. S. Hiremath is Ph.D. in Applied Mathematics , from Karnataka University, Dharwad, Karnataka, India. He is Masters of Science in Applied Mathematics, Karnataka University, Dharwad, Karnataka, India. He had been in the Faculty of Mathematics and Computer Science of various institutions in India, namely, National Institute of Technology, Surathkal (1977-79), Coimbatore Institute of Technology, Coimbatore (1979-80), National Institute of Technology, Tiruchinapalli (1980-86), Karnatak University, Dharwad (1986-1993) and has been presently working as Professor and Chairman of Computer Science in Gulbarga University, Gulbarga(1993 onwards). His research areas of interest are Computational Fluid Dynamics, Optimization Techniques, Image Processing and Pattern Recognition, and Computer Networks. He has published 156 research papers in peer reviewed International Journals and proceedings of International Conferences.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Association Technique based on Classification for Classifying Microcalcification and Mass in Mammogram

Herwanto[1], Aniati Murni Arymurthy[1]

[1]Faculty of Computer Science, University of Indonesia,
Kampus Baru UI, Depok 16424, Indonesia

Abstract

Currently, mammography is recognized as the most effective imaging modality for breast cancer screening. The challenge of using mammography is how to locate the area, which is indeed a solitary geographic abnormality. In mammography screening it is important to define the risk for women who have radiologically negative findings and for those who might develop malignancy later in life. Microcalcification and mass segmentation are used frequently as the first step in mammography screening.

The main objective of this paper is to apply association technique based on classification algorithm to classify microcalcification and mass in mammogram. The system that we propose consists of: (i) a preprocessing phase to enhance the quality of the image and followed by segmentating region of interest; (ii) a phase for mining a transactional table; and (iii) a phase for organizing the resulted association rules in a classification model. This paper also illustrates how important the data cleaning phase in building the data mining process for image classification. The proposed method was evaluated using the mammogram data from Mammographic Image Analysis Society (MIAS). The MIAS data consist of 207 images of normal breast, 64 benign, and 51 malignant. 85 mammograms of MIAS data have mass, and 25 mammograms have microcalcification. The features of mean and Gray Level Co-occurrence Matrix homogeneity have been proved to be potential for discriminating microcalcification from mass. The accuracy obtained by this method is 83%.

Key words: breast cancer, classification, data mining, association rule, Region of Interest (ROI), Gray Level Co-occurrence Matrix (GLCM).

1. Introduction

Breast cancer is a chronic disease, whereas its total healing remains doubtful. It could also take a long time of medical treatment and money consuming. Breast cancer can be diagnosed with various screening mechanisms, including by mammography, a technique for observing X – Ray photo which is proven to be effectively giving a clue of abnormality in the breast. However understanding the mammogram images in the phase of diagnosing is not an easy job. There are many phases to be carried out, such as: the image preprocessing, segmentation, pattern recognition, classification and conclusion. Comprehensive knowledge in many fields of science is needed, including a vast specialized experience, particularly to attain relevant specific characters of breast cancer. By diagnosing breast cancer effectively, it can be detected in an earlier step.

Mass and microcalcification are two confusing features present in a mammogram. Masses are identified by their shape and margin characteristics. Microcalcifications are small calcium deposit and appear as a group of bright spots in mammogram. Most of the breast cancer is detected by the presence of microcalcification [8]. Mammogram readings are usually performed by a radiologist. Variety of factors such as the poor quality of the image, eye fatigue factor, the performance of radiologist could greatly affect the diagnose results. To overcome this problem, several computer aided diagnosis systems are developed for automatic detection of breast cancer.

Breast cancer research has been continuing recently, and the challenge has been shifted from gathering data to finding hidden patterns and trends that are most relevant to cancer diagnosis. This paper presents a method for building a classification system, especially to obtain the different characteristics of mass and microcalcification using association technique based on classification. The classification process mainly consists of two phases: training phase and testing phase. In the training phase the properties of typical image features are isolated and the training samples are having their object class information.

In this study, the statistics features and the Gray Level Co-occurrence Matrix (GLCM) based on texture features are extracted. Feature selection is an important step before the process of any classification scheme. In term of the data used, the importance of data cleaning is considered. The data cleaning is applied in the context of image content mining.

This study finds the different characteristics of microcalcification and mass. The information of centroid and radius provided in MIAS database are used to find and extract the microcalcification and mass. The finding is performed by conducting a data mining process to detect the candidate regions in mammogram into two categories: microcalcification and mass regions. This finding will help radiologists in their tasks to detect subtle abnormalities in a mammogram. Furthermore, it is also helps in making a decision whether a biopsy procedure is necessary.

A decision of the candidate regions were carefully concluded, especially in selecting the threshold value and size of the candidate regions in order to obtain the candidate regions that highly resemble microcalcification or mass. Unfortunately, there are large variations of threshold value from one image to the others, so that a constant threshold will not be good enough.

The paper is organized as follows. Section 1 is an introduction and background. Section 2 presents a review of existing techniques for mammographical feature analysis. Section 3 provides the detail proposed methodology for microcalcification and mass detection. Section 4 discusses the experiment results. Finally, Section 5 presents the concluding remarks.

2. Related Works

Various computer aided diagnosis algorithms have been proposed for analysis mammograms. These algorithms are based on extracting image features from regions of interest and estimating the probability of malignancy for a given mammogram. A variety of features and classification schemes have been used to automatically discriminate between benign and malignant mammogram. GLCM is generally used for extracting features. Khuzi *et al.* have used GLCM to identify the mass region in mammogram [2].

Many research works have used data mining technique to analysis mammogram. Researches that use data mining approach to classify can be found in [11, 12]. Most of them classify a mammogram into benign or malign, and the candidate regions are captured from the whole original image. Luiza *et al.* [11] proposed a classification method based on association rule mining. The original image was split initially in four parts, for a better localization of the region of interest. And the extracted features were discretized over an interval before organizing the transactional data set. Aswini *et al.* [4] proposed an image mining techniques using mammograms to classify and detect the cancerous tissue. The mammogram image is classified into normal, benign and malignant class and to explore the feasibility of data mining approach

Zhang *et al.* [9] have proposed, an automatic segmentation method using a modified region-growing technique. The outputs are a number of segmented regions considered being the most suspicious. Karahaliou *et al.* [3] have investigated whether or not texture properties of the tissue surrounding microcalcification clusters can contribute to breast cancer diagnosis. The steps of the proposed method are as follows: (i) preprocess using a wavelet-based spatially adaptive method for mammographic contrast enhancement; (ii) thresholding technique to exclude microcalcification area; and (iii) texture analysis the tissue surrounding microcalcification area. The study have used four models for textural features extraction, i.e : First Order Statistics (FOS), Grey Level Co-occurrence Matrices (GLCM), Grey Level Run Length Matrices (GLRLM) and Law's Texture Energy Measures (LTEM). The tissue surrounding the microcalcification area are classified using k-Nearest Neighbor (kNN) method. The texture analysis of the tissue surrounding the microcalcification area has shown a promising result in computer-aided diagnosis of breast cancer and may contribute to the reduction of benign biopsies.

Khuzi *et al*. [2] proposed a simple Computer Aided Design (CAD) system to automatically detect areas that have a high probability of masses in digital mammogram. The texture features are extracted using GLCM. The matrices are constructed at a distance of $d = 1$ and for direction of θ given as 0°,45°,90° and 135°. A single direction might not give enough and reliable texture information. For this reason, four directions are used to extract the texture information for each masses and non masses tiles area. The results show that the GLCM at 0º, 45º, 90º and 135º with a block size of 8x8 give significant texture information to identify between masses and non-masses tissues. The detection process of the candidate regions is done based on block processing windows or tiles. The entire mammogram is divided into tiles area before feature extraction is done to each tile. Thus, the segmented image is divided into tiles with a size of 32x32 and a tile would be categorized as suspicious if its average intensity is more than 200. The intensity comparison is applied to each region in the segmented image and regions or tiles that do not fall into this category are rejected. The 32x32 windows are divided into smaller windows with size of 8x8. Furthermore, a tile is considered to be suspicious if its average intensity within 8x8 tiles is more than 210. After that, its texture criteria are evaluated. The tiles are considered as masses if their texture criteria values are within the defined range of masses texture values.

3. Methodology

The overview processes in the proposed methodology is presented in Figure 1.

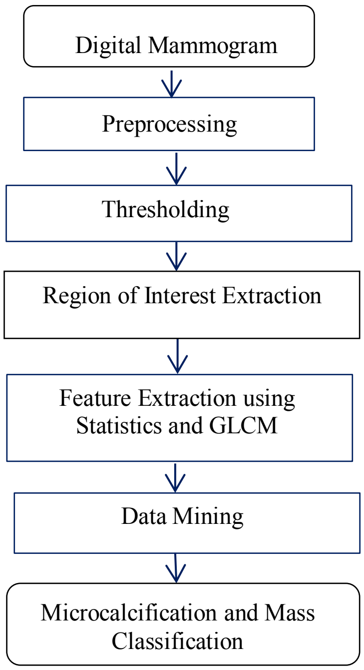


Figure 1. Proposed Methodology

3.1. Digital Mammogram

Mammogram Image Analysis Society (MIAS) database were used for experimental data. The total data consists of 322 images. The database provides some information on abnormal classes types, and the position of individual mass and microcalcification. The mammograms were downloaded from MIAS database.

3.2. Preprocessing

Mammogram is a medical image that difficult to be interpreted, therefore the phase of preprocessing is needed. It is required to increase the quality of the image in order to make the object extraction process easier. The first phase of preprocessing is removing unnecessary parts of mammogram, such as unnecessary background, before improving the quality of image. The image enhancement is used to normalize the extremely bright or dark effect. In this experiment, the method of Contrast Limited Adaptive Histogram Equalization (CLAHE) is used. Figure 2(a-b) shows the original mammogram and the result of cropping process. Figure 2(c-d) shows the morphological opening and the histogram equalization processes.

3.3. Thresholding Value

Thresholding technique is a widely used method for segmentation. It is useful for discriminating objects from background. The simple way is using a technique called global thresholding, where one threshold value is selected for the entire image. All the gray level values below a threshold value will be classified as black (0), and those above threshold value will be white (1). Figure 2(e) shows the thresholding process.

3.4. Region of Interest Extraction

The detection process is important to select suspect area which looks like microcalcification or mass. The election of ROI is based on intensity value and size of ROI, and its texture structures. The ROI election process consists of 2 phases. The first phase is to determine the suspected microcalcification or mass area based on a value of global thresholding and the size of ROI, and mark the connected components. The segmented image is divided into tiles and a tile would be categorized as suspicious if its average intensity is more than 200 [2] and the size of tile is more than 5 millimeter [14]. Figure 2(e) shows the process of marking the connected components.

3.5. Feature Extraction

Texture feature measures the variation in the surface of the image. Two types of measure are the first order and the second order statistics. The ROI feature image using the second order statistic is used as the input for GLCM feature extraction.

GLCM describe the frequency of one gray tone appearing in a specified spatial linear relationship with another gray tone within the area of investigation [1]. The co-occurrence matrix is computed based on two parameters: the relative distance d between the pixel pair d and their relative orientation ф. Normally, ф is quantized in four directions (0^0, 45^0, 90^0 and 135^0).

Several statistics are derived from the co-occurrence matrix. These statistics provide information about the texture of an image. In the classical paper [17], Haralick *et al.* have introduced fourteen textural features from the GLCM and the reference stated that only six of the textural features are considered to be the most relevant. Those textural features are Energy, Entropy, Contrast, Variance, Correlation and Homogeneity. These textural features are computed based on the frequency or repetition of the pixel pair in the co-occurrence matrix.

The feature extraction phase is needed in order to create the transactional database to be mined. The extracted features

were organized in a database, which becomes the input for the mining phase. The features that are used in the experiment consist of four statistical features: mean, variance, skewness, kurtosis, and four GLCM's features: contrast, correlation, energy, and homogeneity. All the extracted features have been computed in 8x8 region of interest as shown in figure 3.
The expressions of the GLCM descriptors are as follows.

$$\text{Contrast} = \sum_{i,j}|i-j|^2 p(i,j) \quad (1)$$

$$\text{Correlation} = \sum_{i,j}\frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j} \quad (2)$$

$$\text{Energy} = \sum_{i,j} p(i,j)^2 \quad (3)$$

$$\text{Homogeneity} = \sum_{i,j}\frac{p(i,j)}{1+|i-j|^2} \quad (4)$$

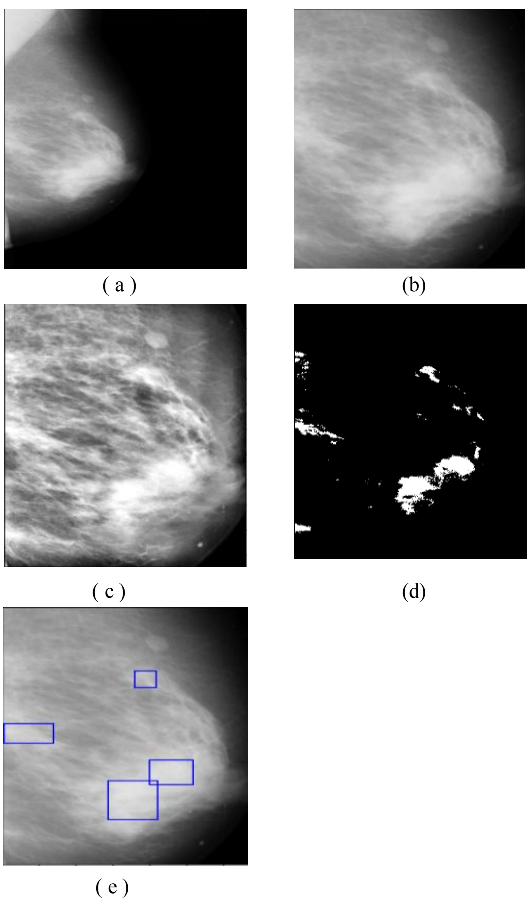


Figure 2. Results of preprocessing phase: (a) original image; (b) crop operation; (c) histogram equalization operation; (d) threshold operation; (e) marking the connected components operation.

3.6. Data Mining

Data mining is the process of discovering meaningful new correlations, patterns, and trends by digging into large amounts of data stored in warehouses. In the experiment the data mining process initially scans the pixels of digital mammograms in the database. This process seeks to discover common and indicative pixel patterns among microcalcification and mass. Using pixel values to search for microcalcification and mass is an important development because the digital mammogram contain subtleties that are not easy to perceive.

The data mining process consists of three steps: (i) the handling of missing data; (ii) the discretization step of data values; (iii) the rule mining and classification.

In this study, the discretization step is done by using k-means clustering algorithm. The k-means clustering algorithm divides a data set into a predetermined number of clusters k. The statistical mean is used as a measure for describing the average location of all members of a particular cluster. Table 1 shows an example of transactional table after the discretization process.

One of the effective algorithm of applying association rules for classification is the Classification based on Predictive Association Rules (CPAR) [6]. The algorithm CPAR is started by reading the data in the form of two-dimensional array which every column represents the attribute where the last attribute represent the object class. The input data is grouped into positive example P and negative example N related to their classes. Weight of positive example | P| and negative example | N| for each attribute is summed up to form PN array. The minimum weight threshold for P is calculated by multiplying the start weight of P by the Total Weight Threshold (TWT) which was set to 0.05 during experimentation. The flowchart of CPAR algorithm is shown in figure 3.

The Laplace accuracy which measures the strength of prediction is calculated using the formula :

$$\text{L.A} = (nc+1) / (ntot+f) \quad (5)$$

where f is the number of classes, $ntot$ is the total number of examples satisfying the rule's body, among which nc example belong to c, which is the predicted class of the rule [6].

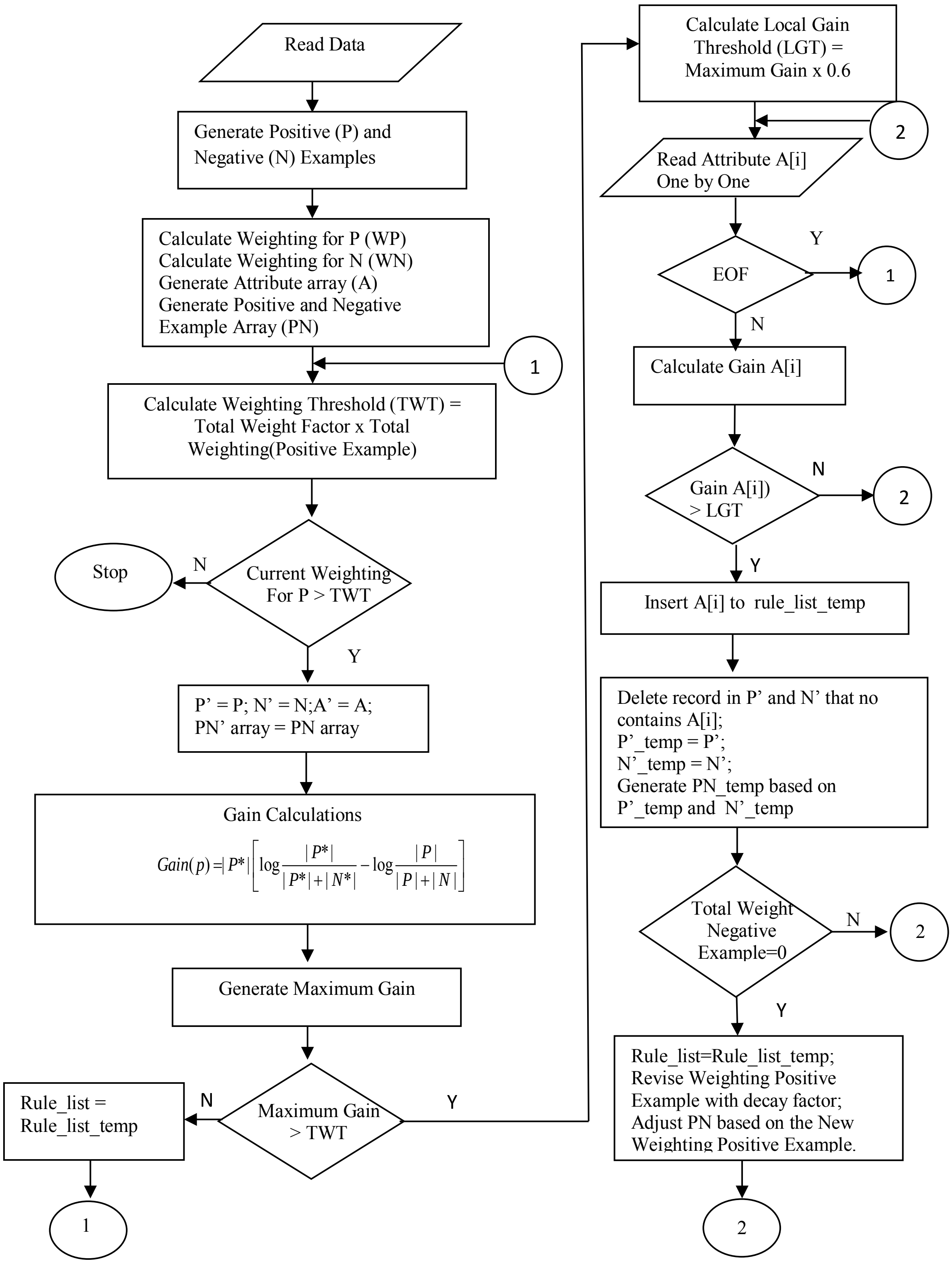


Figure 3. Flowchart of CPAR algorithm

Table 1. Feature's transactional table after discretization

Mean	Median	Variance	Kurtosis	Skewness	Contrast	Correlation	Energy	Homogeneity	Class
01	06	11	18	23	28	31	39	44	46
05	10	11	16	25	28	33	36	44	46
03	08	11	20	25	30	31	36	41	46
04	09	11	16	21	27	34	36	43	46
01	06	11	20	21	28	35	39	44	46
05	10	11	16	23	29	34	36	44	46
04	09	11	20	24	30	35	36	41	46
05	10	11	18	25	27	34	36	43	46
01	06	11	16	25	29	35	39	43	47
03	08	13	20	21	30	33	40	44	47
01	10	11	18	23	29	35	36	44	47
05	10	11	16	24	29	33	39	43	47
05	10	11	20	21	29	34	36	44	47
04	09	11	18	23	29	34	39	43	47
01	06	14	17	23	28	35	36	41	47
03	08	11	20	25	30	31	36	41	47

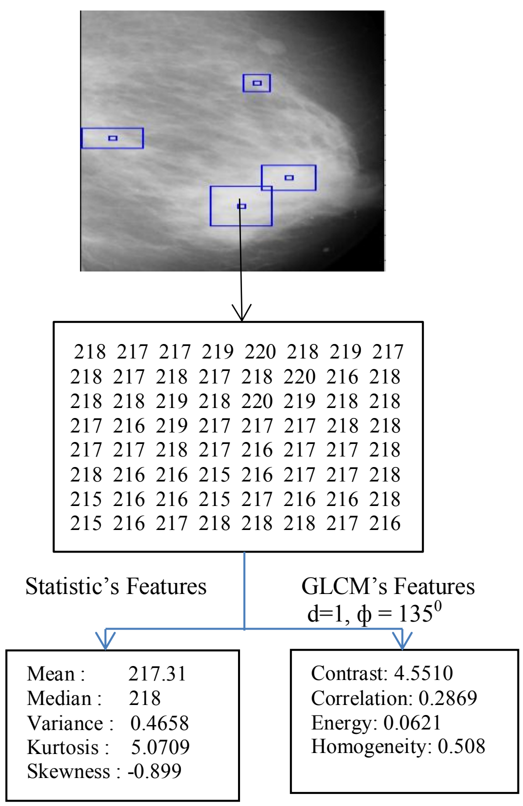


Figure 4. Feature Extraction Process

4. Experimental Results

There are 322 images of MIAS data used in the experiment. The data in the collection consists of the location information of microcalcification, mass, and their radius. It could be summarized that the data processing consists of: preprocessing, a transactional database organization, and mining the transactional database. CPAR algorithm has been used for classifying microcalcification and mass. Table 1 shows a typical example of transactional database after the discretization process. There are ten columns which every column is given a specific attribute. The columns consist of five statistical feature attributes: mean, median variance, kurtosis, skewness, and four GLCM's feature attributes: contrast, correlation, energy, homogeneity. The last attribute is the object class. The positive class of microcalcification and the positive class of mass is determined based on the information found in the MIAS. In this example, the positive class of microcalcification is 47 and the positive class of mass is 46. For the case of using 145 mammogram, the features of homogeneity and mean can be used to classify microcalcification or mass with a maximum accuracy of 83%. The testing result using the association of a selected feature set is shown in Table 2. Table 2 presents five generated rules for classification. There are no contradicting rules in either results. Microcalcification has mean value higher than mass, it means microcalcification are brighter than mass. Mass has a higher homogeneity value than microcalcification, it means mass has more homogeneous texture than microcalcification.

Table 2. The obtained rules for classification

No	Rule	L.A
1	IF Mean between 185.74 and 203.74 Then Positive Microcalcification	0.83
2	IF Homogeneity between 0.61 and 0.65 Then Positive Microcalcification	0.80
3	IF Mean between 167.53 and 184.53 Then Positive Microcalcification	0.69
4	IF Mean between 140.82 and 165.82 Then Positive Mass	0.68
5	IF Homogeneity between 0.65 and 0.71 Then Positive Mass	0.67

5. Proposed Improvement Technique

In the previous section, it was shown how homogeneity in GLCM can be used to distinguish between microcalcification and mass. This section will describe the improvement process to increase accuracy by modifying the homogeneity computation.

The co-occurrence matrix is computed based on two parameter, which are the relative distance between the pixel pair d measured in pixel number and their relative orientation ϕ. Normally, ϕ is quantized in four directions (horizontal: 0^0, diagonal 45^0, vertical: 90^0 and anti-diagonal 135^0). In this experiment, for each d, the resulting values for the four directions are averaged, and counting the number of the pixel pair that occurred in the whole image at relative distance d and direction ϕ. Table 4 shows the GLCM for d=1 and $\phi=0^0$. Table 5 shows the homogeneity matrix, where the resulting homogeneity is diagonal symmetry. Homogeneity matrix is computed with the formula 4:

The method computes weight matrix using Matlab is as follow :

```
mw = zeros(p,q);
[glcm,SI] = graycomatrix(input,'offset',[0
Dthresh],'G',[],'symmetric',true);
mglcm = glcm/sum(sum(glcm));
for j = 1 : p
for k = 1 : q
mw(j,k) = mglcm(j,k)/(1+abs(j-k));
hmat(j,k) = 1/(1+abs(j-k));
end
end
mw = mw;
mwt = triu(mw);
[row col] = find(hmat >= Lthresh);
position = [row col];
position1 = sortrows(position);
[uk bar] = size(position1);
sum = 0;
for su = 1 : uk
sum = sum + mw(position1(su,1),position1(su,2));
end
end
```

Table 3. Input Matrix

198	194	194	193	190	189	187	189
197	196	196	197	195	193	189	188
199	199	200	201	200	196	190	188
199	198	201	201	201	198	194	193
197	197	199	200	200	197	194	194
196	196	196	198	200	197	192	192
195	193	193	195	196	194	190	188
193	191	191	193	194	190	187	186

Table 4. GLCM Matrix

2	2	1	0	0	0	0	0
2	2	3	1	0	0	0	0
1	3	2	3	2	1	0	0
0	1	3	4	6	1	0	0
0	0	2	6	4	4	2	0
0	0	1	1	4	12	2	3
0	0	0	0	2	2	4	5
0	0	0	0	0	3	5	10

Table 5. Homogeneity Matrix

1.00	0.50	0.33	0.25	0.20	0.17	0.14	0.13
0.50	1.00	0.50	0.33	0.25	0.20	0.17	0.14
0.33	0.50	1.00	0.50	0.33	0.25	0.20	0.17
0.25	0.33	0.50	1.00	0.50	0.33	0.25	0.20
0.20	0.25	0.33	0.50	1.00	0.50	0.33	0.25
0.17	0.20	0.25	0.33	0.50	1.00	0.50	0.33
0.14	0.17	0.20	0.25	0.33	0.50	1.00	0.50
0.13	0.14	0.17	0.20	0.25	0.33	0.50	1.00

It is shown in Table 5 that the homogeneity values increase when becoming closer to the diagonal. The proposed technique computed the homogeneity value based on the threshold distance p main diagonal. This means that the computation only involving pixel pair within the threshold distance greater than or equals to p. From Table 3, if the threshold distance equals to 0.2, the homogeneity value is 0.5804, but if it is not use this technique, the homogeneity value is 0,6443. Table 6 presents the accuracy result after modifying homogeneity feature.

Table 6. Improvement Result

No	Rule	L.A
1	IF Mean between 185.74 and 203.74 Then Positive Microcalcification	0.83
2	IF Homogeneity between 0.5 and 0.56 Then Positive Microcalcification	0.82
3	IF Mean between 167.53 and 184.53 Then Positive Microcalcification	0.69
4	IF Mean between 140.82 and 165.82 Then Positive Mass	0.68
5	IF Homogeneity between 0.6 and 0.66 Then Positive Mass	0.73

6. Concluding Remarks

Mammography is one of the best methods in breast cancer detection. The emphasis of using mammography is to define an area with a different appearance from other areas in the breast. The segmentation of microcalcification and mass will help radiologist to judge the likelihood of the present of cancer and to recommend whether a biopsy is necessary. This paper has proposed a method that could be used for a specific application in detection microcalcification and mass in mammogram.

References

[1] A. Baraldi, F. Parmiggiani. “An Investigation Of The Textural Characteristics Associated With GLCM Matrix Statistical Parameters”, *IEEE Trans. on Geoscience and Remote Sensing*, vol. 33, no. 2, pp. 293-304, March 1995.

[2] A Mohd. Khuzi, BEng, R Besar, PhD, WMD Wan Zaki, MEngSc, NN Ahmad, PhD, “Identification of masses in digital mammogram using gray level co-occurrence matrices”, Biomedical Imaging and Intervention Journal, 2009.

[3] Anna N Karahaliou, Ioannis S Boniatis, Spyros G Skiadopoulos, Filippo N, “A Texture Analysis Approach for Charaterizing Microcalcifications on Mammograms”, School of Medicine, University of Patras, Greece, 2006.

[4] Aswini Kumar Mohanty, Sukanta Kumar Swain, Pratap Kumar Champati, Saroj Kumar Lenka, “Image Mining for Mammogram Classification by Association Rule Using Statistical and GLCM features”, International Journal of Computer Science Issue, 2011.

[5] Brijesh Verma, “A Neural Network Based Technique to Locate and Classify Microcalcification in Digital Mammograms”, School of Information Technology, Griffith University, Gold Coast Campus, Australia, 1998.

[6] Coenen F, “The LUCS-KDD Implementations of CPAR (Classification Based on Predictive Association Rules)”, Department of Computer Science The University of Liverpool, 2004.

[7] Gonzalez Rafael C, Richard E.Woods, “Digital Image Processing”, New Jersey, Prentice-Hall, Inc, 2002.

[8] Gunilla Svane, E. James Potchen, Arlene Sierra, Edward Azavedo, “Screening Mammography Breast Cancer Diagnosis in Asymptomatic Women”, St. Louis, Missouri 63146.

[9] Han Zhang, Say Wei Foo, S.M. Krisman & Choon Hua Thng, “Automated Breast Masses Segmentation in Digitized Mammograms”, School of Electrical and Electronics Engineering. Nanyang Technological University, National Cancer Center, Singapore, 2004.

[10] Jay Bhaskar & Poonam Rani Gupta, “Design of an Automated Cancer Detection System”, Proceeding of ASCNT-2009, CDAC, India, pp.9-5, 2009.

[11] Luiza M, O. Zaiane, A Coman, “Application of Data Mining Techniques for Medical Image Clasiification”, Department of Computing Science University of Alberte, Canada, 2001

[12] Luiza M, O. Zaiane, A Coman, “Mammography Classification by an Association Rule-based Classifier”, Department of Computing Science University of Alberte, Canada, 2002.

[13] Juha Koikkalainen, “Image Database in Medical Applications”, Department of Engineering Physics and Mathematics, Hensinki University of Technology, 2006.

[14] Lanyi M, “Diagnosis and differential diagnosis of breast calcification”, Sringer – Verlag, Berlin, 1986.

[15] Le-Gal M, Chavanne G, Pellier D, “Diagnostic value of clustered microcalcifications discovered by mammography (apropos of 227 cases with histological verification and without a palpable breast tumor)”, Bull Cancer 71(1): 57 – 64, 1984.

[16] M. M. Mokji, S.A.R. Abu Bakar, “Adaptive Thresholding Based On Cooccurrence Matrix Edge Information”, Journal of Computers, Vol. 2, No. 8, 2007.

[17] R. Haralick, K. Shanmugam, I. instein, “Texture Features For Image Classifi-cation”, *IEEE Transaction on SMC*, vol.3, no. 6, pp. 610-621, 1973.

[18] Robert J.Pizzutiello, Jr., M.S. & John E. Cullinan, R.T. (R), “Introduction to Medical Radiographic Imaging”, Health Science Division, Eastman Kodak Company, 1993.

A Cross-domain Authentication Protocol based on ID

Zheng Jun[1*], Guo Xianchen[1,2*], Zhang Quanxin[1], Zhang Qikun[1]
[1]Beijing Engineering Research Center of Massive language information processing and cloud computing, School of Computer Science and Technology, Beijing Institute of Technology, Haidian 100081, Beijing, P. R. China

[2]The Sixth Research Institute of China Electronics Corporation, Haidian 100083, Beijing, P. R. China

Abstract
In large distributed networks, many computers must be mutual coordination to complete some works Under the certain conditions, these computers may come from different domains. For ensuring the safety to access resources among these computers in different domains, we propose a cross-domain union authentication scheme. We compute a large prime cyclic group by elliptic curve, and use the direct decomposition of this group to decompose n automorphism groups ,and design an signcryption scheme between domains by bilinear of automorphism group to achieve cross-domain union authentication. this scheme overcome the complexity of certificate transmission and bottlenecks in the scheme of PKI-based, and it can trace the entities and supports two-way entities anonymous authentication, which avoid the authority counterfeiting its member to cross-domain access resources. Analyses show that its advantages on security and communication-consumption .

Key words: *signcryption,cross-domain authentication, elliptic curve,bilinear group.*

Ⅰ. Introduction

Cross-domain alliance is needed in large networks, which services and access points are located in multiple domains. In a distributed network environment where companies and institutions have their own sharing resource, in order to prevent unauthorized users to access these shared resources, every institution will set up a local certification service equipment to provide certification services when users access resources. Therefore, a relatively independent trust domain is formed in every institution, and the users that in a domain trust their certification center, and the certification center provides convenient authentication service for local users to access shared resources. However, in the case of in a large number of demand services, such as the demands of cloud computing, users need anytime and anywhere to access resources .In this case, a single domain is unable to meet the needs of resource requests, therefore it is need many domains mutual cooperation to achieve this requests. For this the requests of shared resource are not only from the internal members of the domain, but also from the other domains. When the foreign entities access to the resources in local domain ,there involve the scheme of cross-domain authentication.

The applications of cross-domain authentication in many fields, such as the authentication among multiple heterogeneous domains within a virtual organization under the grid and cloud computing environment[1][2], the roaming access authentication under the environment of wireless network, etc[3][4]. There are mainly two cross-domain authentication frameworks under specific environments: one is authentication framework (such as Kerberos)[5] [6]based on the symmetric key system. This scheme relates to the complexity of symmetric key management and key consultations, and cannot deal with the anonymous problem effectively. The other is authentication framework based on traditional PKI [7][8][9], The management of credentials under public key cryptography is a heavy burden in this scheme; specifically, the consumptions is caused by the construction of credential paths and the query of the status of credentials and transfer of credentials .It can also cause the network bottleneck of authentication center when under frequent cross-domain accesses. References[10][11][12] proposed an identity-based multi-domain authentication model, which is based on the trust of the authority of the other side, and it requires the key agreement parameters of all domains to be same, this have limitations and it could not avoid the authority faking members in its domain to cross-domain access resources. Reference [13][14] adopt signcryption to implement the authentication when users access resource each other within the same domain, it is confined to a single domain, so it is difficult to meet the needs of large-scale distributed computing. Reference [15] extends the scheme of reference [13],and make it to enable the members from the difference domains to authenticate each other, but the precondition of this solution is the hypothesis that PKG of every domain is honest. PKG possesses the private keys of all the members within its domain, and if PKG is malicious, the truth identity of user and the confidential of private key could not be guaranteed.

The cross-domain authentication alliance protocol proposed in this paper is designed based on inter-domain signcryption, in which each inter-domain authentication centers do not have to set the same parameters for their keys, and the members in a domain register their identities with blind keys other than their private keys to avoid the authentication center faking and cheating his members to access resource from other domains. At the same time it has good anonymity, and it can trace entities when there occurred dispute between two entities for accessing resources, and it has a good defense for various protocol attacks. cross-domain authentication protocol purposed in the paper can achieve the features as follows:

Correctness: a legal user in a domain can be valid verified by all the users when they compute the authentication algorithm of the Cross-domain authentication protocol.

unforgeability: it is infeasible that a faked member generates an algorithm to pass a valid authentication by computing, even if the member is a server of a domain.

Anonymity: except the server of the domain, it should be infeasible that anyone determine the identity of a prover by computing.

Traceability: the KMC of the domain can determine the identity of any prover within its domain.

Anti-attack: Cross-domain authentication protocol should have extensive security and provably secure .

Organization. The rest of paper is organized as follows: In Section II, we introduce the relative knowledge of this paper. In Section III we define the system model. Then, we present our scheme in Section Ⅳ. We provide security analysis, and further analyze the experiment results and performance in SectionⅣ. Finally, we conclude the paper in Section Ⅵ.

2. Preliminaries

2.1 Self-isomorphic group of finite group [16]

Let G be a group, $AutG$ represents self-isomorphic group of G , $C(G)$ is the center of G , $\langle g \rangle$ is an $Abel$ group generated by g . If G is a finite group, and $|G|$ is the order of G and $|G| = p^n (n > 0)$, then G is defined as $p-$group （p is a prime）.

Let Q be a $p-$Subgroup of a finite group G , and if Q is the highest exponentiation of p in the factorization of $|G|$, then Q is defined as $sylow\ p-subgroup$ of G .

Theorem 1[16]: let G be a finite $Abel$ group, $p_1, p_2, ..., p_n$ are all prime factors of $|G|$, $G_{p_i} (1 \le i \le n)$ are the $sylow\ p-$subgroups of G , which gives direct product decomposition: $G = G_{p_1} \times G_{p_2} \times ... \times G_{p_i}$.

Theorem 2[12]: let $G = G_1 \times G_2 \times ... \times G_n$, if K_i is a sub-group of G_i $(1 \le i \le n)$, and $K_1, K_2, ..., K_n$ are isomorphic for each other, and then G has n sub-groups which are isomorphic for each other.

Theorem 3[16]: let $G_1 = \langle g_1 \rangle$ and $G_2 = \langle g_2 \rangle$ be cyclic groups, and m and n are the order of G_1 and G_2 respectively, if $(m, n) = 1$, then $G_1 \times G_2$ is a cyclic group with the order of mn .

2.2. Bilinear group [17].

Firstly ， we give the definition of bilinear map, assuming that G_1, G_2 and G_T are multiplicative groups with same prime order p , $p \ge 2^k + 1, k$ is the security parameter, let $G_1 = \langle g_1 \rangle$ be generated by g_1 and $G_2 = \langle g_2 \rangle$ be generated by g_2 , φ is the isomorphic mapping from G_1 to G_2 : $\varphi(g_1) = g_2$,the solution of discrete logarithm over the G_1 and G_2 and G_T is hard. and e is a computable mapping, and $e: G_1 \times G_2 \to G_T$ has the following properties:

1.Bilinear: For all the $u \in G_1$, $v \in G_2$ and $a, b \in Z_p$,then $e(u^a, v^b) = e(u, v)^{ab}$.

2.Non-degeneracy: There exists $u \in G_1$, $v \in G_2$ such that $e(u, v) \ne 1$.

3. Computable: There is an efficient algorithm to compute $e(u, v)$ for all $u \in G_1$, $v \in G_2$.

Corollary1: for all the $\forall u_1 \in G_1, \forall u_2 \in G_1$, $\forall v \in G_2$,then $e(u_1 u_2, v) = e(u_1, v) e(u_2, v)$.

Corollary2: for all the $\forall u, v \in G_2$,then $e(\varphi(u), v) = e(u, \varphi(v))$.

2.3 Gap Diffie-Hellman Group

We first introduce the following problems in G_1 and G_2 [18].

1. Discrete Logarithm Problem (DLP): if given u and v , to find $n \in Z_p$ from $u = v^n$.

2. Computation Diffie-Hellman Problem (CDHP): Given $(g_1, g_1{}^a, g_1{}^b) \in G_1$, for $a, b \in Z_p$,to compute $g_1{}^{ab}$.

3. Decisional Diffie-Hellman Problem (DDHP:. Give $(g_1, g_1{}^a, g_1{}^b, g_1{}^c) \in G_1$, for $a, b, c \in Z_p$, to decide whether

$c = ab \bmod p$.

We call G_1 and G_2 are GDH groups if DDHP can be solved in polynomial time but no polynomial time an algorithm can solve CDHP or DLP with non-negligible advantage within polynomial time.

3.The cross-domain authentication model

In multi-domain authentication system, the type of authentication is chosen for each domain by themselves demand ,without need a unified authentication model. and inter-domain authentication should try to adopt a common authentication way to achieve cross-domain access interoperability [19]. This cross-domain authentication system model is designed by the paper.

3.1 System Framework.

In this model, the system is composed by multiple domains, each domain is independent and autonomous. Each domain consists of a DAC (domain authority center) and a number of members within the domain, and the domain authority center are similar to traditional CA (Certificate Authority). Every member in a domain not only provides its resources for others but also access resources from others, and they constitute the resource alliance. In the case of collaborative computing, the members of mutual cooperation are not only from a domain, but also from other domains ,for this members in each domain may need to cross-domain cooperation. let $Gset = \{G_k \,|(k = 1, 2, ..., R)\}$ be a large prime set of the automorphism group. In the multi-domain alliance system, each DAC select a different subgroup $G_k (1 \le k \le R)$ from $Gset$ to make key generation parameters for its domain. DAC distributes and manages some keys of their members within its domain, and open the public key of DAC in order to mutual visits and certification. When members join in a domain they need to register with their true identities for entity tracking.

4. Alliance signature scheme between domains

4.1 System initialization.

Let the alliance domain contain R domains,and selects R pairwise relatively prime large prime numbers to form a set of $R_S = \{r_i \,|(i = 1, 2, ..., R)\}$;and choose a big prime p , compute a elliptic curve $E / GF(P)$ that satisfies WDH security hypothesis, G is a sub-group of $E / GF(P)$ with high prime order q ($q = r_1 \times r_2 \times ... \times r_i$),that $|G| = q$. Let $r_1, r_2, ..., r_n$ be all the prime factors of $|G|$,that $q = r_1 \times r_2 \times ... \times r_n$ .Let $G_{r_j} (1 \le j \le n)$ be $sylow_{r_j} - \text{subgroups}$ of G .

From Theorem 1 we known the direct product decomposition of G : $G = G_{r_1} \times G_{r_2} \times ... \times G_{r_i}$,and we can Construct R sub-groups of G that are isomorphism to each other according to the Theorem2,let this set of $sub - groups$ be $Gset = \{G_k \,|(1 \le k \le R)\}$.

Under the multi-domain unite architecture, each domain select a different sub-group $G_k (1 \le k \le R)$ from set $Gset$ as the key generator parameter of the domain.

4.2 inter-domain signatures.

(1) Let D_1 and D_2 be two domains of alliance-domain ,and D_1 selects cyclic group $G_1 = \langle g_1 \rangle$ as the key generation parameter of its domain, D_2 selects cyclic group $G_2 = \langle g_2 \rangle$ as the key generation parameter of its domain , g_1 and g_2 are the generators of the two cyclic groups respectively .and G_1 and G_2 are the isomorphic group in $Gset$,and $e : G_1 \times G_2 \to G_p$ is an efficiently computable bilinear mapping, and $h : \{0,1\}^* \to Z_p$ is a hash function, and the private /public key pairs of the two domains are $(\xi_1, g_1^{\xi_1})$ and $(\xi_2, g_2^{\xi_2})$ respectively($\xi_1, \xi_2, \in Z_p$),and $H = e(g_1^{\xi_1}, g_2^{\xi_2})$ is the mapping value of the two public keys $g_1^{\xi_1}$ and $g_2^{\xi_2}$.

(2) Key distribution and register of members in a domain: assume that domain D_1 has n members within the domain, and DAC_1 (domain authority center) is the domain authority center of the domain D_1 with private key ξ_1 , and the corresponding public key is $P_{D_1} = g_1^{\xi_1}$,. DAC_1 compute $y = g_1^{\frac{1}{\xi_1}}$ and sent y to every member in the domain D_1 ,and each member U_{D_i} in the domain selects $x_i \in Z_p$ as its own private key, and

the corresponding public key is $P_{u_i} = g_1^{x_i}$,and it computes $reg_i = (y)^{x_i}$,and sent reg_i to the DAC_1 as its register key to register. The DAC_1 establishes the relationship between reg_i and identity of U_{D_i} in order to track the certification.

(3) Suppose a member U_{D_1} of the domain D_1 wants to access resources from the member U_{D_2} of the domain D_2. Assume that the private/ public key pair of U_{D_1} is (x_1, P_{u_1}) ,and it's registered key is reg_{u_1}. The private/ public key pair of U_{D_2} is (x_2, P_{u_2}) , and it's registration key is reg_{u_2}.The public key of DAC_1 in domian D_1 is P_{D_1} ,and The public key of DAC_2 in domian D_2 is P_{D_2} ,Certification process is as follows:

1) U_{D_1} Selects $\mu \in Z_p$, and computes

$T_1 = g_1^{\mu}, U_{D_1} \xrightarrow{P_{D_1}, P_{u_1}, reg_{u_1}, T_1} U_{D_2}$;

2) U_{D_2} check whether $e(P_{D_1}, reg_{u_1}) = e(P_{u_1}, g_1)$,if the equation are equal to each other then Selects the message $m \in \{0,1\}^*$, and computes the question value $c \leftarrow h(T_1, m)$, $U_{D_1} \xleftarrow{c} U_{D_2}$;

3) U_{D_1} computes $s_1 \leftarrow \mu + cx_1$

$U_{D_1} \xrightarrow{s_1} U_{D_2}$

4) U_{D_2} verifies the signature on the message m , whether $g_1^{s_1} = T_1 P_{u_1}^{\ c}$

If the signature is correct, it is valid inter-domain signature.

If the verification holds, then the U_{D_2} can prove that U_{D_1} is a number of league domain ,and its the public key is P_{D_1} ,this achieves the results of across multiple domains authentication.

4.3 Session key agreement.

1) U_{D_2} chooses a random number $k_2 \in Z_p$, and compute $f_1 = P_{u_1}^{\ k_2}$. $U_{D_2} \rightarrow U_{D_1} : (P_{u_2}, f_1)$;

2) U_{D_1} can compute $P_{u_1}' = g_1^{k_2}$ from $f_1 = P_{u_1}^{\ k_2}$ with his private key x_1, and then choose a random number $k_1 \in Z_p$, and compute $f_2 = P_{u_2}^{\ k_1}, U_{D_1} \rightarrow U_{D_2} : f_2$;

3) U_{D_2} can compute $P_{u_2}' = g_2^{k_1}$ from $f_2 = P_{u_2}^{\ k_1}$ with his private key x_2;

4) U_{D_1} and U_{D_2} compute their temporary session key $P_{D_1 D_2} = e(P_{u_1}', P_{u_2}') = e(g_1, g_2)^{k_1 k_2}$.

5. Performance analysis

5.1. Correctness analysis.

Cross-domain alliance authentication protocol is established based on inter-domain signature. In order to ensure the safe authentication when the domains access resources each other, the correctness of the signature must be ensured for first time:

(1) DAC that is not in the alliance-domain cannot be valid inter-domain signature;

(2) members that are not in the domains cannot be valid inter-domain signature;

(3) ensure the uniqueness of the internal member in a domain.

$$e(P_{D_1}, reg_{u_1}) = e(g_1^{\xi_1}, g_1^{\frac{x_1}{\xi_1}})$$
$$= e(g_1, g_1)^{x_1} = e(g_1^{x_1}, g_1)$$
$$= e(P_{u_1}, g_1)$$
$$g_1^{s_1} = g_1^{(\mu + cx_1)} = g_1^{\mu} g_1^{cx_1} = T_1 P_{u_1}^c$$

5.2 Anonymity.

There can only determine that a user is a specific member of a certain domain, but the identity of the member can not be determined, and only his DAC can determine the identity of the member through registered identity. The anonymity of cross-domain authentication alliance protocol is designed by two steps:

1) User U_{D_1} sends inter-domain public key $dpk = (g_1, P_{u_i}, reg_i, P_{D_1}, H)$ to U_{D_2} , and U_{D_2} determines U_{D_1} from which domain with the equation $e(P_{D_1}, reg_{u_1}) = e(P_{u_1}, g_1)$.

2) U_{D_1} sends its signature to U_{D_2} , and U_{D_2} can determine U_{D_1} is a specific member that not be faked by

others through verification whether $g_1^{s_1} = T_1 P_{u_1}^{\ c}$, but does not know the identity of the member U_{D_1} .

5.3 Traceability

It is not an ideal method to design cross-domain authentication alliance protocol based on the trust, and it is impractical to let members to trust the DAC that is from different domains, and it is must to provide reliable certification to prove irregularities of a certain entity when the disputes are occurred. This protocol is traceable for that the verifier U_{D_2} verify the expression $e(P_{D_1}, reg_{u_1}) = (P_{u_1}, g_1)$ to ensure the relationship among $P_{D_1}, reg_{u_1} and P_{u_1}$,further to trace the identity of entity U_{D_1} by the registration information in DAC_1 .

5.4 Security analysis

The security of cross-domain alliance authentication protocol has two aspects: one is the security of the inter-domain signature, the other is the security of the authentication protocol. The security of the signature method proposed in this article relies on the elliptic curve discrete logarithmic problem. The security of this authentication protocol as follows:

5.4.1 Against $MITM$. Assume that mediator U_{D_3} attempt to attack this protocol, it can not achieve the consistency session key to U_{D_1} and U_{D_2} ,because U_{D_3} does not have the private key x_1 of U_{D_1} , and he can not compute $P_{u_1}' = g_1^{k_2}$ when $U_{D_2} \to U_{D_1} : (P_{u_2}, f_1)$. Obviously he also can not compute $P_{u_2}' = g_2^{k_1} . U_{D_3}$ and U_{D_1} or U_{D_3} and U_{D_2} can not achieve the consistent session key $P_{D_1 D_2} = e(P_{u_1}', P_{u_2}') = e(g_1, g_2)^{k_1 k_2}$ at last.

5.4.2 unforgeability

Any member or DAC' that is out of the alliance-domain can not fake the DAC that is in the alliance-domain, and any member within a domain can not fake other members to achieve cross-domain access resource.

1) Assume that any DAC' that is out of the alliance-domain can fake the public key P_{D_1} of DAC_1 in domain D_1 .He has not the corresponding private key of DAC_1 ,and the verification $e(P_{D_1}, reg_{u_1}) = e(P_{u_1}, g_1)$ will be fail. If a number U_{D_3} fake the number U_{D_1} to achieve cross-domain access resource, the signature of U_{D_3} will be fail.

2) Assume that the member DAC_1 in the domain D_1 fakes the number U_{D_1} to access the resource of member U_{D_2} within another domain D_2 , because the private key x_1 of U_{D_1} is not published, even if the DAC_1 of domain D_1 can fake the identity of member U_{D_1} with identity U_{D_1}' to send $dpk = (g_1, P_{u_i}, reg_i, P_{D_1}, H)$ to U_{D_2} , and this can only prove that U_{D_1}' is a member in the domain D_1 , but U_{D_1}' do not know the private key x_1 of U_{D_1} , therefore the verification signature of U_{D_1}' will be fail.

5.4.3 Against replay attack

The session key used during the communication between two domains is in one-time key, and thus it can defense replay attack.

5.4.4 Comparative analysis

Compared with the existing cross-domain authentication, our advantages are as follows:

(1) authentication protocol in communication and computation is smaller than SAP scheme,and the efficiency of the certification is higher than SAP scheme .

(2) our scheme greatly simplifies the system architecture compare with the traditional PKI-based authentication framework, and saves system cost.

(3) Compare with the literature [19] in the certification framework, this paper proposed protocol can provide mutual authentication in different trust domains ,and the application is broader, more in line with the actual needs of a distributed network environment.

(4) This paper proposed authentication protocol has forward security, and in the literature [19] the non-interactive authentication session key is static, if an attacker controls a user's private key, he can calculate the session key that between this user and any entity , it does not have forward security.

5.5 consumption analysis.

computation and communication complexity are two important indicators for evaluating the performance of protocols. We analyzed the latest research ,and we also compared the Cross-domain authentication protocol proposed in this paper with the latest research programs in computation complexity and communication overhead . We compared our scheme with the literature [20] [21] in computational complexity, as shown in Table1. These several programs are elliptic curve public key cryptosystem. It is known that 1024-bit keys in conventional cryptosystems offer the same level of security as 160-bit keys in elliptic curve cryptography. In particular, in the case of elliptic curves, we can assume that the exchanged messages have size only 160 bits, since only the x coordinate is necessary for the computation of the point (x, y). We assume that the length of each communication unit is ml = 160 bits in these programs.

Table1 *Complexity analysis of cross-domain authenticated protocols*

authenticated protocols	Number of exponentiations	Number of pairings	Number of scalar multiplications	Number of hash	Number of sent And received messages
literature [31]	0	12	11	8	32ml
literature [33]	0	0	23	10	23ml
Ours scheme	3	2	3	1	6ml

For more intuitive analysis of the energy consumption in each scheme, the literature [22] provided a experiment that on a 133MHZ "Strong ARM" of microprocessor to perform a modular exponentiation arithmetic need to consume 9.1 mJ, to pure scalar multiplications need to consume 8.8 mJ. To perform a Tate Pairing computation need to consumes 47.0 mJ. It use a 100kbps transceiver module to transmit a bit of information need to consume 10.8 μJ and receive a bit of information need to consume 7.51 μJ. as shown in Table 2. We assume that the energy consumption of hash calculation is negligible. The total energy consumption comparison of these three programs is shown in Figure 1.

Table 2 *Energy Costs for Computation and Communication*

Computation cost of Modular Exponentiation	9.1 mJ
Computation cost of Scalar Multiplication	8.8 mJ
Computation cost of Tate Pairing	47.0 mJ
Communication cost for transmitting a bit	10.8 mJ
Communication cost for receiving a bit	7.51 mJ
DSA Signature	9.1 mJ
ECDSA Signature	8.8 mJ
DSA signature verification	11.1 mJ
ECDSA signature verification	10.9 mJ

The energy consumption is shown in figure 1,the scheme of literature[20] is the most in energy consumption, and ours is the minimum in energy consumption .the advantage of ours scheme is that any two entities can mutual authenticate and do key agreement directly, so it needn't the third-party to take part in. The cross-domain authentication scheme in literature [20] and literature[21] when an entity want to access resources from another entity in different domain it must be checked by the third-party, so it is very complex.

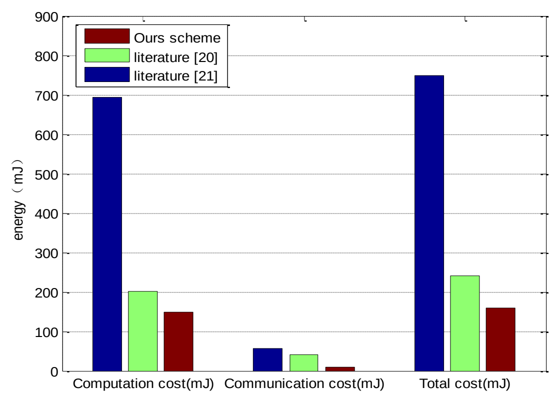


Fig.1 *energy consumption*

Analysis shows that this protocol is correct and can defense attack effectively and is not to need to know the identity of each other, which can achieve the effective

authentication and good anonymous. The entity can be tracked when there have dispute occurs. The computation and communication overhead is relatively low. It has a good security.

6. Conclusion

Multi-domain alliance-authentication is required for security in multi-domain network environment. The scheme of cross-domain alliance-authentication purposed in this article can ensure the security while share the resource among multiple domains. The anonymity can protect the privacy of each entity, and each entity can access cross-domain resources needless the intervention of the key authentication center, which provide good flexibility. It can avoid the bottleneck problem and the complexity of the transfer tickets of the traditional pattern based on PKI. It is safe and practical.

Acknowledgements

This work is supported by National Natural Science Foundation of China under Grant No. 61272511 and Innovation Group Project of Beijing Institute of Technology.

References

[1] Randy Butler, Von Welch, Douglas Engert, Ian Foster, Steven Tuecke, John Volmer, Carl Kesselman, A National-Scale Authentication Infrastructure [J].IEEE Computer, 2000, 33(12):60-66.

[2] Zhang Qikun, Li Yuanzhang, Song Danjie,Tan Yuan. Alliance-Authentication Protocol in Clouds Computing Environment. China Communications,2012,7,pp:42-54.

[3] Qikun Zhang,Yuan Tan, Li Zhang, and Ruifang Wang. A Combined Key Management Scheme inWireless Sensor Networks, SENSOR LETTERS,2011,9(4):1501-1506

[4] Jung-San Lee, Chin-Chen Chang, Pen-Yi Chang, Chin-Chen Chang. Anonymous authentication scheme for wireless communications. International Journal of Mobile Communications 2007, 590 - 601 .

[5] Lv Chao, Li Hui, Ma Jianfeng, Niu Ben. Vulnerability Analysis of Elliptic Curve-Based RFID Protocol[J]. China Communications 2011 Vol. 8 (4): 153-158

[6] Jie Tang, Shaoshan Liu, Zhimin Gu, Chen Liu, Jean-Luc Gaudiot. Prefetching in Mobile Embedded System Can be Energy Efficient[J]. IEEE Computer Architecture Letters, Volume 10, Issue 1 (2011), Page 8-11

[7] Peng Huaxi. An identity-based authentication model for multi-momain[J]. Chinese Journal of Computers, 2006,29(8):1271-1281.

[8] L Chen, K Harrison,D Soldera,N Smart .Applications of multiple trust authorities in pairing based cryptosystems[A].In Proceedings of Infrastructure Security[C].Berlin: Springer-Verlag, 2002,260-275.

[9] Noel McCullagh, Paulo S. L. M. Barreto. A new two-party identity-based authenticated key agreement[OL]. http://citeseerx.ist. psu. edu/viewdoc/download?doi=10.1.1.58.9294 &rep=rep1&type=pdf.

[10] J Malone-Lee. Identity-based signcryption [OL].http:// eprint .iacr.org/2002/098.pdf,

[11] Lu Xiaoming, Feng Dengguo. An identity-ba sed authentication model formulti-doma in grids[J]. Chinese Journal of Electronics 2006,34(4):577-582.) (in Chinese).

[12] Zhu Wen, He Mingxing. On automorphism group of finite groups[J].Journal of UEST of China, 2000,29(5):549-551.

[13] Boneh D. and Franklin M.. Identity based encryption from the Weil pairing [J]. SIAM Journal on Computing. 2003, 32(3): 586-615.

[14] Wenbo Zhang ; Hongqi Zhang ; Bin Zhang ; Yan Yang ; An Identity-Based Authentication Model for Multi-domain in Grid Environment[C]. Computer Science and Software Engineering. 2008, Volume: 3, pp: 165 – 169.

[15] Lu Xiaoming, Feng Dengguo. An identity-based authentication model for multi-domain grids [J]. Chinese Journal of Electronics 2006,34(4):577-582.).

[16] Zhu Wen, He Mingxing. On automorphism group of finite groups[J].Journal of UEST of China, 2000,29(5):549-551) (in Chinese).

[17] David Mandell Freeman, Converting Pairing-Based Cryptosystems from Composite-Order Groups to Prime-Order Groups, Springer Berlin / Heidelberg, 2010(6110):44-61.

[18] J. Cha and J. Cheon. An Identity-Based Signature from Gap Diffie-Hellman Groups. PKC 2003, LNCS 2567, Springer-Verlag, pp. 18–30,2003.

[19] I Foster, C Kesselman, G Tsudik, S Tuecke .A security architecture for computational GRID[A], In Proceedings of the 5th ACM Conference on Computer and Communications Security[C] .New York: ACM press, 1998,83 - 92.

[20] Lu Xiaoming, Feng Dengguo. An identity-ba sed authentication model formulti-doma in grids[J]. Chinese Journal of Electronics 2006,34(4):577-582..

[21] I Foster, C Kesselman, G Tsudik, S Tuecke .A security architecture for computational GRID[A], In Proceedings of the 5th ACM Conference on Computer and Communications Security[C] .New York: ACM press, 1998,83 - 92.

[22] Eleftheria Makri, Elisavet Konstantinou. Constant round group key agreement protocols: A comparative study[J] .computers and security ,2011, 30:643-678.

Supply Chain Dynamic Performance Measurement Based on BSC and SVM

Yan Hong[1], Yang Zhong-Hua[2]

[1] School of Management, Wuhan University of Science and Technology, Wuhan430081,HuBei, China

[2] School of Management, Wuhan University of Science and Technology, Wuhan430081,HuBei, China

Abstract

Now individual contest among enterprises has been turning into collective contest among supply chains. Supply chain management (SCM) has been a major component of competitive strategy to enhance organizational productivity and profitability. In recent years, organizational performance measurement and metrics have received much attention from researchers and practitioners. The foundation of proper supply chain performance assessment system is the basis of its effective operation and management. Most of the traditional supply chain performance evaluation is a static evaluation, while the actual supply chain is a dynamic system, therefore need to adapt with ways to carry out the evaluation. In order to meet the needs of the dynamic alliance's overall performance evaluation, this paper extended the traditional four Balanced Scorecard dimension into five. On this basis, established the five Balanced Scorecard dimension of supply chain, and also established a three-layered of quantitative index system according to this model. Measured then each performance index's value by using the theory of Fuzzy Analytic Hierarchy Process, meanwhile reduced the number of input of the Support Vector Machine (SVM) by using classification method, finally, got performance evaluation's result by using the weighted Least Squares Support Vector Machine (LS-SVM), which provides the basis for rational analysis and decision-making of the supply chain.

Keywords: *Supply Chain Performance, Balanced Scorecard, FAHP, LS- SVM.*

1. Introduction

With the rapid development of economic globalization, knowledge-based, information technology, the competition between enterprises is not a single enterprise in a certain time, a certain space, the competition for certain end markets, customer one-on-one competition, but become a kind of competition based on product development, design, manufacturing, distribution, distribution, sales and service competition across time and space, has been developed into a competition between the supply chain management. Not only as one of the core supply chain performance evaluation of supply chain management is the focus of supply chain management, supply chain management difficulties. Many scholars in-depth study on the performance of the supply chain, and achieved good results.

The need of performance measurement systems at different levels of decision-making, either in the industry or service contexts, is undoubtedly not something new. Kaplan and Norton (1992) have proposed the Balanced Scorecard, as a means to evaluate corporate performance from four different perspectives: the financial, the internal business process, the customer, and the learning and growth [1]. Their Balanced Scorecard is designed to complement "financial measures of past performance with their measures of the drivers of future performance". The name of their concept reflects an intention to keep score of a set of items that maintain a balance "between short long term objectives, between financial and non-financial measures, between lagging and leading indicators, and between internal and external performance perspectives". Since the Balanced Scorecard theory, the industry has a lot of research. In the September 1993 issue of the "Harvard Business Review, Kaplan and Norton published" Putting the Balanced Scorecard to Word "article, the Balanced Scorecard applied to RockWater, Apple and AMD three companies case [2]. Kaplan and Norton proposed a strategic map, marking the Balanced Scorecard performance management system from one jumped as a strategic management tool [3]. A balanced performance evaluation of supply chain such as, Balanced Scorecard not only helps organizations in faster and wider progress monitoring of their operations but can also help them in improving their internal and external functions of business such as engineering and design applications, production, quality improvement, materials management, quick response, gaining lost market shares, proper implementation of business strategies. Therefore, it is clear that for effective supply chain management, measurement goals must consider the overall scenario and the metrics to be used. These should represent a balanced approach and should be classified at strategic, tactical, and operational

levels, and be financial and non-financial measures, as well. Lambert, Cooper and Pagh successful supply chain management requires cross-functional integration, the main challenge is how to successfully integrate. Brewer and Speh proposed the implementation of supply chain performance evaluation method based on Balanced Scorecard and Analytic Hierarchy Process (AHP) [5].
Throughout BSC-based performance evaluation of the status quo, you can see traditional Balanced Scorecard theory ignored supplier factors both internal and external balance of the supply chain system, its inadequacies. Especially in the assessment of the performance of the supply chain based on dynamic alliance, often through the analysis of the value chain of suppliers to help companies take strategic improvement actions to promote the value chain of suppliers and recycling, so you can save production costs and reduce enterprise procurement costs. In view of the lack of traditional Balanced Scorecard theory in the performance evaluation system of dynamic alliance, first extend the traditional thinking balanced scorecard for the five-dimensional dynamic balanced scorecard that the dynamic alliance downstream member satisfaction, alliance within the enterprise supply chain business processes, supplier satisfaction, the economic benefits of the supply chain, supply chain innovation and development capabilities.
Then a dynamic alliance's supply chain performance evaluation, for example, establish a dynamic performance evaluation of supply chain decision-making table, and attribute reduction and value reduction of decision table using Support Vector Machine (SVM), which has been forecast performance evaluation results of the decision-making the rule set. Finally, weighted Least Squares Support Vector Machine (LS-SVM) forecasting overall performance assessment results provide the basis for rational analysis and decision-making of the supply chain.

2. Dynamic evaluation of supply chain performance architecture design

2.1 Five-dimensional BSC Mode

Balanced Scorecard was originally defined as the corporate performance measurement tool. Early 1990s, Kaplan and Norton (1992) have proposed Balanced Scorecard, as a means to evaluate corporate performance from four different perspectives: the financial, the internal business process, the customer, and the learning and growth [1][2][3]. Their Balanced Scorecard is designed to complement "financial measures of past performance with their measures of the drivers of future performance". The emergence of Balanced ScoreCard, changed the status of the pre-financial indicators to dominate the world, the performance indicators extreme imbalance. Four indicators together, constitute the internal and external, results-driven factors, a variety of balanced long-term and short-term, qualitative and quantitative, based on the evaluation of the three-dimensional, forward-looking enterprise performance management. Subsequently, Brewer and Speh made application Balanced Scorecard explore supply chain performance assessment, and explore basic manifestation of a new supply chain performance assessment tools - supply chain performance evaluation, and proposed as a basis for kind of new supply chain performance assessment tool - Supply Chain Balanced Scorecard, which established a framework for links between SCM and Balanced Scorecard [5]. Supply chain Balanced Scorecard to supply chain business processes as a starting point to corporate strategic objectives, performance indicators and corporate strategy linked, comprehensive assessment of business performance, to cultivate enterprise core competitiveness.
However, the traditional supply chain Balanced Scorecard approach in considering the balance of the supply chain system of internal and external factors ignored suppliers. In fact, the suppliers as an important link in the value chain in the production and operation is very important. Only when the vendor for enterprises on time, in ensuring quality and quantity to provide goods and materials required for enterprises to ensure the normal operation and for customers to provide products to meet their needs. In particular, in the performance evaluation of supply chain, supplier value chain analysis can help enterprises improve strategies to promote the recycling value chain of suppliers, in order to save production costs and reduce procurement costs. On the other hand, the supply chain performance evaluation is unique in that is different from the single enterprise evaluation: assessing the indicators of the operating performance of the entire supply chain not only need to assess the operational performance of the node enterprises, but also to consider the operational performance of the node enterprises its upper node enterprise or the entire dynamic alliance. Therefore, reasonable assessment of business process, scientific and objective assessment of the situation of the entire supply chain operations, real-time, dynamic characteristics need to be considered. Accordingly, we believe that creating a Balanced Scorecard framework of supply chain performance evaluation system should be integrated the dynamic alliance downstream member satisfaction alliance within the enterprise supply chain business processes, supplier satisfaction, the economic benefits of supply chain, supply chain five aspects of innovation and development ability. Add Balanced Scorecard classic four dimensions based on the dimensions of suppliers, the formation of the five-dimensional balanced scorecard for

supply chain dynamic alliance structure, as shown in Figure 1.

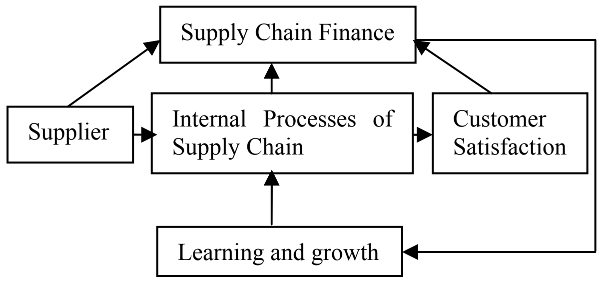


Figure 1: Five-dimensional BSC model of Supply Chain

Balanced Scorecard framework of dynamic alliance of supply chain performance evaluation system includes five aspects as following:

（1）Supplier. This is the part of the traditional supply chain balanced scorecard ignored. The selection and evaluation of supplier is particularly important in dynamic alliance. Therefore, in the framework of the proposed model, the supplier is an important dimension in the Dynamic Balanced Scorecard. Shift from competition to cooperation, to achieve a win-win situation between the manufacturer and supplier to supplier relationships through the assessment of selection and improvement of supplier relationships.

（2）Internal business processes. Internal business process measures which we concern are those internal processes the greatest impact on customer satisfaction and achieve organizational financial goals. Balanced Scorecard method introduced innovative processes to internal business processes, from the point of view of the supply chain considerations, it require companies to create new products and services to meet the current and future target customers demand. These processes can create value in the future to promote the future of corporate financial performance.

（3）Customer satisfaction. Supply Chain Balanced Scorecard is more concerned about the performance of the supply chain in the level of customers and market segments, and clarify how to meet customer needs in order to effectively achieve the financial goals of the entire dynamic alliance. Customer value based on customer perception and therefore requires an assessment of the origin on the customer, including the level of service and customer satisfaction

（4）Learning and growth. Balanced Scorecard goal is revealed in these three aspects of the existing capacity of the system, and the gap between the high performance required capacities. To close these gaps, companies must invest to enable employees to acquire new skills, and straighten out the program and the day-to-day work of the organization. To close these gaps, companies must invest to enable employees to acquire new skills, and straighten out the program and the day-to-day work of the organization.

（5）Supply Chain Finance. On the financial side, the Balanced Scorecard not only to assess the traditional enterprise financial ratios, return on investment, cash flow, profit and other indicators, but also concerned about the financial condition of the entire supply chain indicators. Financial performance measurement method to reveal the overall strategy of the alliance and its implementation and enforcement are to contribute to the improvement of the supply chain.

2.2 The Selection of Indicators and Metrics of Supply Chain Performance Evaluation

Supply chain performance evaluation indicators selection of hot and difficult in the current performance evaluation study. Different research institutions and personnel have different views on supply chain performance index system. We believe that the selection of indicators should be in a number of ways to achieve a balance, so as to build a good performance evaluation system. A good performance indicator system includes not only reflect the short-term and long-term goals, the level of internal and external indicators, including the balance between leading indicators and lagging indicators, quantitative indicators and qualitative indicators. This paper extends the traditional Balanced Scorecard theory, the formation of a five-dimensional Balanced Scorecard. The indicators measuring dynamic alliance upstream suppliers, including on-time delivery rate, production flexibility and other indicators. We select time delivery rate, the rate of qualified products and supply chain flexibility and other indicators to reflect the performance of the supply chain in supplier dimension. On-time delivery rate is one of the most important indicators of the Alliance selection of suppliers, the high and low values reflect the supplier delivery performance, supplier performance considerations. Product qualification rate refers to the number of products of acceptable quality percentage of total product output, and it reflects the quality level of the suppliers of goods. For flexible indicators, Supply Chain Council defines it as the ability to respond to supply chain. On-time delivery rate is one of the most important indicators of the supply chain alliance selection of suppliers, the high and low values reflect the supplier delivery performance, supplier performance considerations. Product qualification rate refers to the number of products of acceptable quality percentage of total product output, which reflects the quality level of the suppliers of goods. For the other four dimensions, we refer to the standardized indicators SCOR model based on the operation of the internal processes. According to the characteristics of dynamic alliance in the supply chain, the paper selected 15 classic sample indicators to build a performance evaluation system. As shown in Table 1.

Table 1 Supply Chain Balanced Scorecard performance assessment indicators and metrics

Dimension	KPIs	Metrics
Financial	Profitability(F1)	Net profit / total revenue (%)
	Asset turnover(F2)	Total sales / total net assets (%)
	Inventory turnover rate(F3)	Average proportion of cost of goods sold / inventory
	Cash turnaround time(F4)	Supply of inventory days + receivables aging - payables aging
Customer	Customer Satisfaction(C1)	Fuzzy Evaluation
	The expansion of the market rate(C2)	(Current market share - previous period market share) / previous period market share
	Market share(C3)	The total number of sales / industry sales (%)
Business processes	Response time(P1)	The time required to meet sudden demand
	The level of information systems (P2)	The level of information systems
	Technological advance(P3)	Fuzzy Evaluation
	Total cost of quality(P4)	The total cost of quality cost / product
	Reliability(P5)	Fuzzy Evaluation
Learning and growth.	Quality SystemL1(L1)	Fuzzy Evaluation
	Employee Satisfaction(L2)	Fuzzy Evaluation
	Profit growth(L3)	(Profit for the period - the previous period profit) / Profit
	New product development cycle(L4)	Statistical average
Suppliers	On-time delivery rate(S1)	The number of on-time delivery / delivery (%)
	Flexible(S2)	Fuzzy Evaluation
	The rate of qualified products(S3)	The number of acceptable quality / total number

3. The dynamic evaluation method based on Fuzzy Analytic Hierarchy Process and LS-SVM forecast

3.1 Fuzzy Analytic Hierarchy Process

Fuzzy Analytic Hierarchy Process (FAHP) is a fuzzy comprehensive evaluation method and analytic hierarchy process evaluation method, it contains the evaluation of the complex system of multiple indicators (or factors, evaluation factors) objects total evaluation. A wide range of applications in the evaluation system, performance assessment, and system optimization, is a combination of qualitative and quantitative evaluation model, usually is the first to determine the factors set chromatography analysis, and then judge the effect of fuzzy comprehensive evaluation to determine.

1）Determining the weights

The key of FAHP lies in the establishment of judgment matrix. The importance of a factor than the other factors to be quantified, get fuzzy complementary judgment matrix $A=(a_{ij})_{n\times n}$. Commonly used in Table 1 shown in the 0.1 to 0.9 scale their relative importance of the number of scale the a_i's weight value fuzzy complementary judgment matrix A and w_i is

$$w_i = \frac{\sum_{i=1}^{n} a_{ij} + \frac{n}{2} - 1}{n(n-1)} \quad i=1,2,...,n \qquad (1)$$

Formula (1) contains the excellent characteristics of fuzzy consistency judgment matrix and its judgment information, a small amount of calculation and easy computer programming.

Table 2 0.1 to 0.9 scale law and its meaning

Scale	Definition
0.5	Two factors compared equally important.
0.6	Comparison of two factors, one factor is slightly important than the other factors.
0.7	Comparison of two factors, one factor is obviously important than the other factors.
0.8	Comparison of two factors, one factor is much more important than the other factors.
0.9	Comparison of two factors, one factor is extremely important than the other factors.
0.1,0.2,0.3,0.4	If factors a_i and a_j compared to get a_{ij}, then the factor a_i and a_j comparison phase determines $a_{ij}=1-a_{ji}$.

Set up $w=\left(w_1,w_2,...w_n\right)^T$ sort of fuzzy complementary judgment matrix $A=\left(a_{ij}\right)_{n\times n}$ vector, if $a_{ij}=w_i-w_j+0.5$, then $A=\left(a_{ij}\right)_{n\times n}$ for the of fuzzy consistency of judgment matrix, Which is w_i exactly the same sort of judgment matrix vector weights for each index weight value.

2）Data processing and calculation

Table 1 performance indicators, both quantitative also include qualitative, its dimension is also different for each indicator should be dimensionless, convert it to a number in [0, 1], it has comparable. Qualitative indicators are generally obtained by fuzzy assessment methods, and in both cases the normalization process with quantitative indicators:

Normalized for positive indicators (index value is the bigger the better, such as market share), can be carried out in accordance with the equation (2):

$$v^{'}=\left(x-x_{\min}\right)/\left(x_{\max}-x_{\min}\right) \qquad (2)$$

Normalized for reverse indicators (the smaller the index value the better, for example, response time), can be carried out in accordance with the equation (3):

$$v^{'}=\left(x_{\max}-x\right)/\left(x_{\max}-x_{\min}\right) \qquad (3)$$

Finally, the comprehensive evaluation value of the supply chain five-dimensional Balanced Scorecard can be obtained through the formula $R_i=\sum_{i=k-1}^{n} w_{ik}\times X_{ik}$ (Where n each one indicator the two indicators number, X_{ik}, for each index values obtained after pretreatment.), so as to achieve the drop dimension and simplify the calculation.

3.2 Predict the performance of the weighted least squares support vector machine

Conditions of the supply chain, each node enterprise operation strategy dynamically adjust its operational behavior of a random change in trend is non-linear model of time. Support Vector Machine (SVM) can solve the nonlinear law of supply chain performance evaluation and the problem of inadequate samples, to be able to predict the overall performance of the supply chain of the future at a certain moment.

The realization of the Support Vector Machine (SVM) is mapped to high dimensional feature space through some kind of selected nonlinear mapping (kernel function) the input vector construct optimal separating Hyperplane in this space. Least Squares Support Vector Machine (LS-SVM) inherited the Support Vector Machine (SVM)structural risk minimization criterion and use of nuclear function converted to high-dimensional feature space to solve ideological, and the Support Vector Machine (SVM) converted into solving linear quadratic programming equations to avoid insensitive loss function greatly reduces the computational complexity. Weighted Least Squares Support Vector Machine (WLS-SVM) weighted them according to the degree of importance of the different samples, to eliminate sample different impact, has important implications on the predicted results. Therefore, using the Weighted Least Squares Support Vector Machine WLS-SVM [10] to predict and analyze the performance of the supply chain, the process is shown in Figure 2.

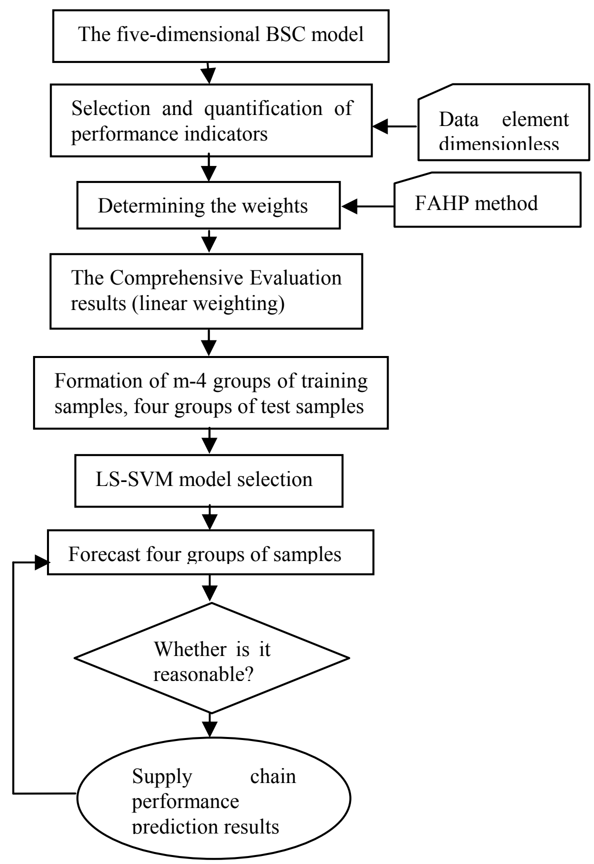


Figure 2: The flow chart of supply chain performance prediction

4. Case study

In this paper, a dynamic supply chain, for example, in the above five-dimensional balanced scorecard supply chain performance evaluation. Based on historical experience and fuzzy complementary judgment matrix of establishing rules to get the judgment matrix as follows:

$$A=\begin{bmatrix}0.5 & 0.6 & 0.6 & 0.8 & 0.9\\ 0.4 & 0.5 & 0.6 & 0.6 & 0.7\\ 0.4 & 0.4 & 0.5 & 0.6 & 0.7\\ 0.2 & 0.4 & 0.4 & 0.5 & 0.6\\ 0.1 & 0.3 & 0.3 & 0.4 & 0.5\end{bmatrix}$$

The use of (1) Extended Performance Balanced Scorecard five weights can be obtained: W_F=0.3, W_C=0.22, W_P=0.22, W_L=0.22, W_S=0.22.
Similarly weights can be obtained for each two indicators：

W_{Fi}=(0.37,0.28,0.16,0.19);

W_{Ci}=(0.43,0.21,0.43);

W_{Pi}=(0.22,0.19,0.16,0.15,0.27);

W_{Li}=(0.17,0.23,0.28,0.32);

WSi=(0.48,0.18,0.34).

In the past two years, the monthly value of dynamic supply chain performance indicators can be obtained after pretreatment of the support vector machine prediction sample value, the first 20 sets of data for the training sample, after four sets of test samples, output Y is actual monthly supply chain performance situation, the result is divided into four grades of the G_1、G_2、G_3、G_4 respectively correspond to the excellent performance in the differential state, and the corresponding value of 0.9,0.8,0.7,0.6, such as shown in table 3.

Table 3 Supply Chain Performance Indicators Quantization

F	C	B	L	S	P
0.185	0.072	0.132	0.107	0.085	0.7(G_3)
0.111	0.106	0.168	0.063	0.095	0.7(G_3)
0.218	0.132	0.191	0.104	0.101	0.9(G_1)
0.248	0.114	0.201	0.112	0.102	0.9(G_1)
0.174	0.101	0.127	0.129	0.093	0.8(G_2)
0.112	0.174	0.110	0.108	0.075	0.7(G_3)
0.127	0.157	0.111	0.050	0.055	0.7(G_3)
0.080	0.044	0.066	0.067	0.079	0.6(G_4)
⋮	⋮	⋮	⋮	⋮	⋮

RBF Radial Basis weighted support vector machines in the selection of the kernel function for high precision radial basis parameters δ = 0.45, to take balance factor C = 500, the fitting precision ε = 0.01, tuning constants γ = 500. Weighted through FAHP calculated Supply Chain Performance Balanced Scorecard five weights. Finally, the four groups of the predicted value and the actual value shown in Table 4.

Table 4 Supply Chain Performance Predictive Value of Contrast with the Actual Value

F	C	B	L	S	Predictive value	Actual value
0.14	0.183	0.125	0.078	0.062	0.783	G3(0.7)
0.201	0.095	0.184	0.117	0.100	0.868	G_2(0.7)
0.204	0.123	0.167	0.085	0.090	0.819	G_2(0.7)
0.135	0.156	0.127	0.162	0.067	0.714	G_3(0.7)

As can be seen from Table 3, the results of the following four groups of test data is 0.78313, 0.86805, 0.81998, 0.71378. The sets of data belongs to the interval [0.7, 0.8], the performance level is moderate. The other two sets of data belong to the interval [0.7, 0.9], the performance level is good. This is entirely consistent with the actual situation. And this also proves the effectiveness of the weighted least squares support vector machine model of dynamic supply chain performance prediction.
Finally, we input the performance evaluation of five aspects the value (0.19876, 0.14254, 0.16734, 0.09242, 0.08194), and call the above model to obtain a prediction value is 0.87392. That forecast next month to assess the performance is good. Therefore, by entering the Balanced Scorecard five values, call the dynamic evaluation model to predict supply chain performance assessment results and trends, and provides a basis for rational analysis and decision-making of the supply chain to develop, but also for the supply chain the performance evaluation provides a new idea.

5. Conclusion

Quantifiable indicators system to evaluate the performance of dynamic supply chain plays a central role in the day-to-day operations and management of the supply chain. In this paper, we consider a real-time and dynamic nature of the dynamic supply chain, and classic extended supply chain performance for the five-dimensional Balanced Scorecard Balanced Scorecard, to build a three-tier evaluation index system, and then apply the FAHP theory get the weight of the performance indicators input to support vector machine classification method dimensionality reduction, thereby reducing the amount of computation to increase the accuracy of the forecast. Finally, weighted least squares support vector machine prediction method of the future results of the assessment, and to provide a basis for rational analysis and decision-making of the supply chain.

Acknowledgements

This work was supported by Science Foundation for Young Scholars of Wuhan University of science and Technology under Grant 2010XZ041, by Humanities and

Social Science Foundation of Hubei Provincial Department of Education under Grant 2011jyte267.

References

[1] Kaplan, R. S., and Norton, D. P., "The balanced scorecard-measures that drive performance", Harvard Business Review, Vol. 70, No. 1, 1992, pp. 71-79.

[2] Kaplan, R. S., and Norton, D. P., "Linking the balanced scorecard to strategy", California Management Review, Vol. 39, No. 1, 1996, pp. 53-79.

[3] Kaplan, R. S., and Norton, D. P., "Using the balanced scorecard as a strategic management system", Harvard Business Review, Vol. 74, No. 1, 1996, pp. 75-85.

[4] Kaplan, R. S., and Norton, D. P., "Transforming the balanced scorecard from performance measurement to strategic management", Accounting Horizons, Vol. 15, No. 1, 2001, pp. 87-104.

[5] Brewer, P.C., and Speh, T.W., "Using the Balanced Scorecard to Measure Supply Chain Performance", Journal of Business Logistics, Vol. 21, No. 1, 2000, pp. 87-104.

[6] Ma, S.H., and Li, H.Y., and Lin, Y., "Study on application of Balanced Scorecard to performance measurement of supply chain", Industry Engineering and Management, Vol. 38, No. 4, 2002, pp. 5-10.

[7] Lohman, C., and Fortuin, L., and Wouters, M., "Designing a performance measurement system: A case study", European Journal of Operational Research, Vol.156, No. 2, 2004, pp. 267-286.

[8] Beamon, B.M, "Measuring Supply Chain Performance", International Journal of Operations & Production Management, Vol.19, No. 1, 1999, pp. 275–292.

[9] Cortes C, and Vapnik V., "Support-Vector Network", Machine Learning, Vol.20, No. 3, 1995, pp. 273-297.

[10] Suykens J A K, and Vandewalle J., "Least Squares Support Vector Machines Classifiers", Neural Processing Letters, Vol.9, No. 3, 1999, pp. 293-300.

[11] Suykens J A K, and De Brabanter J, and Lukas L., "Weighted Least Squares Support Vector Machines: Robustness and Sparse Approximation", Neuro computing, Vol.48, No. 1, 2002, pp. 85-105.

[12] Ma J S, and James Theiler, and Simon Perkins, "Accurate On-line Support Vector Regression", Neural Computation, Vol.15, No. 11, 2003, pp. 2683-2703.

[13] YuGang Fan, and Ping Li, and ZhiHuan Song，"Danamic Weightded Least Squares Support Vector Machines", Control and Decision, Vol.21, No. 10, 2006, pp. 1129-1133.

[14] Ekici S., "Classification of power system disturbances using support vector machines", Expert Systems with Applications, Vol.36, No. 6, 2009, pp. 9859–9868.

[15] Baccarini L M R, and Rochae Silva V V, and de Menezes B R, Caminhas W M., "SVM practical industrial application for mechanical faults diagnostic", Expert Systems with Applications, Vol.38, No. 6, 2011, pp. 6980–6984.

[16] Teti R, and Jemielniak K, and O'Donnell G, and Dornfeld D., "Advanced monitoring of machining operations", CIRP Annals-Manufacturing Technology, Vol.59, No. 2, 2010, pp. 717–739.

YanHong is currently an Instructor in the Information Management Department, School of Management, Wuhan University of Science and Technology (WUST), China. He received the Bachelor's degree in Computer Science from the Hubei University of Technology in 2000, and Master's degrees in Management Science and Engineering from Wuhan University in 2003. Her research interests include supply chain management, information system engineering and intelligent information systems.

Yang ZhongHua is currently an Associate Professor and tutor of graduate students in the Information Management Department, School of Management, Wuhan University of Science and Technology (WUST), China. He received the Bachelor's degree in Business management from the Hubei University of Technology in 1998, and Master's degrees in Computer Application Technology from Wuhan University of Technology in 2002. His research interests include information system engineering, decision support systems and intelligent information systems.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Study on the Distribution of the Magnetic Field of Circular and Square Exiting Coils in Electromagnetic Flow Meter

Ling-fu Kong [1], Sheng-xue Du [2] and Ying-wei Li [3]

[1] **College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei, China**

[2] **College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei, China**

[3] **College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei, China**

Abstract

Exciting coil is an important part of the electromagnetic flow meter, its reasonable design has significant influence on the performance of electromagnetic flow meter. This paper discusses the magnetic field distributions of circular and square exciting coils in electromagnetic flow meter, using Biot-Savart Law and superposition principle, the simulation of magnetic field distribution is done, in the cross section that containing electrodes inside measurement pipe, and two indicators the magnetic induction intensity parallel degree in direction and magnetic induction intensity uniform degree in size is proposed, the analysis and comparison of induced magnetic field generated circular and square exciting coils is done based on above two indicators. This paper provides a reference for the optimal design of exciting coils in electromagnetic flow meter.

Key words: *circular, square, exiting coil, magnetic field distribution, electromagnetic flow meter.*

1 Introduction

Electromagnetic flow meter is an instrument measuring the flow volume of conductive liquid by Faraday's law of induction, it is widely used in metallurgy, drainage, chemical and petroleum, food-making, medical, environmental protection, aviation, agriculture irrigation, and so on. It is mainly composed of a sensor and a converter. Electromagnetic flow sensor is installed in fluid transmission process pipe, it converts the flow volume of conductive liquid into induction voltage signal linearly, the converter provides excitation current to the sensor for generating magnetic field, accepts the induction voltage signal, and processes it to standard current or voltage signal.

The magnetic field distribution of electromagnetic flow meter causes the attention of researchers. In 1998, A. Michalski etc established 2D model of electromagnetic flow meter by finite element method [1], studied the optimal design of exciting coil, in order to get the uniform weighting function, making induction voltage signal measured by electrode only related with the mean flow velocity of fluid. In 2002, A. Michalski etc established the 3D hybrid mathematical model of exciting coil of electromagnetic flow meter [2]. XiaoZhang Zhang studied calculation of magnetic field of electromagnetic flow meter with large diameter and multi-electrodes by idealized magnet model [3]. Chen Zhao etc proposed a approximate calculation method for magnetic field distribution of saddle shape exciting coil of electromagnetic flow meter [4]. Through the finite element method [5], Jingzhuo

Wang done numerical simulation and verification of weighting function of electromagnetic flow meter [6]. The measurement principle of the electromagnetic flow meter is based on Faraday's law of induction. As shown in figure 1, while conductivity fluid cut magnetic field line inside magnetic field $\vec{B}$ of the sensor, induced potential E that is proportional to the fluid velocity v is generated at two electrodes, usually it is expressed as following formula [7]:

$$E = kBDV \tag{1}$$

In above formula, k is coefficient of instrument, D is the inner diameter of sensor pipeline.

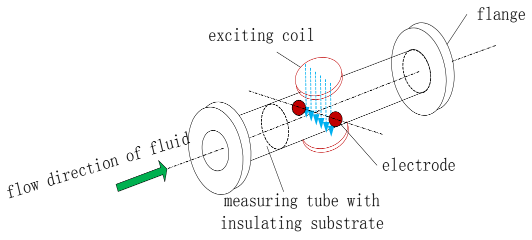


Fig. 1 Measurement principle of electromagnetic flow meter.

The volume flow is determined by the average flow velocity, in a circular pipe the flow volume is as following:

$$Q = \frac{\pi}{4} D^2 V \tag{2}$$

Gained by (1) (2)

$$Q = \frac{\pi D}{4kB} E \tag{3}$$

When the magnetic induction $\vec{B}$ and the inner diameter D of the pipeline are constants, the flow volume Q of fluid is proportional only to induction potential E, and has nothing to do with other physical parameters (such as density, conductivity, etc), that is the significant advantage of electromagnetic flow meter.

Above equations simply illustrate the working principle of electromagnetic flow meter, they are set up only when the following conditions are met: (1) in the infinite range, the magnetic induction B is evenly distributed; (2) the speed of the fluid is as solid conductor, the internal particle speed is the same everywhere, and equals to the average velocity.

According to the required magnetic induction in condition (1), to realize the optimal design of exciting coil of the electromagnetic flow meter, this paper focuses on the analysis of magnetic field distribution for circular and square exciting coils.

2 Calculation and simulation of magnetic field distribution of electromagnetic flow meter

2.1 Calculation of magnetic field distribution of exciting coil

Basing on Biot-Savart Law, the basic law of magnetic field produced by a current carrying conductor is: Any current element $Id\vec{l}$ produce magnetic induction $d\vec{B}$ at any point P in space is as following:

$$d\vec{B} = \frac{\mu_0}{4\pi} \bullet \frac{Id\vec{l} \times \vec{r}}{r^3} \tag{4}$$

In above formula, $\vec{r}$ is the vector from current element to point P, $d\vec{l}$ is the vector of wire element. The total magnetic induction at point P could be got through the integral of the magnetic field generated along the current-carrying conductor.

If $d\vec{B}$ is magnetic field generated by a small section of wire at place $\vec{r}$, above formula could be wrote as:

$$\vec{B} = \frac{\mu_0}{4\pi} \frac{I}{r^3} \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ l_x & l_y & l_z \\ r_x & r_y & r_z \end{vmatrix} = B_x \vec{i} + B_y \vec{j} + B_z \vec{k} \tag{5}$$

To calculate magnetic field generated by n small sections of wire at place $\vec{r}$, it is

$$\sum_{i=1}^{n} \vec{B} = \sum_{i=1}^{n} \frac{\mu_0}{4\pi} \frac{I}{r^3} \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ l_{xi} & l_{yi} & l_{zi} \\ r_x & r_y & r_z \end{vmatrix} = \sum_{i=1}^{n} (B_{xi}\vec{i} + B_{yi}\vec{j} + B_{zi}\vec{k}) \tag{6}$$

Above formula can be applied to calculate the magnetic induction produced by arbitrary shape current at any place. l_{xi}, l_{yi}, l_{zi} mean the components of current element along the axis in rectangular coordinate system. According to formula (6), calculation and simulation of the magnetic field distribution of the exciting coil of electromagnetic flow meter could be done.

2.2 Simulation of magnetic field distribution of exciting coil

First of all the coordinate system is established, center axis of measuring pipe for z axis, the connection of geometry centers of two field coils for x axis, the connection of two point electrodes for y axis. In the rectangular coordinate system, calculation and simulation of magnetic field produced by field coil in measurement pipeline's cross section containing electrodes, could be done according to the following steps:

1) In x-y plane, the part between two field coils is meshed, and then coordinate (*x*, *y*, *0*) of each grid point is determined;

2)The current carrying conductor is divided into many small elements, and coordinate (*xc*, *yc*, *zc*) of every small element and coordinate (l_x, l_y, l_z) of every small element vector are determined;

3)Calculating the vector $\vec{r}$ (r_x, r_y, r_z) and corresponding distance *r* from each grid point to small wire elements;

4) Respectively in the x, y direction, calculating the component B_{xi} and B_{yi} of magnetic induction produced at a certain point as following:

$$B_{xi} = \frac{\mu_0}{4\pi}\frac{I}{r^3}(l_{yi}r_z - l_{zi}r_y), \quad B_{yi} = \frac{\mu_0}{4\pi}\frac{I}{r^3}(l_{zi}r_x - l_{xi}r_z);$$

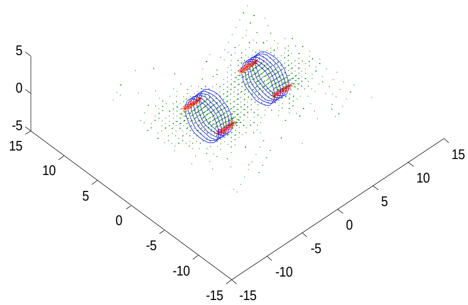
5) Calculating the summation of the magnetic induction at a certain point produced by all the small elements $B_x = \sum_{i=1}^{n} B_{xi}$, $B_y = \sum_{i=1}^{n} B_{yi}$;

6) At each grid point in the given cross section, calculating magnetic induction according to the above steps;

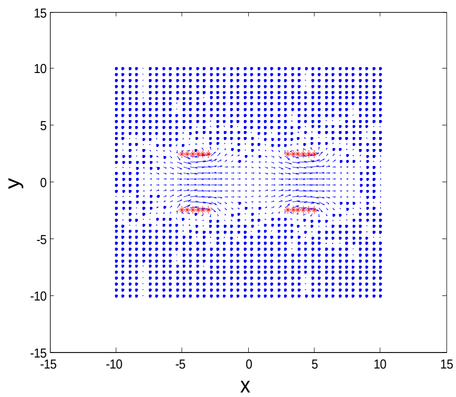
7) Doing simulation, describing the graphics of magnetic strength in given cross section.

In the following content, the part of section between the two field coils is divided into 40 by 40 grids, at each grid point, the simulation of magnetic induction produced respectively by circular and square field coils is done.

When the shape of magnet coil is circle, the radius of magnet coil is 2.5cm, the turn of each field coil is 6, and the thickness of each field coil is 2cm, the distance between two field coils is 6cm, current strength is 10mA in coils. In the cross section of measuring tube containing electrodes, the magnetic field distribution of above circular magnet coils is as shown in figure 2, the red dot mean intersection of field coils and the cross section.



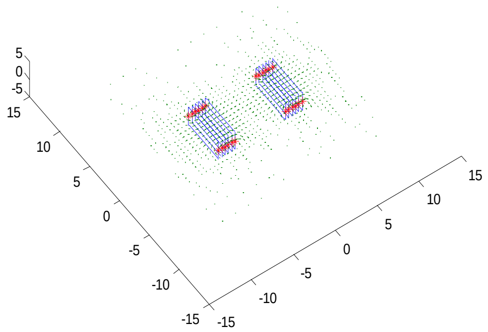
(a) Circular magnet coil



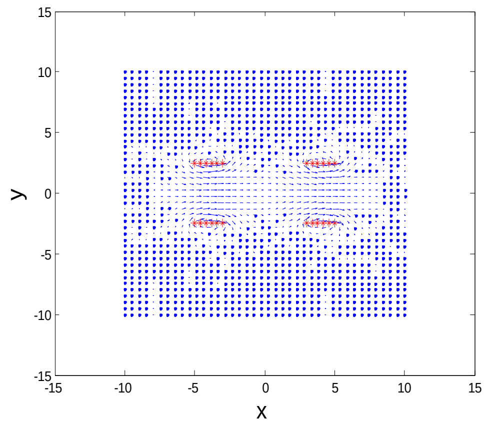
(b) Sectional view of magnetic field distribution

Fig. 2 Magnetic field distribution of circular magnet coil.

When the shape of magnet coil is square, suppose the length of square side is 5cm, the turn of each field coil is 6, and the thickness of each field coil is 2cm, the distance between two field coils is 6cm, current strength is 10mA in coils. In the cross section of measuring tube containing electrodes, the magnetic field distribution of above square magnet coils is as shown in figure 3.



(a) Square magnet coil



(b) Sectional view of magnetic field distribution

Fig. 3 Magnetic field distribution of square magnet coil.

3 Magnetic field analysis of field coil

Analysis for the magnetic field distribution of field coil is done in cross section of measuring tube containing electrodes. In the cross section, the vector of produced magnetic induction is: $\vec{B} = cx\vec{i} + cy\vec{j}$

3.1 Two indicators for analysis

Because the magnetic induction is a vector, so naturally the indicators for analysis and comparison of magnetic field include: uniform degree of strength and parallel degree of direction. In the considered cross section including electrodes, the direction of magnetic induction vector could be represented by the angle θ between magnetic induction vector and positive direction of x axis as following:

$$\tan\theta = \frac{cx}{cy}, \quad \theta = \arctan\frac{cx}{cy} \tag{7}$$

Describing function for parallel degree of field direction is proposed, supposing that at each grid point, the average of the Angle θ_k between magnetic induction vector and positive direction of x axis is θ_0, then $\theta_0 = \frac{1}{n}\sum_{k=1}^{n}\theta_k$. Defining as follows:

$$RM_1 = MAX(|\frac{\theta_k - \theta_0}{\theta_0}|) \tag{8}$$

$$RD_1 = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(\frac{\theta_k - \theta_0}{\theta_0})^2} \tag{9}$$

RM_1 reflects the maximum deviation of field direction in the area; RD_1 reflects the whole parallel degree of field direction in the region. The smaller value of RD_1, the more ideal the whole parallel degree of field direction is.

Vector size of magnetic induction could be represented by length d of vector as following:

$$d = \sqrt{cx^2 + cy^2} \qquad 10)$$

Describing function for uniform degree of field strength is proposed, supposing that at each grid point, the average of the length d_k of magnetic induction vector is d_0, then $d_0 = \frac{1}{n}\sum_{k=1}^{n} d_k$. By the same token, defining as follows:

$$RM_2 = MAX(|\frac{d_k - d_0}{d_0}|) \tag{11}$$

$$RD_2 = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(\frac{d_k - d_0}{d_0})^2} \tag{12}$$

RM_2 reflects the maximum deviation of field strength in the area; RD_2 reflects the whole uniform degree of field strength in the region. The smaller value of RD_2, the more ideal the whole uniform degree of field strength is.

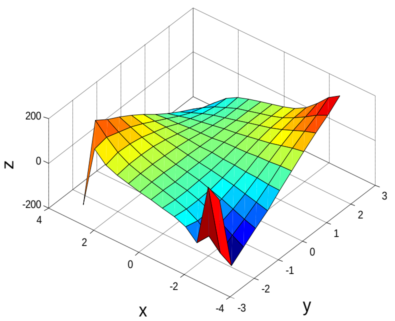
3.2 Magnetic field analysis of circular and square magnet coils

Based on the above indicators, calculation and analysis is done for magnetic field distribution of the circular and square magnet coils described in cross section containing electrodes, and the specific situation is as shown in table 1.

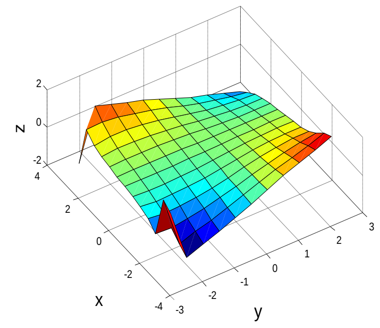
Table 1: Magnetic field distribution of circular and square magnet coils

Shape of coil	direction of magnetic induction		
	θ_0	RM_1	RD_1
circle	0.0080	195.1070	68.3793
square	0.0078	198.8714	66.3584
Shape of coil	strength of magnetic induction		
	d_0	RM_2	RD_2
circle	5.8677e-005	1.4594	0.5036
square	6.2802e-005	1.7527	0.4455

It is known from table 1, that for the whole parallel degree in direction of magnetic induction , the square magnet coil is better than circular magnet coil, but the maximum deviation is slightly bigger; for the overall degree of uniformity of field strength, square field coil is better than circular magnet coil, but the maximum deviation is also slightly bigger.
The specific distribution of magnetic induction direction for circular and square magnet coils is respectively as shown in figure 4 and figure 5.

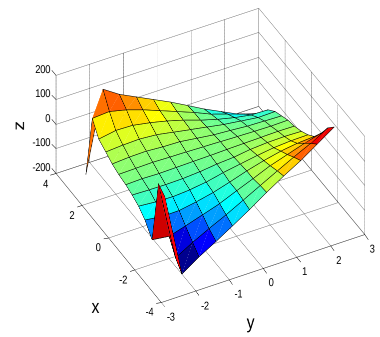


(a) Relative direction of magnetic induction $\frac{\theta_k-\theta_0}{\theta_0}$

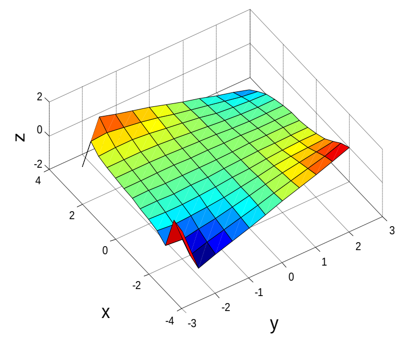


(b) Direction of magnetic induction θ_k

Fig. 4 Distribution of magnetic induction direction of circular magnet coil.



(a) Relative direction of magnetic induction $\frac{\theta_k-\theta_0}{\theta_0}$

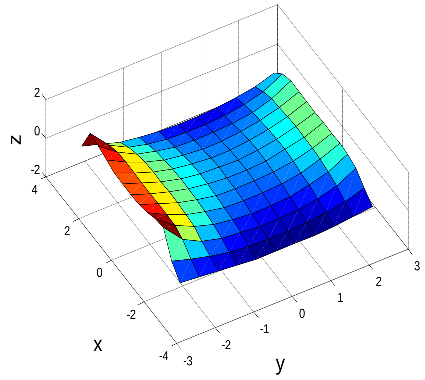


(b) Direction of magnetic induction θ_k

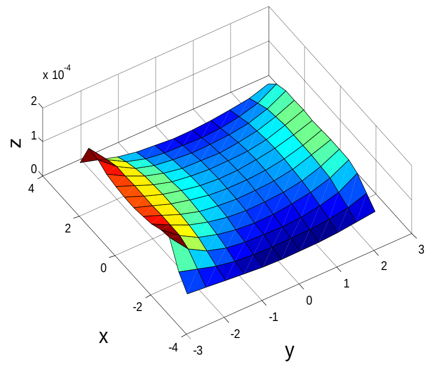
Fig. 5 Distribution of magnetic induction direction of square magnet coil

Contrasting figure 4 and figure 5, it also reflects square field coil is better than circular magnet coil for the whole parallel degree of magnetic induction direction, which is consistent with the results in table 1.

The specific distribution of field strength for circular and square magnet coils is respectively as shown in figure 6 and figure 7.

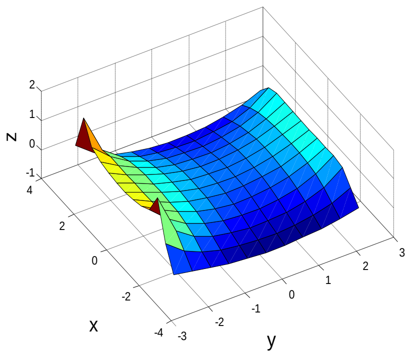


(a) Relative strength of magnetic induction $\frac{d_k - d_0}{d_0}$

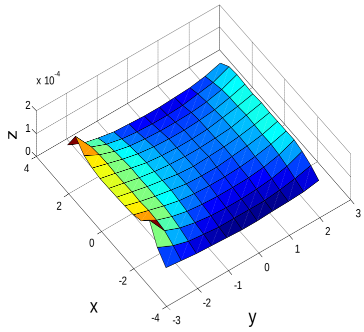


(b) Strength of magnetic induction d_k

Fig. 6 Strength distribution of magnetic induction of circular magnet coil



(a) Relative strength of magnetic induction $\frac{d_k - d_0}{d_0}$



(b) Strength of magnetic induction d_k

Fig.7 Strength distribution of magnetic induction of square magnet coil.

Contrasting figure 6 and figure 7, it also reflects square field coil is better than circular magnet coil for the overall uniform degree of magnetic induction strength, which is also consistent with the results in table 1.

4 Conclusion

In the infinite range, the magnetic induction B is constant and uniform, which is one of the ideal working conditions for the electromagnetic flow meter. In order to realize reasonable design of field coil of electromagnetic flow meter, and improve its performance, in this paper basing on Biot-Savart Law and superposition principle, calculation and simulation for the magnetic field distribution of circular and square magnet coils is done, in the cross section containing electrodes of measuring tube. Two indicators for analysis and comparison of magnetic field distribution are proposed: uniform degree of magnetic induction strength and parallel degree of field direction. Based on above two indicators, the magnetic field distribution of circular and square magnet coils is analyzed, it shows that: square field coil is better than circular magnet coil both for the whole parallel degree of magnetic induction direction and for the overall uniform degree of magnetic induction strength. It could provide certain reference for optimal design of field coil of electromagnetic flow meter.

References

[1] A. Michalski, and J. Starzynski, "Optimal design of the coils of an electromagnetic flow meter", IEEE Transactions on Magnetics, Vol.34,No. 5,1998, pp. 2563-2566.

[2] A. Michalski, J. Starzynski, and S. Wincenciak, "3-D approach to designing the excitation coil of an electromagnetic flow meter", IEEE Transactions on Instrumentation and Measurement, Vol. 51,No. 4, 2002, pp. 833-839.

[3] Xiaozhang Zhang, "Calculation and measurement of the magnetic field in a large diameter electromagnetic flow meter", ISA Transactions, Vol. 42, 2003, pp. 167-170.

[4] Chen Zhao, Bin Li, and Wenjian Chen, "An Algorithm for Analyzing Magnetic Field Distribution of Saddle Excitation Coils in EM Flowmeter Sensor", Journal of Shanghai University(Natural Science), Vol. 14, No. 1, 2008, pp. 31-35.

[5] Wadhwa, and Rhythm-Suren, "Electromagnet Gripping in Iron Foundry Automation Part II: Simulation", International Journal of Computer Science Issues, Vol. 9, No. 22-2, 2012, pp. 238-243.

[6] Jingzhuo Wang, "Numerical simulation and verification of weight function of electromagnetic flow meter", Chinese Journal of Scientific Instrument, Vol. 30, No. 1, 2009, pp. 132-137.

[7] R. C. BAKER, Flow measurement handbook, Cambridge: Cambridge University Press, 2000.

Multi- view Video Coding Scheme based upon enhanced Random Access capacity

Xiaoxing Lv[1] Lini Ma [2] Jingjing Guo [3]

[1] School of Automation,Beijing Institute of Technology, Beijing, China 100081

[2] Computer Science School ,Beijing Information Science and Technology University,Beijing, China 100192

[3] Computer Science School ,Beijing Information Science and Technology University,Beijing, China 100192

Abstract

Due to the multi-view video coding scheme using inter-view prediction stucture, increased coding complexity, and reduced multi-view video random access performance, so one proposed multi-view video coding prediction scheme is proposed on the basis of analysis and study of several typical multi-view video coding schemes in this paper. This coding prediction scheme calculates the location of base-view by global disparity, and introduces a rational inter-view prediction structure, as well as, according the relationship between the length of GOP and random access performance to select the number of frames in a GOP. Experimental results show that the proposed multi-view video coding scheme can significantly improve the random access performance, while maintaining high coding efficiency.

Keywords: *Multi-view Video Coding; Global Disparity; Random Access; Hierarchical B Frames; Temporal Layer Identification*

1. Introduction

Multi-view video consists of several video sequences captured by multiple cameras which are aligned in a parallel, which includes the depth information of images and provides users with three-dimensional and interactive features to meet them watch video images from different angles. The MVC will be applied to a number of emerging multimedia services including free-viewpoint video (FVV), free-viewpoint television (FVT), three-dimensional video (3DV) and three-dimensional television (3DTV) [1].

Compared with the traditional single-view video coding, the multi-view video technology as an important area of research, which to be more comprehensive consideration to the dynamic scene and gives the immersive feel. Therefore, the amount of datas that need to be processed is also multiplied and which reduces the efficiency of video coding. However, there is a considerable temporal and spatial correlation among the various views, which offers the possibility for the coding efficiency. So, how to remove the correlation of intra-view and inter-view has become one of the hot issues in the field of multi-view video technology.

Presently, there are a variety of multi-view video coding schemes have been proposed, In which the use of hierarchical B frame prediction structure coding scheme can significantly improve the coding efficiency, but most of the current coding scheme mainly for compression efficiency, less considerations for coding complexity and random access performance. This paper studies a number of typical multi-view video coding schemes, then selects the location of base-view, and changes inter-view prediction structure, while decreaces the GOPlength in a Hierarchical B frame to improve the random access performance while maintaining high coding efficiency.

2.Typical MVC Coding Scheme

2.1 Hierarchial B Frames Coding Scheme

Hierarchical B frames can be better to remove the temporal redundancy, which can improve the image quality, while maintaining low bit rate. Experiments show that hierarchical B frames coding structure can achieve more coding efficiency than traditional IPPP structure [2]. Coding structure shown in Fig. 1: I denotes intra code frame; P denotes single direction of inter prediction farame; B1 and B2 are double direction inter prediction frame, can use them as reference frames; B3 is also double direction inter prediction frame, but can't use as reference frame.

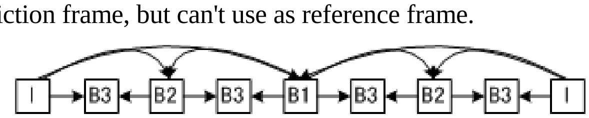


Fig. 1 Hierarchical B frames coding structure

2.2 MVC Coding Scheme

The 3DAV special group did unified subjective and objective tests for MVC schemes which all companies and research institutes proposed, and then got the test results [3]. The test results show that Fraunhofer-HHI proposed the coding prediction structure with both spatial and temporal references based on the hierarchical B frames that obtains higher coding efficiency [4]. Compared with the simulcast scheme, this coding scheme achieves up to1.4~1.6dB in coding efficiency [5], so this scheme was chosen as a reference prediction structure for MVC by the Joint Video Team (JVT). This scheme removes temporal redundancy with hierarchical B frames prediction coding structure in intra-view direction and removes intra-view redundancy with IBPBP prediction structure in inter-view direction.

The MVC encoding prediction scheme includes 8-channel views, and the length of each view's GOP is 8, the last picture is called anchor picture, which helps improve the random access performance and synchronization, the other pictures are called no-anchor picture. Where, View V0 only uses temporal references, but don't use inter-view references. So V0 is called base view, and the other views are called no-base views. In this scheme, non-base view is classified P-view (V2, V4, V6 and V7) and B-view (V1, V3, V5).Fig. 2 shows that Compared with simulcast scheme, this coding scheme greatly improved the coding efficiency and data transmission capacity, but it included prediction relation in inter-view direction, which reduced random access performance and increased decoding delay. The prediction structure is illustrated in Fig. 2.

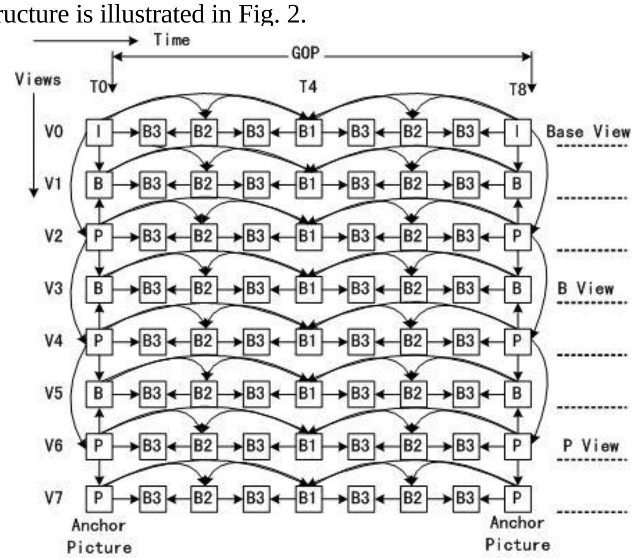


Fig. 2 MVC coding scheme

3. Proposed MVC Coding Scheme

3.1Coding Scheme based on Selection of Base-view

The base-view is reference of other views, so the selection of base-view can improve coding efficiency and random access performance. This paper calculates the location of base-view by global disparity. Park defines mean absolute global disparity (MAGD) as follows:

$$MAGD(v) = \frac{1}{N}\sum_{w=0}^{N-1} | g(v,w) | \quad (1)$$

Where N is the number of views and g (v, w) is the global disparity between views v and w. MAGD (v) means the average of absolute global disparity between view number v and all the other views. TABLEI lists MAGD values of test sequences and the views location with minimum MAGD values are marked with grey boxes.

TABLE 1:MAGD VALUES FOR VARIOUS SEQUENCES

View	0	1	2	3	4	5	6	7
Ballroom	38.9	27.4	24.0	21.5	23.8	33.9	28.9	38.8
Exit	101.5	79.4	64.9	58.0	57.0	64.4	80.9	104.8
Race1	54.9	27.6	34.4	20.8	19.9	16.9	35.5	40.4

However, for N-view sequences require $(N-1)^2/2$ times of global disparity calculations, this process has significant complexity [6]. So Park proposes a simplified method to select the location of base-view, S_l as follows:

$$S_l = \lfloor (N-1)/2 \rfloor [6] \quad (2)$$

The MAGD values show that, base-view mainly locates in the middle of views in the 8-channel multi-view video sequences. Compared with the MVC coding prediction scheme, this scheme significantly improves the random access capability.

In this proposed multi-view encoding scheme, V4-view be chosen as base-view, and others are no base-view, while using the hierarchical B pictures prediction structure in the temporal direction, e.g.Fig. 3.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

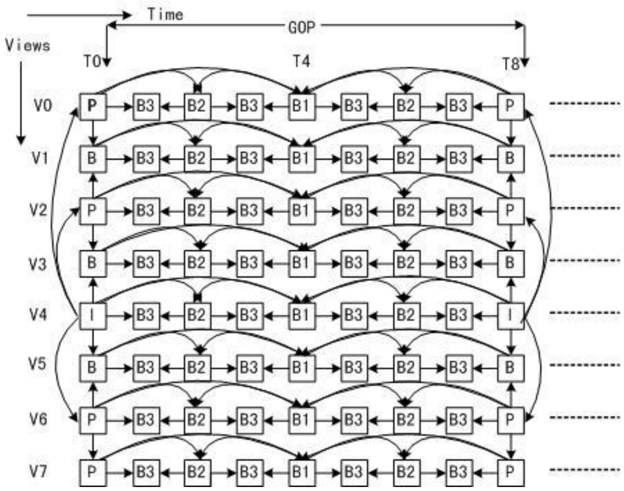


Fig. 3 Proposed coding prediction scheme MVC-1 (GOP=8)

3.2Coding Scheme based on Decreasing the GOPlength

According to the relation difference of the hierarchical B frames prediction structure in the temporal direction, images can be divided into different temporal layers, the temporal layer can be marked by temporal layer identification (TLID) [7]. In one GOP of MVC reference prediction structure, when TLID is equal to i, the number of images Ni is:

$$N_i = \begin{cases} 1 & i = 0 \\ 2^{i-1} & i = 1 \sim TL_{\max} - 1 \\ GOPlength - (1 + \sum_{j=1}^{TL_{\max}-1} 2^{j-1}) & i = TL_{\max} \end{cases} \quad (3)$$

Where, GOPlength is the length of a GOP, and is the maximum value of TLID, TLmax is given by:

$$TL_{\max} = \lceil \log_2 GOPlength \rceil \quad (4)$$

Where, $\lceil x \rceil$ denotes the minimum integer that is more than or equal to x. If the GOPlength is equal to 8, TLmax is equal to 3. According to the formula (3) and (4), the number of images is 1, 1, 2 and 4, when TLID is equal to 0, 1, 2 and 3 respectively. If getting frame of TLID is equal to 0, 1, 2 and 3, the frame number that to be decoded is 0, 2, 3 and 4 respectively, as shown in Fig. 5. If GOPlength is equal to 4, TLmaxis equal to 2. According to the formula (3) and (4), the number of images is 1, 1 and 2, when TLID is equal to 0, 1 and 2 respectively. If getting frame of TLID is equal to 0, 1 and 2, the frame number that to be decoded is 0, 2 and 3 respectively, as shown in Fig. 4 and Fig. 5.

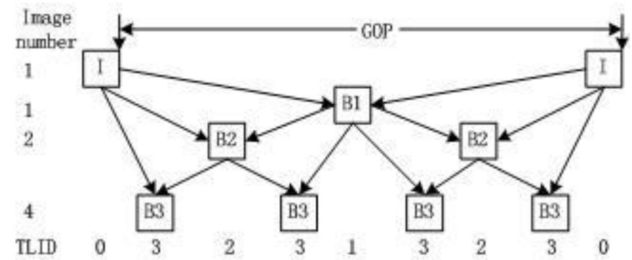


Fig. 4 Image number of the different temporal layer (GOP=8)

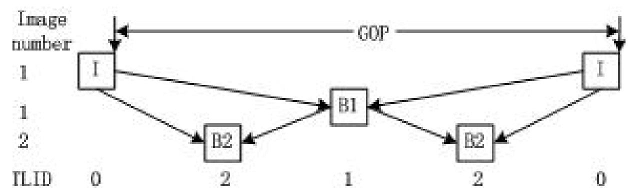


Fig. 5Image number of the different temporal layer (GOP=4)

This figure can be seen the length of GOP has direct influence on maximum number of frames which need to be decoded to access a frame, the random access performance of hierarchical B frame prediction structure decreases with the increase of GOP length, therefor, the improved random access performance can be got by decreasing the GOP length of hierarchical B frame. Based on the above mentiond points, this paper puts forward another coding prediction scheme that V4-view be chosen as base-view and GOPlength is equal to 4, which as shown in Fig. 6.

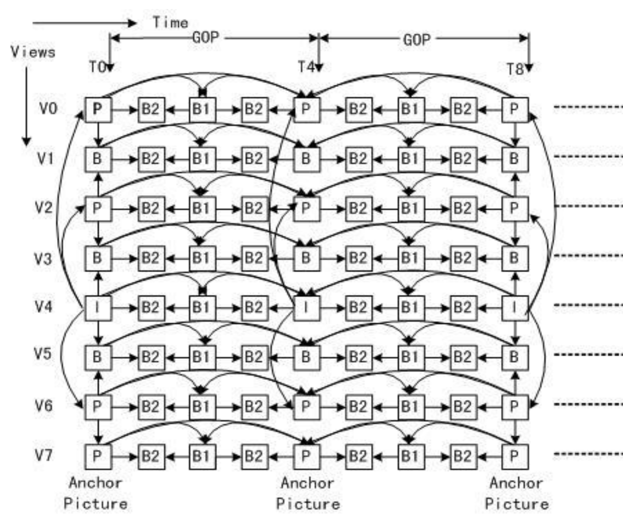


Fig. 6 Proposed coding prediction scheme MVC-2 (GOP=4)

3.3Random Access Performance

Random access performance is cost for accessing any frame in one video sequence, which is important indicator of evaluating prediction structure. In this paper random access performance measured by the number of frames which need to be decoded to access a frame in one GOP.

When access a frame, the number of frames which need to be decoded are marked as n_{ij} .Where, i denotes the serial number of the frame accessed, j denotes the relative serial number of the frame in the GOP.

For evaluating synthetically random access performance of the prediction structure, two parameters are defined [8]: average number of frames marke as N_{avg} ; maximum number of frames to access a frame are marked as $N_{\max}$. Where N_{avg} denotes the average random access performance, and $N_{\max}$ denotes the worst random access performance, the smaller the value of $N_{\max}$ and N_{avg} , the better random access performance of prediction structure. The formulas are given by:

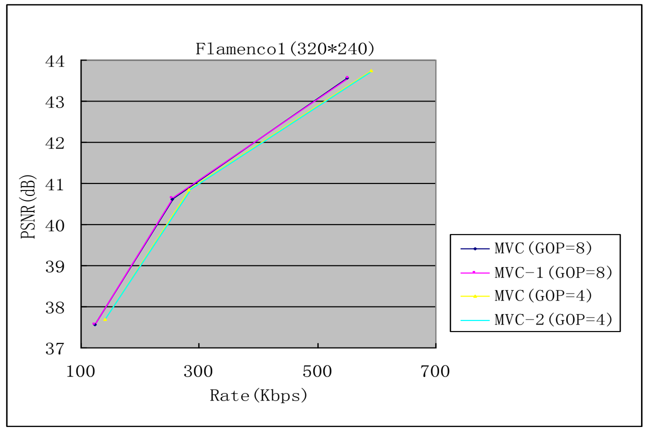
$$N_{avg} = \frac{1}{ViewNum \times GOPlength} \sum_{i=0}^{ViewNum} \sum_{j=0}^{GOPlength} n_{ij} \qquad (5)$$

$$N_{\max} = \max(n_{ij}) \quad i = 0 \sim \text{ViewNum - 1, j} = 0 \sim \text{GOPlength - 1} \qquad (6)$$

Where, ViewNum denotes the number of views, GOPlength denotes the length of one GOP.

4.Experiment Results

To test the coding efficiency of MVC scheme and proposed scheme, some experiments has been done, which bases on H.264/AVC MVC codec JMVC 7.0 with the sequences “Flamenco1” and “Race1”, which consists of 8 views captured by KDDI Corp. and Nagoya University. Because scene change slowly in sequence “Flamenco1”, and scene change quickly in sequence “Race1”, so we select the two typical sequences for test. The spatial resolution is 320×240 for Flamenco1, and 640×480 for Race1. The frame rate is 30 fps for both Flamenco1 and Race1. BasisQP of encoder JMVC is 32, 27 and 22 respectively. The GOPSize is 4 and 8 respectively. The SearchMode is FastSearch. The rate-distortion performance about different coding schemes in two MVC sequences as shown in Fig. 7. The MVC-1 (GOP=8) scheme changes inter-view prediction structure, while the MVC-2 (GOP=4) scheme decreases the GOPlength and changes inter-view prediction structure, experiment rusults show that The PSNR of MVC-1 (GOP=8) scheme was slightly higher than MVC (GOP=8) scheme, which less than 0.05 dB. The PSNR of MVC-2 (GOP=4) scheme was slightly lower than MVC (GOP=8) Scheme, which less than 0.15 dB, but The PSNR of MVC-2 (GOP=4) scheme was slightly higher than MVC (GOP=4) Scheme, which less than 0.03 dB.



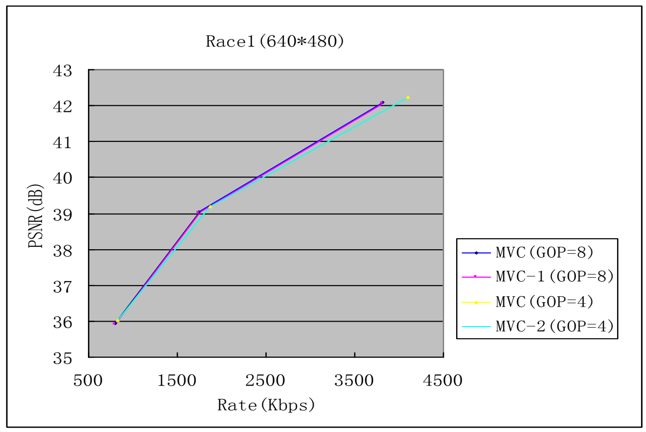


Fig. 7 Comparing the proposed scheme with MVC coding prediction structure

Random accsee performance as shown in TABLE II, TABLE III, TABLE IV, and TABLE V.

TABLE 2:NUMBER OF REFERENCE FRAMES NEED TO BE DECODED IN MVC (GOP=8) SCHEME

V/T	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
V_0	4	3	4	2	4	3	4	0
V_1	8	7	8	6	8	7	8	2
V_2	6	5	6	4	6	5	6	1
V_3	10	9	10	8	10	9	10	3
V_4	8	7	8	6	8	7	8	2
V_5	12	11	12	10	12	11	12	4
V_6	10	9	10	8	10	9	10	3
V_7	12	11	12	10	12	11	12	4

N_{avg} =7.45 $N_{\max}$ =12

TABLE 3:NUMBER OF REFERENCE FRAMES NEED TO BE DECODED IN MVC-1 (GOP=8) SCHEME

V/T	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
V_0	6	5	6	4	6	5	6	1
V_1	8	7	8	6	8	7	8	2
V_2	6	5	6	4	6	5	6	1
V_3	8	7	8	6	8	7	8	2
V_4	4	3	4	2	4	3	4	0
V_5	8	7	8	6	8	7	8	2
V_6	6	5	6	4	6	5	6	1
V_7	8	7	8	6	8	7	8	2

$N_{avg} = 5.57$ $N_{\max} = 8$

TABLE 4:NUMBER OF REFERENCE FRAMES NEED TO BE DECODED IN MVC (GOP=4) SCHEME

V/T	T_1	T_2	T_3	T_4
V_0	3	2	3	0
V_1	7	6	7	2
V_2	5	4	5	1
V_3	9	8	9	3
V_4	7	6	7	2
V_5	11	10	11	4
V_6	9	8	9	3
V_7	11	10	11	4

$N_{avg} = 6.15$ $N_{\max} = 11$

TABLE 5:NUMBER OF REFERENCE FRAMES NEED TO BE DECODED IN MVC-2 (GOP=4) SCHEME

V/T	T_1	T_2	T_3	T_4
V_0	5	4	5	1
V_1	7	6	7	2
V_2	5	4	5	1
V_3	7	6	7	2
V_4	3	2	3	0
V_5	7	6	7	2
V_6	5	4	5	1
V_7	7	6	7	2

$N_{avg} = 4.40$ $N_{\max} = 7$

It can be seen that random access performance of MVC-1 (GOP=8) scheme were higher than MVC (GOP=8)one, average number of frame(N_{avg}) to be referred decreases 1.88, more than 25%, and maximum number of frame($N_{\max}$)to be referred lower 4; Compare the coding scheme of MVC-2 (GOP=4) with the MVC(GOP=8) , the average number of frame(N_{avg}) to be referred decreases 1.17, more than 21%, and the maximum number of frame($N_{\max}$)to be referred decreases 1; Compare the encoding scheme of MVC-2 (GOP=4) with the MVC (GOP=4) , the average number of frame(N_{avg}) to be referred decreases 1.75,more than 28%, and the maximum number of frame($N_{\max}$) to be referred decreases 4.

5. Conclusion

On the basis of the analysis of the several typical Multi-view Video Coding schemes, this paper puts forward a proposed coding scheme, which selects the location of base-view and the prediction structure of inter-view by calculating global disparity; while improving the random access performance by decrease the length of GOP. Experiment results show that the proposed scheme with more performance of view random access while maintaining high coding efficiency.

References

[1] Ho Yo-Sung, Oh Kwan-Jung, "Overview of Multi-view Video Coding" ,Systems, Signals and Image Processing, 2007 14th Intl Conf On Held With Speech and Image Processing, Multimedia Comm&Svcs, 2007 6th Eurasip Conference Focused On, Maribor, Slovenia. Maribor, Slovenia: IWSSIP, 2007, 5-12.

[2] Lini Ma and Feng Pan，Efficient Compression of Multi-View Video Using Hierarchical B Pictures, Proceedings of MUE 2008, 118-121, Busan, Korea.

[3] Jens-Rainer Ohm,"Submissions received in Cfp on Multiview Video Coding",75th MPEG meeting,M12969, 2006.

[4] MPEG Test and Video subgroup, "Subjective test results for the Cfp on Multi-view Video Coding",75th MPEG meeting, W7779, 2006.

[5] PhiliPPMerkle,AljisehaSmolie,KarstenMuller,etal,"Efficient Predictio Structures for Multiview Video Coding".IEEE Transaetions on Circuits and Systems for Video Teehnology, 2007, Vol.17, No.ll，pp.1461-1471.

[6] Park P K, Oh K J, and Ho Y S, "Efficient view-temporal prediction structures for multi-view video coding", Electronics Letters, Vol.44, No.2,2008, pp.102-103.

[7] P.Merkle, A.Smolic, K.Muller,and T.Wiegand,"Efficient Prediction Structures for Multiview Video Coding", IEEE Trans. On Circuits and Systems for Video Technology,Vol.17, No.11,2007, pp.1461-1473.

[8] Heiko Schwarz, Detlev Marpe, Thomas Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard" IEEE Trans, On Circuits and Systmes for Video Technology, Vol.17, No.9,2007,pp.1103-1120.

Secure and Verifiable (2, 2) Secret Sharing Scheme for Binary Images

Sonali Patil[1], Sandip Sathe[2], Pravin Mehetre[3], Deepak Shinde[4], Kiran Bhalerao[5], Pawankumar Pandey[6]

[1] Assistant Professor, Computer Department, Pimpri Chinchwad Colege of Engineering, Pune, Maharashtra, India

[2] Student, Computer Department, Pimpri Chinchwad Colege of Engineering, Pune, Maharashtra, India

[3] Student, Computer Department, Pimpri Chinchwad Colege of Engineering, Pune, Maharashtra, India

[4] Student, Computer Department, Pimpri Chinchwad Colege of Engineering, Pune, Maharashtra, India

[5] Student, Computer Department, Pimpri Chinchwad Colege of Engineering, Pune, Maharashtra, India

[6] Student, Computer Department, Pimpri Chinchwad Colege of Engineering, Pune, Maharashtra, India

Abstract

Visual Cryptography is a technique in which a secret is encrypted into several image shares and then decrypted later using a human visual system to stack all the share images. Conventional visual cryptography methods divide a secret image into n shares (shadows) and distribute these shares to n participants. But in network while transmission the shadows can be changed by attackers or damaged. To remedy such kind of vulnerabilities verifiability of shadows can be a solution. Watermarking can add verifiability to secret sharing, but the shadows are meaningless which can attract the attacker's attention. The proposed scheme embeds created shadows in cover images which make them more secure. In this paper we have explained how a low computational complexity visual secret sharing scheme is verifiable and more secure by combining Watermarking and Steganography.

Keywords: *Network Security, Visual Cryptography, Secret Sharing, Verifiable Secret Sharing, Steganography.*

1. Introduction

Secure transmission of data is more and more needed in the worldwide computer network environment. The effective and secure protections of sensitive information are primary concerns where only encrypting data is not a solution. Secret Sharing Schemes refers to method for distributing a secret amongst a group of participants, each of whom is allocated a share of the secret. The secret can be reconstructed only when a sufficient number of shares are combined together; individual shares are of no use on their own. Shamir [1] introduced a secret sharing in 1979. Visual cryptography (VC) is a secret-sharing scheme that uses the human visual system to perform the computations. Naor and Shamir [2] introduced Visual Cryptography (VC) in 1994.

Very few researchers have proposed the combination of secret image sharing and hiding techniques. These techniques give higher reliability and security at the same time compared to only sharing or only hiding techniques. Chin-Chen Chang and Duc Kieu [3] have proposed a novel secret sharing and information-hiding scheme by embedding a secret image and a secret bit stream into two shadow images. It has limited reliability and shadow image size is more. Y.S. Wu, C.C. Thien, and J.C. Lin [4] have proposed sharing and hiding of secret images but with size constraint. Here in proposed scheme each shadow is individually embedded into cover image using BPCS (Bit Plane Complexity Segmentation) [5] method. Wang's [6] verifiable secret sharing method is used to create the shares/shadows for binary images.

2. Review

2.1 Review of Shamir's [1] Secret Sharing Scheme

Shamir developed the idea of a (k, n) threshold-based secret sharing technique ($k \leq n$). The technique allows a polynomial function of order ($k - 1$) constructed as,

$f(x) = d_0 + d_1x_1 + d_2x_2 + \ldots + d_{k-1}x_{k-1} \pmod p$,
where, the value d_0 is the secret and p is a prime number.
The secret shares are the pairs of values (x_i, y_i),
Where, $y_i = f(x_i)$, $1 \le i \le n$ and $0 < x_1 < x_2 \ldots < x_n \le p - 1$.
The polynomial function f(x) is destroyed after each shareholder possesses a pair of values (x_i, y_i) so that no single shareholder knows the secret value d_0. In fact, no groups of $(k - 1)$ or fewer secret shares can discover the secret d_0. On the other hand, when k or more secret shares are available, then we may set at least k linear equations $y_i = f(x_i)$ for the unknown d_i's. The unique solution to these equations shows that the secret value d_0 can be easily obtained by using Lagrange interpolation [1].
Shamir's SSS is regarded as a perfect secret sharing scheme because knowing even $(k - 1)$ linear equations doesn't expose any information about the secret.

2.2 Review of Image Embedding Schemes

Image embedding hides a secret message in a cover image, this process is usually parameterized by a hide-key, and the detection or reading of embedded information is possible only by having this key.

2.2.1 Least Significant Bit Insertion [7]
In this method the secret message is embedded into the least significant bit plane of the image. Since this only affects each pixel by +/- 1, if at all, it is generally assumed with good reason that the degradation caused by this embedding process would be perceptually transparent. Hence there are a number of LSB based Steganography techniques available in the public domain. The problem with this method is that it does not provide protection against small changes resulting from lossy compression or image transformations. The other disadvantage of this method is that it is having very less data hiding capacity. Therefore, improvements as suggested by R. J. Anderson and F. A. P. Petitcolas [8] are urged for LSB.

2.2.2 Adaptive MELSBR Method [9]
To avoid changing the properties of cover-images, the message must be embedded in "random texture" areas of each bit-plane. For taking advantage of local characteristics, an adaptive Steganography method based on the Minimum Error LSB Replacement (MELSBR) method is proposed. First, the upper bound of embedding capacity for each pixel in the cover-image is evaluated. If the amount of message to be embedded is less than the total embedding capacity provided by the cover-image, whole secret message will be embed in a local area and it can be easier for the attacker to extract the secret. To treat this scattering method is provided.

3. Proposed Method

The proposed method is based on verifiable (2, 2) secret sharing for binary images proposed by Wang [6]. Wang used watermarking for verifiability of shares as well as reconstructed secret. The receiver end receives the share and then extract watermark image from the original image. If watermark image is same as what is sent by sender, the received secret image is verified. But in network the created meaningless shadows can attract attacker's attraction. In proposed method BPCS Steganography is added to make those meaningless created shadows in meaningful images which add more security to the scheme.
BPCS Steganography:
Suggested technique to embed secret data into a dummy cover image is based on BPCS. The key idea to this approach is that a binary image can be categorized as "informative" and "noise-like" regions, which are segmented by a "complexity measure". If the embedding data is noise-like, we can hide it in the noise-like region of the dummy image. If a part of embedding data is simple, then we apply "image conjugate" operation to it. This operation transforms a simple pattern into a complex pattern.
Following steps describes the algorithm for embedding:

a. Segment each bit-plane of the cover image into informative and noise-like regions by using a threshold value (α). A typical value is $\alpha = 0.3$.
b. Group the bytes of the secret file into a series of secret blocks.
c. If a block (S) is less complex than the threshold (α), then conjugate it to make it a more complex block (S*). The conjugated block must be more complex than α.
d. Embed each shadow image block into the noise-like regions of the bit-planes (or, replace all the noise-like regions with a series of secret blocks). If the block is conjugated, then record this fact in a "conjugation map."
e. Also embed the conjugation map as was done with the secret blocks.

When the stego image is ready for transmission it is transmitted over the network. This transmission is more secure and reliable in comparison to any other technique.
The proposed scheme is divided into 2 parts as follows:

Construction Phase and Reconstruction Phase

The flowcharts for construction phase and reconstruction phase are given below:

Construction phase

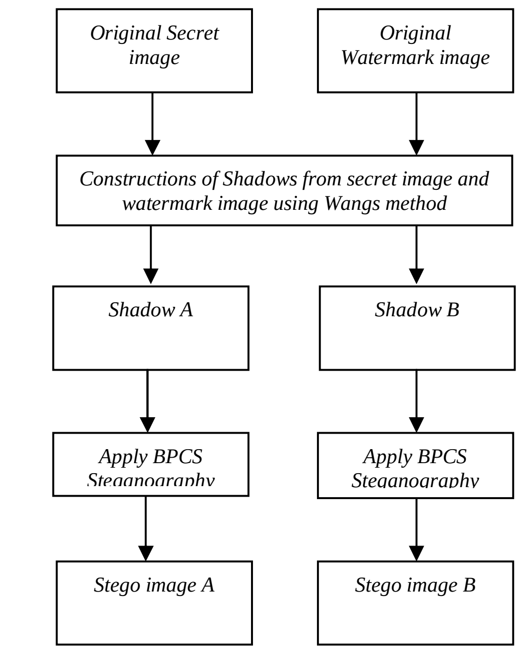


Fig. 1. Construction of secret image

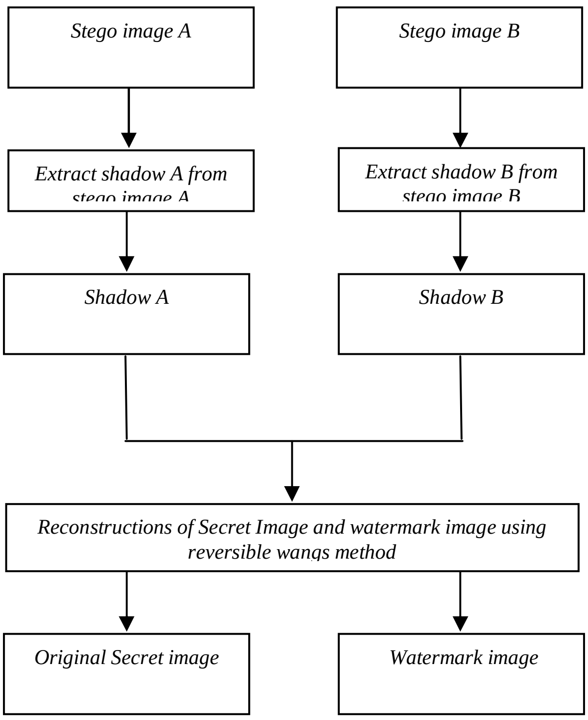


Fig. 2. Reconstruction of secret image

4. Experimental Results

The experimental results are produced to compare used BPCS Steganography with LSB method.

Table 1. Comparison Between LSB and BPCS Steganography methods

Cover Image	Share Image	Using LSB		Using BPCS	
		PSNR in db	DHC in %	PSNR in db	DHC in %
Brain.bmp 128×128	Share1.bmp 32×32	103.26	12.5	121.47	52.6
Baboon.bmp 512×512	Share2.bmp 32×32	97.23	12.5	117.83	57.13

PSNR: *Peak to Signal Noise Ratio* ***DHC:*** *Data Hiding Capacity*

5. Conclusion

By adding embedding in verifiable secret sharing a more secure secret sharing for binary images is obtained. The used embedding method provides high PSNR and also the data hiding capacity is more. The experimental results shows that the scheme is secure, verifiable and with low computation complexity. In future the proposed secret sharing scheme can be extended to color images and (2, 2) secret sharing can be extended to threshold (t, n) secret sharing scheme.

References

[1] Shamir, How to share a secret, Communications of the Association for Computing Machinery, vol. 22, no. 11, pp. 612-613, 1979.

[2] M. Naor and A. Shamir, Visual cryptography, Lecture Notes Computer Science, vol. 50, pp. 1-12, 1995.

[3] Chin-Chen Chang, The Duc Kieu “Secret Sharing and Information Hiding by Shadow Images”, 2006.

[4] C.Thien, and J. C. Lin, Secret Image Sharing, Computers and Graphics, vol. 26, no. 1, pp. 765-770,2002.

[5] Michiharu Nimmi, Hideki Noda and Eiji Kawaguch, An image embedding in image by a complexity based region segmentation method, Proceedings of the 1997 International Conference on Image Processing (ICIP '97).

[6] Zhi-hui Wang, Chin-Chen Chang, Huynh Ngoc Tu, Ming-Chu Li, Sharing a Secret Image in Binary Images with Verification, Volume 2, Number 1, January 2011.

[7] N.F. Johnson, Z. Duric, and S. Jajodia, “Information hiding: Steganography and watermarking- attacks and countermeasures”, Kluwer Academic Publishers, 2000.

[8] R. J. Anderson and F. A. P. Petitcolas, “On the limits of steganography,” IEEE J. Select. Areas Commun, vol. 16, no. 4, pp. 474–481, May 1998.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Optimization Of IPv4 Packet's Headers

Fahim A. Ahmed Ghanem[1], Vilas M. Thakare[2]

[1]Research Student, School of Computational Sciences,
Swami Ramanand Teerth Marathwada University, Nanded, India,

[2]Professor and Head of Computer Science, Faculty of Engineering & Technology
Post Graduate Department Of Computer Science, SGB, Amravati University, Amravati.

ABSTRACT

This paper aims to optimize the current IP packet structure by reducing the size of the headers, increasing the size of the data, reducing packet processing time and speed up the data transmission.

KEYWORDS: TCP/IP, UDP, IPv4, Ethernet Frame.

1. Introduction

The current IPv4 Packet's scheme has several fields to be carried and transmitted in each packet even if the source, the destination and the payload of the packets are same. These fields utilize packet space and thus decrease the size of user data. Therefore the processing of each packet requires more time and delay of end-to-end transmission. While data transmission, data has to be broken down into similar structures of the packets, which are reassembled to the original data chunk once it reaches the destination. A packet is also called a datagram, a segment, a block, a cell or a frame, depending on the protocol.

2. Objective Of This Study

This paper aims to optimize data transmission by redesigning OSI packet structure by dividing packets into master packet and slave packet, decreasing the size of packet's header and increase the size of the data while maintaining the same current packet size. This optimization aims to reduce the processing time as a result of reducing the contents of processed headers.

3. OSI Model Layers

Before we start with packet optimization, it will be helpful to have a quick review of the OSI model packet's construction. In this section we will have a look through OSI model. There are two types of header fields:

a. Variable Fields: fields that it is information from packet to packet and has to be repeated in each packet.
b. Invariable Fields (Tagged with Inv.): field that it is information doesn't vary from packet to other and doesn't have to be repeated in each packet.

Open Systems Interconnection (OSI) model (ISO/IEC 7498-1) is a product of the Open Systems Interconnection effort at the International organization for Standardization. It is a prescription of characterizing and standardizing the functions of a communications system in terms of abstraction. Similar communication functions are grouped into logical layers. A layer serves the layer above it and is served by the layer below it.

There are seven interconnection layers start from up to down (layer 7 to layer 1) prospectively:

Application forwards user data to Presentation layer which forwards it to Session layer. Session layer forwards data user to Transportation layer. Transport layer adds the header and sent the segment to Network layer which adds header and forward the packet to Data link layer. Data link adds the header and forward the frame to Physical layer.

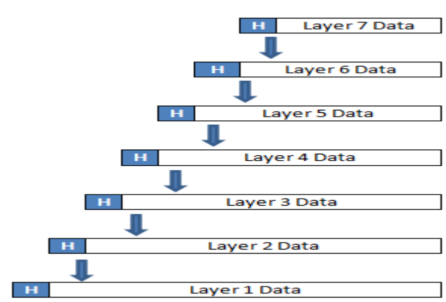


Fig. 1 OSI 7 layers' hierarchy

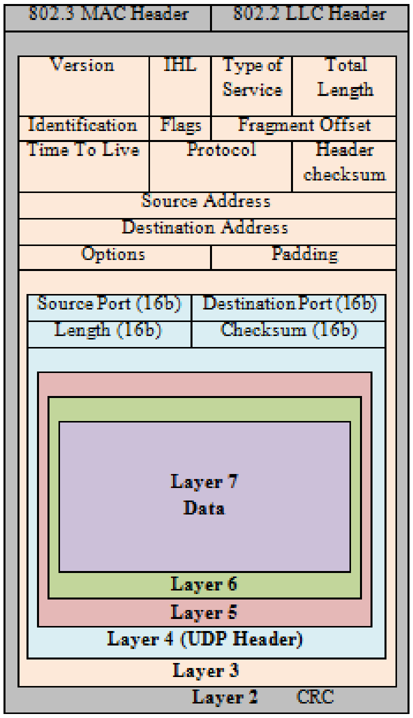


Fig. 2 full frame (UDP)

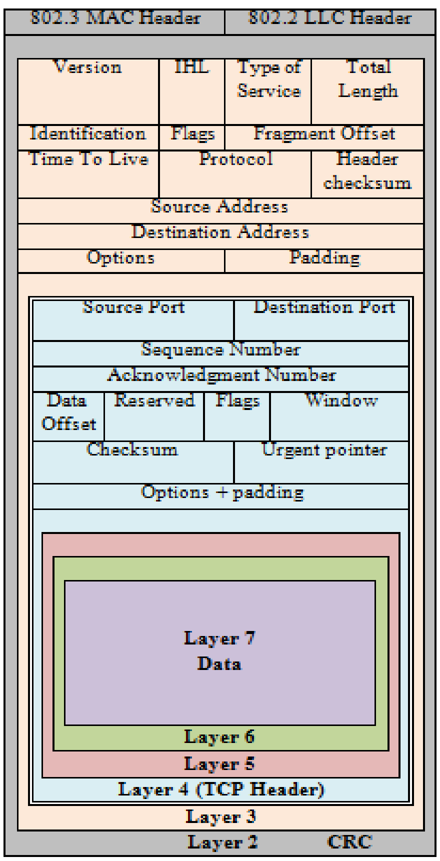


Fig. 3 full frame (TCP)

In this paper, we will concentrate on layer 4, layer 3 and layer 2 headers. We will ignore layer 7, layer 6 and layer 5 headers as it is variables from application to other.

3.1 Transport Layer Headers

Transport layer contains either TCP or UDP headers:

3.1.1 TCP Header

The Transmission Control Protocol (TCP) is one of the core protocols of the Internet Protocol Suite provides reliable, ordered delivery of a stream of bytes from a program on one computer to another program on another computer. TCP is the protocol that major Internet applications such as the World Wide Web, email, remote administration and file transfer rely on.

Below description displays the contents of TCP header contents:

a. Source Port (inv.): the source port number.
b. Destination Port (inv.): the destination port number.
c. Sequence Number: the sequence number of the first data octet in this segment.
d. Acknowledgment Number: if the ACK control bit is set this field contains the value of the next sequence number the sender of the segment is expecting to receive.
e. Data offset (inv.): this indicates where the data begins.
f. Reserved (inv.): reserved for future use. Must be zero.
g. Control Bits:
   1 URG: Urgent Pointer field significant
   2 ACK: Acknowledgment field significant
   3 PSH: Push Function
   4 RST: Reset the connection
   5 SYN: Synchronize sequence numbers
   6 FIN: No more data from sender
   7 Window: the number of data octets beginning with the one indicated in the acknowledgment field which the sender of this segment is willing to accept.
h. Checksum: the checksum field is the 16 bit one's complement of the one's complement sum of all 16 bit words in the header and text.

i. Urgent Pointer (inv.): this field communicates the current value of the urgent pointer as a positive offset from the sequence number in this segment.
j. Options (inv.): variable, Options may occupy space at the end of the TCP header and are a multiple of 8 bits in length.
k. Padding (inv.): variable, the TCP header padding is used to ensure that the TCP header ends.

TCP header varies from 20 bytes up to 60 bytes.

3.1.2 UDP Header

UDP is a connectionless protocol that provides an unreliable data services [1]. UDP is one of the core members of the Internet Protocol Suite, the set of network protocols used for the Internet. With UDP, computer applications can send messages to other hosts on an Internet Protocol (IP) network without requiring prior communications to set up special transmission channels or data paths. UDP applications must generally be willing to accept some loss, errors or duplication. Some applications such as TFTP may add rudimentary reliability mechanisms into the application layer as needed [2]. UDP provides application multiplexing (via port numbers) and integrity verification (via checksum) of the header and payload [3]. UDP is suitable for purposes where error checking and correction is either not necessary or performed in the application, avoiding the overhead of such processing at the network interface level [4].

Below description displays the contents of UDP header contents:

a. Source port (inv.): It identifies the sender's port.
b. Destination port (inv.): it identifies the receiver's port.
c. Length, it specifies the length in bytes of the entire datagram: header and data.
d. Checksum: It is used for error-checking of the header and data. If no checksum is generated by the transmitter, the field uses the value all-zeros [5].

UDP header varies from 4 to 8 bytes.

3.2 Network Layer Header

Network layer header consists of 13 fields, which 12 are required, the 13th field is optional. Network layer header size is varies from 20 bytes to 60 bytes.

Below description explains the contents of network layer header contents:

a. Version (inv.): indicates the version of IPv4 currently used.
b. IPv4 Header Length (IHL): indicates the datagram header length in 32-bit words
c. Type-of-Service (inv.): specifies how an upper-layer protocol would like a current datagram to be handled, and assigns datagrams various levels of importance.
d. Total Length: specifies the length, in bytes, of the entire IPv4 packet, including the data and header.
e. Identification: contains an integer that identifies the current datagram.
f. Flags (inv.): Consists of a 3-bit field of which the two low-order (least-significant) bits control fragmentation. The low-order bit specifies whether the packet can be fragmented. The middle bit specifies whether the packet is the last fragment in a series of fragmented packets. The third or high-order bit is not used.
g. Fragment Offset (inv.): indicates the position of the fragment's data relative to the beginning of the data in the original datagram, which allows the destination IPv4 process to properly reconstruct the original datagram.
h. Time-to-Live (inv.): maintains a counter that gradually decrements down to zero, at which point the datagram is discarded. This keeps packets from looping endlessly.
i. Protocol (inv.): indicates which upper-layer protocol receives incoming packets after IPv4 processing is complete.
j. Header Checksum: helps ensure IPv4 header integrity.
k. Source Address (inv.): specifies the sending node.
l. Destination Address (inv.): specifies the receiving node.
m. Options (inv.): Allows IPv4 to support various options, such as security.

The minimum IPv4 header is 20 byte and the maximum is 40 byte.

3.3 Data Link Header and Tailor (802.3 IEEE)

Data link frame header and Tailor consist of 4 fields. It is chunk of data that is packaged for transmission over network [6].

Ethernet II framing (also known as DIX Ethernet) named after DEC, Intel and Xerox, the major participants in its design. Below description display the contents of Ethernet header:

a. Destination Address (Inv.): It specifies either a single recipient node (unicast mode), a group of recipient nodes (multicast mode), or the set of all recipient nodes (broadcast mode).
b. Source Address (Inv.): sender's globally unique node address. This may be used by the network layer protocol to identify the sender, but usually other mechanisms are used (e.g. arp). Its main function is to allow address learning which may be used to configure the filter tables in a bridge.
c. LLC Header (Inv.): it contains DSAP, SSAP and control byte.
d. CRC: this field is added at the end of the frame (trailer) and provides error detection in the case where line errors (or transmission collisions in Ethernet) result in corruption of the MAC frame. Any frame with an invalid CRC is discarded by the MAC receiver without further processing. The MAC protocol does not provide any indication that a frame has been discarded due to an invalid CRC.

The size of Ethernet header along with CRC is 21 bytes.

As per the above overview of OSI models and based on the traffic type, Ethernet frame can contains two types of segment headers: TCP header or UDP header. The accumulated headers within Ethernet frame (TCP header + L3 header + L2 header and tailor) varies from 81bytes to 121 bytes.

In sample case, if we have a TCP/IP sentence contains of 50,000 bytes, this sentence will be divided into small chunks called frames and the maximum size of each frame will not exceed 1518 bytes. Headers will consume 121 byte out of 1518 and the remaining 1379 bytes will be available for data, so 50,000 bytes will be transmitted over 37 packets as a result of dividing 50,000 by 1379.

The accumulated headers within Ethernet frame (UDP header + L3 header + L2 header and tailor) varies from 45 bytes to 69 bytes.

The maximum size of Ethernet packet will not exceed 1518 bytes. Headers consume 69 bytes out of 1518 and the remaining 1449 bytes will be available for data. According to the above data, we observe that we are facing two obstacles: first one is the size of the headers and the other one is the processing of those headers. Headers consume some packet's size and most of the data in the header sections is repeated in all packets and it is processed in each middle device till it reaches the destination.

4. Steps Of Optimization

Optimization steps are divided into two stages:

1- Headers Optimization:
   It contains creation of master and slave packet.
2- Packet Transmission:
   Packet's processing divided into two steps
   a. Processing of Master packet.
   b. Processing of Slave packet.

4.1 Headers Optimization

4.1.1 Proposed Master Packet

The proposed Master packet contains all header fields (Variables, Invariable, along with ID and Tag Fields) but without user data field.

4.1.1.1 Transport Layer Master Segment

The proposed transport layer Master segment contains all Transport layer header fields (Variables, Invariable, along with ID and Tag Fields) but without user data field.

4.1.1.1.1 TCP Master Segment

The proposed TCP Master segment is created when transport layer receives first packet in the session from session layer in order to be transmitted to the desired destination.

The contents of Master TCP segment are same contents of OSI TCP header with extra two fields:

a. ID: 1 bit indicates the type of the packet, as value 1 indicates Master segment and value 0 indicate slave packet.
b. Tag: 31 bits, is a unique ID for each session, the

value of tag will be maintained same till session is closed.

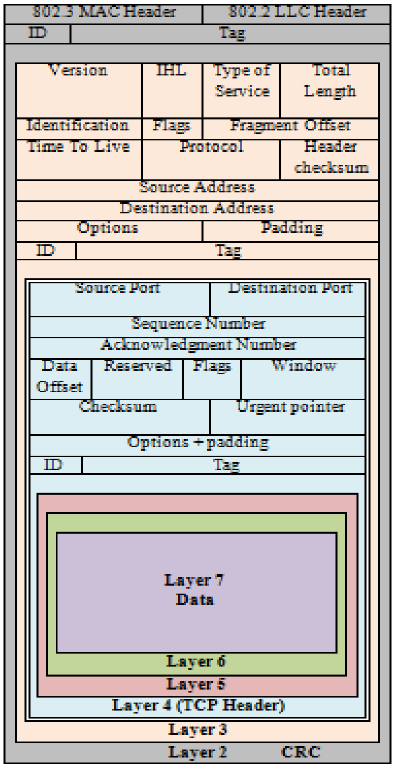


Fig. 4 Proposed full master TCP frame

4.1.1.1.2 UDP Master Segment

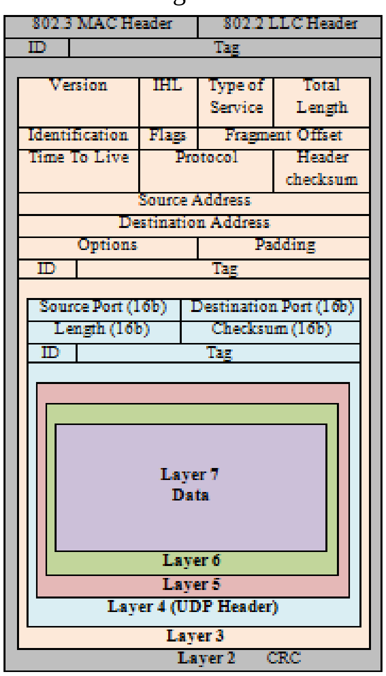


Fig. 5 Proposed full master UDP frame

The proposed UDP Master segment is created when transport layer receives first packet for each session from session layer in order to be transmitted to the desired destination. The contents of Master UDP segment are same as the contents of OSI UDP header with extra ID and Tag fields. The roles of these two fields are same as the role of TCP master segment.

4.1.1.2 Network Layer Master Packet

The proposed Network layer header is added to the segment received from transport layer. The contents of Master packet are same the contents of network layer header of OSI model with extra ID and Tag fields. The roles of these two fields are same as the role of transport layer for Master segment.

4.1.1.3 Data Link Master Frame

The proposed Ethernet header is added to the packet received from network layer. The contents of Master frame header are same the contents of Ethernet header of OSI model with extra ID and Tag fields. The roles of these two fields are same as its role in transport layer Master segment.

The size of Master Frame (TCP header) is the size of the layer 4, layer 3, layer 2 headers and tailor only without user Data which varies from 96 bytes (without optional fields) to 133 bytes (with optional fields). The size of Master (UDP headed) is the size of the layer 4, layer 3, layer 2 headers and tailor only without Data which varies from 57 bytes to 81 bytes.

4.1.2 Proposed Slave Packet

Slave packet contains only invariable header fields in addition to ID and Tag fields along with user data.

4.1.2.1 Transport Layer Slave Segment

The proposed transport layer packet contains invariable header fields in addition to ID and Tag fields along with user data.

4.1.2.1.1 TCP Slave Segment

The size of the header is 134 bits equals to 17 bytes.

4.1.2.1.2 UDP Slave packet

The proposed UDP packet header contains only invariable header fields in addition to ID and Tag Fields along with data. The size of the header is 64 bits equals to 8 bytes.

4.1.2.2 Network Layer Slave packet

The proposed layer 3 header contains only invariable header fields in addition to ID and Tag Fields along with user data. The size of the layer 3 header is 84 bits equals to 11 bytes.

4.1.2.3 Ethernet Slave Frame

The proposed layer 2 Frame contains only invariable header and Tailor fields in addition to ID and Tag Fields along with user data. The size of Ethernet header and tailor is 46 bits equals to 6 bytes.

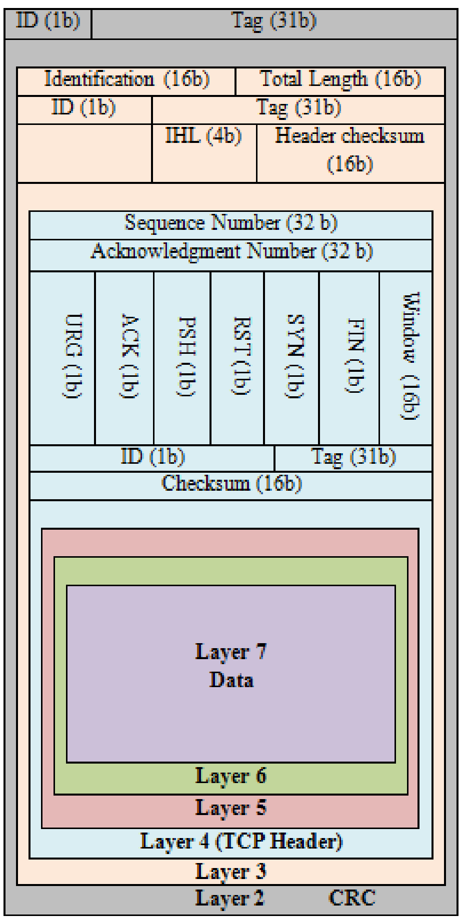


Fig. 6 Proposed full slave TCP frame

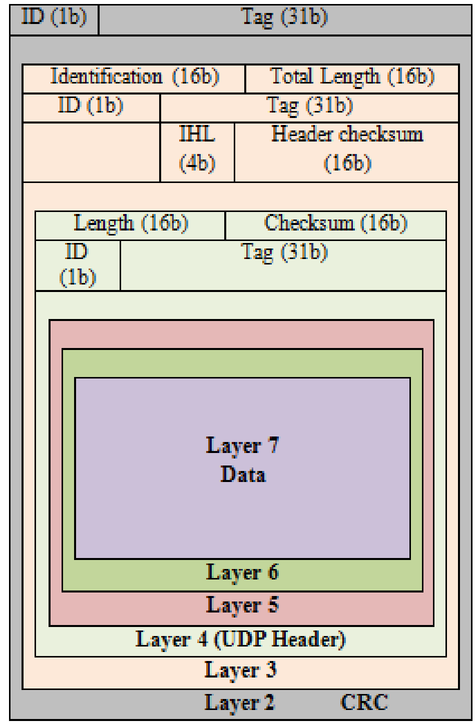


Fig. 7 Proposed full slave UDP frame

4.2 Packet Transmission

Packet transmission includes the process of transmission of master and slave packets from source to destination.

4.2.1 Master Packet's Transmission

4.2.1.1 Master TCP Segment, Packet and Frame Transmission

a. Application layer of the source station forwards user data to Presentation which forwards it to Session which also forwards it to transport layer.
b. Transport layer buffers the received user data, creates one master TCP segment and forward it to network layer.
c. Network layer adds Network layer master header to the received segment and send the packet to Ethernet layer.
d. Ethernet layer adds Ethernet master header and tailor and sends the frame to physical layer.
e. Physical layer adds it is header sends it to another station or middle devices towards destination.

f. Ethernet layer in layer 2 device checks the packet ID and identify whether it is Master packet or Slave packet by value of ID (1= master and 0 = slave).
g. If the value ID is 1, Ethernet layer processes Master packet as the processing of OSI model packet based on source, destination and content etc. and decide the next hop MAC.
h. Ethernet layer cashes the frame processing result, save it in the forwarding table along with frame's Tag and send the frame towards the destination.
i. If the next hop is layer 2, steps *f, g, and h* will be repeated.
j. If the next hop is layer 3 device, layer 3 checks the packet ID and identify whether it is Master Packet or Slave Packet by value of ID (1= master and 0 = slave).
k. If value ID is 1, layer 3, layer 3 processes master packet as OSI model processing based on source, destination and content etc. and decide the next hop IP.
l. Layers 3 cashes the processing result, save it in the forwarding table along with packet's Tag and forwards the packet towards the destination.
m. If the next hop is layer 3, steps *j, k,* and *l* will be repeated.
n. If the next hop is layer 4, layer 4 processes the segment as per layer 2 and layer 3 and forwards it to next hop or layer.
o. Steps f – n will be repeated till packet reaches the destination.
p. Once Transport layer of source station receives the acknowledgement about receiving Master packet from the destination, it starts creating and sending slave segments.
   Note: master TCP packet was sent only one time per each session.

4.2.1.2 Master UDP Segment, Packet and Frame Transmission

Transmission of master UDP packet takes steps from *a - n* of TCP packet but not step o, because UDP packet doesn't rely on acknowledgment.

Note: master UDP packet being transmitted every 5 seconds in order to insure continuously delivering of packets in case there is any transmission or processing issue in the middle devices.

4.2.2 Slave Packet's Transmission

4.2.2.1 Slave TCP Segment, Packet and Frame Transmission

a. Source transport layer creates either TCP or UDP slave header and sends the segment to Network layer.
b. Network layer adds it is slave header and sends the packet to Ethernet layer.
c. Ethernet layer adds it is slave header and tailor and sends it to physical layer.
d. Physical layer adds it is header and forwards it towards the destination.
e. Ethernet layer of the destination station or middle devices checks the packet ID and identify whether it is Master Frame or Slave Frame by value ID (1= master and 0 = slave).
f. If the value ID is 0, layer 2 matches the Tag with forwarding table information and forwards it to the next hope without any extra processing.
g. If there is no information about the Tag in the forwarding table, layer 2 simply drops the packet.
h. If the next hope is layer 2, step *e* and *f* will be repeated.
i. If the next hope is layer 3 device of the destination station, layer 3 will check the packet ID and identify whether it is Master Frame or Slave Frame by value ID (1= master and 0 = slave).
j. If the value ID is 0, layer 3 matches packet Tag with forwarding table information and forwards it to the next hope without any extra processing.
k. If there is no information about the Tag in the forwarding table, layer 3 simply drops the packet.
l. If the next hope is layer 3, step *i* and *j* will be repeated.
m. If the next hop is layer 4, layer 4 processes the segment like layer 2 and layer 3 and forwards it to next layer.
n. Steps *e* to *m* were repeated till packet reaches the destination.
o. Packet acknowledgment was sent as OSI model.

4.2.2.2 Slave UDP Segment, Packet and Segment Transmission:

Transmission of slave UDP segment, packet and frame takes is same as steps from *a - n* of TCP packet but not step *o*, because UDP packet doesn't rely on acknowledgment.

Note: master UDP packet was sent every 5 seconds in order to insure continuously delivering of the packets in case there is any transmission or processing issue in the middle devices.

Slave Ethernet packet (TCP header + L3 header + L2 header and tailor) is fixed to 282 bits equals to 36 bytes whether master packet contains optional fields or not.

Slave Ethernet packet (UDP header + L3 header + L2 header and tailor) is fixed to 212 bits equals to 7 bytes whether master packet contains optional fields or not.

5. Verification

Optimization of Packet header and packet transmission optimization gives ability to transmit more user data along with small headers and less processing time which increase the efficiency of the existing and feature network infrastructure.

Below two tables and four charts illustrate the comparison between ISO model and new proposed optimized packet where we observe the decreased size of Ethernet packet headers using TCP layer 4 and using UDP layer 4. The total size of Ethernet headers for optimized Ethernet (TCP header + IP header + Ethernet header and tailor) varies from 28.09% to 55.73% of OSI headers. The total size of Ethernet headers for optimized Ethernet (UDP header + IP header + Ethernet header and tailor) varies from 36.2 till 55.55% of OSI headers.

Table 1: comparison between OSI headers (TCP) and optimized packet headers

	IOS (TCP) without optional	Optimized	IOS (TCP) without optional	Optimized
TCP Header	20	17	60	17
IP Header	20	11	40	11
Frame Header	21	6	21	6
Total headers (bytes)	61	34	121	34

Table 2: comparison between OSI headers (UDP) and optimized packet headers

	IOS (TCP) without optional	Optimized	IOS (TCP) without optional	Optimized
TCP Header	4	8	8	8
IP Header	20	11	40	11
Frame Header	21	6	21	6
Total headers (bytes)	45	25	69	25

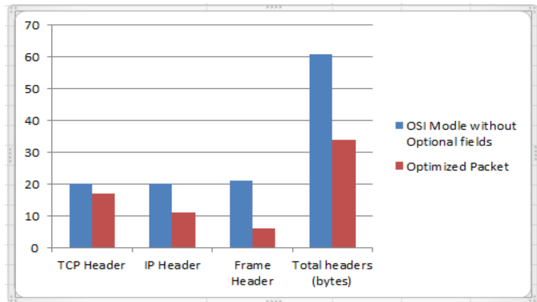


Fig. 8 comparison chart between OSI headers (TCP) without optional fields and optimized packet headers

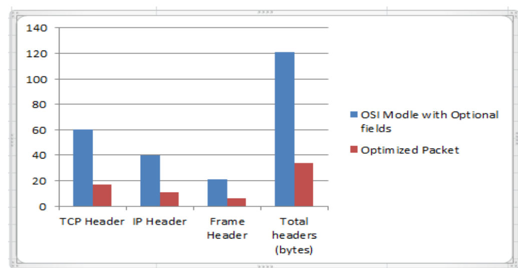


Fig. 9 comparison chart between OSI headers (TCP) with optional fields and optimized packet headers

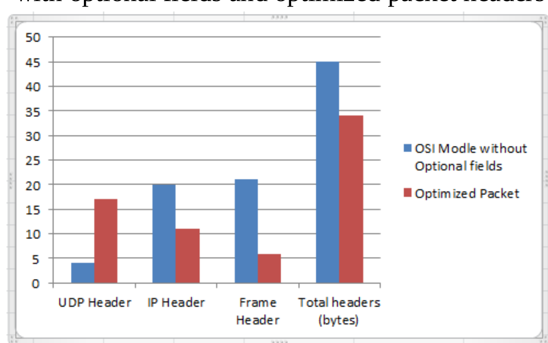


Fig. 10 comparison chart between OSI headers (UDP) without optional fields and optimized packet headers

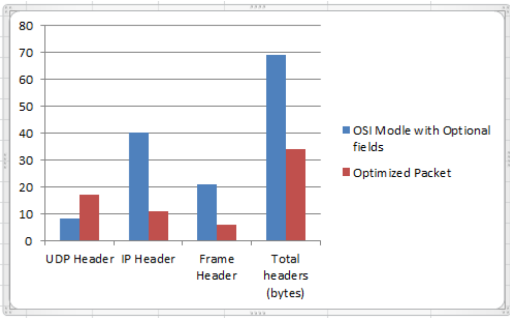


Fig. 11 comparison chart between OSI headers (UDP) with optional fields and optimized packet headers

6. Conclusion

Referring to figures 8 through 11, the new proposed IP packet structure and IP packet transmission decreases the size of headers between 44% up to 71%. Decreasing the headers' space give an opportunity to increase the size of user data while maintaining same frame size. Moreover, reducing the size of the headers will reduce the processing times as middle devices or layers will only process Master packet, cash the processing result and apply the result to all Slaves packets which belong to same session.
The proposed optimization solution can be applied to the current and new network infrastructure in order to increase the efficiency of the networks.

7. References

[1] Michael A. Gallo, William M. Hancock, "Computer Communications and Networking Technologies", New Delhi, India, Cengage Learning India Private Limited.
[2] Forouzan, B.A. "TCP/IP: Protocol Suite, 1st edition", New Delhi, India, Tata McGraw-Hill Publishing Company Ltd.
[3] Clark, M.P., "Data Networks IP and the Internet, 1st edition", West Sussex, England, John Wiley & Sons Ltd.
[4] Andrew S. Tanenbaum, "Computer Networks 4th Edition", Amsterdam, the Netherland, Prentice Hall PTR.
[5] Barrie Sosinsky, "Networking", New Delhi, India Wily India Pvt. Ltd. 2010.
[6] Kurose, J. F., Ross, K. W. "Computer Networking: A Top-Down Approach (5th ed.)", Boston, MA, Pearson Education. ISBN 978-0-13-136548-3

Mr. Fahim A. Ahmed Ghanem is a Ph.D. student at School of Computational Sciences, SWAMI RAMANAND TEERTH MARATHWADA UNIVERSITY, Nanded, India. He is doing research in field of IP Network Optimization. He received MSC-IT degree from SMU, India 2005 – 2007. He is a dual CCIE R&S and SP. He works as a Senior Network Engineer in mobily co. KSA.

(Supervisor) Mr. Vilas Mahadeorao Thakare, has received his M.E. degree in field of Advance Electronics, received his M.Sc. degree in field of Applied Electronics, Diploma in Computer Management. He was obtained his PhD in field of Computer Science.

Backstepping Adaptive Fuzzy Control for two-link robot manipulator

Yongqiao Wei[1], Jingdong Zhang[2], Li Hou[1], Fenglan Jia[1],Qinglin Chang[1]

[1]**School of Manufacturing Science and Engineering,Sichuan University,**
No.24 South Section 1Yihuan Road,Chengdu,610065,China

[2]**School of Transportation and Automobile Engineering,Panzhihua University,**
No.10 Airport Road,Panzhihua,617000,China

Abstract

In this paper, based on Lyapunov method, a backstepping adaptive fuzzy control scheme is presented for the two-link robot manipulator system. The control strategy consists of the traditional backstepping control and adaptive fuzzy control to cope with the model unknown and parameter disturbances.The simulation is presented to verify the effectiveness of the proposed control scheme. From the simulation results, fast response, strong robustness, good disturbance rejection capability and good angle tracking capability can be obtained. The output tracking error between the actual position output and the desired position output can asymptotically converge to zero. It is also revealed from simulation results that the proposed control strategy is valid for the two-link robot manipulator.

Keywords: *Backstepping control, Adaptive Fuzzy control, Two-link robot manipulator,MATLAB simulation.*

1. Introduction

In recent decades, the robot research has been paid great attention. Robotic is a vast research field, mainly because of the many potential applications. The basic problem in controlling robot is to make manipulator to perform preplanned trajectory. Therefore, it must be controlled properly to track some trajectory because there exist the uncertainties, nonlinear, strong coupling and time-varied of the robot system, and external disturbances. In order to achieve this goal, many schemes which were PID control, optimal control, sliding mode control(SMC), adaptive control, fuzzy control and so on, have been presented[1].

In this paper, based on Lyapunov method, an adaptive fuzzy control scheme combining backstepping is presented for the two-link robot manipulator system which has two rotational joints on a horizontal plane and one translational joint on the vertical axis. It is proved that the closed-loop system is globally stable in the Lyapunov sense. If all the signals are bounded, the system output can track the desired reference output asymptotically with uncertainties and disturbances[2-3].

2、Model of two-link manipulator

2.1 Model Description

In Engineering, robots not only can improve productivity but also can achieve high-strength, highly difficult and hazardous jobs. Manipulators are the usual plants in robotics. Using the lagrangian formulation, the dynamic equation of a planner n degrees of freedom rigid manipulator can be expressed as follows.

$$M\left(q_k(t)\right)\ddot{q}_k(t)+C\left(q_k(t),\dot{q}_k(t)\right)\dot{q}_k(t)+G\left(q_k(t)\right)$$
$$=\tau_K(t)+d_k(t) \quad (1)$$

Where $M\left(q_k(t)\right)\in R^{n\times n}$ is the inertia matrix. $q_k(t)\in R^n$ is the angle vector, $\dot{q}_k(t)\in R^n$ is the velocity vector, also the $\ddot{q}_k(t)\in R^n$ is acceleration vector. t denotes the time and is the nonnegative integer. $k\in Z_+$ denotes the operation or iteration times. $C\left(q_k(t),\dot{q}_k(t)\right)\in R^n$ is a vector resulting from the centrifugal and coriolis forces. $G\left(q_k(t)\right)$ is the gravity. $\tau_K(t)\in R^n$ is the control moment applied to the joints,and $d_k(t)\in R^n$ is the vector containing the unmodeled dynamics and other unknown external disturbances.

The characteristics of the kinetic model of a robot manipulator[4]:

(1) Kinetic mode contains more number of items: The number of items included in the equation increases with the increase in the number of robot joint.
(2) Highly nonlinearity: Each item of the equations contains non-linear factors such as sine and consine, et al.
(3) High degree of coupling.
(4) Model uncertainty and time-variant: Because the objects are not similar, the load will vary when the robot moves the objects. Also the joint friction torque will also change over time.

Suppose that the parameters of the system are unknown, and the following properties are written as:

Property 1

$M(q_k(t)) \in R^{n \times n}$ is a positive-definite symmetric, and bounded matrix. There exists positive constant,

$\sigma_0 > 0, \sigma_0 \in R, 0 < M(q_k(t)) \le \sigma_0 I$

Property 2

$C(q_k(t), \dot{q}_k(t))$ is bounded. there exists known $C_b(q)$ such that

$$\left|C(q_k(t), \dot{q}_k(t))\right| \le C_b(q)\left\|\dot{q}_k(t)\right\|$$

Property 3

Matrix $\dot{M}(q_k(t)) - 2C(q_k(t), \dot{q}_k(t))$ is a symmetric matrix, and

$$X^T\left(\dot{M}(q_k(t)) - 2C(q_k(t), \dot{q}_k(t))\right)X = 0 \qquad (2)$$

where X is a vector.

Property 4

The known disturbance is satisfied with $\|d_k\| \le d_m$, where d_m is a known positive constant.

2.2 Mathematical model

The two-link robot manipulator generally has two revolute joints and prismatic joint. The schematic diagram of the two-link manipulator is shown in Fig.1. The robots transport a load horizontally by actuating the two revolute joints. The robots transport a load vertically by actuating the prismatic joint[5].

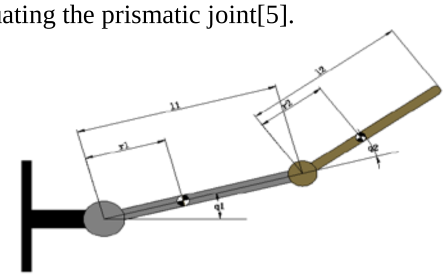


Fig.1 The schematic diagram of the two-link manipulator

The dynamic equation of the model for two-link robot manipulator can be expressed as follows:

$$M(q_k(t))\ddot{q}_k(t) + C(q_k(t), \dot{q}_k(t))\dot{q}_k(t) + d_k(t) = \tau_K(t)$$

$$y = q_k(t) \qquad (3)$$

According to Lagrangian method, the inertia matrix $M(q_k(t))$ can be written as:

$$M(q_k(t)) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \qquad (4)$$

Where

$$M_{11} = J_1 + J_2 + 2m_2r_2l_1\cos\theta_2,$$

$$M_{12} = J_2 + m_2r_2l_1\cos\theta_2,$$

$$M_{21} = J_2 + m_2r_2l_1\cos\theta_2,$$

$$M_{22} = J_2,$$

$$J_1 = \frac{4}{3}m_1r^2{}_1 + m_2l^2{}_1,$$

$$J_2 = \frac{4}{3}m_2r^2{}_2$$

Also the $C(q_k(t), \dot{q}_k(t))$ is given by:

$$C(q_k(t), \dot{q}_k(t)) = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \qquad (5)$$

Where

$$C_{11} = -2m_2r_2l_1\dot{\theta}_2\sin\theta_2,$$

$$C_{12} = -m_2r_2l_1\dot{\theta}_2\sin\theta_2,$$

$$C_{21} = m_2r_2l_1\dot{\theta}_2\sin\theta_2,$$

$$C_{22} = 0,$$

$$q = [\theta_1 \quad \theta_2]^T, \tau = [\tau_1 \quad \tau_2]^T$$

Where m_1 is the mass of the 1th link, m_2 is the mass of the 2th link, l_1 and l_2 are the length of each link. r_1 and r_2 are distances between the gravity center position and rotational position of the each links. θ_1 and θ_2 are the angle of each links. J_1 and J_2 are the inertia matrix of two links. In addition, $G(q_k(t))$ terms are ignored in this paper because the absence of the $G(q_k(t))$ terms in the equation of motion could be interpreted as assuming the robot is statically balanced. On the other hand, the absence of gravity is interpreted by simply assuming that the robot is working in outer space.

2.3 Design of Controller and Analysis of the Stability

In order to apply Backstepping method, define

$$x_1 = q_k(t), x_2 = \dot{q}_k(t).$$

Using (1),we can obtain :

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = M^{-1}(x_1)\tau - M^{-1}(x_1)C(x_1, x_2)x_2 - M^{-1}(x_1)d \qquad (6)$$

$$y = x_1$$

STEP 1:

Assuming y_d is the expected angle and has second order derivative. Define $z_1 = y - y_d$. α_1 is the estimation of the x_2. Define $z_2 = x_2 - \alpha_1$. According to selecting the appropriate α_1, making $z_2 \to 0$, we can obtain $\dot{z}_1$ following as:

$$\dot{z}_1 = \dot{x}_1 - \dot{y}_d = z_2 + \alpha_1 - \dot{y}_d \tag{7}$$

Select the virtual control item as:

$$\alpha_1 = -\lambda_1 z_1 + \dot{y}_d \left(\lambda_1 > 0\right) \tag{8}$$

Select the Lyapunov function for the first subsystem as :

$$V_1 = \frac{1}{2} z_1^T z_1 \tag{9}$$

There

$$\begin{aligned} \dot{V}_1 &= z_1^T \dot{z}_1 = z^T_1(\dot{y} - \dot{y}_d) \\ &= z^T_1(z_2 + \alpha_1 - \dot{y}_d) \\ &= -\lambda_1 z^T_1 z_1 + z^T_1 z_2 \end{aligned} \tag{10}$$

If z_2 is zero, the first subsystem is stable.

STEP 2:

Using (3) (5), we can obtain :

$$\dot{z}_2 = \dot{x}_2 - \dot{\alpha}_1 = -M^{-1}Cx_2 - M^{-1}d + M^{-1}\tau - \dot{\alpha}_1 \tag{11}$$

Select the control rule as following:

$$\tau = -\lambda_2 z_2 - z_1 - \phi \tag{12}$$

Select the Lyapunov function for the second subsystem:

$$V_2 = V_1 + \frac{1}{2} z^T_2 M z_2 \tag{13}$$

Therefore

$$\begin{aligned} \dot{V}_2 &= \dot{V}_1 + \frac{1}{2}\dot{z}_2^T M z_2 + \frac{1}{2} z^T_2 \dot{M} z_2 + \frac{1}{2} z^T_2 M \dot{z}_2 \\ &= -\lambda_1 z^T_1 z_1 + z^T_1 z_2 + z^T_2 M(\dot{x}_2 - \dot{\alpha}_1) + \frac{1}{2} z^T_2 \dot{M} z_2 \\ &= -\lambda_1 z^T_1 z_1 + z^T_1 z_2 + z^T_2 M(-M^{-1}Cx_2 - M^{-1}d + M^{-1}\tau - \dot{\alpha}_1) + z^T_2 C z_2 \\ &= -\lambda_1 z^T_1 z_1 + z^T_1 z_2 + z^T_2(f - \lambda_2 z_2 - z_1 - \phi) - z^T_2 d \\ &= -\lambda_1 z^T_1 z + -\lambda_2 z^T_2 z_2 + z^T_2(f - \phi) \end{aligned}$$

(14)

Where $f = -C\alpha_1 - M\dot{\alpha}_1$

From the expression of the f, we can obtain the modeling information of robotics system. In order to realize the control without model information, we make fuzzy system approximate the f. If ϕ is used to approximate the fuzzy system of the nonlinear function f, the single value fuzzification, the product inference engine and the center average defuzzifier are adopted [3,6,7,8].

If fuzzy system is consisted of N fuzzy rules, the ith fuzzy rule is expressed as:

R^i :IF x_1 is μ^k_1 and...and x_n is μ^k_n, then y is B^k （k=1,2，…N）,Where μ^k_i is the membership function of the x_i (i=1,2，…n).

The output of fuzzy system is written as:

$$y = \frac{\sum_{k=1}^{N} \theta_K \prod_{i=1}^{n} \mu^k_i\left(x_i\right)}{\sum_{k=1}^{N} \prod_{i=1}^{n} \mu^k_i\left(x_i\right)} = \xi^T \theta \tag{15}$$

Where $\xi = \left[\xi_1(x), \xi_2(x), \xi_3(x), \ldots, \xi_N(x)\right]$,

$$\xi_K(x) = \frac{\prod_{i=1}^{n} \mu^k_i\left(x_i\right)}{\sum_{k=1}^{N} \prod_{i=1}^{n} \mu^k_i\left(x_i\right)},$$

$$\theta = \left[\theta_1, \theta_2, \theta_3, \ldots, \theta_N\right]^T. \tag{16}$$

Based on the fuzzy approximation of the f, the fuzzy system for $f(1)$ and $f(2)$ is designed as:

$$\phi_1(x) = \frac{\sum_{k=1}^{N} \theta_{1K} \prod_{i=1}^{n} \mu^k_i\left(x_i\right)}{\sum_{k=1}^{N}\left[\prod_{i=1}^{n} \mu^k_i\left(x_i\right)\right]} = \xi_1^T \theta_1$$

$$\phi_1(x) = \frac{\sum_{k=1}^{N} \theta_{2K} \prod_{i=1}^{n} \mu^k_i\left(x_i\right)}{\sum_{k=1}^{N}\left[\prod_{i=1}^{n} \mu^k_i\left(x_i\right)\right]} = \xi_2^T \theta_2 \tag{17}$$

Define

$$\Phi = \left[\phi_1, \phi_2\right]^T = \begin{bmatrix} \xi_1^T & 0 \\ 0 & \xi_2^T \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \xi^T \theta \tag{18}$$

θ^* is defined the optimal approximative constant. The following inequality is established for a given any small constant $\varepsilon(\varepsilon > 0)$ [9].

$\left\|f - \Phi^*\right\| \le \varepsilon$。令 $\tilde{\theta} = \theta^* - \theta$.

The self-adaptive control law is designed as:

$$\dot{\theta} = \gamma(z_2^T \xi^T(x))^T - 2k\theta \tag{19}$$

STEP 3:

The Lyapunov function is selected for the whole system:

$$V = \frac{1}{2} z_1^T z_1 + \frac{1}{2} z_2^T M z_2 + \frac{1}{2\gamma} \tilde{\theta}^T \tilde{\theta} (\gamma > 0) \tag{20}$$

Therefore

$$\dot{V} = -\lambda_1 z_1^T z_1 - \lambda_2 z_2^T z_2 + z_2^T(f - \xi(x)\theta) - z_2^T - \frac{1}{\gamma}\tilde{\theta}^T\dot{\theta} = -\lambda_1 z_1^T z_1 - \lambda_2 z_2^T z_2 + z_2^T(f - \xi(x)\theta^*) + z_2^T(\xi(x)\theta^* - \xi(x)\theta) - z_2^T d - \frac{1}{\gamma}\tilde{\theta}^T\dot{\theta} \quad (21)$$

Then

$$\dot{V} \le -\lambda_1 z_1^T z_1 - \lambda_2 z_2^T z_2 + \left\|z_2^T\right\|\bullet\left\|(f - \xi(x)\theta^*)\right\| + z_2^T(\xi(x)\tilde{\theta}) + \left\|z_2^T\right\|\bullet\left\|d\right\| - \frac{1}{\gamma}\tilde{\theta}^T\dot{\theta}$$
$$\le -\lambda_1 z_1^T z_1 - \lambda_2 z_2^T z_2 + \frac{1}{2}\left\|z_2^T\right\|^2 + \frac{1}{2}\varepsilon^2 + \frac{1}{2}\left\|z_2^T\right\|^2 + \frac{1}{2}\left\|d\right\|^2 + \tilde{\theta}^T\left[(z_2^T\xi(x))^T - \frac{1}{\gamma}\dot{\theta}\right] \quad (22)$$

Because of $\dot{\theta} = \gamma(z_2^T\xi^T(x))^T - 2k\theta$, we can obtain:

$$\dot{V} \le -\lambda_1 z_1^T z_1 - (\lambda_2 - 1)z_1^T z_2 + \frac{k}{\gamma}(2\theta^{*T}\theta - 2\theta^T\theta) + \frac{\varepsilon^2}{2} + \frac{1}{2}d^T d \quad (23)$$

Using (22)and $\left(\theta - \theta^*\right)^T\left(\theta - \theta^*\right) \ge 0$,
we can obtain:

$$\dot{V} \le -\lambda_1 z_1^T z_1 - (\lambda_2 - 1)z_2^T z_2 + \frac{k}{\gamma}(-\theta^{*T}\theta^* - \theta^T\theta) + \frac{2k}{\gamma}\theta^{*T}\theta^* + \frac{\varepsilon^2}{2} + \frac{1}{2}d^T d \quad (24)$$

Due to $\left(\theta + \theta^*\right)^T\left(\theta + \theta^*\right) \ge 0$

Then

$$\tilde{\theta}^T\tilde{\theta} = (\theta^{*T} - \theta^T)(\theta^* - \theta) \le 2\theta^{*T}\theta^* + 2\theta^T\theta - \theta^T\theta - \theta^{*T}\theta^* \le -\frac{1}{2}\tilde{\theta}^T\tilde{\theta} \quad (25)$$

Namely

$$\dot{V} \le -\lambda_1 z_1^T z_1 - (\lambda_2 - 1)z_2^T M^{-1}Mz_2 - \frac{k}{2\gamma}\tilde{\theta}^T\tilde{\theta} + \frac{2k}{\gamma}\theta^{*T}\theta^* + \frac{\varepsilon^2}{2} + \frac{1}{2}d^T d \quad (26)$$

The optimal parameter is defined as: $\lambda_2 > 1$,

$$M \le \sigma_0 I,\ -M^{-1} \le -\frac{1}{\sigma_0}I \quad (27)$$

We can obtain:

$$\dot{V} \le -\lambda_1 z_1^T z_1 - (\lambda_2 - 1)\frac{1}{\sigma_0}z_2^T Mz_2 - \frac{k}{2\gamma}\tilde{\theta}^T\tilde{\theta} + \frac{2k}{\gamma}\theta^{*T}\theta^* + \frac{\varepsilon^2}{2} + \frac{1}{2}d^T d \quad (28)$$

Define

$$\frac{c_0}{2} = \min\left\{\lambda_1, (\lambda_2 - 1)\frac{1}{\sigma_0}, \frac{k}{2}\right\} \quad (29)$$

Then

$$\dot{V} \le -c_0\left(\frac{1}{2}z_1^T z_1 + \frac{1}{2}z_2^T Mz_2 + \frac{1}{2\gamma}\tilde{\theta}^T\tilde{\theta}\right) + \frac{2k}{\gamma}\theta^{*T}\theta^* + \frac{\varepsilon^2}{2} + \frac{1}{2}d^T d \le -c_0 V + \frac{2k}{\gamma}\theta^{*T}\theta^* + \frac{\varepsilon^2}{2} + \frac{1}{2}d^T d \quad (30)$$

Because the interference $d \in R^n$ is bounded, then there exists the $D > 0$ and meets the $d^T d \le D$.

Then

$$\dot{V} \le -c_0 V + \frac{2k}{\gamma}\theta^{*T}\theta^* + \frac{\varepsilon^2}{2} + \frac{D}{2} = -c_0 V + c_{V\max} \quad (31)$$

Where $c_{V\max} = \frac{2k}{\gamma}\theta^{*T}\theta^* + \frac{\varepsilon^2}{2} + \frac{D}{2}$

By solving the inequality(30), we can obtain:

$$V(t) \le V(0)\exp(-c_0 t) + \frac{c_{V\max}}{c_0}\left[1 - \exp(-c_0 t)\right] \le V(0) + \frac{c_{V\max}}{c_0},\ (t \ge 0) \quad (32)$$

Where $V(0)$ is the initial value of the V. Defining the compact set, $\Omega_0 = \left\{XV(X) \le V(0) + \frac{c_{V\max}}{c_0}\right\}$, then $\left\{z_1, z_2, \tilde{\theta}\right\} \in \Omega_0$, we can obtain the conclusion that V is bounded, and all the signals of closed-loop system are bounded.

3、Simulation

In order to verify the effectiveness of the Backstepping Adaptive Fuzzy controller, MATLAB is used to make simulation for the two-link robot manipulator. The parameters for simulation are shown in table 1.

Table 1. parameters

Description	parameter
m_1	0.1kg
m_2	0.1kg
l_1	0.25m
l_2	0.25m
r_1	0.15
r_2	0.15

The initial state value of system is as:

$$x(0) = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}^T$$

The external disturbance is selected as:

$$d = \begin{bmatrix} 0.5\sin t & 0.5\cos t \end{bmatrix}$$

Design parameter are selected as:

$$\lambda_1 = 10, \lambda_2 = 15, k = 1.5, \gamma = 2,$$

$$\lambda_1 = 2, \lambda_2 = 2.5, k = 1.5, \gamma = 2,$$

Desired trajectory is as: $y_d = \sin(2\pi t)$.

The membership function is seclected as:

$$\mu_{F_i}^{1} = \exp[-0.5((x_i + 1.25)/0.6)^2];$$

$$\mu_{F_i}^{2} = \exp[-0.5((x_i)/0.6)^2];$$

$$\mu_{F_i}^{3} = \exp[-0.5((x_i - 1.25)/0.6)^2];$$

Simulation results are shown from Fig.2 to Fig.6.

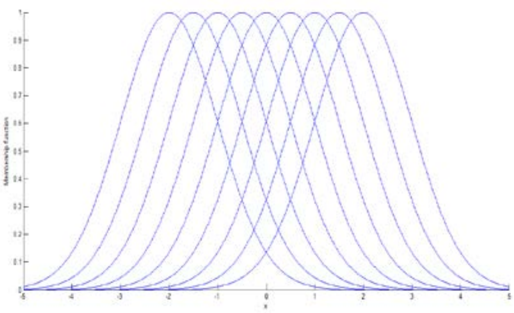


Fig.2 The member function of x_i

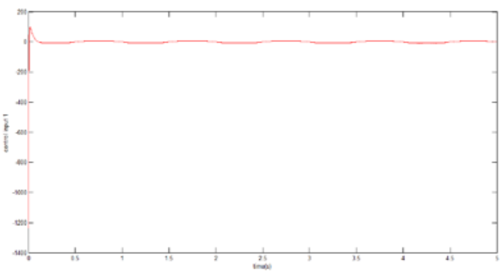


Fig.3 Control input of 1th link

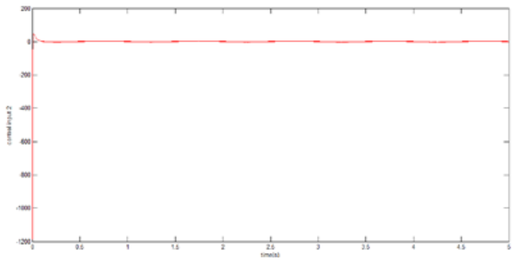


Fig.4 Control input of 2th link

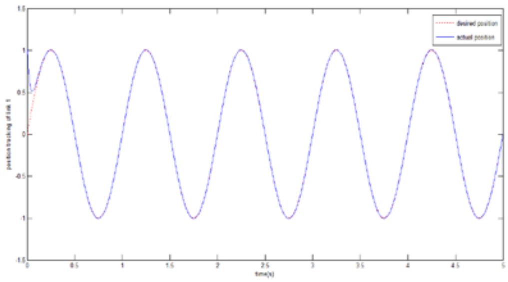


Fig.5 Response of the angle q_1

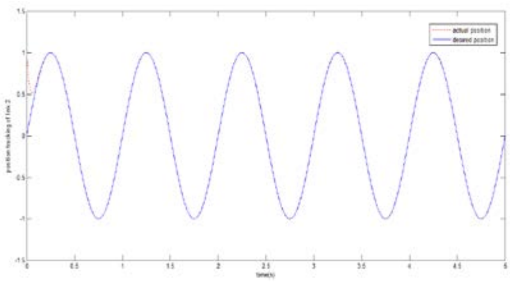


Fig.6 Response of the angle q_2

4、Conclusions

In this paper, an adaptive fuzzy control algorithm combining backstepping control algorithm is proposed for the two-link robot manipulator. Based on the above control algorithm, the robust tracking performance of the two-link robot manipulator can be guaranteed without needing an accurate robot model. The simulation results show that the adaptive controller can achieve desired performance and the algorithm is suitable for an inaccurate robot system. Simulation results also show the precise angle control, which is obtained in spite of disturbance and uncertainties in the system. These results also prove that the proposed control schemes are effective for the two-link robot manipulator.

Reference

[1]Jinkun Liu,The design of robot control system and MATLAB simulation[M], Tsinghua press.2008

[2]Jing Li,Guodong Li,Robust Adaptive Trajectory Control for Robotic Manipulators[J],Journal of Jiangnan University(Natural Science Edition),2008, 7(4):448-452

[3]D.Nganga-Kouya,M.Saad,L.Lamarche,Backstepping Adaptive hybrid Force/Position Control for Robot Manipulators[J]. Proceedings of the American Control Conference Anchorage, 2002,6(10),4595-4600

[4]Jinkun Liu, Advanced Sliding Mode Control for Mechanical systems[M], Tsinghua press.2011

[5]Chun-Yi Su,Yury Stepanenko, Sadik Dost,Hybrid Integrator Backstepping Control of Robotic Manipulators Driven By Brushless DC Motors[J]. IEEE/ASME transatctions on mechatronics,1996, 1(4),266-277

[6]R. Kussub, M. Zohdy, Feedback linearizng versus integrator backstep for trajectory tracking of three link cylindrical-type robot[J], Proceedings of the American Control Conference Chicago,Illinois 2000,2844-2848

[7]D. Nganga-Kouy, M.Saad, L. Lamarche , C. Khairallah,Backstepping adaptive position control for robotic manipulators [J], Proceedings of the American Control Conference Arlington, 2001,636-640

[8]Shoji Takagi , Naoki Uchiyama, Robust Control System Design for SCARA Robots Using Adaptive Pole Placement[J], IEEE transactions on industrial electronics, 2005,52(3),915-921

[9]Che-Min Ou,Jung-shan Lin.Nonlinear Adaptive Backstepping Control Design of Flexible-Joint Robotic Manipulators[J],ASCC,2011,1352-1357

Yongqiao Wei received the B.Sc. degree from Sichuan University,Chengdu,China, in 2012 and he will receive a M.S. degree from Sichuan University, in 2015, in Mechanical design and theory.Now he is a student in Mechanical design and theory of Sichuan University.

Jingdong Zhang received an M.S. degree from Kunming University of Science andTechnology,Kunming ,China,in 2012.He is now a Associate professor.His research interests cover mechanical transmission, intelligent controland CAD/CAPP/CAM/PMD.

Li Hou received an M.S. degree from Chongqing University,Chongqing,China, in 1992 and a Ph.D. Degree from Sichuan University in 2000.He is now a professor and PhD supervisor at Sichuan University.His research interests cover mechanical transmission, intelligent control and embedded systems and CAD/ CAPP/CAM/PMD.

Fenglan Jia received the B.Sc. degree from Qingdao University of technology,Qingdao,Chian,in 2012 and she will receive a M.S. degree from Sichuan University,in 2015,in Mechanical design theory.Now she is a student in the Mechatronics of Sichuan University.

Qinglin Chang received the B.Sc.degree from Sichuan University, Chengdu, China,in 2012.Now he is a student In the mechatronic engineering of Sichuan University.

Multi-Attribute Decision Making Scholarship Selection Using A Modified Fuzzy TOPSIS

Gusti Ayu Made Shinta Wimatsari[1], I Ketut Gede Darma Putra[2] and Putu Wira Buana[3]

[1] **Department of Information Technology, Udayana University**
Bali, 80361, Indonesia

[2] **Department of Information Technology, Udayana University**
Bali, 80361, Indonesia

[3] **Department of Information Technology, Udayana University**
Bali, 80361, Indonesia

Abstract

This research demonstrates the scholarship selection with cases in Udayana University, Bali, Indonesia by using Fuzzy Multi Attribute Decision Making (FMADM) with the method TOPSIS. TOPSIS method is a method of support decision that is based on the concept with the best alternative that is not only has the shortest distance from the positive ideal solution but also it has the longest distance from the negative ideal solution. Selection of recommended students who have the highest level of eligibility for the scholarship based on the value preferences held. The final results of this study provide the ranking of the largest to the smallest value of the calculation method FMADM TOPSIS that can help decision-making selection of scholarship at Udayana University.

Keywords: *Criteria, Distance, Scholarship Selection, Eligibility, Topsis.*

1. Introduction

Scholarship finance is not sourced from its own funding or parents, but provided by the government, private companies, embassies, universities, and non-educators or researchers. Scholarships are given to the right receiver, especially based on classification, quality and competence of the recipients. [2]

Scholarships are also awarded to students at the University of Udayana, Bali, Indonesia. Selection of the scholarship will be difficult and takes a long time because of a lot of scholarship applicants and the criteria that is used to determine scholarship decisions as expected. Data management in the selection of the scholarship has not been fully optimized leading to difficulties in processing the data, and the length of the delivery of information results from the scholarship selection.

This research is expected to assist in the selection of scholarship at the University of Udayana, and assist students in determining the type of scholarships that match the values of criteria. Assessment of selection based on the value of criteria from the student.

This study uses fuzzy MADM (Multi Attribute Making Decision) with Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method. TOPSIS is a decision support method which is based on the concept the best alternative and not only has the shortest distance from the positive ideal solution but also it has the longest distance from the negative ideal solution. Decision support system of scholarship selection will recommend the type of scholarships for student at capacity scholarship and value of criteria.

2. Previous Research

Research related to decision support systems have been done by using one methods such as Fuzzy MADM (Multi Attribute Making Decission) method Technique For Order Preference by Similarity to Ideal Solution (TOPSIS), with several research objects as follows:

Ji-Feng Ding develops a model of an integrated fuzzy topsis method for ranking alternatives and its application. The main purpose is to develop an integrated Fuzzy Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method to improve the quality of decision making for ranking alternatives. The proposed fuzzy TOPSIS method mainly accounts for the classification of criteria, the integrated weights of criteria and sub-criteria, and the performance values of decision matrix. The criterion are classified into subjective criteria and objective ones [6].

K.Savitha and DR.C.Chandrasekar develop a model of network selection using topsis in vertical handover decision schemes for heterogeneous wireless networks. Topsis is used to choose the best network from the available Visitor networks (VTs) for the continuous connection by the mobile terminal. In their work they mainly concentrated to the handover decision Phase and to reduce the processing delay in the period of handover[12].

Kamran Shahanaghi and Seyed Ahmad Yazdian develop a model vendor selection using a new fuzzy group TOPSIS approach. The research presented a new Fuzzy Multi Criteria Group Decision Making (FMCGDM) approach for vendor (supplier) selection problem. The proposed approach is based on TOPSIS method under fuzzy environment to account for vagueness and uncertainty of the real-world situations [13].

Pragati Jain and Manisha Jain develop a model fuzzy TOPSIS method in job sequencing problems on machines of unequal efficiencies. The research presented the way of making sequence of a finite number of jobs on a finite number of machines of unequal efficiencies by using the technique of order preference by similarity to ideal solution or simply TOPSIS Method in fuzzy environment. The order of machines is random. The time taken by the machines for conducting jobs is assumed as imprecise processing time or fuzzy numbers [14].

A.R. Karimi, N. Mehrdadi, S.J. Hashemian, Gh.R. Nabi-Bidhendi and R. Tavakkoli-Moghaddam develop a model by using the fuzzy TOPSIS and fuzzy AHP methods for wastewater treatment process selection. The research investigated five different anaerobic wastewater treatment processes operated in Iranian industrial estates. The criteria evaluation and priorities of alternatives have been done by fuzzy TOPSIS and fuzzy AHP methods by the use of triangular fuzzy numbers. Finally, selection of these five processes is ranked by these foregoing methods, in which their differences are discussed [15].

3. The Advantages of Scholarship Decision Support System

Technological developments are adopted to facilitate the selection process of the scholarship recipients at the University of Udayana in particular. Decision support system allows grantee selection process in a short time and in accordance with the requirements of the scholarship recipients are offered a scholarship provider.

Candidates inserts the value of the criteria into the system. The next step is processing the value in the method of Fuzzy MADM TOPSIS. The output of the system is the closeness coefficients of the alternatives ranked from the largest value to the smallest value.

The ability of the system provides assessment of results the scholarship selection using by FMADM TOPSIS. The results display the name of the candidates / ID candidates, the closeness coefficients of the alternatives, the rank of the closeness coefficients of the alternatives, and the type of scholarships received by the candidate. Mistakes in the selection process will be reduced by the use of a decision support system compared to manually selection.

4. Methodology

4.1 Fuzzy Set Theory

The fuzzy set theory [5] is designed to deal with the extraction of the primary possible outcome from a multiplicity of information that is expressed in vague and imprecise terms. Fuzzy set theory treats vague data as probability distributions in terms of set memberships. Once determined and defined, sets of memberships in probability distributions can be effectively used in logical reasoning.

4.2 Fuzzy Multi Attribute Decision Making (MADM)

Determination of the decision in the case of Multi-Attribute Decision Making (MADM), resolved by selecting the best alternative out of several alternatives. However, because the data are used appropriately, it can’t be expressed in crisp; the method used is an advanced development of the MADM methods.

This development method called the Fuzzy Multi Attribute Decision Making (FMADM), where the application of fuzzy logic method is applied. The essence of FMADM is to determine weights for each attribute, followed by a ranking process will select the alternative that has been given.

The method can be used to solve the problem FMADM [4]:

1. Simple Additive weighting method (SAW)
2. Weighted Product (WP)
3. ELECTRE
4. Analytic Hierarchy Process (AHP)
5. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

4.3 Linguistic Values

In fuzzy decision environments, two preference ratings can be used. They are fuzzy numbers and linguistic values characterized by fuzzy numbers [7]. Depends on practical needs, DMs may apply one or both of them. In this paper, the rating set is used to analytically express the linguistic value and describe how good of the alternatives against various criteria above the alternative level is. The rating set is defined as S = {VP, P, F, G, VG}; where VP = Very Poor, P = Poor, F = Fair, G = Good, and VG = Very Good. Here, we define the linguistic values [8] of VP = (0, 0, 0.25), P = (0, 0.25, 0.5), F = (0.25, 0.5, 0.75), G = (0.5, 0.75, 1), and VG = (0.75, 1, 1), respectively.

4.4 TOPSIS Procedure

The following is the procedure of TOPSIS method are: [4]

1. Normalized decision matrix
   Each element of the matrix D is normalized to obtain the normalization matrix R. Each normalized value r_{ij} can be calculated as follows:

   $$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \text{ .... (1)}$$

   where i=1,2,3,…,m; j=1,2,3,…,n

2. Weighted normalized matrix has given weight W = (w_1, w_2, ..., w_n), so the weighted normalized matrix V can be calculated as follows:

$$Y = \begin{bmatrix} w_{11}r_{11} & \cdots & w_{1n}r_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1}r_{m1} & \cdots & w_{nm}r_{nm} \end{bmatrix} \text{............. (2)}$$

   where i=1,2,3,…,m and j=1,2,3…,n

3. Determine the positive ideal solution and negative ideal solution
   Positive ideal solution A^+ and A^- negative ideal solution can be determined based on normalized weighted rating (Y_{ij}):
   $A^+ = (y_1^+, y_2^+, \ldots, y_n^+)$; ....(3)
   $A^- = (y_1^-, y_2^-, \ldots, y_n^+)$; ....(4)
   Where,

   $$y_j^+ = \begin{cases} \max_i y_{ij}, & \text{if } j \text{ is an benefit attribute} \\ \min_i y_{ij}, & \text{if } j \text{ is an cost attribute} \end{cases}$$

   $$y_j^- = \begin{cases} \min_i y_{ij}, & \text{if } j \text{ is an benefit attribute} \\ \max_i y_{ij}, & \text{if } j \text{ is an cost attribute} \end{cases}$$

4. Calculating Separation Measure
   Separation measure is a measurement of the distance of an alternative to the positive ideal solution and negative ideal solution. Mathematical calculation is as follows:
   Separation measure to the positive ideal solution

   $$D_i^+ = \sqrt{\sum_{j=1}^{n}(Y_{ij} - Y_j^+)^2} \text{ where } i = 1,2,3,\ldots,n \text{......... (5)}$$

   Separation measure to the negative ideal solution

   $$D_i^- = \sqrt{\sum_{j=1}^{n}(Y_{ij} - Y_j^-)^2} \text{ where } i = 1,2,3,\ldots,n \text{......... (6)}$$

5. Calculating the relative closeness to the positive ideal. Relative closeness of the alternative A^+ to A^- ideal solution represented by:

$$V_i = \frac{D_i^-}{D_i^- + D_i^+}, \text{where } 0 < V_i < 1 \text{ and } i = 1,2,3,\ldots m \text{.......... (7)}$$

6. Sorting Preference
   Alternatives can be ranked based on the order of A_i.The best alternative is the one that is shortest to the longest and is the ideal solution to the negative ideal solution.

4.5 Requirement Analysis for Fuzzy Multi Attribute Decision Making

Problem-solving and computation scholarship selection with TOPSIS method described as:

Phase 1: Collect the number of alternatives that will be used and some of the attributes or criteria. There are four criterias used as a basis for making decisions in the selection of scholarship that is used at Udayana University, Bali, Indonesia. The criteria are:

C1 = GPA (Grade Point Average)
C2 = quotient of income parents by the number of dependents
C3 = The Usage of Electrical Power
C4 = Student Activities

Scholarship recipients at Udayana University, Bali Indonesia selected from alternatives that fulfill all the administrative requirements of the type of scholarship that is being selected. One example of scholarship that is selected Academic Improvement Scholarship (PPA). Administrative requirements of the PPA scholarships to students in order to qualify for the scholarship selection are:

1. Achievement will be given by considering the background of the economic capacity of parents to their students,
2. Student minimum was in the second semester and the highest was in eighth semester,
3. Attach files as Student Identity Card Copy (KTM) and Card Study Plan (KRS) or similar as proof of active student,
4. Copy of last electric bill and proof of payment or the United Nations of a parent / guardian,
5. Statement that is not currently receive any scholarships from other sources in the environment known to the Ministry of National Education Leadership Education high Student Affairs,
6. Copy of family card, recommendation from the head of the Faculty / Department,
7. Copy of academic transcript with a grade point average (GPA) of at least 3.00 were endorsed by the leadership of the college,
8. Income certificate of parents and approved by the authorities. [9]

Other scholarships offered a scholarship at the University of Udayana is Underprivileged Scholarship. The requirements Underprivileged Scholarship almost equal to the achievement Scholarship Requirements, but there are differences in the number of GPA. Total minimum GPA for Underprivileged Scholarship is 2.75.

The purpose of the scholarship selection is to get some of the candidates who fulfill the administrative requirements specification. For the example in this research, use 8 students (alternatives) that are listed in the scholarship selection. Of 8 students (alternatives) will be selected by FMADM TOPSIS method.

In this research, the decision support system would recommend the appropriate types of scholarship based on the requirements and quotas from each scholarship.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Phase 2: Assessment of Fuzzy Sets

The research uses the linguistic variables which are developed by Chen & Hwang [10]. This research also uses triangular fuzzy number to express the importance of each criterion. All criteria use data fuzzy in scholarship selection.

Table 1: Assessment of Fuzzy Sets

Low	0, 0.3, 0.6
Medium	0.3, 0.6, 0.9
High	0.6, 0.9, 1

Table 2: Linguistic variables for the importance weight of each criteria for Achievement scholarship

Criteria	Linguistic Variable	Fuzzy Number
C1	High	0.6, 0.9, 1
C2	Low	0, 0.3, 0.6
C3	Low	0, 0.3, 0.6
C4	High	0.6, 0.9, 1

Table 3: Linguistic variables for the importance weight of each criteria for Underprivileged scholarship

Criteria	Linguistic Variable	Fuzzy Number
C1	Low	0, 0.3, 0.6
C2	High	0.6, 0.9, 1
C3	High	0.6, 0.9, 1
C4	Low	0, 0.3, 0.6

GPA criteria (C1) by using suitability degree with several alternative decisions: T (suitability) = {VL, SL, L, M, SH, H, VH}. Membership functions of each element represented by using triangular fuzzy value with VL = Very Low, SL= Slightly Lower, L=Low, M=Medium, SH= Slightly Higher, H=High, VH=Very High. The value of each membership function will be showed in Fig. 1 and Table 4.

Table 4: Fuzzy linguistic terms and their correspondent fuzzy numbers for C1

Range GPA	**Linguistic Variable**	**Fuzzy Number**
0 – 2.5	VERY LOW	0, 0, 0.5
2.51 – 2.75	SLIGHTLY LOWER	0, 0.5, 0.6
2.751 – 3.0	LOW	0.5, 0.6, 0.7
3.01 – 3.25	MEDIUM	0.6, 0.7, 0.8
3.251 – 3.5	SLIGHTLY HIGHER	0.7, 0.8, 0.9
3.51 – 3.75	HIGH	0.8, 0.9, 0.1
3.751 – 4	VERY HIGH	0.9, 1, 1

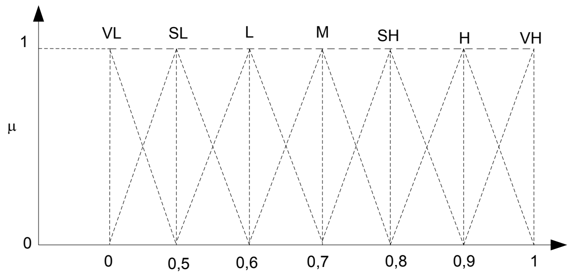


Fig. 1 The fuzzy linguistic variables for C1

Quotient of parent income by the number of dependents criteria (C2) by using suitability degree with several alternative decisions: T (suitability) = {S, M, RB, B, VB, VVB}. Membership functions of each element represented by using triangular fuzzy value with S=Small, M=Moderate, RB=Rather Big, B=Big, VB=Very Big, VVB=Very Very Big. The value of each membership function will be showed in Figure 2 and Table 5.

Table 5: Fuzzy linguistic terms and their correspondent fuzzy numbers for C2

Range C2 (IDR)	**Linguistic Variable**	**Fuzzy Number**
≤ 300.000	SMALL	0, 0.1, 0.4
> 300.000 - 500.000	MODERATE	0.1, 0.4, 0.5
> 500.000 – 1.000.000	RATHER BIG	0.4, 0.5, 0.6
>1.000.000 – 1.500.000	BIG	0.5, 0.6, 0.8
>1.500.000 – 2.000.000	VERY BIG	0.6, 0.8, 1
> 2.000.000	VERY VERY BIG	0.8, 1, 1

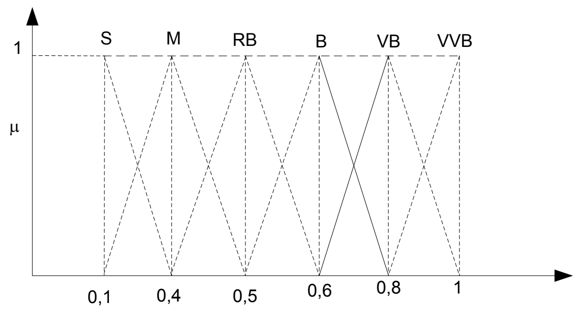


Fig. 2 The fuzzy linguistic variables for C2

The usage of electrical power criteria (C3) by using suitability degree with several alternative decisions: T (suitability) = {L, M, RB, B, VL}. Membership functions of each element represented using triangular fuzzy value with L=Low, M=Moderate, RB=Rather Big, B=Big, VB=Very Big. The value of each membership function will be showed in Figure 3 and Table 6.

Table 6: Fuzzy linguistic terms and their correspondent fuzzy numbers for C3

Range Use of electrical power(VA)	Linguistic Variable	Fuzzy Number
0 – 450	LOW	0, 0.2, 0.4
451 – 900	MODERATE	0.2, 0.4, 0.6
901 – 1300	RATHER BIG	0.4, 0.6, 0.8
1301 – 2200	BIG	0.6, 0.8, 1
2201 - 10000	VERY BIG	0.8, 1, 1

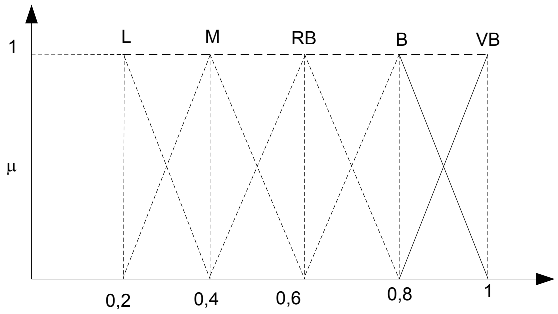


Fig. 3 The fuzzy linguistic variables for C3

Student activities (C4) by using suitability degree with several alternative decisions: T (suitability) = {VL, L, M, H, VH}. Membership functions of each element represented using triangular fuzzy value with VL=Very Low, L=Low, M=Medium, H=High, VH=Very High. The value of each membership function will be showed in Figure 4 and Table 7.

Table 7: Fuzzy linguistic terms and their correspondent fuzzy numbers for C4

Student activities	Linguistic Variable	Fuzzy Number
0 – 64	VERY LOW	0, 0.2, 0.4
65 – 129	LOW	0.2, 0.4, 0.6
130 – 194	MEDIUM	0.4, 0.6, 0.8
195 – 295	HIGH	0.6, 0.8, 1
260 – 320	VERY HIGH	0.8, 1, 1

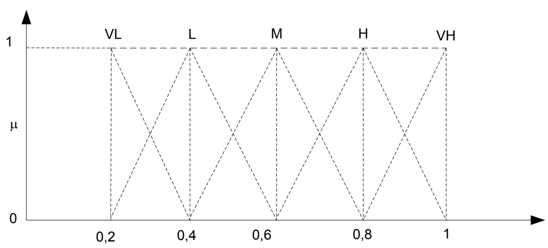


Fig. 4 The fuzzy linguistic variables for C4

5. Experiments and Results

The output of this research is result of scholarship selection which is sorted from highest value of the closeness coefficient of alternative to lowest value of the closeness coefficient of alternative and based on quota of recipients the type of scholarship offered. The final results were implemented in a decision support system scholarship selection based on input the values of criteria are and processed using TOPSIS FMADM.

The scholarship selection process in this research using two types of scholarships are Achievement Scholarship and Underprivileged Scholarship. The number of candidates is used as an example of the data is 8 students. Scholarship capacity of Achievement Scholarship is 5 students and Underprivileged scholarship is 3 students.

First step, decision support system will process and recommend the Achievement Scholarship to 5 students which is have high value of the closeness coefficient of alternative. The candidates who do not fulfill capacity and requirements of Achievement Scholarship will be recommended to Underprivileged Scholarships.

Table 8 shows the criteria values of each alternative. Value criterion is used as the input of the scholarship selection.

Table 8: Criteria values of each alternative selection

A	C1	C2	C3	C4
001	3.44	662.500	450	24
002	3.15	750.000	900	20
003	3.57	1.500.000	900	23
004	3.2	600.000	450	33
005	3.87	700.000	900	25
006	3.02	300.000	450	65
007	2.80	600.000	900	48
008	2.79	400.000	450	50

Table 9 shows the suitability rating of each alternative on each criterion. These values are used for the decision matrix normalization process in the next step.

Table 9: Value rating compatibility based on value criteria

A	C1	C2	C3	C4
001	0.8	0.4	0.2	0.2
002	0.7	0.4	0.4	0.2
003	0.9	0.6	0.4	0.2
004	0.7	0.4	0.2	0.2
005	1	0.4	0.4	0.2
006	0.7	0.1	0.2	0.4
007	0.6	0.4	0.4	0.2
008	0.6	0.2	0.2	0.2

Figure 5 shows the results of the normalized matrix using equation 1. The results of the calculation criteria matrix obtained results in the form of normalization. The next step in this research is the weighting matrix using equation 2.

NIM	C1	C2	C3	C4
001	0.3714	0.3636	0.2236	0.3015
002	0.325	0.3636	0.4472	0.3015
003	0.4178	0.5455	0.4472	0.3015
004	0.325	0.3636	0.2236	0.3015
005	0.4642	0.3636	0.4472	0.3015
006	0.325	0.0909	0.2236	0.603
007	0.2785	0.3636	0.4472	0.3015
008	0.2785	0.1818	0.2236	0.3015

Selanjutnya »

Fig. 5 The result of normalized decision matrix calculation

Figure 6 shows the results of the normalized weighted matrix. Each criterion to get a certain weight value in which case the GPA (C1) criteria and student activities (C4) using a higher weight than the other criteria.

NIM	C1	C2	C3	C4
001	0.3343	0.1091	0.0671	0.2714
002	0.2925	0.1091	0.1342	0.2714
003	0.376	0.1637	0.1342	0.2714
004	0.2925	0.1091	0.0671	0.2714
005	0.4178	0.1091	0.1342	0.2714
006	0.2925	0.0273	0.0671	0.5427
007	0.2507	0.1091	0.1342	0.2714
008	0.2507	0.0545	0.0671	0.2714

Selanjutnya »

Fig. 6 The result of calculate normalized matrix has been weighted

Calculation of positive ideal solution using equation 3, and the negative ideal solution using equation 4. The result of this process produces the value as shown in figure 7.

POSITIVE & NEGATIVE IDEAL SOLUTION

POSITIVE IDEAL SOLUTION		NEGATIVE IDEAL SOLUTION	
GPA (C1)	: 0.4178	GPA (C1)	: 0.2507
Income Parent / The Number of Dependents (C2)	: 0.0273	Income Parent / The Number of Dependents (C2)	: 0.1637
The Usage of Electrical Power (C3)	: 0.0671	The Usage of Electrical Power (C3)	: 0.1342
Student Activities (C4)	: 0.5427	Student Activities (C4)	: 0.2714

Selanjutnya »

Fig. 7 The result of calculated positive and negative ideal solutions

The next step as showed in figure 8 are positive and negatives to determine the distance of each alternative. Determination positive distance using equations 5 and negative distance using equation 6.

NIM	Positive Distance	Negative Distance
001	0.2954	0.1203
002	0.317	0.0688
003	0.3138	0.1253
004	0.3098	0.0961
005	0.2912	0.1758
006	0.1253	0.3138
007	0.3357	0.0546
008	0.3198	0.1282

Fig. 8 The result of calculated the positive and negative distance

The next step is by calculating the closeness coefficients of the alternatives using equation 7. The results of the five alternatives are obtained as shown in Figure 9.

NIM	The Closeness Coefficients of The Alternatives
001	0.2894
002	0.1782
003	0.2854
004	0.2367
005	0.3764
006	0.7146
007	0.1399
008	0.2861

Fig. 9 The result of calculated the closeness coefficients of the alternatives.

THE RESULT OF THE RANK

NIM	Name	The Closeness Coefficients of The Alternatives	Rank
006	Shelina	0.7146	1
005	Hendra Wijaya	0.3764	2
001	Jeshika	0.2894	3
008	Kurniawan	0.2861	4
003	Nurul Savitri	0.2854	5
004	Fransiska	0.2367	6
002	Justin	0.1782	7
007	Dwi Saputra	0.1399	8

Finish »

Fig. 10 The result of ranks from the value of closeness coefficients of the alternatives.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

NO.	NIM	Name	Value of Closeness of Cefficients of the alternatives	Rank
1	006	Shelina	0.7146	1
2	005	Hendra Wijaya	0.3764	2
3	001	Jeshika	0.2894	3
4	008	Kurniawan	0.2861	4
5	003	Nurul Savitri	0.2854	5

Fig. 11 The final result of TOPSIS methods for Achievement Scholarship

The final step in the selection of Achievement Scholarship is adjusting capacity of recipients. In this research Achievement Scholarships capacity is 5 receiver. So, the five candidates who have five highest score was selected as the recipient of a scholarship achievement. While, the lowest 3 candidates will be recommended to the Underprivileged Scholarship.

Fig. 10 show the highest rank owned by the student ID/NIM 006 with value 0.7146 and the lowest rank value is 0.2854 that is owned by student ID/NIM 003.

First rank was achieved by student ID/NIM 006 with value of the closeness coefficient of alternative is 0.7146. The values of criterion are C1 = 3.02, C2 = IDR 300.000, C3 = 450VA and C4 = 65. This alternative has the highest eligibility for Achievement Scholarship selection when compared to other alternatives.

Second rank was achieved by student ID/NIM 005 has value of the closeness coefficient of alternative is 0.3764. The values of criterion are C1 = 3.87, C2 =IDR 700.000, C3=900VA and C4 = 25. Assessment achievement scholarship is prioritized to criteria C1 and C4.

Student ID/NIM 001 has value of criterion C1 = 3.44, C2 = IDR. 662.500, C3 = 450VA and C4 = 24. The value of the closeness coefficient of alternative is 0.2894 and this alternative is the third rank.

Selection of Achievements Scholarship have been completed. Then the system will process the Underprivileged Scholarship. From the previous selection, three students failed in the selection of Achievement Scholarship. Three students who are recommended by the the decision support system in the next scholarship selection.

Table 8 will be showed the value of the criteria of the three candidates who failed the selection of achievement scholarships and will go to the Underprivileged Scholarship selection.

Table 8: Criteria values of each alternative selection

A	C1	C2	C3	C4
002	3.15	750.000	900	20
004	3.2	600.000	450	33
007	2.80	600.000	900	48

Table 9: Value rating compatibility based on value criteria

A	C1	C2	C3	C4
002	0.7	0.4	0.4	0.2
004	0.7	0.4	0.2	0.2
007	0.6	0.4	0.4	0.2

Figure 12 shows the results of the normalized matrix using equation 1. As the phase in the selection of Achievement Scholarship, the results of the calculation criteria matrix obtained results in the form of normalization. The next step in this research is the weighting matrix using equation 2.

No	NIM	C1	C2	C3	C4
1	002	0.6047	0.5774	0.6667	0.5774
2	004	0.6047	0.5774	0.3333	0.5774
3	007	0.5183	0.5774	0.6667	0.5774

Selanjutnya »

Fig. 12 The result of normalized decision matrix calculation the Underprivileged scholarship

Figure 13 shows the results of the normalized weighted matrix. Each criterion to get a certain weight value in which case criteria C2 and C3 using a higher weight than the other criteria.

No	NIM	C1	C2	C3	C4
1	002	0.1814	0.5197	0.6	0.1732
2	004	0.1814	0.5197	0.3	0.1732
3	007	0.1555	0.5197	0.6	0.1732

Selanjutnya »

Fig. 13 The result of calculate normalized matrix has been weighted the Underprivileged scholarship

Calculation of positive ideal solution using equation 3, and the negative ideal solution using equation 4. The result of this process produces the value as shown in figure 14.

POSITIVE & NEGATIVE IDEAL SOLUTION

POSITIVE IDEAL SOLUTION		NEGATIVE IDEAL SOLUTION	
GPA (C1)	: 0.1814	GPA (C1)	: 0.1555
Income Parent / The Number of Dependents (C2)	: 0.5197	Income Parent / The Number of Dependents (C2)	: 0.5197
The Usage of Electrical Power (C3)	: 0.3	The Usage of Electrical Power (C3)	: 0.6
Student Activities (C4)	: 0.1732	Student Activities (C4)	: 0.1732

Selanjutnya »

Fig. 14 The result of calculated positive and negative ideal solutions the Underprivileged scholarship

The next step as showed in figure 15 are positive and negatives to determine the distance of each alternative. Determination positive distance using equations 5 and negative distance using equation 6.

NIM	Positive Distance	Negative Distance
002	0.3	0.0259
004	0	0.3011
007	0.3011	0

Fig. 15 The result of calculated the positive and negative distance the Underprivileged scholarship

The next step is by calculating the closeness coefficients of the alternatives using equation 7. The results of the five alternatives are obtained as shown in Figure 16.

NIM	The Closeness Coefficients of The Alternatives
002	0.0795
004	1
007	0

Fig. 16 The result of calculated the closeness coefficients of the alternatives the Underprivileged Scholarship

THE RESULT OF THE RANK

NIM	Name	The Closeness Coefficients of The Alternatives	Rank
004	Fransiska	1	1
002	Justin	0.0795	2
007	Dwi Saputra	0	3

Finish »

Fig. 17 The result of ranks from the value of closeness coefficients of the alternatives the Underprivileged scholarship

Figure 17 shows the results of the ranks selection. In this research the case of types of Underprivileged Scholarships have a capacity of three recipients. Therefore, all candidates in fig. 17 awarded the Underprivileged Scholarships.

NO.	NIM	Name	The Closeness Coefficients of The Alternatives	Rank
1	004	Fransiska	1	1
2	002	Justin	0.0795	2
3	007	Dwi Saputra	0	3

Fig. 18 The final result of TOPSIS methods for achievement scholarship the Underprivileged scholarship

Figure 18 shows the the final result of Underprivileged Scholarship selection. First rank was achieved by student ID/NIM 004 with value of the closeness coefficient of alternative is 1. The values of criterion are C1 = 3.2, C2 = IDR 600.000, C3 = 450VA and C4 = 33. This alternative has the highest eligibility for scholarship selection achievement when compared to other alternatives.

Assessment Underprivileged Scholarship is prioritized to criteria C2 and C3. Second rank was achieved by student ID/NIM 002 has value of the closeness coefficient of alternative is 0.0795. The values of criterion are C1 = 3.15, C2 =IDR 750.000, C3=900VA and C4 = 20. The third rank was achieved by student ID/NIM 007 has value of the closeness coefficient of alternative is 0. The values of criterion are C1 = 2.80, C2 =IDR 6000.000, C3=900VA and C4 = 48.

4. CONCLUSION

This study has successfully obtained a scholarship recipient selection results by using the Fuzzy Multi Attribute Decision Making Technique for using Order Preference by Similarity to Ideal Solution (TOPSIS). In this study, several criteria were used GPA (Grade Point Average), quotients of income parents by the number of dependents, number of dependents parents, the usage of electrical power and student activities.

Scholarship selection process is done by normalizing the value of each criteria in a decision matrix, multiplies by weight according to the degree of influence of each criteria in the selection process, the calculation of the positive and negative ideal solution of each of the criteria, calculating the distance from the positive alternative negative, calculate the relative closeness to the positive ideal, and the last is the rank of the selection results. The selection recommend an alternative that has the highest level of eligibility to the most low to get a scholarship based on value preferences held.

Acknowledgments

Our thank goes to Department of Information Technology Udayana University, Bali, who has helped organize this research's Indonesia.

REFERENCES

[1] Hwang, C.L. and Yoon, K. (1981). Multiple Attribute Decision Making Methods and Applications. Springer, New York, NY.

[2] Gafur, Abdul. 2008. Cara Mudah Mendapatkan Beasiswa. Jakarta, Indonesia. Penebar Plus Publisher

[3] Uyun, Shofwatul, and Imam, Riadi. A Fuzzy Topsis Multiple-Attribute Decision Making for Scholarship Selection. TELKOMNIKA, Vol.9, No.1, April 2011, pp. 37~46.

[4] Kusumadewi, Sri., Hartati, S., Harjoko, A., dan Wardoyo, R. (2006). Fuzzy Multi-Attribute Decision Making (FUZZY MADM). Yogyakarta, Indonesia. Graha Ilmu Publisher.

[5] Zadeh, L. A. "Fuzzy sets. Information and Control", Vol. 8, pp. 338-353. 1965.

[6] Ding, Ji-Feng. An Integrated Fuzzy Topsis Method For Ranking Alternatives And Its Application. Journal of Marine Science and Technology, Vol. 19, No. 4, pp. 341-352 (2011)

[7] Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part 1, 2 and 3," Information Sciences, Vol. 8, No. 3, pp. 199-249 (1975); Vol. 8, No. 4, pp. 301-357 (1975); Vol. 9, No. 1, pp. 43-80 (1976).

[8] Chang, P. L. and Chen, Y. C., "A fuzzy multi-criteria decision making method for technology transfer strategy selection in biotechnology," *Fuzzy Sets and Systems*, Vol. 63, No. 2, pp. 131-139 (1994).

[9] Dikti. 2010. Pedoman Beasiswa Peningkatan Prestasi Akademik (PPA) dan Bantuan Belajar Mahasiswa (BBM). Jakarta, Indonesia.

[10] S. J. Chen and C. L. Hwang, *Fuzzy Multiple Attribute Decision Making*, (Lecture Notes in Economics and Mathematical System Series 375). Springer-Verlag. New York, 1992.

[11] TIEN-CHIN, WANG. HSIEN-DA, LEE. CHUAN-CHENG, WU., "A Fuzzy TOPSIS Approach with Subjective Weights and Objective Weights," *Proceedings of the 6th WSEAS International Conference on Applied Computer Science, Hangzhou, China*, April 15-17, 2007

[12] K.Savitha, DR.C.Chandrasekar., "Network Selection Using TOPSIS in Vertical Handover Decision Schemes for Heterogeneous Wireless Networks". *IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011*

[13] Kamran Shahanaghi. Seyed Ahmad Yazdian. "Vendor Selection Using a New Fuzzy Group TOPSIS Approach". Journal of Uncertain Systems Vol.3, No.3, pp.221-231, 2009

[14] Pragati Jain and Manisha Jain. "Fuzzy Topsis Method In Job Sequencing Problems On Machines Of Unequal Efficiencies". *Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine Vol. 2 No. 6, June 2011*

[15] A.R. Karimi, N. Mehrdadi, S.J. Hashemian, Gh.R. Nabi-Bidhendi, and R. Tavakkoli-Moghaddam "Using Of The Fuzzy Topsis And Fuzzy Ahp Methods For Wastewater Treatment Process Selection". *International Journal of Academic Research vol. 3. No. 1. January, 2011, part III*

Gusti Ayu Made Shinta Wimatsari studied Information Technology in Department of Information Technology Udayana University since August 2008, and now working her research for S.Ti. degree in Information Technology.

Dr. I Ketut Gede Darma Putra, S.Kom., MT received his S.Kom degree in Informatics Engineering from Institut Teknologi Sepuluh Nopember University, his MT. degree in Electrical Engineering from Gajah Mada University and his Dr. degree in Electrical Engineering from Gajah Mada University. He is lecturer at Electrical Engineering Department (major in Computer System and Informatics) of Udayana University, lecturer at Information Technology Department of Udayana University.

Putu Wira Buana, S.Kom., MT received his S.Kom degree in Informatics Engineering from Institut Teknologi Sepuluh Nopember University and his MT. degree in Electrical Engineering from Brawijaya University. He is lecturer at Electrical Engineering Department (major in Computer System and Informatics) of Udayana University, lecturer at Information Technology Department of Udayana University.

A Novel Approach to Query Modification Based on User's Why-not Question

Jianfeng Zhang[1], Weihong Han[1], Yan Jia[1], Peng Zou[2], Hua Fan[1]

[1]**School of Computer Science, National University of Defense Technology, Changsha, Hunan, P.R.China, 410073**

[2]**Academy of Equipment Command and Technology, Beijing, P. R. China**

Abstract

Top-k query is an efficient way to show the most important objects to user from massive amounts of data. After huge effort working on database performance, recently, an explain capability has attract more attention in recent years. In top-k query, since people may not specify his/her accurate preference, he may feel frustrated with the top-k results and propose a question such that "why my expecting tuple is not appeared in top-k results as long as tuple p has been appeared in top-k results". Based on this motivation, in this paper, we propose a new method to approach this problem. Given the inputs as the original top-k query, the expecting tuple and the comparable tuple, our algorithm returns a new query to the user which makes the smallest change in the original top-k results. Finally, an extensive performance study using both synthetic and real data set is reported to verify its effectiveness and efficiency.

Keywords: *Top-k Query; Refined Query; Result Explaining; Why Not.*

1. Introduction

Recently, the support of rank aware query processing has attracted much more attentions in database research area. Top-k queries return only a limited of k objects that best match the user's preference, thus avoiding the huge and overwhelming result sets.

Although database system researchers have made tremendous advanced on functionality and performance related issues in the past decades, research on improving database usability has not attracted as much as it deserves. Recent years, the feature of explaining missing objects in database queries, or the so-called "why-not" questions has received growing attentions [1]. For example, users often feel frustrated when they find their expected tuple m is not in the query results but their unexpected tuple p is in the query results without any explanation. So they may propose such a question: "since tuple p is in top-k results, why not tuple m?" If the database system can give a good explanation for it, it would be very useful for uses to understand and modify the query.

Example 1. Take Fig.1 for example. There are five network security alerts in this Table, and each alert is composed of three attributes: Asset, Threat and Occurrence. All these attributes are normalized to a range from 0 to 1. In order to retrieve the Top-2 risky alert, user need to use the scoring function to rank all the alerts. Here we use the simplest but very common scoring function ($f_{\vec{w}}(p) = \vec{w} \cdot \vec{p} = \sum_{i=1}^{d} w[i] \times p.A_i$) to aggregate score of each alerts. Suppose user initiates the query with weight setting [0.333, 0.333, 0.334], then the Top-2 alerts are {5, 4}. Unfortunately, user may be confused by this Top-2 results. In his instinctive thinking, he may consider alert-1 is more risky than alert-4. Now alert 4 is in top-2 results, why not alert-1.

Alert ID	Asset	Threat	Occurence
1	0.7	0.3	0.1
2	0.2	0.4	0.7
3	0.5	0.2	0.2
4	0.4	0.8	0.5
5	0.8	0.6	0.8

```
QUERY:

SELECT * FROME Alert
ORDER BY W[1]*Asset
         +W[2]*Threat
         +W[3]*Occurence
DESC LIMIT K
```

Fig.1 A Set of Some Alerts

In this paper, we present methods to answer user's questions by modifying the original top-k query. Generally speaking, a Top-k query depends on two parameters: the number of objects to be shown in result (i.e. the k value) and user's preference on each attribute (i.e. the weight vector $\vec{w}$). In order to answer user's questions, we need to modify the original query to get a new query. To solve this problem, we first define an evaluation model to evaluate the difference between the original query and the new query. The difference is estimated by the changes of weight vectors and the top-k results. Next, we find a new query with the least change of weight vectors as the temporary optimal new query. Based on this new query, we use a sample method to get some optimal weight vectors and evaluate the quality of query under each weight vector. Finally, we return the new query with the least changes of original query as the answer.

Our contributions can be summarized as follows:

1. We formulate our problem in explaining user's why not questions on the top-k query. We compactly represent the problem and formally define it.
2. We propose a greedy approach to find the approximate solution to answer user's why not questions, and present several efficient techniques to improve the efficiency of our algorithm.
3. We present an implementation and evaluation of the proposed algorithm in synthetic and real data sets. Extensive experimental results on synthetic and real data sets demonstrate the effectiveness and efficiency of our algorithm

The remainder of this paper is organized as follows. In Section 2, we review the related works. In Section 3, we present a formal problem definition. In Section 4 we propose a greedy algorithm to search the approximate solution to answer user's why not questions. An extensive empirical evaluation using both synthetic data sets and real data sets is reported in Section 5. Finally, our work of this paper is summarized in Section 6.

2. Related Works

There has been some different ways to answer why-not questions. To the best of our knowledge, the major existing approaches to explain why-not questions on top-k query can be summarized as follows.

The first approach explains why-not question by modifying some objects in the database which will include both the original results and the specified missing objects. For example, in [2], the author answers user's why-not question on Select-Project-Join (SPJ) queries by telling him/her which query operators eliminated his/her desired objects.

The second approaches explain missing objects by identifying the manipulation operations in the query plan that is responsible for excluding the missing objects. In [3], the missing answers of Select-Project-Join-Union-Aggregation (SPJUA) queries are explained by telling the user how to modify the data. In [4], the author proposes novel algorithms to generate good quality refined queries that not only similar to the original query but also produce precise query results with minimal irrelevant objects, and these algorithms can answer the SPJ queries both the basically and complex why-not questions with aggregation that involve comparison constraints.

The third typical approaches proposed in the literature to handle the many answers problem are to utilize scoring functions and return only the Top-k ranked results. A represented work that is also related to us is reserve Top-k query[1], which is defined by a given product p and returns the weighting vector for which p in the top-k set. Two versions of reverse top-k queries, namely monochromatic and bichromatic, are presented in this paper. Based on the geometrical properties of the result set, an algorithm for evaluating monochromatic reverse top-k queries is proposed. Thereafter, the author presents an efficient threshold based algorithm for computing bichromatic reverse top-k queries. Another related work is answering why-not questions on top-k query [5]. The authors present methods to answer why-not questions on Top-k queries through modifying both the k value and the set of weighting together. By returning the user a refined query with approximate minimal changes to the k value and their weightings, the user could get not only his/her desired query, but also learn what was/were wrong with her initial query. However, it only considers the situation that makes the missing objects be appeared in top-k results, and ignore the effects on the original top-k results.

3. Problem Statement

In this section, we present some basics regarding to top-k query, and then we formally define the problem how to modify a query based on quality function. Table I summarizes some notions frequently used in this paper.

Table I: The summary of frequently used notions

Notion	*Meaning*
D	A multidimensional data set
d	The dimensionality of D
p,q,m	Objects in data set
$\prec$	A preference relationship
$\|S\|$	The number of objects in S
$\vec{w}$	The weight vector for the scoring function
$w[i]$	The i-th coordinate value of $\vec{w}$

3.1 Preliminaries

We have a data space D of n objects. Each object is described by d attributes. We use $p.A_i$ to refer to the value of an attribute A_i for an object p. For ease of discussion, we assume that all of these numerical attributes are normalized to range from 0 to 1. Furthermore, without loss of generality, we assume that greater score values are more preferable.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Top-k queries are defined based on the scoring function f that aggregates the individual scores into an overall scoring value, which enables the ranking of the data points in data space D. In the simplest but very common case, a linear aggregation function is adopted, which is specified as a weighted sum of scores. Each dimension d_i has an associated query-dependent weight $w[i]$ indicating d_i 's relative importance for the query. The result of a top-k query is a ranked list of the k objects with the best scoring values, where $score(\vec{p},\vec{w}) = \vec{p}\cdot\vec{w}$.

In the Euclidean space a linear top-k query can be represented by a vector $\vec{w}$. As discussed in [6] the magnitude of the query vector does not influence the query result, as long as the direction remains the same, i.e. representing the relative importance between different dimensions. Therefore, we make the assumption that $\sum_{i=1}^{d} w[i] = 1$.

3.2 Problem Definition

In this section, we formally define the problem how to modify a query based on quality function.

Definition 1 (Top-k query): Given a positive integer k and a weight vector $\vec{w}$, the result of $Q(k,\vec{w})$ is a ranked list of objects such that $Q(k,\vec{w}) \in D$, $|Q(k,\vec{w})| = k$ and $\forall p_1, p_2 : p_1 \in Q(k,\vec{w}), p_2 \in D - Q(k,\vec{w})$ it holds that $score(\vec{p}_1,\vec{w}) > score(\vec{p}_2,\vec{w})$.

Definition 2 (The k-th object): Given a positive integer k and a weight vector $\vec{w}$, the result of $Tuple(k,\vec{w})$ is an object p such that $p \in Q(k,\vec{w}) - Q(k-1,\vec{w})$.

At the beginning, user gives a top-k query $Q_0(k_0,\vec{w}_0)$. Based on the returned results, she/he may propose a question such that: Since object p is in the top-k results, why not miss object m in the top-k results? To solve this problem, we try to define *a new query answer* $Q'(k',\vec{w}')$ which satisfies the conditions as follows.

- object m is in the top-k results of $Q'(k',\vec{w}')$
- $score(\vec{m},\vec{w}') > score(\vec{p},\vec{w}')$

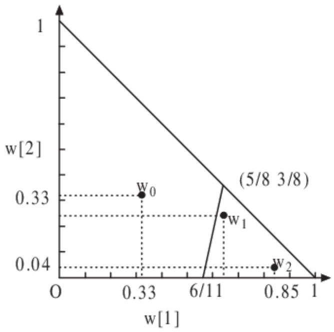
Generally, there exist too many new queries which satisfy the above conditions. So in order to evaluate the quality of the new query, we define an evaluation model as follows.

$$Eval(Q(k',\vec{w}')) = \lambda_w \Delta W + \lambda_c \Delta C \qquad (1)$$

where λ_w, λ_c are the user's tolerance to the changes of w and top-k results on her/his original query, and $\lambda_w + \lambda_c = 1$. $\Delta w = \| \vec{w}' - \vec{w}_0 \|_2$ is used to calculate the changes of w, and ΔC is used to measure the changes of the initial top-k results, which is calculated by the following equation.

$$\Delta C = |Q(k_0,\vec{w}_0) \cup Q(k',\vec{w}')| - |Q(k_0,\vec{w}_0) \cap Q(k',\vec{w}')|$$

Since ΔC is much greater than ΔW, we normalize them in Section 4.3. Through this definition, we can see that the new query with the smaller evaluated value is better than these queries with larger values, because it makes a litter change on the original query.



Ranking	1	2
Alert	5	4

$Q(2,w_0)$

Ranking	1	2	3	4	5
Alert	5	1	4	3	2

Ranking under w_1

Ranking	1	2	3	4	5
Alert	5	1	3	4	2

Ranking under w_2

Fig.2 A 3-Dimension example

Example 2. Let us give an example. We use Fig. 2 to explain how to answer user's question that is proposed in example 1. First the results of initial query $Q(2,\vec{w}_0)$ are alert-5 and alert-4. In order to satisfy user's preference $score(\vec{m},\vec{w}') > score(\vec{p},\vec{w}')$, we get some weight vectors as candidate, such as $\vec{w}_1$ and $\vec{w}_2$. Take $\vec{w}_1$ for example, both $Q(2,\vec{w}_1)$ and $Q(3,\vec{w}_1)$ satisfy the conditions of a new query answer. However, we think that $Q(3,\vec{w}_1)$ is a better new query answer than $Q(2,\vec{w}_1)$, because it only adds a new alert-1 into the original top-k results ($\Delta C = 1$). In the same way, $Q(2,\vec{w}_2)$ is an better new query answer than any other queries $Q(k',\vec{w}_2)$ when $k' > 2$. Finally, using evaluation model defined in Eq. (1), we find that $Q(3,\vec{w}_1)$ is better than $Q(2,\vec{w}_2)$ because they have the same ΔC, and $\Delta W_1 < \Delta W_2$.

3.3 Problem Analysis

As discussed above, we try to define a new query which makes m be ranked before p (formally described as $score(\vec{m},\vec{w}') > score(\vec{p},\vec{w}')$). In the data space, we say if an object p domains an object m, and then $score(\vec{p},\vec{w}') > score(\vec{m},\vec{w}')$ is always hold, so this situation is out of our consideration. Otherwise if an object p is incomparable with m, then a hyper plane

$H:(\vec{m}-\vec{p})\cdot\vec{w}=0$ partitions weight space into two part $W_>$ and $W_<$. The $\vec{w}$ located in $W_>$ makes $score(\vec{m},\vec{w})>score(\vec{p},\vec{w})$. The $\vec{w}$ located in $W_<$ means that $score(\vec{m},\vec{w})<score(\vec{p},\vec{w})$. So if we want to get a new query that satisfies the condition $score(\vec{m},\vec{p})>score(\vec{p},\vec{w})$, we only need to consider the weight space $W_>$. Based on each $\vec{w}$ in $W_>$, we can get a new query to satisfy the conditions. Furthermore, based on the above quality function, we can evaluate all the new queries, and return the best new query to user as the refined query. However, there are infinite $\vec{w}$ in $W_>$, it's very hard to enumerate all the possible $\vec{w}$. Therefore, we propose a new optimal algorithm to solve this problem.

4. **Detailed Techniques**

In this section, we present our method to solve the problem how to refine a query based on the quality function. First we give a basic solution to explain our main idea. Then base on this basic solution, we give some techniques to improve the performance of our algorithm.

4.1 Basic Solution

Let us start the discussion from the basic solution. Basically, we solve this problem by four core phases. First, we get the weight space that satisfies $(\vec{m}-\vec{p})\cdot\vec{w}>0$, and sample s weight vectors from this space, which is donated by $S=\{\vec{w}_1,\vec{w}_2,\cdots,\vec{w}_s\}$. Second, based on S, we use the progressive top-k algorithm [7-9] to get the top-k results and the terminal conditions are discussed in Section 4.2. Third, for each $\vec{w}_i\in S$, we terminate the progressive top-k algorithm after executing k_i steps. Then we can evaluate the quality of each query $Q(k_i,\vec{w}_i)$ by Eq. (1). Finally, we choose the query with the least value as the new query which is returned to user as the answer.
In the following sections we try to improve the efficiency of the basic solution.

4.2 The Terminal Conditions of Progressive Top-k Results

In this section, we try to explain how to terminate the progressive top-k algorithm. As discussed above, we want to get the new query with the least value which is calculated by Eq. (1). So for each $\vec{w}_i\in S$, we need to minimize the ΔC. However through the definition of $\Delta C=|Q(k_0,\vec{w}_0)\cup Q(k',\vec{w}_i)|-|Q(k_0,\vec{w}_0)\cap Q(k',\vec{w}_i)|$, we can see that ΔC depends on $Q_0(k_0,\vec{w}_0)$ and $Q(k',\vec{w}_i)$. Since k_0, $\vec{w}_0$ and $\vec{w}_i$ are certain, the terminal conditions of the progressive top-k algorithm under $\vec{w}_i$ plays a crucial role.

For the sake of brevity, we denote ΔC_k instead of ΔC under the new query $Q(k,\vec{w}_i)$. First, we will give the algorithm (as shown in Algorithm 1) on how to calculate the terminal condition. Then we will give a detailed analysis of this algorithm.

Algorithm 1: GetOptimalK

Input: Dataset D, Weight vector $\vec{w}$, User's initialized query $Q_0(k_0,\vec{w}_0)$
Output: new query $Q(k,\vec{w})$
1. $Q(k_m,\vec{w})\leftarrow$ Running progressive top-k algorithm until to see m;
2. $a=|q\mid q\in Q(k_m,\vec{w})\cap Q_0(k_0,\vec{w}_0)|$;
3. $UpperBound\leftarrow k_m+2(k_0-a)$;
4. $\Delta C=\Delta C_{min}=k_0+k_m-2a$;
5. $k_{min}=k_m$;
6. $k=k_m+1$;
7. while $k\le UpperBound$ do
    7.1 if $Tuple(k,\vec{w})\in Q_0(k_0,\vec{w}_0)$ then
        $\Delta C=\Delta C-1$;
    7.2 else
        $\Delta C=\Delta C+1$;
    7.3 if $\Delta C<\Delta C_{min}$ then
        $\Delta C_{min}=\Delta C; k_{min}=k$;
    7.4 $k=k+1$;
return $Q(k_{min},\vec{w})$;

To prove the correctness of the above algorithm, in the following, we need to prove that the optimal answer that minimizes Eq. (1) in terms of ΔC has an upper bound (as shown in line 3 Algorithm 1).

Theorem 1: The progressive top-k algorithm could find the optimal answer in $k_m+2(k_0-a)$ steps, where k_m is the rank of m under the weight vector $\vec{w}$ (line 1), a is the number of objects which is both in $Q_0(k_0,\vec{w}_0)$ and $Q(k_m,\vec{w})$ (line 2).

Proof: we proof this theorem by two cases: (i) $Q(k_m,\vec{w})$ may not be the optimal answer. (ii) k' in the optimal answer $Q(k',\vec{w})$ has an upper bound $k_m+2(k_0-a)$. In case (i), we can construct an example to show that $Q(k_m,\vec{w})$ may not the optimal answer. For example,

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

if $Tuple(k_m+1,\vec{w}) \in Q_0(k_0,\vec{w}_0)$, then ΔC decreases after we increase k_m. So this case shows that the missing object m has appeared in top-k results, but we cannot stop progressive top-k algorithm because of $Eval(Q(k_{m+1},\vec{w})) < Eval(Q(k_m,\vec{w}))$. In case (ii), we try to prove $\forall k'' > k_m + 2(k_0 - a)$, $\Delta C_{k''} > \Delta C_{k_m}$. The proven is as follows.

$$\begin{aligned}
\Delta C_{k''} &= | Q(k_0,\vec{w}_0) \cup Q(k'',\vec{w}) | - | Q(k_0,\vec{w}_0) \cap Q(k'',\vec{w}) | \\
&= (k_0 + k'' - a'') - a'' = k_0 + k'' - 2a'' \\
&\quad \because\ k'' > k_m + 2(k_0 - a) \\
&> k_0 + k_m + 2(k_0 - a) - 2a'' \\
&\quad \because a'' \le k_0 \\
&\ge k_0 + k_m - 2a = k_0 + (k_m - a) - a \\
&= | Q(k_0,\vec{w}_0) \cup Q(k_m,\vec{w}) | - | Q(k_0,\vec{w}_0) \cap Q(k_m,\vec{w}) | \\
&= \Delta C_{k_m}
\end{aligned}$$

where $a'' = | q \mid q \in Q(k'',\vec{w}) \cap Q_0(k_0,\vec{w}_0) |$. Through this case, we can see that $Eval(Q(k_m,\vec{w})) < Eval(Q(k'',\vec{w}))$. So we can see that the progress top-k algorithm could be terminated after $k_m + 2(k_0 - a)$ steps. Combining these two cases together, we can guarantee that our algorithm can get the optimal answer correctly.

4.3 Pruning the weight space

In this section, we will discuss where to get the candidate weight vectors. As discussed above, based on the top-k results returned by the original query, user may propose a question such as p is in top k results, why not object m. In the basic idea, to answer this question, we sample some weight vectors from the weight space that constrained by some inequalities $\Gamma = \{(\vec{m} - \vec{p}) \cdot \vec{w} > 0, w[i] \in [0,1], \sum w[i] = 1\}$. Actually, we only need sample a subspace of Γ. First we will give main steps about how to get the subspace (As shown in Algorithm 2). Then we will try our best to explain why we only need this subspace to be sampled.

Algorithm 2: GetSubRegion

Input: User's initialized query $Q_0(k_0,\vec{w}_0)$, Parameters λ_w, λ_c, Missing object m, Comparable object p

Output: Sampling Region S_{region}

1. $\Gamma = \{(\vec{m} - \vec{p}) \cdot \vec{w} > 0, w[i] \in [0,1], \sum w[i] = 1\}$
2. $\vec{w}_1 \leftarrow$ Project $\vec{w}_0$ to Γ.
3. $Q(k_1,\vec{w}_1) \leftarrow$ GetOptimalK($\vec{w}_1$)
4. $\Delta W_1 = \| \vec{w}_1 - \vec{w}_0 \|_2$
5. $\Delta C_1 = | Q(k_0,\vec{w}_0) \cup Q(k_1,\vec{w}_1) | - | Q(k_0,\vec{w}_0) \cap Q(k_1,\vec{w}_1) |$
6. $\Delta W_{max} = \Delta W_1 + \lambda_c \Delta C_1 / \lambda_w$
7. $S_{region} = \{ \| \vec{w} - \vec{w}_0 \|_2 < \Delta W_{max} \} \cap \Gamma$

return S_{region};

Now we will give a brief walkthrough of Algorithm 2. First we project $\vec{w}_0$ to Γ, and get the projection point $\vec{w}_1$ [10-13]. Second, base on the $\vec{w}_1$, we use Algorithm 1 to get the optimal query $Q(k_1,\vec{w}_1)$. Then we evaluate the quality of $Q(k_1,\vec{w}_1)$, and get the maximal value of ΔW. Finally, we can prune the weight space, and only reserve the weight space which is described as S_{region} (Theorem 2).

Theorem 2. For any $\vec{w}_i \in W - S_{region}$, it holds that $Eval(Q(k_1,\vec{w}_1)) < Eval(Q(k_i,\vec{w}_i))$, which means that the quality of any query $Q(k_i,\vec{w}_i)$ whose weight vector $\vec{w}_i$ locates out of region S_{region} is worse than the quality of $Q(k_1,\vec{w}_1)$.

Proof:

$$\begin{aligned}
Eval(Q(k_i,\vec{w}_i)) &= \lambda_w \Delta W_i + \lambda_c \Delta C_i \\
&> \lambda_w \Delta W_{max} + \lambda_c \Delta C_i \\
&= \lambda_w \Delta W_1 + \lambda_c \Delta C_1 + \lambda_c \Delta C_i \\
&= Eval(Q(k_1,\vec{w}_1)) + \lambda_c \Delta C_i
\end{aligned}$$

So when $\vec{w}_i \in W - S_{region}$, $\Delta W_i = \| \vec{w}_i - \vec{w}_0 \|_2 > \Delta W_{max}$, $Eval(Q(k_1,\vec{w}_1)) < Eval(Q(k_i,\vec{w}_i))$ is always hold.

Based on Theorem 2, we can improve our algorithm from three parts. First we try to improve our evaluation model. Second, we try to improve the efficiency of sampling. Finally, we do our best to stop the progressive top-k algorithm as earlier as possible.

1) Normalizing the evaluation model

As we mentioned in Section 3.2, ΔC is much greater than ΔW, so we normalize each of them by their maximal values. We normalize Δw using ΔW_{max}, because for any $\vec{w}_i \in S_{region}$, $\Delta W = \| \vec{w}_i - \vec{w}_0 \|_2 \le \Delta W_{max}$. In a similar way, we normalize ΔC by ΔC_1 in Algorithm 2. Because for any $\vec{w}_i \in S_{region}$, $\Delta W_i > \Delta W_1$ always holds, if $\Delta C_i > \Delta C_1$, the quality of new query under $\vec{w}_i$ must be worse than the query under $\vec{w}_1$. In this case, we discard this $\vec{w}_i$. Now we have a normalized evaluation model as follows.

$$Eval(Q(k',\vec{w}')) = \lambda_w \frac{\Delta W}{\Delta W_{max}} + \lambda_c \frac{\Delta C}{\Delta C_1} \qquad (2)$$

2) How to get the weight vector candidate S?

Instead of sampling the whole weight space, we sample the candidate weight vectors from the S_{region}. Now another problem is how many weight vectors we should sample. In a straightforward thinking, the bigger the $|S|$, the more approximate the answer. However, the bigger size of S means we need to execute more progressive top-k query which increases the running time. To solve this problem, we adopt the method which is discussed in Section III.C in [5]. We say a new query is the best $T\%$ query if its quality is better than $(1-T)\%$ new queries in the whole answer space, and we hope the probability of getting at least one such new query is larger than a certain threshold $\Pr$:

$$1-(1-T\%)^{|S|} \geq \Pr \tag{3}$$

This Equation is general. The sample size is independent of the data size and the dimension but controlled by two parameters $T\%$ and $\Pr$.

3) Improve the terminal condition in Section 4.2

After getting the candidates $S=\{\vec{w}_1,\vec{w}_2\cdots,\vec{w}_n\}$ for weight vectors, as shown in Section 4.2, we can use Algorithm 1 to compute the optimal k for each weight vector $\vec{w}\in S$. We can see that for specific $\vec{w}_i$, we use the progressive top-k algorithm to get the top-k objects one by one, and the optimal k under $\vec{w}_i$ is calculated until meeting the upper bound. In order to decrease the execution time, in this section, we try to terminate the progressive top-k algorithm before meeting the upper bound of k. The main idea is that we terminate the progressive top-k algorithm and discard this $\vec{w}_i$ if the quality under this $\vec{w}_i$ will not be better than the quality we calculate before (Suppose that the optimal answer we have gotten is $Q_{opt}(k_{opt},\vec{w}_{opt})$).

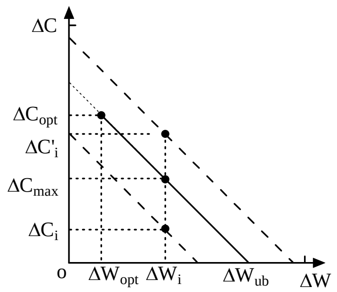


Fig.3 Terminal Conditions

According to Fig. 3, now we will give details about how to improve the performance of progressive top-k algorithm. First, we have an optimal query $Q_{opt}(k_{opt},\vec{w}_{opt})$ (At the beginning, this optimal query is set to be the new query under $\vec{w}_1$). Based on this optimal query, we can evaluate the quality of this query through the Eq. (2) .

$$E_{min}=Eval(Q(k_{opt},\vec{w}_{opt}))=\lambda_w\frac{\Delta W_{opt}}{\Delta W_{max}}+\lambda_c\frac{\Delta C_{opt}}{\Delta C_1}$$

Given a w_i, we consider these two situations.

The first situation is $\Delta W_i>\Delta W_{ub}=E_{min}\Delta W_{max}/\lambda_w$. In this situation, we can see that the quality of new query under $\vec{w}_i$ will never be better than the query $Q_{opt}(k_{opt},\vec{w}_{opt})$. This property could be proven using the same methods in Theorem 2.

The second situation is $\Delta W_i\leq\Delta W_{ub}$. First, we get the upper bound of ΔC_{max}. When using progressive top-k algorithm to get the top-k results, we can get a lower bound of ΔC (represented by ΔC_{lb}) at each step. In case (1), if the missing object m does not appear in top-k results when $\Delta C_{lb}>\Delta C_{max}$, we can discard this weight vector since it could not get a better new query than $Q_{opt}(k_{opt},\vec{w}_{opt})$. In case (2), when the missing object appears in the top-k results, $\Delta C_{lb}<\Delta C_{max}$ is satisfied. As mentioned in Theorem 1, we continue to execute progressive top-k algorithm to find the optimal k which minimizes ΔC. During this procedure, when the terminal condition $\Delta C_{lb}>\Delta C_{max}$ or $k>UpperBound$ (As discussed in Algorithm 1) is satisfied, we can terminate the top-k algorithm under this weight vector.

4.4 Algorithm

The pseudo code of our complete idea is presented in Algorithm 3. It is self explain and mainly discussed above, so we do not give it a walkthrough here.

Algorithm 3: GetOptimalQuery

Input: Dataset D, User's initialized query $Q_0(k_0,\vec{w}_0)$, Parameters λ_w,λ_c, Missing object m, Comparable object p

Output: new query $Q_{opt}(k_{opt},\vec{w}_{opt})$

1. $S_{region}\leftarrow$ *GetSubRegion*
   //Sample S_{region} and sort them by ΔW
2. $S\leftarrow\{\vec{w}_2,\cdots,\vec{w}_s\}$
   // $\vec{w}_1$ is the projection point of $\vec{w}_0$
3. $Q_{opt}(k_{opt},\vec{w}_{opt})\leftarrow$ *GetOptimalK*($\vec{w}_1$)
4. For each $\vec{w}_i\in S$
   4.1 $\Delta W_i=\|\vec{w}_i-\vec{w}_0\|_2$
   4.2 $E_{min}=Eval(Q_{opt}(k_{opt},\vec{w}_{opt}))$

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

4.3 If $\Delta W_i > E_{min}\Delta W_{max} / \lambda_w$ then
Continue
4.4 $\Delta C_{max} = (E_{min} - \lambda_w \frac{\Delta W_i}{\Delta W_{max}})\Delta C_1 / \lambda_c$
//running progressive top-k algorithm
4.5 $k = 1$; $\Delta C_{lb} = 0$
4.6 While $Tuple(k, \vec{w}_i) \neq m$ do
if $Tuple(k, \vec{w}_i) \notin Top(k_0, \vec{w}_0)$ then
$\Delta C_{lb} = \Delta C_{lb} + 1$
If $\Delta C_{lb} > \Delta C_{max}$ then
continue
$k = k + 1$;
4.7 $a = | q \mid q \in Q(k, \vec{w}_i) \cap Q_0(k_0, \vec{w}_0) |$
4.8 $UpperBound \leftarrow k + 2(k_0 - a)$;
4.9 $\Delta C = \Delta C_{min} = k_0 + k - 2a$;
4.10 $k_{min} = k$; $k = k + 1$;
4.11while $k \le UpperBound$ and $\Delta C_{lb} \le \Delta C_{max}$ do
if $Tuple(k, \vec{w}_i) \in Q_0(k_0, \vec{w}_0)$ then
$\Delta C = \Delta C - 1$;
else
$\Delta C = \Delta C + 1$; $\Delta C_{lb} = \Delta C_{lb} + 1$
if $\Delta C < \Delta C_{min}$ then
$\Delta C_{min} = \Delta C; k_{min} = k$;
$k = k + 1$;
//evaluate the quality of new query
4.12 If $Eval(Q(k_{\min}, \vec{w}_i)) < E_{min}$
$E_{min} = Eval(Q(k_{min}, \vec{w}_i))$
$Q_{opt}(k_{opt}, \vec{w}_{opt}) \leftarrow Q(k_{min}, \vec{w}_i)$
return $Q_{opt}(k_{opt}, \vec{w}_{opt})$;

Now we give a brief complexity analysis of our algorithm. We use the algorithm in [10] to compute the distance from a point to a simplex, the complexity of this algorithm is $O(n^4)$, where n is the number of vertices of a simplex. In our case, since the simplex is defined by $\Gamma = \{(\vec{m} - \vec{p}) \cdot \vec{w} > 0, w[i] \in [0,1], \sum w[i] = 1\}$, which makes n be small. Then in the loop to compute the new query under each $\vec{w}_i$, the time cost mainly comes from executing the progressive top-k. As discussed in [8], the cost of the progressive top-k algorithm is $k + |skyline(D)|$. The complexity of our algorithm is $|S|(k + |skyline(D)|)$.

5. Experimental Study

We conduct a thorough performance evaluation on the efficiency and effectiveness of our techniques. Since this work is the first work in query modification (as discussed in related works), our performance evaluation is conducted against our techniques only. Specifically, we focus on evaluating our GetOptimalQuery algorithm in Section 4.4. As there is no existing work, we compare our algorithm under different quality function in Eq. (2) (TMW stands for "tolerate modifying $\vec{w}$" which sets $\lambda_c = 0.9$ and $\lambda_w = 0.1$. TMR stands for "tolerate modifying top-k results" which sets $\lambda_c = 0.1$ and $\lambda_w = 0.9$. NM stands for "Never mind" which sets $\lambda_c = 0.5$ and $\lambda_w = 0.5$).

5.1 Experiment setup

All the experiments are implemented by Java and compiled by JDK 1.7, and we run all the experiments on Ubuntu Linux Operating system with Intel Core-2 Duo Processor and 2GB memory. We use both real and synthetic data sets in our evaluation process.

Real dataset is extracted from NBA players' game-by-game statistics (http://www.nba.com), containing 16916 game statistics of all NBA players from 1973 to 2009 [14]. Each record represents a NBA player by regular season, player name, points per game (PTS), rebounds per game(REB), assists per game(AST), steals per game(STL), blocks per game(BLK), field goal percentage (FGP), field throw percentage (FTP), and three points percentage (TPP). Based on this real dataset, we give an interesting case to show the meaning of our method.

Table II: Parameters setting

Parameter	***Ranges***
Data size	1K, **10K**, 50K, 100K,
Dimension	2, **3**, 4, 5
Original k_0	5, **10**, 15, 20

Synthetic datasets are generated using classical method in [15] with respect to the following parameters. Table II summarizes parameter ranges, and the default values are in bold font. Note that the sample size is determined by the Eq. (3), here we set the default T to be 0.2 and Pr to be 0.8 (resulting the sample size of 800 weight vectors). In the experiments below, these parameters use default values unless otherwise specified. Based on these datasets, we test the efficiency of our algorithm.

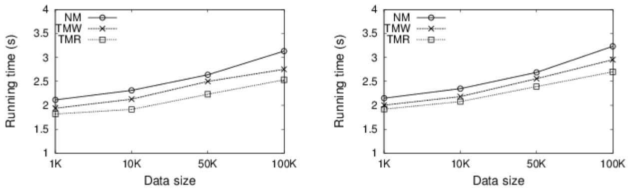
IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

5.2 Experiment Results

The first experiment is done on the NBA data set. In this experiment, we use three attributes PTS, FGP and FTP to find the top-5 players in NBA history. Therefore, we issue an original top-5 query with the preference $\vec{w}_0$ = (0.333, 0.333, 0.334). The result of this query is {Michael Jordan (1986), Michael Jordan (1987), Michael Jordan (1989), Michael Jordan (1988), George Gervin (1979)}. Then we may be confused with these top-5 results and propose such a question: "since George Gervin is in top-5 results, why not Kobe Bryant". In order to answer this question we use our proposed algorithm with the parameters $Q_0(5,\vec{w}_0)$, missing player Kobe Bryant and the comparable player George Gervin. Using TMW mode, in 3422ms, we get a new query with k=6 and $\vec{w} = (0.379,\ 0.272,\ 0.349)$. The new query indicates us to put more weights on PTS ability if we wish to see that Kobe Bryant is ranked before George Gervin. The corresponding result of $Q(6,\vec{w})$ is {Michael Jordan (1986), Kobe Bryant (2005), Michael Jordan (1987), Bob Mcadoo (1974), Michael Jordan (1989) George_Gervin (1979)}. From this case, we can see that we answer the user's why-not questions through little change of the original top-k query results, only add two new players and remove a player.

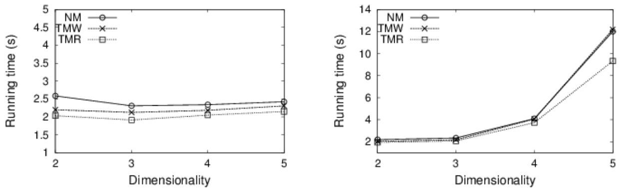
Statement. In the following experiments, we test the performance of our algorithm on the synthetic data sets. Given an original query $Q_0(k_0,\vec{w}_0)$, we set the object which is ranked at last of Q_0 to be the comparable object p . Then the missing object m is random selected from $D-Q_0$. As discussed above, there are two case relationships between m and p . If p dominates m (m will never dominates p), we drop this m , and do not run our method under this situation. Otherwise we get the running time of our algorithm and computer the average time cost.

The second experiment tests the running time of our algorithm under different data size. As depicted in Fig. 4, this experiment is conducted against the 4 data size and the 3 tolerant models. It shows that our algorithm scales linearly with the data size. The time cost under TMR (tolerate modifying top-k results is) model is the best because ΔW_{ub} in Fig. 3 is closer to ΔW_{opt} than the other models. Therefore, many weight vectors are discarded at the beginning of Algorithm 4 because they cannot get a better new query answer than the existing optimal answer.



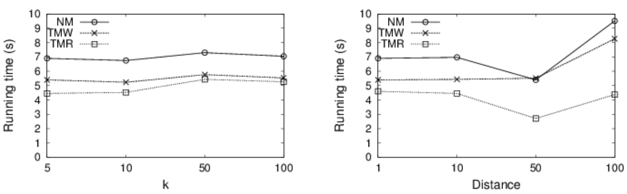
(a) Uniform Data (b) Anti-correlated Data
Fig.4 Varying Data Size

The third experiment, conducted against Uniform and Anti-correlated datasets with dimension range from 2 to 5, evaluates effects of dimension. The results are reported in Fig.5. It demonstrates that the time cost scales linearly with the dimension. We also notice that the running times do not increase under the Uniform Data, but at a faster rate on Anti-correlated Data. Generally speaking, executing a progressive top-k algorithm in high dimension on anti-correlated data set needs more time because of the more attribute and more top-k candidate.



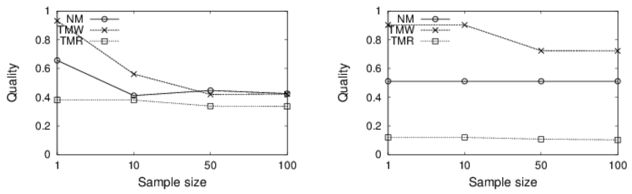
(a) Uniform Data (b) Anti-correlated Data
Fig.5 Varying Dimensionality

The fourth experiment is to test running time under different original k_0 and missing object m . We find that there is almost no relationship between the running time and k_0 or m . So we only show results on anti-correlated data set. In Fig. 6(a), we vary k_0 (This is the same means as we vary the object p), and we select the missing object which is ranked at k_0+10 in original query (If m is dominated by p , we use k_0+11 instead). In Fig. 6(b), distance=10 means there at least exists 10 objects between p and m in the original query.



(a) Varying k (b) Varying missing object
Fig.6 Varying Original Query

The final experiment is to evaluate the effects of sample size. We try to analyze the effects which sample size makes on the quality of the new query. Since the sample size is controlled by T and Pr , we do not show the results of varying T or Pr separately. Fig. 7 shows the quality of optimal new query under different sample sizes. We can see that the quality of the optimal new query under larger sample size is better than the one under smaller sample size. Note that in Fig. 7(b), the quality of NM model is keep the same under different sample size because the quality of new query under the sample weight is worse than the quality of new query under $\vec{w}_1$.



(a) Uniform Data (b) Anti-correlated Data

Fig.7 Varying Sample Size

6. Conclusions

In this paper, we have introduced a new model for solving the problem how to modify the query to answer user's question. At the beginning, we explain the motivation of our problem. Then we have presented this problem and outlined the challenges. Based on the formally definition, we give our solutions for practically realizing such an algorithm. Our experimental evaluation of an implementation of this algorithm in real and synthetic data sets demonstrates the efficiency of our approach.

Acknowledgments

This work is supported by the National 863 Program of China (2010AA012505, 2011AA010702, 2012AA01A401 and 2012AA01A402), National Natural Science Foundation of China (60933005), National Science Technology Support Program of China (2012BAH38B04) and National 242 Information Security Program of China (2011A010). The authors would like to thank the reviewers for their detailed comments which significantly improved this paper.

References

[1] A. Vlachou, C. Doulkeridis, Y. Kotidis et al., "Reverse top-k queries," ICDE, 2010, pp. 365-376.
[2] A. Chapman, and H. V. Jagadish, "Why not?," SIGMOD Conference, 2009, pp. 523-534.
[3] D. Nanongkai, A. D. Sarma, A. Lall et al., "Regret-Minimizing Representative Databases," PVLDB, vol. 3, no. 1, pp. 1114-1124, 2010.
[4] Q. T. Tran, C.-Y. Chan, and C.-Y. Chan, "How to ConQueR why-not questions.," SIGMOD Conference, 2010, pp. 15-26.
[5] Z. He, and E. Lo, "Answering Why-not Questions on Top-k Queries.," ICDE, IEEE Computer Society, 2012, pp. 750-761.
[6] P. Tsaparas, T. Palpanas, Y. Kotidis et al., "Ranked Join Indices," ICDE, 2003, pp. 277-288.
[7] Y.-C. Chang, L. D. Bergman, V. Castelli et al., "The Onion Technique: Indexing for Linear Optimization Queries," Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, May 16-18, 2000, Dallas, Texas, USA, ACM, 2000, pp. 391-402.
[8] L. C. Lei Zou, "Dominant Graph: An Efficient Index Structure to Answer Top-K Queries," ICDE, 2008, pp. 536-545.
[9] R. Fagin, A. Lotem, and M. Naor, "Optimal aggregation algorithms for middleware," in Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, Santa Barbara, California, United States, 2001, pp. 102-113.
[10] V. M. Oleg Golubitsky, and S. M. Watt, "An Algorithm to Compute the Distance from a Point to a Simplex," East Coast Computer Algebra Day, 2012.
[11] R. A. Waltz, J. L. Morales, J. Nocedal et al., An interior algorithm for nonlinear optimization that combines line search and trust region steps, Mathematical Programming 107, 2006.
[12] P. Vincent, and Y. Bengio, "K-Local Hyperplane and Convex Distance Nearest Neighbor Algorithms," Advances in Neural Information Processing Systems, The MIT Press, 2001, pp. 985-992.
[13] C. Michelot, "A finite algorithm for finding the projection of a point onto the canonical simplex of R^n," Journal of Optimization Theory and Applications, vol. 50, pp. 195-200, 1986.
[14] "Databasebasketball_2009_v1.zip," http://www.databasebasketball.com/stats_download.htm.
[15] S. Börzsönyi, D. Kossmann, and K. Stocker, "The Skyline Operator," ICDE, 2001, pp. 421-430.

Jianfeng Zhang is a Ph.D student in School of Computer Science, National University of Defense Technology. He received the B.S. and M.S. degrees in Computer Science from National University of Defense Technology in 2006 and 2009 respectively. His current research is in network security, database system and data mining.

Weihong Han received the PhD degree from National University of Defense Technology. Currently she is an associate professor, and her research interests include network & information security, database and data mining.

Yan Jia received the PhD degree from National University of Defense Technology in 2001. She is a doctoral tutor at National University of Defense Technology. Her research interests include massive database system, information security and data mining.

Peng Zou is a professor and Ph.D. supervisor at National University of Defense Technology. His research interests include network & information security, distributed computing.

Hua Fan is currently a Ph.D student at National University of Defense Technology. He received the B.S. and M.S. degrees in Computer Science from National University of Defense Technology in 2006 and 2009, respectively. He His research interests include stream data management and Sensor network.

Using Arabic Wordnet for semantic indexation in information retrieval system

Mohammed Alaeddine Abderrahim[1], Mohammed El Amine Abderrahim[2]
Mohammed Amine Chikh[2]

[1] **Department of Computer Science, University Abou Bekr Belkaid, Faculty of Science, PoBox 119, Tlemcen 13000, Algeria**

[2] **Department of Electrical Engineering and Electronics, University Abou Bekr Belkaid, Faculty of Engineer Science, PoBox 230, Tlemcen 13000, Algeria**

Abstract

In the context of arabic Information Retrieval Systems (IRS) guided by arabic ontology and to enable those systems to better respond to user requirements, this paper aims to representing documents and queries by the best concepts extracted from Arabic Wordnet. Identified concepts belonging to Arabic WordNet synsets are extracted from documents and queries, and those having a single sense are expanded. The expanded query is then used by the IRS to retrieve the relevant documents searched. Our experiments are based primarily on a medium size corpus of arabic text. The results obtained shown us that there are a global improvement in the performance of the arabic IRS.

Keywords: *Information Retrieval System, Disambiguation, Arabic WordNet, ontologies, Semantic indexing.*

1. Introduction

The ontologies are known as tools able to manipulate the knowledge behind the concepts. We can used them in several fields such as informations search, the automatic translation..,. The ontologies can be used at different levels in the IRS. The orjectives of our study is to see the effects of the ontologies in process of indexing documents and queries, we are talking about the semantic indexing. In the literature, there are many definitions of the semantic indexing. The semantic indexation (indexation by the sense of words) aims to correct the problems of the lexical matching by using the semantic indexes rather than the simple keywords. The semantic indexation method aims to retrieve the correct sense of the word in the text from different possibility senses word as defined in dictionaries, ontologies and other language resources [1]. It is based on algorithms of the word sense disambiguation (WSD). Among the disambiguation methods : those combining the disambiguated word with words taken from the context of a document witch help to determine their appropriate sense, more advanced approaches of disambiguation are using hierarchical representation to calculate the semantic distance or the semantic similarity between the compared words[1]. According to Sanderson [2] the successful of disambiguation improves the performance of the IRS, particularly in the case of the short queries (title only). Within the context of using the ontologies for the indexation, we found several works for English language cited in [3], the idea is to built an structure representing the document (respectively query) by using the semantic of the ontologies, this structure is called a semantic core of document (respectively query). Therefore, This is the first work of the semantic indexation of the documents (respectively query) with arabic texts.

In this paper we have implemented the method of semantic indexing of the documents and query for the information retrieval where are use Arabic Wordnet as a semantic resource to exploring the impact of passage from an indexation based on single words to an indexation based on concepts.

This paper is organized as follows. First, we describe the architecture with a discription of the operating process of our system. Then we present the experimentation with a discussion of results achieved and we have finished with a conclusion and prospects.

2. Architecture of our System

In this section, we describe the semantic indexing method based on Arabic Wordnet. This approach start with extracting the concepts of wordnet from the documents (respectively query). Then we retrieve the senses of those concepts from the synsets of arabic wordnet and with the

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

method of disambiguation[1] based on calculation of the semantic distances between those senses, we identify the appropriate sense (having only one sense) for every concept from proposed senses. For terms that don't belong to the vocabulary of WordNet, the system extracts their basic form before passing by the semantic indexing method described above. For example, the arabic wordnet does not contain the concept "أسباب", but it contains their basic form "سبب". Formally, let consider: D a document of collection composed of n words.

D= {w1, w2,…, wn}

The result of the concept detection process will be a document Dc. It corresponds to: Dc= {C1, C2,…, Cm, W'1, W'2,…, W'm'}. Where C1, C2,…, Cm are the concepts extracted from the document and identified like wordnet entries. If they are terms that do not belong to the WordNet vocabulary, they are not replaced like the case of words W'1, W '2, ..., W'm'. However, they will be added to complete the representation of the information expressed by the document in order to be used at the search stage.

2.1 Details of Our Approach with Example

Let consider the following text of document :

" سواء كانت حالة فقدان الذاكرة بشكل مؤقت أو دائم، أو جاءت بشكل مفاجئ أو ببطء فذلك يعتمد علي أسباب حدوث فقدان الذاكرة. إن عملية تقدم العمر قد ينتج عنها صعوبة في تعلم أو إدراك الأشياء الحديثة علي الشخص أو يمكن أن تتسبب في استغراق وقت أطول من قبل الشخص المسن في تذكر أو استدعاء الأشياء الحديثة عليه (ولكن التقدم في العمر لا يكون سبب في فقدان الذاكرة إلا إذا كان هذا التقدم مصحوباً بمرض معين ساعد في حدوث هذه الحالة). "

Table 1 presents the terms to be indexed after the elimination of the stop words. As well as the segmentation process that is used to link the terms that distinguished only with inflectional mark. Finally, the text is represented by an index of lemmatized words:

Table 1 : List of terms to index

فقدان	وقت	مؤقت	دائم	مفاجئ	بطء	يعتمد
مرض	ينتج	اشياء	اطول	حديثة	استغراق	سبب
جاءت	حدوث	ذاكرة	شخص	تذكر	مصحوب	ساعد
تقدم	ادراك	حالة	صعوبة	مسن	استدعاء	عمر

After omit the stop words, for example: {بعض ,سواء}. The process of extracting concepts recognized all the terms of the documents that belong to the Arabic Wordnet. Then, the method of the searching synonyms retrieved all senses of the concepts extracted, and the disambiguation method is used to select the right sense for every concept. The terms that do not belong to the vocabulary of the Arabic Wordnet, they are passed by the module of racine extraction in order to restart the search of the senses with the root. Or else, the words of text will be added to the final index for complete the representation of the information contained in the documents. Table 2 gives an example of selecting indexes to some concepts identified in the text:

Table 2: Example of selecting concepts from Arabic Wordnet

Dr/terms	Example of Synset Corresponding	index choice
حدوث	{حَدَث, حُصُول, حُدُوث, ظُهُور, وُقُوع} {حُدُوث, حُصُول, حَادِثَة, حَدَث, وَاقِع}	حُصُول
استدعاء	{ذِكْرَى, اِسْتِدْعاء, تَذَكُّر} {اِسْتِدْعاء, طَلَب حُضُور}	تَذَكُّر
تذكر	{ذاكِرَة, تَذَكُّر} {ذِكْرَى, اِسْتِدْعاء, تَذَكُّر}	ذَاكِرَة
جاء	{أَتَى, جَاءَ} {جَاءَ, ظَهَرَ} {أتى,حَضَرَ,جَاءَ,قَدِم}	أَتَى
ذاكرة	{ذاكِرَة, فِكْر} {ذاكِرَة,تَذَكُّر}	تَذَكُّر

For search step, the user queries are expanded with the same method as the documents using the synonyms of those terms to retrieve more relevant results and reduce the silence. Table 3 shows examples of queries before and after semantic indexing method.

Table 3: Examples of queries Expanded

N° query	Query	Proximate concepts
1	إِثْم	خَطِيئَة
……	……	……
12	بَحْث	دراسة
13	إِسْتِخْدَام	اِسْتِعْمَال
18	اِسْتِثْمار	تَوْظِيف

The detailed of our system are discribed with figure 1:

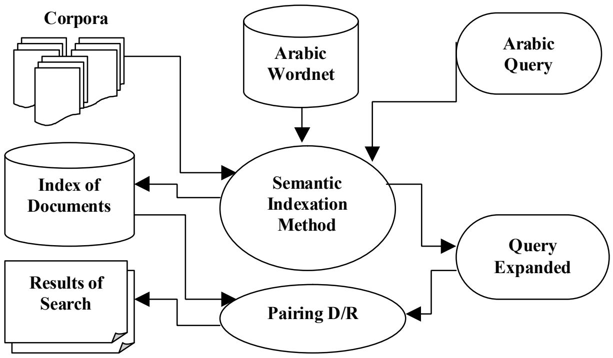


Fig. 1 Architecture of our system

[1] this method choose the appropriate sense (concept) from the proposed senses witch has most linked with other concepts of the same document, the similarity is calculated between senses that belongs to the different sets (synsets).

In the following, we have described our experimentation and discussion the results obtained.

3. Description of the experimentations

For our experimentation we have used a corpora of over 22,000 arabic documents (approximately 180 MB) in different areas (health, sport, politic, science, religion, ...). This corpora has approximately 17 millions words with 612,650 are differents word. A set of 70 keywords queries in various fields are chosen for evaluation.

Arabic Wordnet is a lexical database free available for standard arabic. This database follows the conception and methodology of Princeton Wordnet for English and Euro-WordNet for European languages. Its structure is like a thesaurus, it is organized around the structure of synsets, that is to say, sets of synonyms and pointers describing relations to other synsets. Each word can belong to one or more synsets, and one or more categories of discourse. These categories are organized in four classes: noun, verb, adjective and adverb. Arabic WordNet is a lexical network whose nodes are synsets and relations between synsets are the arcs. it currently counts 11,269 synsets (7,960 names, 2,538 verbs, adjectives, 110 adverbs 661), and 23,481 words [4], [5], [6], [7].

To evaluate the semantic indexing method we have segmented our experimentation to four search types and we will study them individually in order to estimate the augmentation of each type to improving the search performance.

The types of search are cited below:

- Simple search or research before semantic indexing (R0): we have used a list of 70 simple queries like keywords with a simple indexation of documents.
- Total Semantic Search (R1): we have indexed semantically a list of 70 queries and the collection of documents used for search.
- Expansion of query (R2): we have indexed semantically only a list of 70 queries and we have used a single word to index the documents.
- Semantic representation of the documents (R3): we have indexed semantically only the database of the documents and we have used a list of 70 simple queries like keywords.

The tables above describe search results:

- The number of documents found.
- The number of relevant documents found.
- The precision at 5 documents (P @ 5).
- The precision at 10 documents (P @ 10).
- The precision at 20 documents (P @ 20).
- The precision at 100 documents (P @ 100).
- The precision at 1000 documents (P @ 1000).
- The median average precision.

Table 4 presents : the number of documents found and the number of relevant documents found.

Table 4: The documents found and the relevant documents for each type of indexation

N° query	before semantic indexation		After semantic indexation					
	R0		R1		R2		R3	
	Nb Doc Founds	Nb Doc Relevants	Nb Doc Founds	Nb Doc Relevants	Nb Doc Founds	Nb Doc Relevants	Nb Doc Founds	Nb Doc Relevants
1	405	164	11588	6287	518	329	8937	6092
2	674	272	9332	5071	2579	1630	1914	1265
3	366	96	4237	2225	3560	2163	357	95
4	3539	361	17687	10985	9825	5564	3781	2438
…	…	…	…	…	…	…	…	…
49	681	423	6652	3161	4860	1414	663	423
50	1578	1129	6163	5267	1938	1154	3077	1451
…	…	…	…	…	…	…	…	…
70	170	50	7176	3071	573	297	155	49

A simple comparison of the results obtained before and after using the semantic indexation method to representing the documents and queries, enables us to deduce that this method (for any types) improves in most cases the number of documents and the number of relevant documents returned. In other words, semantic indexing can improve the recall.

Concretely:

- NDTB = The number of documents found before the semantic indexing method.
- NDTA = The number of documents found after the semantic indexing method.
- D = NDTA - NDTB (1)
- NDTPB = The number of Relevant Documents found before the semantic indexing method.

- NDTPA = The number of Relevant Documents found after the semantic indexing method.
- DP = NDTPA - NDTPB (2)
- If (D> 0 or DP> 0) then we can say that semantic indexation improves the performance of IRS in terms of recall.
- In contrast, if (D = 0 or DP = 0), in other words we have the same number of documents returned after the semantic indexing. So, we can say that there are no improvements in the quality of IRS of a recall viewpoint.

Counting The number of queries in terms of D and DP enabled us to establish the results (see table 5):

Table 5: Contribution of semantic indexing based on the documents found and the relevant documents found

Documents Found						
	Total queries (R1)		Total queries (R2)		Total queries (R3)	
D<0	0	0%	0	0%	35	50%
D=0	0	0%	9	12.85%	0	0%
D>0	70	100%	61	87.15%	35	50%
Relevant Documents Found						
	Total queries (R1)		Total queries (R2)		Total queries (R3)	
DP<0	0	0%	2	2.85%	10	14.29%
DP=0	0	0%	4	5.72%	9	12.85%
DP>0	70	100%	64	91.43%	51	72.86%

As shown on table 5, we notice that increasing the number of documents and the relevant documents found covers pratically all queries in R1. Moreover, R2 and R3 are the less appropriate methods for semantic indexing (D <0) and (DP <0) because the use of semantic indexation method modify the vocabulary in documents only (R3) or the queries only (R2). For Example: the term « إثم » it replaced in the semantic index of corpora by « خطيئة » and if we search by using this term query « إثم », the result will be negative.

Based on Table 4, we have established a comparison between the three search types (R1, R2 and R3) in order to identify the best method of semantic indexing of a viewpoint the documents found and the relevant documents found. Table 6 presents the results of this comparaison.

Table 6: Comparison between the various search types (R1, R2 and R3)

Documents found			
Percentage of queries which R1 has sent more documents than the others systems	Percentage of queries which R2 has sent more documents than the others systems	Percentage of queries which R3 has sent more documents than the others systems	Percentage of queries which the three systems (R1, R2, R3) sent the same number of documents
85.71%	4.29%	0%	0%
Relevant documents found			
Percentage of queries which R1 has sent more relevant documents than the others systems	Percentage of queries which R2 has sent more relevant documents than the others systems	Percentage of queries which R3 has sent more relevant documents than the others systems	Percentage of queries which the three systems (R1, R2, R3) sent the same number of relevant documents
90%	1.43%	0%	0%

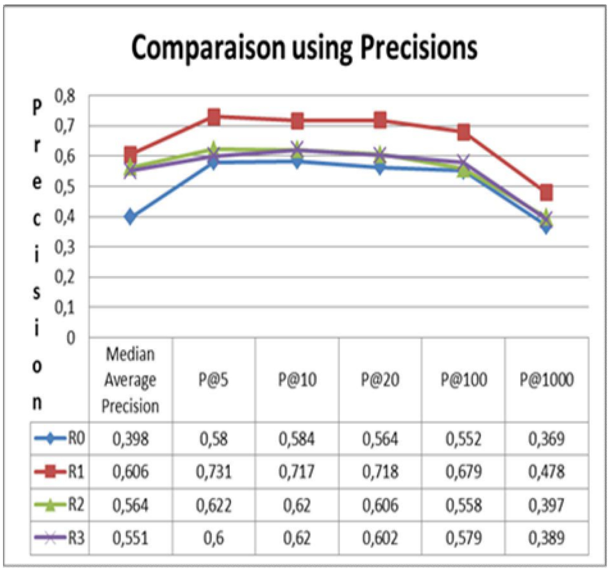
The results described in Table 6 preferred the system R1 so we can say that the semantic indexing of documents and queries together present the best system of search of a viewpoint the number of documents found and number of relevant documents found. This result affirms first consequent which was given in the table (5). Table 7 describes the different values of precision obtained in both systems before and after the use of the semantic indexing method.

Table 7: Different precision values obtained by both systems

		Median Average Precision	**P@5**	**P@10**	**P@20**	**P@100**	**P@1000**
Before Semantic Indexation	**R0**	0,398	0,580	0,584	0,564	0,552	0,369
After Semantic Indexation	**R1**	0,606	0,731	0,717	0,718	0,679	0,478
	R2	0,564	0,622	0,620	0,606	0,558	0,397
	R3	0,551	0,600	0,620	0,602	0,579	0,389

The comparison of three experimentations using the following graphic (see Figure 2) showed us that the semantic indexing method of documents and queries together (R1) give the best rate of precisions in all the measures taken into accounts (P@5, P@10, P@20, P@100, P@1000) as well as the median average precision. whereas, the semantic indexing of documents and queries separately (R2, R3) give inappropriate results for all the measures considered.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org



	Median Average Precision	P@5	P@10	P@20	P@100	P@1000
R0	0,398	0,58	0,584	0,564	0,552	0,369
R1	0,606	0,731	0,717	0,718	0,679	0,478
R2	0,564	0,622	0,62	0,606	0,558	0,397
R3	0,551	0,6	0,62	0,602	0,579	0,389

Fig. 2 Comparison of precision obtained by different systems

4. Discussions

In these experimentation we were interested by testing the semantic indexing strategy to represent the documents and the queries, the implementation of our system is organized as follows: we have started with indexing semantically the collection of documents which is considered as a preparation step for search, by using a semantic resource (as Arabic Wordnet). Then, we have tested different methods of searches started with (R1) which is based on the semantic indexing of documents and queries together. Another way to search, is to index semantically the queries (R2) or choosen to index semantically the collection of documents and use a simple query for search (R3). The objective behind the study of all these methods (R1, R2, R3) is to determine at what level in the IRS, the use of the semantic (in indexation of documents or queries, or together) produces best results.

From the viewpoint documents found and relevant documents found we can say that the use of semantic indexing method to represent both documents and queries together improves the performance of IRS. From the precision viewpoint, (R1) has good values for all the measures considered, consequently, it can be chosen as a method to represent (indexing) information in IRS.

If we must classified the other methods (R2) and (R3), we can said that R2 has the advantage to be more precise for 5 and 10 and 20 and 1000 firsts documents, and the median average precision. Contrary, it presents low values for 100 firsts documents as compared to R3.

The evaluation of the contribution of the arabic ontologies to IRS deduced by this experiment confirms the following characteristics:

- Reducing the silence in response of user queries.
- reduce the noise from responses of queries.
- facilitate the expression of query (assistance in the reformulation of query).
- Increasing the recall and precision.

In this context, we must emphasize that using concepts in the place of terms allows of:

- Provide a good representation of document collections by exploiting the semantics of concepts.
- Facilitating the reformulation of the user query.
- Provide a real support for matching process query/ documents by exploiting the semantic distance existing between the concepts.

5. Conclusion

In this paper we have developed an approach that have been proved its force for the English language. The idea of this article is to exploit a lexical resource (Arabic Wordnet) to index the documents as well as the user query in order to improve the retrieval results. Our experiments based on a medium corpus of Arabic language, we have proved that semantic resources (in our case: Arabic Wordnet) improves the quality of IRS and achieving our aims fixed at the beginning. We have remarked that the use of semantic indexing method to represent the documents and the queries together gives better results than using separately. The contribution of the ontologies in information retrieval system with arabic language is very interesting but it requires complete lexical resources witch are not available at present.

It therefore remains many things to do in the future, and the the most imminent extension of our research is to built a semantic core to represent the documents using Arabic Wordnet, as well as the study of the effect of every semantic relationship used in this process like (synonymy, hyponymy,…).

References

[1] R. Mihalcea, D.I. Moldovan, Semantic Indexing using WordNet Senses, in Proceedings of ACL Workshop on IR & NLP; Hong Kong, 2000, pp 35-45.

[2] M. Sanderson, Word Sense Disambiguation and Information Retrieval, PhD Thesis, Technical Report (TR-1997-7) of the Department of Computing Science at the University of Glasgow, Glasgow G12 8QQ, UK.10(1);1997.

[3] M Baziz: Indexation conceptuelle guidée par ontologie pour la recherche d'information. Thèse de doctorat, université Paul Sabatier; 2005.

[4] S. Elkateb, W. Black, H. Rodríguez, M. Alkhalifa, P. Vossen, A. Pease, C. Fellbaum: Building a WordNet for Arabic, in Proceedings of The fifth international conference on Language Resources and Evaluation; Genoa-Italy, 2006, pp 29-34.

[5] S. Elkateb, W. Black, P. Vossen, D. Farwell, A. Pease, C. Fellbaum: Arabic WordNet and the Challenges of Arabic. The Challenge of Arabic for NLP/MT. International conference at the British Computer Society, London, 2006; pp 15-24.

[6] W. Black, S. Elkateb, H. Rodriguez, M. Alkhalifa, P. Vossen, A. Pease, C. Fellbaum: Introducing the Arabic WordNet project, in Proceedings of the 3rd Global Wordnet Conference, Jeju Island, Korea, 2006, pp 295-299.

[7] W. Black, Sabri, El-Kateb, A Prototype English-Arabic Dictionary Based on WordNet, proceedings of the Second International WordNet Conference, Brno, Czech Republic, 2004, pp 67-74.

Mohammed Alaeddine Abderrahim is a research teacher at the University of Tiaret, Algeria. His research interests are natural language processing, information retrieval, information extraction, data mining and ontology. Med Alaeddine has a Magister in computer science from the University of Tlemcen, Algeria. He is a Doctorate candidate and a member of the Laboratory of Arabic Natural Language Processing in the University of Tlemcen.

Mohammed El Amine Abderrahim is a research teacher at the University of Tlemcen, Algeria. His research interests are natural language processing, information retrieval, information extraction, databases and data mining. Med El Amine has a Magister in computer science from UST Oran, Algeria, and a Doctorate in computer science from the University of Tlemcen, Algeria. He is a member of the Laboratory of Arabic Natural Language Processing, university of Tlemcen.

Mohammed Amine CHIKH received his Ph.D from the University of Tlemcen, Algeria and is currently a professor at the University of Tlemcen, Faculty of Technologie. His research interests include: knowledge engineering, knowledge discovery, data mining and medical semantic Web. Prof. Med Amine is a project manager in the Laboratory of Génie BioMédical (GBM), university of Tlemcen.

Design and Development of an Intelligent Mobile Health Tele-assistance System in Ambulance Practice Service.

Rachid MERZOUGUI
Faculty of Engineering Science of Tlemcen
STIC Laboratory, Chetouane
Tlemcen, Algeria

Abstract— **The development and the design of telemedicine services have taken a great consideration and care in the domain of wireless communication nowadays. The set of these researches is linked to the aging population and the persons exposed to accident risks in their daily life or the decline of their health in short time. In this case, it is indispensible to make a diagnostic in a real time and well manage the patient's computerized data between the various medical actors with the permanent security insurance of highly risky persons. Furthermore, the need to make a speed diagnostic of patients and to detect their health state, their parameters (medical information) of analyses with efficiency, allows us to gain time while transmitting the patient to the hospital (before his/her admission to the emergency service). It concerns the conception and development of services on mobile terminals for transferring medical information (of evacuated patients) in a real time with ensuring the mobility, the permanent security and the reliability insurance in covered zone by the mobile network, PLMN (GSM/GPRS…).**

Our attention has focused on the choice of a relevant work. It concerns an implemented and conceived service of medical Tele-assistance for monitoring risky persons brought by ambulance service. Thus, a medical bulletin is automatically filled on mobile terminal, immediately transmitted to the service of the emergency (on doctor's mobile terminal, data base of clinic…). This paper recalls a complete architecture of an economic wireless transmission system with the implementation of an effective application, adapted to the portable phone, allowing the doctor to have the medical information of patients who have a risk of accident.

Thus, the stakes of setting up such systems are numerous, so much for patients, medical staff and the society in general.

Keywords-component; Remote Monitoring; Services of Telemedicine; Medical Tele-assistance; Mobile; J2ME; Wireless sensors network.

I. INTRODUCTION

The recent technological progress of mobile telecommunications networks have contributed to the development of telemedicine [1]. It appears to be a medical reality and the use of portable devices as mobile phones is already imposed. These progresses applied to the medical field (medical imaging, transmission rate, confidentiality of data, the conviviality of systems…) and the miniaturization of devices open perspectives for medical development of Tele-assistance, remote monitoring in terms of a better quality of care and a reduction of public health cost [2]. These new technologies led to the emergence of a vast diversity of new ways for users to access and use information anywhere and anytime.

The majority of work and the systems undertaken in this field are carried out, implemented and installed in out of ambulance medical service. Moreover, these systems on one hand would have required large means and a large infrastructure in their deployment (installation, configuration…), which generates the disadvantage of the excessive expenditure and, on the other hand, they are directed to patients who need the aid of an ambulance.

Our contribution within the framework of work of this article is to introduce the medical Tele-assistance through the ambulances medical service and to handle the portable phones as basic tools. It is the reason that we are attached towards the medical Tele-assistance by mobile terminals.

Therefore, this research orientation exploits the mobility of wireless networks to transfer remotely the medical bulletin of persons brought in ambulance medical service. The suggested solution is an implementation of an algorithm which transmits the medical information received (from sensors networks set on person's body) and the calculated parameters (ECG parameters, Medical signals…) via a wireless communication in the purpose of exploiting a mobile phone for medical Tele-assistance.

II. Problem

In this work, we particularly focused on the design and the implementation of an open flexible system of service supplying for mobile environment. This system on one hand must satisfy the needs of the injured persons, and on the other hand, it is based on the technologies and the defined standards in this context. They notably concern the conception and development of communication architectures between the actors of these systems, monitoring, transfer in real time and improvement of human's quality life, the storage units of the collected data remotely, analytical tools and processing of large quantities of parameters.

The problem also is posed on the implementation level of an efficient algorithm intended to solve all the preceding constrains and adapted to mobile phones. This implementation requires many constraints (low resource calculation, size of screen, memory capacity, resolution and too simple software) to run properly.

In this context, the considered study leads to an inexpensive solution, efficient and comfortable for patients brought in ambulance, at anytime and anywhere, provided that they have a mobile terminal. Indeed, they could benefit from medical Tele-assistance and monitoring security, without the hospital care inconvenience and excessive expenditures.

III. Platform Description of Tele-assistance

The considered platform allows a person intended to the emergency (ambulance service) to be in contact, at any time, with his doctor to simplify the work in providing him means to be more effective in term of a rapid data transmission. So, the patient's problem is treated as soon as he reaches the emergency (Figure 1).

The purpose of this paper is to design and develop an Intelligent Mobile Assistant for medical Tele-assistance, allowing a transfer of medical information in real time between the various medical actors with mobility, security and reliability guarantees. Thus, an application of medical Tele-assistance allows the transferred medical file of a patient at home or in the emergency car before his admission to the emergency service.

The implemented application on an Intelligent Assistant (a mobile phone) of medical Tele-assistance functions on all mobile terminals or PDAs equipped with a KVM J2ME virtual machine. This algorithm allows not only PAN connections on limited perimeters (wireless sensors network), but also of WAN connections with medical actors. In this case, the transmitted medical file relates to the parameters, biological signals and medical information characterizing the current evacuated patients. The procedure consists of implementation and integration on a mobile device the following operations:

- Collection, treatment and organization of the transmitted measured parameters from wireless sensors network for the generation of a medical bulletin.
- Transfer in real time of the medical bulletin (all necessary information: ECG signal, Temperature, Tension…) to the emergency.
- Exploitation and activation of all transmission modes between the heterogeneous networks (Internet, GSM/GPRS…).

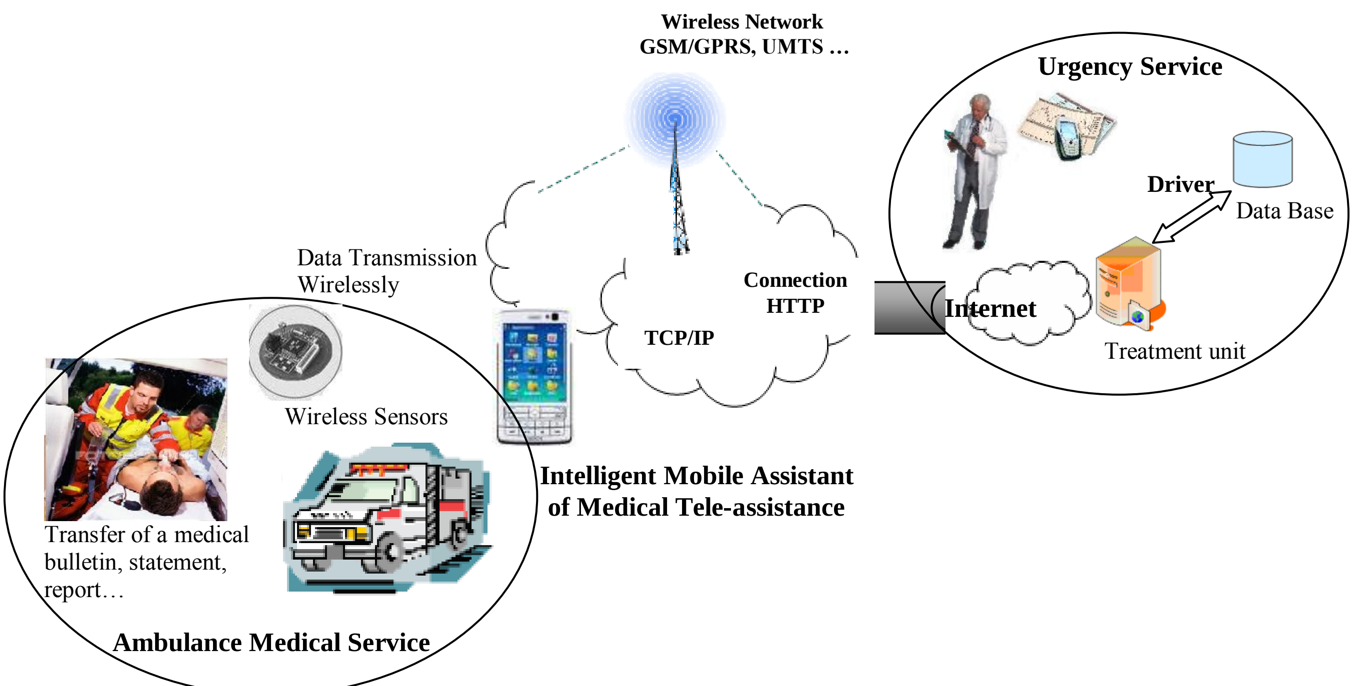


Figure 1. Architecture of the platform system.

The first part which must be realized, concerns the collection and automatic recuperation of remote medical data on the mobile terminal. These data are generated by sensors installed on person's body. The anticipation of the assistant intervenes for some information such as: Last Name, First Name…, before the immediate transfer of medical bulletin.

Then it is necessary to keep permanent interconnection of Intelligent Assistant of Tele-assistance which is in this case the portable phone with the sensors network and medical actors, so that they can exchange their data.

An adaptation strategy to medical context was followed to manage and gather the heterogeneity of medical data on the mobile terminal (detail in the following sections). Therefore, the patient will be evacuated at that moment and his reception will be done in the shortest period of time with the best conditions.

IV. Assessment of the Analysis

This part, most significant which lies in studying various technologies and protocols used in the world of wireless communications. The tackled subjects are related to the data exchange between the various parts of the designed platform including a mixture of networks.

The analysis carried out made it possible to a better understand of the main protocols which can intervene in the development of works of this article.

A. Selected technology

After studying the different technologies, in terms of exploitation of the data sent by a sensors network on a mobile phone using a specific algorithm to treat medical data and transmit the medical bulletin to the medical actors (heterogeneous devices). The most adapted solution relies on the usage of a simple mobile phone linked by two different systems, a wireless support WPAN (Bluetooth technology or ZigBee) and a GSM/GPRS system. Otherwise, these kinds of technologies are simple and rapid to implement.

This implies the exploitation of a pallet of integrated network protocols to establish connections that are described before, the mobile application, wireless medicals sensors networks and the doctor's tools. The choice of this pallet is dictated by the following characteristics: [3]

- *Protocols* in question are obligatorily implemented on all terminals MIDP (J2ME).
- *Technologies* must simple, effective and more easily installed.
- *A reduced cost* of the implementation, deployment and installation.

Concerning the devoted study to the transmission of medical information of the patient remotely, one learns that this transfer is based on the following communications:

1) Transmission between two mobile terminals: To transmit a data between two mobile phones, there is not an enormous existing possibilities. Indeed, it is possible to send SMS, MMS and e-mail. On these three modes of transmitted data, two are available only on the last generation phones; they are e-mail and MMS. Moreover, these two possibilities appear more sophisticated than the others [3], [4]. They allow sending all kind of electronic documents (text, photo, sound…).
For our project, the choice is related to the MMS service which has the following characteristics:

- *MMS Protocol is implemented in the optional packages of J2ME.*
- *A large range of use.*
- *An important content of multi-media which can be transmitted.*

2) Transmission between a mobile and a server

This implies the existence of an https connection (exchange of protected information) between intelligent assistant of Tele-assistance and the data base server via WAP gateway to transfer medical bulletin. This choice is dictated by the following characteristics:

- HTTPS is obligatorily implemented on all terminals MIDP (J2ME).
- HTTPS is independent of the network.
- The port of the https protocol is more easily working on the firewall.
- HTTPS protocol is implemented by default in J2ME package. Other protocols are not necessarily available [5].

This transfer is based on a communication WSP/https. As it suggests Wireless Session Protocol (WSP), layer session allows the connection setting to make transactions. Thus it allows the layer application to benefit from two different types of sessions:

- Connected session mode which the layer session will interact with the layer transaction.
- No-Connected session mode where the session layer will act directly at the transport layer for sending brutes' datagram.

WSP is equivalent to the https protocol, and we find moreover many identical implementations to https in WSP [6].

3) Transmission between Sensors Network and a mobile terminal: The sensors, which are placed on the patient body, use a wireless support of WPAN technologies: Bluetooth or ZigBee. They transmit on a short perimeters measured data of a patient through these kinds of technologies.

The part concerning the reception of these data on a mobile phone does not require a particular study; it's the research of library which allows such a handling of a sensor. It's on the level of the implemented application that we open a simple tunnel (Input Stream: for the reading, Output Stream: for the writing [7]) with a Buffer to recover and to store the transmitted data.

The particularity of sensors networks is located in the routing and economy of the network layer. The current protocols of routing use the metric (a number of hops, stability of the bonds) which inevitably do not optimize the energy of the nodes like that of the network and this by the use of some nodes more than others. Indeed, the protocols of routing with energy conservation must determine the optimal roads while being based on the metric related to the energy state of the nodes. In this context, many protocols are proposed.

For that, we proposed an improvement of the protocol DSR to include the aspect of energy economy in establishment of the roads. The original version of DSR does not take in consideration this aspect and chooses as road that having the minimum of hop which is not always effective in the sensors networks having constraints major of energy.

Our named proposal TMM-DSR (Taux Min-max Dynamic Source Routing) saves energy during the establishment of road, as in the remainder of the lifespan of the sensors because the metric used by TMM-DSR does not support the roads having the minimum of hop but uses each time the road having the best energy rate. What balances the use of the sensors for the routing and saves their battery and consequently gives a long lifespan for the network [8].

TMM-DSR preserves the nature of DSR as a reactive protocol, based on two operations: discovering and maintenance of road. To implement this technique each node must have information concerning energy level and the energy rate of its battery at anytime during the lifespan of network [9].

a- Technique of road selection (Max-Min)

In the conventional networks, the metric used is the number of hop which separates a source node from the destination. This metric is adapted to the wire networks but for wireless networks, the number of hop as parameters to evaluate a road is insufficient, following the imposed constraints by these networks such as the mobility of the nodes, the limitation of the band-width as well as the energy constraints.

New suggested metric is based on consumption rate of nodes' batteries in order to improve the power consumption of the network. It is called "Max-Min" technique.

b- Calculate rate of energy consumption

The rate of energy consumption or battery discharge can be defined as being effective energy (the remainder of energy) divided by maximum energy (initial).

$$T = (E_{ini} - e_c) / E_{ini}$$

$T \cong 1$: Low consumption rate.

$T \cong 0$: Very significant rate consumption.

e_c : Power consumption.

The implemented algorithm, based on the new metric, is responsible for the choice of the roads. This algorithm will proceed in the following way:

1- Each node when it receives a new request, will insert the rate pre-calculated on the request heading of discovered road until the arrival of each node to the recipient.
2- It waits a time D after the reception of each new request. Then, it determines the minimum of the rates of each received road:

$$T_k = Min(T_i)$$

i : The number of the nodes on the road K.

After the determination of the minimum rates of each road, the recipient will choose principal road that has the maximum rates.

$$T = Max(T_k)$$

The algorithm Max-Min does not give any guarantee on the time from the beginning to the end. To solve this problem, we have introduced a factor of rate energy differences that makes it possible to switch between the ways which have performance indexes very close in such manner to choose the shortest way then the smallest time from the beginning to the end. This parameter is given by the following relation:

If $[Min(T_j) - Min(T_i)] < \varepsilon$ and $hR_j > hR_i$. So to use the road R_i

B. Development environment

Java applications have been implemented under NetBeans IDE environment.

A simulation tool Sun Java ™ Wireless Toolkit was exploited to examine all the possible wireless communication. It allows applications on devices with low calculation resources such as a mobile phone.

The choice of Java is justified by the different problems associated to coding in C++ on Symbian operating system:

- Management of the memory: for the majority of applications, java system seems to be sufficient.
- Environment of execution: the proposed options on executable Java as protections for downloading or secure execution are free, whereas in C++, it is

necessary to develop them, test them and maintain them.

- Perpetuity: Java seems to have been accepted for the development of applications on mobile phone. The future developments will make Java perhaps as fast as C++.

Therefore, Java was chosen in our project, but it is necessary to mention that both environments can be used [7].

V. PROCESS OF THE IMPLEMENTATION OF PROPOSED MODEL

As mentioned before, our implementation achieves the medical service which provides the continuity and the constant of a remote medical Tele-assistance, the transfer immediate then the interactions in real time between the various medical actors for a rapid and effective taking off of patient, an immediate evacuation, a reception within shortest time, to consult the relative details and the automatic release of the treatments.

The schedule of conditions of our project consists of:

- The implementation of this service requires the development of two distinct applications:

- A first to be installed on the mobile phone (Intelligent assistant) to collect, taken measures and gather the received medical data of patient and the immediate sending of the medical bulletin to the medical actor via wireless support.
- A second function on the doctor's devices in order to receive, record and analyze the bulletin that is related to the patient.

- To program the application in a language which is most portable possible, the algorithm must be simple to use and install.

- To program a user interface of high quality.

The process of suggested model is also based on techniques and methods of programming adapted not only to the limited resources but also to the generated heterogeneous parameters. What allows in particular showing the diversity of persons' profiles and types of generated situations, including the simulation of "normal" modifications and disturbing of behavior. Moreover, it is possible to show, to organize the diversity of the data resulting from various sensors and these regroupings with the distinction, the re-assembly and recovery on the level of reception.

The following paragraphs present, (I) Development on mobile phone, (II) The process of proposed model.

A. *Development on mobile phone*

J2ME is a collection of technologies and specifications which are conceived for numerous parts of the market of the small devices [5]. The principal part of platform J2ME is composed of two different configurations (Figure 2): Connected Device Configuration (CDC) and Connected Limited Device Configuration (CLDC).

A configuration defines the central libraries of Java technology and virtual storage capacities of the device. CLDC is adapted to recent mobile phones. This configuration is useful for our application. To still note, that in the case of J2ME, the virtual machine is called KVM for Kilobyte Virtual Machine.

At the top of the configurations (Figure 2), there are the profiles which define the functionalities in each specific category of devices. The "Mobile Information Device Profile" (MIDP) is a profile for the mobile devices using configuration CLDC, like the mobile phones. Profile MIDP specifies the functionalities like the use of the interface user, the persistence of storage, the setting in network and the model of application.

On the majority of the current phones, J2ME is composed of configuration CLDC and profile MIDP.

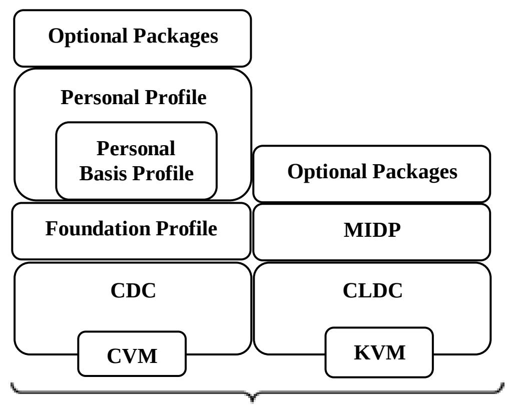


Figure 2. Architecture of J2ME.

In addition to standard MIDP (Figure 2), additional (optional) packages can be added according to the devices, allowing the use of their specificities.

As these options are typically reserved to mobile phones, it was natural to not integrate them directly in the profile MIDP.

As a result, the development of our application on mobile phones is based on the use of configuration CLDC and profile MIDP [10]. Besides to these two standard elements, we have exploited some optional packages such as WMA for the management of services SMS / MMS and Web Services API. The libraries necessary for the implementation for each component of J2ME are as follows:

API MIDP: is currently that which one finds on the compatible mobiles:

- javax.microedition.Icdui: For the graphic components necessary to the creation of applications.

- javax.microedition.midlet: It provides the component application as well as the primitives managing the life of the application.

- javax.microedition.rms: A possibility of storing information on the terminal.

API CLDC:

Javax.microedition.io: It contains the classes allowing connection via TCP/IP or UDP. The main class of this package is the class Connector.

This network part determines on which means is used to communicate medical information.

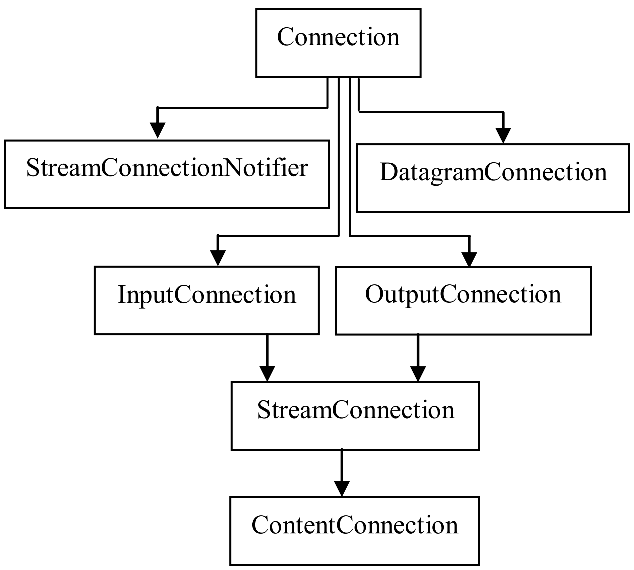


Figure 3. Arborescence of the classes javax.microedition.io [5].

The previous diagram corresponds to the different classes of javax.microedition.io. Thanks to these implementations for allowing us to make http requests.

This algorithm, called *MIDlet*, is carried out with the virtual machine J2ME (KVM) on the mobile terminal. It has the role to receive measurements of the sensors of the patient, to treat these data, to transmit or to store them if necessary. It also allows to the doctor to be able to evacuate the patient at that moment and his reception in shortest time with the best conditions.

B. Simulation of proposed model

The networks GSM/GPRS are useful to transmit information concerning the person taken a way at that moment.

Currently, mobile phones of last generations are able to send and receive all sorts of messages (text, image, sound…). They offer in addition to the voice communication, a supply of services on a large scale, which allows multi-applications for these devices.

Our investigation, thus consists to integrate and grouping in a medical assessment a whole of the heterogeneous data (ECG signal, parameters of ECG, Temperature, Tension…) on mobile telephone (Intelligent assistant) with a medical Tele-assistance service insurance of remotely person taken a way by the ambulance service.

1) *Cost of the developed project*

The cost of the health represents a considerable weight in the economic balance sheet on international scale. Also, in many countries, aging or psychological shocks tend to increase the number of people who are in an urgent need of medical monitoring even more or less intensive care, and based on that fact this affects the global cost of medical care.

As all the technologies, the mobile telephony is evolved, and actually the offered possibilities are more important than the last decade. But indeed the majority of users usually use only the basic functions, phoning and sending messages, what have already allowed imaged the multi-applications.

In our research paper, we propose the exploitation of the mobile phone in order applications apart from the vocal communications. The idea is to divert these devices of their basic function to make them useful for the telemedicine and the medical Tele-assistance.

Whereas ten years before, such a taking off would have required large means as well as a large infrastructure. Today a simple mobile phone can effectively contribute to the protection of human lives.

2) *Application*

This section describes a medical application which exploits the portable telephone for an application of Tele-assistance in urgent case. Our application depends on a development of the *MIDlet* to take the reception and the collection of medical information (measures taken, ECG…) of a patient remotely before his arrival at the hospital or on any other place.

The implementation of process of the proposed model for transmission simulating, storage and medical data processing is realized with the J2ME environment. It is always preferable to avoid carrying out specific applications to a type of mobile equipment owner (Windows, Symbian, Palm OS) for reasons that is related to the rapid evolution of technology. J2ME allows the development of applications which can be executed on all compatible mobiles [11]. It brings to the portable systems the power and the modularity of the programming JAVA and this in a way is adapted to the characteristics of the embarked terminals.

Concerning the study of mobility (the key factor) as part of this work is bound by the variations which can be observed in a mobile environment. They are low-level phenomena. These phenomena directly influence the protocol layers used by the applications. To make transparent the movements and disconnections, it is therefore necessary to adapt the protocols for mobile environments. IP addressing, TCP, NFS and HTTP are the main protocols that have been the subject of research in this direction.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

The following paragraphs present the principle of the implementation, the global structure of the implementation of process of simulation and finally the sending/the reception of medical assessment.

a) The principle of implementation

Being given the complexity between the portable telephone's technology, the number of parameters of the proposed model which must be defined in priori and the heterogeneity of the communicating systems, an adaptation of data to this context consists to use a set of medicals files in the format text for the definition of the current exploited parameters for the simulation and the default information. The adapted algorithm undertakes to recover, to collect and analyze the associated flow to the transmitted medical files through the radio interface between wireless sensors network and telephone of Tele-assistance. It implements an intermediate graphic interface on intelligent assistant allowing the display, the transfer in a sufficient time; not only the contained parameters in the files and measurements taken, but also biological signals obtained (ECG…).

After execution of the application, one can then recover the values and the information in files. At the end of the application, the generated medical bulletin (measurements, medical signals obtained, information…) are remotely stored on a data base of hospital, a doctor's mobile terminal or doctor's personal assistant and possibly displayed on his screen.

Figure 4 presents a diagram synthesizing the general principle of the process implementation of simulation:

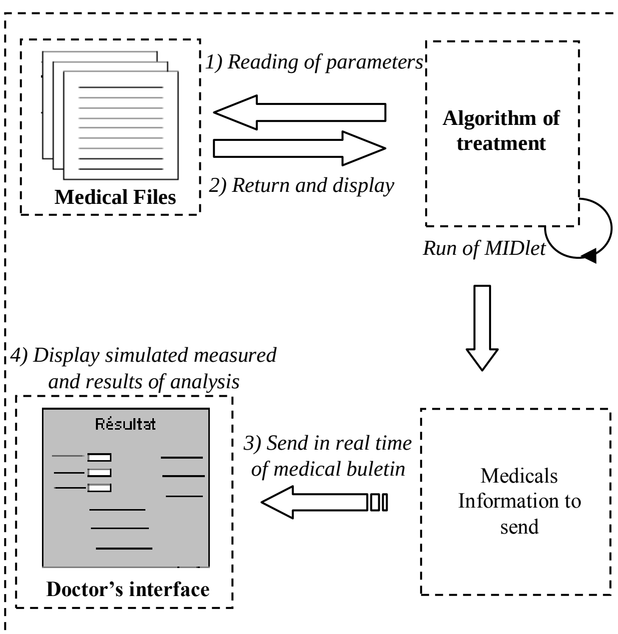


Figure 4. Principle of the implementation of the process of simulation.

The result interface is presented on Figures 5, 6 and Figure 7. The medical signals and the parameters are generated according to the simulation model sequences on which they involve. They are integrated into the medical bulletin and possibly sent in a possible time to the hospital's data base via an intermediate Servlet and on the doctor's telephone portable.

b) Global structure of the implementation of simulation

The global structure of the program of the assistant of Tele-assistance is completely sequential. It calls one after the other, the functions realizing the principle stages of treatment and successively corresponding to the generation of the distinct medical signals. These signals and these parameters respectively relate to the ECG and measurements taken. Each called function of treatment takes as entering parameters the results of the call of the previous function and provides the results of its execution to the following function [12].

In order to optimize the storage memory on the assistant of Tele-assistance (which is in this case the portable telephone), we exploited vector of the fixed dimensions to fill them during periods then will be affected to others parameters or values so on. The stages of the *MIDlet* execution are detailed in the following section.

c) Result

As we have already seen before, j2me wireless development was exploited to implement the proposed model of mobile Tele-assistance. In this section, we present and discuss the various stages of execution of the algorithm. All this series of tests was made thanks to the phone emulator.

During the launching of the application, *MIDlet* allows to the assistant to activate the mode of the medical Tele-assistance.

At the beginning, the application operates and communicates in autonomous mode with patient's measurement sensors network. Then, the phone collects and recovers periodically the measurement samples, generated by these wireless sensors of patient (Figure 5, 6). These data are stored in a sequential way in a whole of the tables (several boxes). Part of these data will be treated by this algorithm to generate the desired signals and consequently to integrate them in the medical bulletin to be sent. The remainder of these data will be to organize and gather in a filled form.

The unit carries out the medical assessment of the person taken a way by emergency car.

Figure 5. Medical assessment.

This interface of the intelligent assistant of Tele-assistance represents the organization of the medical assessment (different parameters) on a form automatically filled by various measurement sensors of person transferred in the ambulance.

NB: For some information concerning the Last name, First name…, requires the intervention of the assistant at the time of the sending of the Medical Assessment.

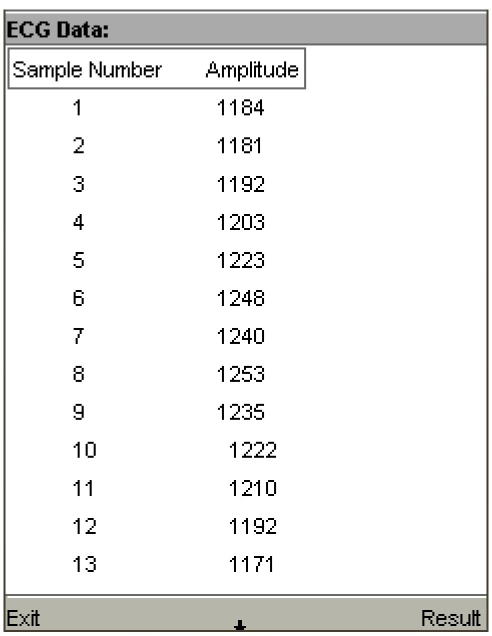


Figure 6. ECG Data transmitted to the portable telephone.

Figure 6 shows the organization in vectors (10000 samples) of ECG values transferred via wireless PAN technologies (Bluetooth, ZigBee…) to the internal memory of the phone.

Such a medical application proposes a set of services to the professional users (List of the patients, the display of the medical profile of a patient, Digital processing of the signals…). These services make treatments with variable complexities (management of data via a wireless sensors network set by the person or a data base, numerical calculation…) and exchange data with the user through a graphical interface on doctor's display device (Mobile phone, Server with data base…). This environment type presents important heterogeneous information, a great variability and numerous possibility of evolution, as well on the level of the means of execution as the means of communication. Indeed, the offered resources on the level of terminal can be extremely different according to the use of a personal assistant, a laptop or a workstation. Thus it is necessary to implement an adaptation strategy to conceive and develop the algorithm by respecting these required constraints.

The intermediate graphic interface of the assistant of Tele-assistance which is in this case the screen of mobile phone, allows to collect, display, store, calculate (treat) and to transmit the bulletin in a possible time to the hospital or clinic, after an adaptation of medical data to the context.

Our adaptation strategy leans on to exploit a vector of storage of medical files, of fixed dimension relating to the size memory available on the mobile phone (32 KB). This contained is sequentially transmitted by sensors network what leads with time to increase the space of storage on the intelligent assistant's terminal.

For this reason it is necessary to adapt these measurements to the mobile context. For example, in the case of data ECG, this adaptation consists to recover and treat regularly each 10000 samples in a vector to be periodically erased, in order to optimize the memory for the storage of the other parameters.

The medical data recovered on wireless Tele-assistance tool, are exploited to generate the medical signals, or are organized in specific fields before sending them. Thus, the doctor is invited to consult the medical assessment of his patient (Figure 7) in order to take the adapted decision and makes it possible to re-examine the files

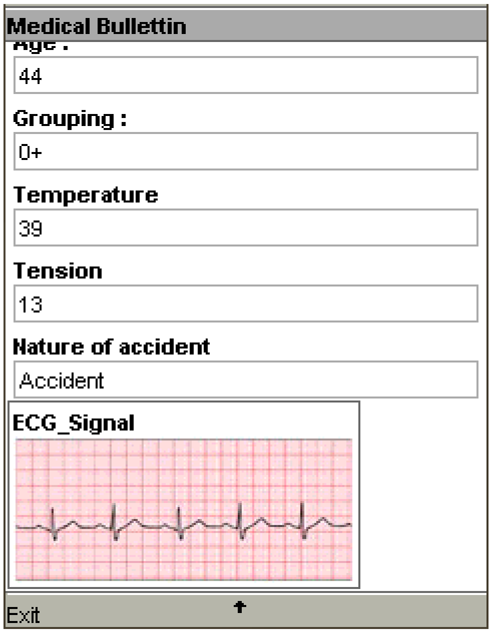


Figure 7. The received medical assessment.

The transfer and the complete treatment could be done by using the implementation of the calculation algorithm on the mobile. It makes it possible to calculate the most significant parameters necessary to the characterization and the accuracy and continuous location of all the waves characteristic of a signal ECG (Figure 7, 8).

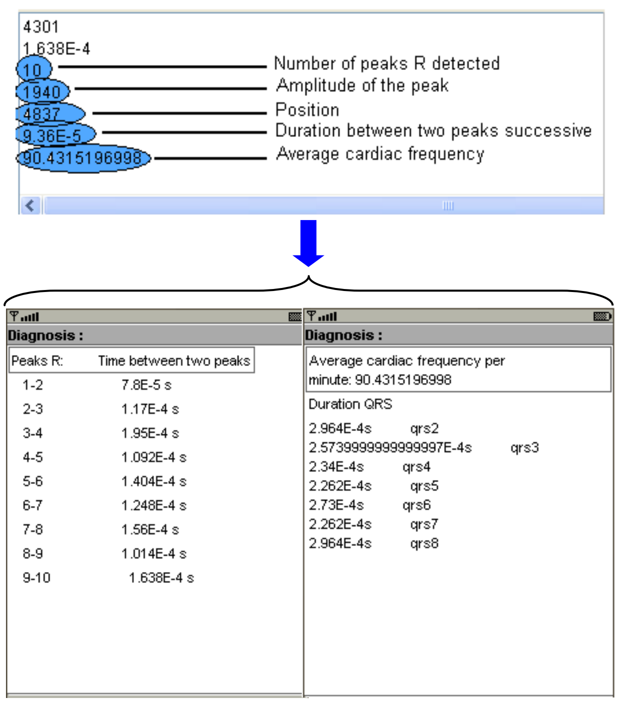


Figure 8. Calculation of the significant parameters of the ECG.

Thus, the doctor can observe the taken measurements and the ECG signal in real time on his screen. A more importantly option allowing the zoom of the part which presents an ambiguity and an anomaly, is implemented in our application.

It is enough to the doctor to introduce a start and end point (time interval) in order to widen the part in question.

VI Evaluation

The proposed simulation algorithm is articulated on two fundamental points. The first one relates to the simulation step in the respect of the complexity and objectives of the medical Tele-assistance context at distance. The second links to the global vision in the cycle of resolution of the problems: time of transport (to make travel information rather than the patients), the construction of the behavior profile of person to ensure a critical situation, immediate evacuation of the patient and his reception in the most shortest time and with the best conditions.

The suggested Tele-assistance consists to monitor and diagnose the state of an evacuated patient using the methodology developed in this project. Thus, the doctor treating a person at the moment of his injury can at any time receive the medical information and control the state of his patient by consulting in a real time the transferred medical signals on his device (Mobile terminal, Personal assistant, Server…) and the measurements taken by the sensors networks.

The development times of the achieved functionality using J2ME were considerably reduced. For this, time measurements carried out on the principal operations of a Teleassistance system such as, the collection and the regrouping of data, the treatment and the transfer of medical bulletin represent a shortest time and effective to contribute in the protection of human lives. Tests were carried out on a Smartphone: Nokia Series N, N96, memory of stockage intern 16 Go, Processor CPU-ARM9, given rhythm of 264 MHz with 128 MB of RAM.

Relative to the capability of setting near real-time transmission of the data received from medical sensors and transmit them (with treatment) to the emergency, one remarkable design decision has been the evolution from a traditional synchronous request/response mode of HTTP to pipelining request mode using persistent TCP connections.

VII Conclusion

This article refers to a collection, an organization, a treatment of taken measurements and to a construction of biological signals of the evacuated person, transmitted by the sensors on portable telephone then to send them remotely in a good time to the doctor. This technique of medical Tele-assistance allows an evacuated of an injured person remotely via the ambulance medical service such as the cardiac, the hypertensive patients…

Also, the stages of identification, by the developed algorithm, of the medical profile of a remote patient and the construction of the medical signals cannot cover all medical indicators corresponding to each patient. Thus, the improvements of this algorithm must be added as one is interested to the new profile (parameters, medical signals…).

This solution is adapted to the portable devices ensuring medical Tele-assistance anytime and anywhere and above all it is cheap and easily realizable. It is in this vision that other services, associated to mobiles and intended for the telemedicine and the house automation are under development.

ACKNOWLEDGMENT

I deeply thank my supervisor and the person in charge of the laboratory STIC, Mohamed FEHAM, Professor at the university Abou Bekr Belkaid of Tlemcen, Algeria. The correctness of his advice, the motivation and the project financing, were very precious and brought a successful conclusion to this work. A special thank to the researchers of the laboratory who gave their help to the realization of this project.

REFERENCES

[1] O. Fouial, & I. Demeure, " Fourniture de services adaptables dans les environnements mobiles ". Proceedings of European Conference of Systems with adaptable and extensible components. Grenoble, France, 2004.

[2] A. Nemo, "La télémédecine : Faire voyager les informations plutôt que le malade ", Journal du Téléphone, Grenoble pp. 4, 1994.

[3] K. Juha, "Introduction to 3G Mobile Communications", 2nd edition, Artech House Publishers Norwood USA, 2003.

[4] D. HARDY, G. MALLEUS, & J_N. MEREUE, "Réseau Internet, Téléphonie, Multimédia ", 2003.

[5] B. Delb, "J2ME, Application java pour terminaux mobiles", EYROLLES, Paris, France, 2002.

[6] G. PUJOLLE, " Les Réseaux ", Editions EYROLLES, 2000.

[7] A. F. Quintas, "Bluetooth J2ME Java 2 micro edition", manual de usuario y tutorial, Madrid: Ra-Ma, 2004.

[8] M. Tommaso, P. Dario, F. A. Ian, "Optimal Local Topology Knowledge for Energy Efficient Geographical Routing in Sensor Networks", INFOCOM, Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, page(s): 1705-1716 vol.3, 2004.

[9] B. BOUYEDDOU, "Implémentation d'un protocole d'économie d'énergie EMM-DSR pour les réseaux ad hoc 802.11", These of Magister, University of Tlemcen, Algeria, June 2007.

[10] J. Knudsen, "Wireless Java Developing with J2ME", Second Edition, Apress, Berkeley, United States, 2003.

[11] H. Mahmoud, "Learning Wireless Java", O'Reilly, Sebastopol, USA, 2002.

[12] R. Merzougui, M. Feham, "Algorithm of remote monitoring ECG using mobile phone: Conception and implementation", Proc. Third International Conference on Broadband Communications, Information Technology & Biomedical Applications, 2008.

AUTHORS PROFILE

Rachid. Merzougui received the Master degree in Systems and Networks of Telecommunications from the University of Tlemcen (Algeria) in 2006. Since this year, he has been Assistant Professor of Mobile Networks and Services. He has served on the scientific committees of the Telecommunication Department of the University of Tlemcen. He is interested now in mobile services.

Review of Load Balancing in Cloud Computing

Suriya Begum
Research Scholar.
Visvesvaraya Technical University
Belgaum, India

Dr. Prashanth C.S.R
Prof. and Head of Department
Dept. of Computer Science and Engineering
New Horizon College of Engineering
Bangalore, India

Abstract— **With the exponential rise in the demands of the clients worldwide, a large scale distributed systems have been introduced as a computing environment. Cloud computing has paved a revolutionary path in this direction of distributed environment for accomplishing optimized performance, shortest response time, network resource utilization, and adaptability of service level agreement. Cloud computing has multiple benefits as well as it is also accompanied with certain serious technical loopholes. The proposed paper has focused on one such issue of load balancing. The consequences of inefficient load balancing may lead to detritions of an organization business performance on cloud environment. Hence this paper illustrates various aspects pertaining to domain of cloud computing, its evolution, its generic issues, and particularly to issues related to load balancing. Various techniques adopted in the past research work have been analyzed and the findings were illustrated in this paper.**

Keywords-; Cloud computing, Grid Computing, Load balancing

I. INTRODUCTION

Cloud computing can be illustrated as the operation of heavy computing resources that includes hardware and software package that are delivered to clients as a service over an outsized scale network [1]. The name „cloud" has been originated from the employment of a cloud-shaped representation as associate abstraction for the advanced infrastructure that is contained in a very system diagrams. The origination of the term „cloud computing" is very vague, however it appears to be derived from the practice of using schematic diagrams of computing and communications systems. The word cloud is employed as a allegory for the big scale network that is supported by the standardized use of a cloud-like form to denote a network on telephone schematics associated later to depict the web in network diagrams as an abstraction of the underlying infrastructure it represents. The cloud image [2] was used to represent the web as early as 1994. The elemental thought of cloud computing essentially dates back to the Nineteen Fifties once there was associate availableness of huge scale mainframe in establishments and companies. Owing to the pricey nature of mainframe, there arise a necessity to search out another answer for allowing the multiple users for accessing and sharing equivalent central processor time thereby truncating the chance of periods of inactivity (also termed as time-sharing) [3]. As computers became additionally prevailing, scientists and technologists explored ways to create large-scale computing power on the market to cater additional users through sharing, experimenting with algorithms to produce the best use of the infrastructure, platform and applications with prioritized access to the computer hardware and potency for the tip users [4]. The high value of those powerful computing systems has forced several prime organizations to require Associate in Nursing initiative to explore higher value effective answer exploitation sharing. The prime organization can embody IBM, GE, National CSS etc who took the initiative to launch and marketed time sharing. With the ascension of net technology and standards, varied merchandise and demands of distributed computing were on high increase. The presence of pervasive high computing network, value effective computing devices, storage devices together with massive scale use of virtualization of hardware, service headed design has paved the trail for top demands in new technology, therefore known as „cloud computing."

The domain of cloud computing is still surfaced by many issues which will be discussed in this paper in later section. The prime focus of the paper will be to analyze the research issues in load balancing protocols or understanding its requirement in cloud platform. Since, with the scene of rapid use of cloud computing resources, the demand along with provisioning of the cloud resources has to be effectively design to claim better SLA (Service Level Agreement) with zero downtime claims. Currently, there is a presence of multiple vendors offering cloud services (Amazon, Microsoft, IBM, Google, Salesforce, HP, Oracle, Citrix, EMC etc.) and there are growing numbers of clientele too. Hence, it is quite obvious that catering the massive and dynamic requirements of such exponentially growing clients will become one of the most challenging issues. And to mitigate this issue, an effective load balancing technique should be explored. The proposed paper will introduce a thorough analysis of the evolution of cloud platform right from the origination of the initial distributed computing system. The paper will mainly focus on the research issues of load balancing and will attempt to analyze the prior work done in this field. Section 2 will highlight about the evolution of cloud computing followed by Section 3 for Issues in cloud platform. Section 4 will discuss about the load balancing issues along with the review of the past research work focused on load balancing in cloud platform. Section 5 will discuss about the research issues finally followed by concluding remarks in Section 6

II. Evolution of Cloud Computing

A distributed computing system enables the sharing, selection, and aggregation of distributed heterogeneous computational and storage resources, which are under the control of different sites or domains. The Idea-phase of cloud computing started in 1960 followed by pre-Cloud phase in and around 1999-2006 and final cloud phase was seen on 2007. The evolution of cloud is discussed in this paper from cluster computing, grid computing and then cloud computing analyzing the merits and demerits of every constituent.

A. Cluster computing

The concept of clustering in computer system is the use of multiple computers, typically PCs or UNIX workstations, multiple storage devices, and redundant interconnections, to form what appears to users as a single highly available system [5]. Cluster computing can be used for load balancing as well as for high availability. Advocates of clustering suggest that the approach can help an enterprise achieve 99.99% availability in some cases. One of the main ideas of cluster computing is that, to the outside world, the cluster appears to be a single system. The standard cluster architecture as defined in the work of Buyya [5] is as shown below:

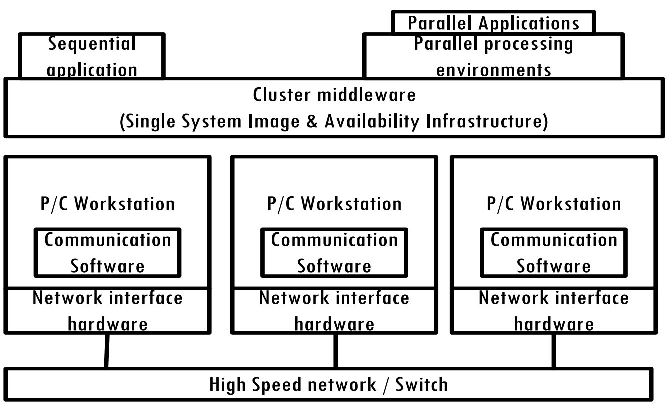


Figure 1 Cluster Computing Architecture

A frequent use of cluster computing is to load balance traffic on high-traffic internet sites. An internet page request is forwarded to a "manager" server, that then determines that of many identical or extremely similar internet servers to forward the request to for handling. Having an internet farm (as such a configuration is usually called) permits traffic to be handled additional quickly. clump has been offered since the Nineteen Eighties once it absolutely was employed in DEC's VMS systems [6]. Cluster computing may also be used as a comparatively inexpensive kind of multiprocessing for scientific and different applications that lend themselves to parallel operations. associate early and well-known example was the Beowulf project [7] within which variety of off-the-peg PCs were wont to kind a cluster for scientific applications

- *Fail-over Cluster Computing*: In this type of cluster computing, the machines‟ work is continuously monitored and when one of the two host stops working the other machine takes over. The aim is to ensure a continuous service.
- *Cluster with load balancing*: In this type of cluster computing, the work requests are sent to the machine with fewer loads.
- *HPC Cluster Computing*: In this type of cluster computing, the computers are configured to provide extremely high performance. The machines break down the processes of a job on multiple machines in order to gain in performance.

The main precincts of Cluster Computing are:

- Complicated to manage and organize a large number of computers
- Poor performance in the case of non-parallelizable applications.
- Physical space requirement is considerably greater than that of a single server
- Maximized energy consumption compared to a single server

B. Grid Computing

Grid computing permits virtual organizations to share geographically distributed resources as they follow universal goals, forward the absence of central location, central management, state, associate degreed an existing trust relationship [8]. Grid computing may be utilized in a spread of how to handle varied styles of application necessities. Often, grids square measure categorized by the sort of solutions that they best address. The 3 primary styles of grids square measure summarized below [9]. Of course, there aren't any hard boundaries between these grid varieties and infrequently grids could also be a mixture of 2 or a more of those. However, as one think about developing applications which will run in an exceedingly grid atmosphere, keep in mind that the sort of grid atmosphere that one can simply are exploitation can have an effect on several of one‟s selections.

- Computational grid: A computational grid is focused on setting aside resources specifically for computing power. In this type of grid, most of the machines are high-performance servers.
- Scavenging grid: A scavenging grid is most commonly used with large numbers of desktop machines. Machines are scavenged for available CPU cycles and other resources. Owners of the desktop machines are usually given control over when their resources are available to participate in the grid.
- Data grid: A data grid is responsible for housing and providing access to data across multiple organizations. Users are not concerned with where this data is located as long as they have access to the data. For example, you may have two universities doing life science research, each with unique data. A data grid would allow them to share

their data, manage the data, and manage security issues such as who has access to what data.

A Computational grid states that it is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities. Whereas, [10] Desktop grid refers to a grid infrastructure that is confined to an institutional boundary, where the spare processing capacity of an enterprise‟s desktop PCs are used to support the execution of the enterprise‟s applications.

Table 1: Key components of Computational and Desktop Grids

KEY COMPONENTS	
Computational Grids	Desktop Grids
Infrastructure	Physical node Management
Dependable service	Resource Scheduling
Consistency of service	Job scheduling
Pervasive access	

Computational and Desktop Grids can be classified on the basis on Infrastructure, Models and Software applications and platforms etc. some of them are given in Table 2.

Table 2: Classifications of Computational and Desktop Grids

CLASSIFICATIONS	
Computational Grids	Desktop Grids
gLite	Distributed.net
NorduGrid/ARC	Entropia
UNICORE	SETI@home
MiG	Bayanihan
WebCom-G	Condor
Office Grid	BOINC

However, the concept and utilization of grid computing is also explored with multiple limitations:

- For the application that cannot properly use Message Passing Interface (MPI), the user may by bound to work on large symmetric multiprocessor (SMP).
- The network connection requirement is quite higher (minimum GB Ethernet)
- Some specific application running on grid environment may need to be tweaked in order to completely utilize the new model.
- Licensing in grid across multiple servers will make it eventually resistive operationally for certain application running on grid environment.
- Although Grid environments offer the privilege of many smaller servers across various administrative domain, but it becomes tremendously challenging task to manage efficient tools for change and controlling server configuration in proper synchronization.
- Standard and benchmarked resource sharing policy with better SLA and resource provisioning is still missing

C. Cloud Computing

A Cloud is a form of parallel and distributed system possessing a group of inter-connected and virtualized computers that are dynamically scheduled and highlighted as one or more unified computing resources based on service-level agreements established through conciliation between the service provider and consumers. [11]. The cloud service model is shown below.

Software-as-a-Service (SaaS)
Platform-as-a-Service (PaaS)
Infrastructure-as-a-Service (IaaS)
Hardware-as-a-Service (HaaS)

Figure 2 Cloud Service Models

Cloud schematic design is that the blueprint of the software package applications that uses internet-accessible on demand service. Cloud architectures are underlying on infrastructure that is used only when it is required that draw the necessary resources on-demand and perform a specific job, then relinquish the unneeded resources and often dispose them after the job is done. SaaS is a multi-tenant platform that uses common resources and a single instance of both the object code of an application as well as the underlying database to support multiple customers simultaneously. The prime global players of SaaS are SalesForce, IBM, Microsoft, and Oracle. PaaS provides developers with platform including all the systems and environments comprising the end-to-end life cycle of developing, testing, deploying, and hosting of complex web applications. Microsoft Azure is one example of PaaS. Iaas Is the delivery of computer infrastructure as a service. The usage based payment factor is one key benefit of PaaS. GoGrid, Flexiscale, Rackspace etc are to name a few examples for PaaS.

Table 3 Cloud Deployment models

CLOUD COMPUTING DEPLOYMENT MODELS	
Private Cloud	The cloud infrastructure is owned or leased by a single organization and is operated solely for that organization.
Community Cloud	Several organizations that have similar polices, objectives, aims and concerns share the cloud infrastructure.
Public Cloud	A large organization owns the cloud infrastructure and sells cloud services to industries or public.
Hybrid	It is combination of two or more clouds. It

Cloud	enables data and application probability.

Each deployment model is either internal or external. Internal clouds are covered under organizations security policy but external clouds are not. Cloud computing offers lots of advantages [12]:

- *Cost*- As in the clouds the user need not own the resources, it just need to pay as per the usage in terms of time, storage and services. This feature reduces the cost of owning the infrastructure [13], [14].
- *Performance*-the performance is improved because the cloud is not a single computer but a large network of powerful computers resulting in high processing power [15], [11], [16].
- *Freedom from up gradation and maintenance*- the cloud infrastructure is maintained and upgraded by the cloud service provider [14], [11].
- *Scalability*- The user is can request to increase the resources if the area of application grows or new functionality is added. On the other hand if requirement shrinks the user can request to reduce the resources as well [11], [15].
- *Speedy Implementation*- Time of Implementation of cloud for an application may be in days or sometimes in hours. You just need a valid credit card and need to fulfill some online registration formalities [16].
- *Green*- The cloud computing is a green technology since it enable resource sharing among users thus not requiring large data centers that consumes a lot of power [16].
- *Mobility*- We don‟t need to carry our personal computer, because we can access our documents anytime anywhere [11].
- *Maximized Storage Capacity*- In Cloud computing we have extreme resources for storing data because our storage consists of many bases in the Cloud. Another thing about storing data in the Cloud is that, because of our data in the Cloud can automatically duplicated, they will be more safety [11].

Along with list of advantages, various constraints found in the usage of cloud computing are as follows:

- When the applications, processes and data are tightly coupled or interdependent.
- When there are not well defined points to share the data, process and behavior within an application.
- When the application require a very high level of security.
- When you want total control on your processes and data and thus cannot outsource your application or its critical components.
- When the core internal architecture of the organization is not functioning well, then first make it strong so that it can be easily mapped to cloud architecture.
- When one need native APIs, since the cloud does not provide native APIs.
- When someone is already using a legacy system, since older systems posses number of difficulties to move to cloud architecture.

Table 4 Evidence of Cloud Service issues

Service Issue	Duration	Date
Microsoft Azure: Malfunction in Windows Azure [17]	22 Hrs	March, 2008
Gmail and Google Apps Engine [18]	2.5 Hrs	July, 2009
Google search outage: Programming error [19]	40 min	Jan, 2009
Gmail: Site unavailable due to outage in contacts system [20]	1.5 Hrs	Aug, 2008
Google AppEngine partial outage programming Error [21]	5 Hrs	June, 2008
S3 outage: authentication service overload leading to unavailability [22]	2 Hrs	Feb, 2008
FlexiScale: core network failure [23]	18 Hrs	Oct, 2008
Salesforce.com: zero connectivity due to server disruption stopping all data from Japan, Europe, & North America [24]	1-2 Hrs	Jan 2009
Microsoft: zero connectivity for design issue [25]	2 Hrs	Sep-2010
Skype: network outage due to unstable and overloaded clusters [26]	1 day	Dec-2010
Amazon Web Service: zero connectivity due to incorrect traffic shift [27]	4 days	April-2011
Amazon EC2 Cloud: Transformer exploded, caught fire in datacenter and caused connection outage [28]	2 days	Aug, 2011
Amazon: Connectivity Issues in EC2 cloud computing service [29]	3 Hrs	June, 2012

The above table represents only a few evidence of the connection outage issue from some of the reputed Cloud service providers.

III. ISSUE IN CLOUD PLATFORM

Apart from the issues specified in the previous section, the other set of the technical issues in cloud computing will include load balancing, security, reliability, ownership, data back-up, data portability, multiplatform support, and intellectual property. Here is a rundown on most of the current issues concerning cloud computing:

- *Load balancing* [30]: Load leveling is usually mechanized to implement failover-the continuance of a service when the failure of 1 or additional of its parts. The components are monitored frequently and once one becomes non-responsive, the load balancer is up on and now not sends traffic to it. This can be an inherent feature from grid-based computing for cloud-based platforms. Energy conservation and resource consumption don't seem to be continuously attentiveness once discussing cloud computing; but with correct load leveling in place of resource consumption are often unbroken to a minimum. This is not only serves to keep cost low and enterprise „greener", it also puts less stress on the circuits of each individual design making them more potentially last longer.

- *Security* [31]: While a leading edge cloud services provider will employ data storage and transmission encryption, user authentication, and authorization (data access) practices, many people worry about the vulnerability of remote data to such criminals as hackers, thieves, and disgruntled employees. Cloud providers are enormously sensitive to this issue and apply substantial resources to mitigating concern.

- *Reliability* [32]: Some people worry also about whether a cloud service provider is financially stable and whether their data storage system is trustworthy. Most cloud providers attempt to mollify this concern by using redundant storage techniques, but it is still possible that a service could crash or go out of business, leaving users with limited or no access to their data. A diversification of providers can help alleviate this concern, albeit at a higher cost.

- *Ownership* [33]: Once data has been relegated to the cloud, some people worry that they could lose some or all of their rights or be unable to protect the rights of their customers. Many cloud providers are addressing this issue with well-crafted user-sided agreements. That said, users would be wise to seek advice from their favorite legal representative. Never use a provider who, in their terms of service, lays any kind of ownership claim over your data.

- *Data Backup* [34]: Cloud providers employ redundant servers and routine data backup processes, but some people worry about being able to control their own backups. Many providers are now offering data dumps onto media or allowing users to back up data through regular downloads.

- *Data Portability and Conversion* [35]: Some people are concerned that, should they wish to switch providers, they may have difficulty transferring data. Porting and converting data is highly dependent on the nature of the cloud provider‟s data retrieval format, particular in cases where the format cannot be easily discovered. As service competition grows and open standards become established, the data portability issue will ease, and conversion processes will become available supporting the more popular cloud providers. Worst case, a cloud subscriber will have to pay for some custom data conversion.

- *Multiplatform Support* [36]: More an issue for IT departments using managed services is how the cloud-based service integrates across different platforms and operating systems, e.g. OS X, Windows, Linux and thin-clients. Usually, some customized adaption of the service takes care of any problem. Multiplatform support requirements will ease as more user interfaces become web-based.

- *Intellectual Property* [37]: A company invents something new and it uses cloud services as part of the invention. Is the invention still patentable? Does the cloud provider have any claim on the invention? Can they provide similar services to competitors? All good questions and answerable on a case-by-case basis.

Once someone understands that cloud computing potentially suffers from much of the same fate as proprietary systems, the question becomes "do the advantages of using the cloud outweigh my concerns?" For low-risk operations and for insensitive information, the answer can easily be "yes." Realize that cloud-based services can be backed-up, verified, double-checked, and made more secure by combining them with traditional non-cloud IT processes.

IV. LOAD BALANCING IN CLOUD PLATFORM

Load balancing [38] is a process of reassigning the total load to the individual nodes of the collective system to make resource utilization effective and to improve the response time of the job, simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded. A load balancing protocol is dynamic in nature doesn't contemplate the previous state or behavior of the system, that is, it depends on the current behavior of the system. It is common these days in redundant high-availability computer systems that incoming network traffic is distributed on network level by deploying one of the frequently used network load balancing algorithms like:- random-allocation, round-robin allocation, weighted round-robin allocation, etc). These algorithms use solely network parameters of incoming traffic to create selections wherever to forward traffic, with none data from different elements of database system, like current load of application or info servers. Since these days it is extremely common to possess internet servers acting as application servers, it is usual that load balancers use session-switching technique, which suggests that once a user opens website on one server, it will stay on it server whereas the session lasts.

Depending on who initiated the process, load balancing algorithms can be of five categories:

- Sender Initiated: If the load balancing algorithm is initialized by the sender
- Receiver Initiated: If the load balancing algorithm is initiated by the receiver
- Symmetric: It is the combination of both sender initiated and receiver initiated
- Static: It doesn‟t depend on the current state of the system. Prior knowledge of the system is needed.
- Dynamic: Decisions on load balancing are based on current state of the system. No prior knowledge is needed. So it is better than static approach.

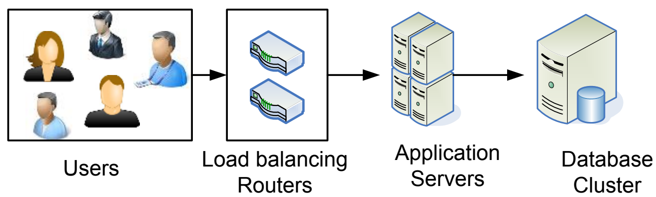


Figure 3 Schematics of typical high-availability computer system with hardware load balancers.

Table 5 Metrics in existing LB techniques in cloud computing

LOAD BALANCING METRICS	
Metric	Illustration
Throughput	It is used to calculate the no. of tasks whose execution has been completed. It should be high to improve the performance of the system
Overhead	It determines the amount of overhead involved while implementing a load-balancing algorithm. It is composed of overhead due to movement of tasks, inter-processor and inter-process communication. This should be minimized so that a load balancing technique can work efficiently.
Fault Tolerance	It is the time to migrate the jobs or resources from one node to other. It should be minimized in order to enhance the performance of the system.
Response Time	It is the amount of time taken to respond by a particular load balancing algorithm in a distributed system. This parameter should be minimized.
Resource Utilization	It is used to check the utilization of re-sources. It should be optimized for an efficient load balancing.
Scalability	It is the ability of an algorithm to perform load balancing for a system with any finite number of nodes. This metric should be improved.
Performance	It is used to check the efficiency of the system. This has to be improved at a reasonable cost, e.g., reduce task response time while keeping acceptable delays

Central to many other issues lies the establishment of an effective load balancing algorithm. The load can be CPU load, memory capacity, delay or network load. Load balancing is the process of distributing the load among various nodes of a distributed system to improve both resource utilization and job response time while also avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or doing very little work. Load balancing ensures that all the processor in the system or every node in the network does approximately the equal amount of work at any instant of time. This technique can be sender initiated, receiver initiated or symmetric type (combination of sender initiated and receiver initiated types).

V. EXISTING TECHNIQUES IN LOAD BALANCING

This review aims at summarizing the current state of the art of existing load balancing techniques in cloud computing. Here inspite of quantity of work done, the focus in given to only names of distinctive techniques used to mitigate load balancing issue in cloud computing), (load balancing techniques in cloud computing), (load balancing in clouds) and (load balancing in datacenters). Only papers written in English were included. Following load balancing techniques are currently prevalent in clouds:

A. Technique-1: Event-Driven

V. Nae et al. [39] bestowed event-driven load balancing algorithmic for real time Massively Multiplayer on-line Games (MMOG) that is characterized by a new type of large-scale distributed applications with a real-time virtual world entertaining massive volumes of players in the network. In order to cater up the variable computational and latency-aware resource demands, the MMOG machinist over-provision an own multi-server infrastructure with adequate competence for assuring the Quality of Service (QoS) requirements and a smooth game play at all times. This statically scheduled infrastructure has two prime challenges: it has maximum functional costs and is susceptible to capacity shortages in case of unanticipated increases in demand. In contrast to uniform provisioning, the new cloud computing technology based on resource virtualization has the potential to provide an on-demand infrastructure for MMOGs, where resources are provisioned and paid for only when they are actually needed. Conversely, this technology can introduce virtualization overheads which may cancel out the benefits. In prior work, the authors have studied the outcomes of deploying virtualized resources regarding the technology-incurred overheads and their impact on the QoS offered to the clients and the cost-effective results, while considering ideal resources in terms of accessibility. It consists of three modules with unique and distinct roles, e.g.1) *Client*: it is the module that connects to the game operator‟s sessions. It can join the sessions offered by game machinist on the basis of its subscriptions. A subscription represents a contract between a client and a game machinist based on which the client is allowed, under certain terms and with certain QoS guarantees, to join a session managed by the game machinist. 2) *Machinist*: This module provisions resources from the resource provider and ensures the proper execution of the sessions. The game machinists interact with clients and offer them a selection of games, usually by contracting new games from game development companies. The machinists execute distributed sessions with guaranteed QoS comprising interconnected application servers. 3) *Resource Provider*: This module offers the physical or virtualized machines on which the game servers will run. The resource providers lease virtual machines with fuzzy definitions of their characteristics, but with much more precise guaranties in terms of resource availability. This algorithmic once receiving capability events as input, analyzes its elements in context of the resources and also the global state of the game session, thereby generating the game session load reconciliation actions. It is capable of scaling up and down a game session on multiple resources in keeping with the variable user load however it also has occasional QoS breaches. The findings of the work are i) it is capable of scaling a game session on multiple resources in keeping with the variable user load and ii) it conjointly provides rise to occasional QoS breaches

B. Technique-2: VectorDot

A. Singh et al. [40] planned a unique load balancing rule referred to as VectorDot. In this work, the author has described a novel VectorDot algorithm that takes into account challenging hierarchical and multi-dimensional constraints while load balancing a system. The VectorDot algorithm is inspired by the Toyoda method for multi-dimensional knapsacks [41] and has several interesting aspects. While the original Toyoda heuristic was limited to selecting items into a single knapsack (item-comparison only), here we extend it to a collection of knapsacks (node-comparisons also) and dynamic load balancing among them in hierarchical SAN topologies. It was also seen that another use of Toyoda heuristic for resource management in storage systems, however it neither addresses the hierarchy, nor has been evaluated for performance and quality. It handles the hierarchal quality of the data-center and multidimensionality of resource hundreds across servers, network switches, associate degreed storage in an agile knowledge center that has integrated server and storage virtualization technologies. VectorDot mechanizes dot product to differentiate nodes supported by the item needs and assists in removing overloads on servers, switches and storage nodes. The findings of the work are i) it handles hierarchal and third-dimensional resource constraints and ii) data size need more validation testing for proving reliability of the model.

C. Technique-3: LBVS

H. Liu et al. [42] projected a load balancing on virtual memory strategy (LBVS) that gives an outsized scale web information storage model and Storage as a Service model supported Cloud Storage. This work basically highlights a load balancing virtual storage strategy and use Fair-Share Replication (FSR) and a executing a balancing algorithm to archive it. Since bandwidth and CPU speed are usually expensive to change, the another alternative way is by placing replicas of data objects closer to clients is the cheapest way. The main idea of FSR is to identify best candidate nodes for replica placement primarily based on access load. Storage virtualization is achieved using a schema of three-layered and load balancing is achieved using 2 load balancing modules. It helps in rising the potency of coincident access by using duplicate balancing any reducing the interval and enhancing the capability of disaster recovery. This strategy additionally helps in raising the utilization rate of storage resource, flexibility and strength of the system. The findings of the work are i) it improves task interval and ii) the process components & cumulative data center load are not considered throughout the work.

D. Technique-4: Server-based LB for Internet distributed services

A. M. Nakai et al. [43] projected a novel server-based load balancing policy for internet servers that are distributed everywhere the world. The main contributions of this work are: (i) the evaluation of client-based server selection schemes in scenarios where several clients use the same schemes; and (ii) the proposal of a new solution that outperforms existing ones by dynamically adapting the fraction of load each client submits to each server. In order to evaluate the solution, the author have implemented in a discrete event simulator framework using Java. The author has used the PackMime Internet traffic model [44] to generate HTTP traffic in the simulations. PackMime allows the generation of both HTTP/1.0 and HTTP/1.1 traffic. The intensity of the traffic is controlled by the rate parameter, which is the average number of new connections started per second. The implementation provides a set of random variable generators that drive the traffic generation. Each random variable follows a specific distribution. In this paper, we argue that if clients collaborate in order to balance server load they can obtain better response times. The solution adaptively changes the fraction of load each client sends to each server giving higher priorities to nearby servers. Although this less greedy strategy of sending fractions of the load to worsen servers seems to be counterintuitive, our experiments have shown that our solution overcomes the two types of policies proposed so far, even an in scenario that favors one type or another. It assists in reducing the service response times by employing a protocol that limits the redirection of requests to the closest remote servers while not overloading them. A middleware is represented to implement this protocol. It conjointly uses a heuristic to assist net servers to endure overloads. The finding of the work can embody the actual fact that reliability of the simulation result doesn‟t return up with real time situation.

E. Technique-5: Fuzzy Logic

Srinivas Sethi et. al. [45] has introduced the novel load equalization technique using fuzzy logic in cloud computing, within which load equalization could be a core and difficult issue in Cloud Computing. In this work, the authors have designed a new load balancing algorithm based on round robin in Virtual Machine (VM) environment of cloud computing in order to achieve better response time and processing time. The load balancing algorithm is done before it reaches the processing servers the job is scheduled based on various parameters like processor speed and assigned load of Virtual Machine (VM) and etc. It maintains the information in each VM and numbers of request currently allocated to VM of the system. It identify the least loaded machine, when a request come to allocate and it identified the first one if there are more than one least loaded machine. The authors have tried to implement the new load balancing technique based on Fuzzy logic. Where the fuzzy logic is natural like language through which one can formulate their problem. In this architecture, the fuzzifier performs the fuzzification process that converts two types of input data like processor speed and assigned load of Virtual Machine (VM) and one output like balanced load which are needed in the inference system. The design also considers the processor speed and load in virtual machine as two input parameters to make the better value to balance the load in cloud using fuzzy logic. These parameters are taking as inputs to the fuzzifier, which are used to measure the balanced load as the output. Two parameters named as the processor speed and assigned load of Virtual Machine (VM) of the system are jointly used to evaluate the balanced load on data centers of cloud computing environment through fuzzy logic. The results obtained with performance evaluation can balance the load with decreases the processing time as well as improvement of overall response time, which are leads to maximum use of resources. The processor speed and allotted load of Virtual Machine (VM) are deployed to balance the load in cloud computing through fuzzy logic. The finding of the work can embrace the very fact that solely Round Robin is taken into account that doesn't cater the important real time demand of the many large files over network.

F. Technique-6: Task Scheduling

Tayal [46] has proposed an optimized protocol based on Fuzzy-GA improvement that makes a programming call by evaluating the whole cluster of task within the job queue. In this work, the authors have described and assessed Fuzzy sets to model inexact scheduling parameters and also to symbolize satisfaction grades of each objective. Genetic algorithms with different components are designed on the based technique for task level scheduling in Hadoop MapReduce. In order to achieve a better balanced load across all the nodes in the cloud environment, the scheduler is enhanced by forecasting the execution time of tasks assigned to certain processors and making an optimal decision over the entire group of tasks. This work also assumes centralized scheduling policy where a master processor unit in cloud, collecting all tasks, will take charge of dispatching them to other process units. The system architecture illustrates the data store and computing cluster that jobs could be allocated to the cluster includes machines arranged in a general tree-shaped switched network. The nodes are commodity PCs. Data are distributed through these nodes. There are several replicas for each data block in the distributed file system. By default, the number of replicas is set as 3 in Hadoop. Map tasks generate the intermediate data stored the same node. The design also assumes the communication overhead exits when the data does not locate in the same node as the computing node. However it will require further improvement as this whole rule is predicated on the accuracy of the anticipated execution time of every task. The work is sensible however includes advanced rule and incorporates a tendency to extend network overhead too.

G. Technique-7: Particle Swarm Optimization

Wu et. al [47] has experimented with a set of workflow applications by varying their data communication costs and computation costs according to a cloud price model. First, the algorithm starts with swarm initialization using greedy randomized adaptive search procedure to guarantee each particle in the initial swarm is a feasible and efficient solution. Then, compute the potential exemplars, pbest and gbest, for particles to learn from while they are moving. The stop condition is considered as the user‟s QoS requirements, such as deadline, the budget for computation cost or data transfer cost. The particle‟s new position generation procedure has three steps: 1) select elements from the promising set of pairs with larger probability, that is, the particle learns from gbest and pbest; 2) due to the discrete property of scheduling, there are usually not enough feasible pairs in gbest to generate new position, so the particle will learn from its previous position; 3) all the unmapped tasks should choose resources from other feasible pairs. Finally, gbest will be return as optimal solution. The authors have also compared the total computation cost optimization ratio by varying the tasks number. The result shows that when the task number of the workflow becomes large, their technique optimization ratio increases relatively dramatic. It means the technique can actually achieve lower cost for executing the workflow. Experimental results show that the proposed algorithm can achieve much more cost savings and better performance on makes pan and cost optimization. Result could be better if SLA was considered. The goal of this study was to determine whether the literature on load balancing techniques in cloud computing provides a uniform and rigorous base. The papers were initially obtained in a broad search in four databases covering relevant journals, conference and workshop proceedings. Then an extensive systematic selection process was carried out to identify papers describing load balancing techniques in cloud computing. The results presented here thus give a good picture of the existing load balancing techniques in cloud computing.

VI. Conclusion

The random arrival of load in such an environment can cause some server to be heavily loaded while other server is idle or only lightly loaded. Equally load distributing improves performance by transferring load from heavily loaded server. Efficient scheduling and resource allocation is a critical characteristic of cloud computing based on which the performance of the system is estimated. The considered characteristics have an impact on cost optimization, which can be obtained by improved response time and processing time. Load balancing is one of the main challenges in cloud computing. It is required to distribute the dynamic local workload evenly across all the nodes to achieve a high user satisfaction and re-source utilization ratio by making sure that every computing re-source is distributed efficiently and fairly. With proper load balancing, resource consumption can be kept to a minimum which will further reduce energy consumption. Therefore, there is a need to develop an energy-efficient load balancing technique that can improve the performance of cloud computing by balancing the workload across all the nodes in the cloud along with maximum resource utilization, in turn reducing energy consumption and carbon emission to an extent which will help to achieve Green computing.

REFERENCES

[1] http://www.ibm.com/cloud-computing/us/en/. Accessed on 26th Sep, 2012

[2] Steve Gibbs, Cloud Computing, International journal of Innovative Research in Engineering & Science, issue 1, Volume 1, July, 2012

[3] C. L. Liu, Computing and Communication: There's so much that we share, ICESA 2007

[4] Corbato, Fernando J., "An Experimental Time-Sharing System". SJCC Proceedings. MIT. AIEE-IRE '62 (Spring) Proceedings of the May 1-3, 1962, spring joint computer conference, Pages 335-344, 1962.

[5] Mark Baker, Amy Apon, Rajkumar Buyya, Hai Jin, Cluster Computing and Applications, Encyclopedia of Computer Science and Technology, Allen Kent and Jim Williams (eds), pp.87-125 (Chapter 4), Volume 45 (Supplement 30), ISBN: 0-8247-2298-1, Marcel Dekker, Inc., New York, USA, Jan. 2002.

[6] David Cepeda, Ran Ding, Mohammed Ahmed, Ruth Diaz, Clustered Environments, Report, 2010.

[7] Stephanie Tanner, Todd Frederick, Jeremy Gustafson, Richard Brown, Beowulf cluster computing for all: The HiPerCiC project, 2009

[8] Komal Vashisht, Shefali, Karishma Shukla, A Survey On Grid Computing Approach, IJCST Vol. 1, Iss ue 2, December 2010

[9] Bart Jacob, Grid computing: What are the key components?, ITSO Redbooks, 2003

[10] Bill Allcock, Joe Bester, John Bresnahan, Ann L. Chervenak, Ian Foster1, Carl Kesselman, Sam Meder, Veronika Nefedova, Darcy Quesnel1 Steven Tuecke1, Data Management and Transfer in High-Performance Computational Grid Environments, DOI:10.1016/S0167-8191(02)00094-7

[11] Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal, Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities, HPCC 2008: 5-13

[12] Rajnish Choubey, Rajshree Dubey, Joy Bhattacharjee, A Survey on Cloud Computing Security, Challenges and Threats, International Journal on Computer Science and Engineering (IJCSE), ISSN : 0975-3397, Vol. 3 No. 3 Mar 2011

[13] David S. Linthicum, Cloud Computing and SOA Convergence in your Enterprise, Pearson, 2010.

[14] Mehrdad Mahdavi Boroujerdi, Soheil Nazem, Cloud Computing: Changing Cogitation about Computing, World Academy of Science, Engineering and Technology 58 2009.

[15] Top Threats to Cloud Computing V1.0, Cloud Security Alliance, March 2010.

[16] Amazon web service, [Online]. Available: http://aws.amazon.com/

[17] Mary Jo Foley, Microsoft's Windows Azure has a meltdown, February 29, 2012. Retrieved from http://www.zdnet.com/blog/microsoft/microsofts-windows-azure-has-a-meltdown/12076

[18] Chris Talbot, Google Apps Cloud Outage Lessons Learned, Published on 20th July, 2011. Retrieved from http://www.channelinsider.com/c/a/Cloud-Computing/Google-Apps-Cloud-Outage-Lessons-Learned-392917/

[19] Nick Antonopoulos, Lee Gillam, Cloud Computing: Principles, Systems and Applications, Springer, 03-Aug-2010 - 379 pages

[20] http://blog.cloudimpact.com/?p=78. Accessed on 25th Sep, 2012

[21] Z. Kalbarczyk, Toward Resilient Cloud Environment: Virtualization of Error Monitoring and Recovery, March 2012

[22] Neha Thakur, Performance Testing in Cloud: A pragmatic approach, White Paper Submitted for STC 2010

[23] Anastasia Tubanos, FlexiScale Suffers 18-Hour Outage, theWHIR.com on October 31, 2008. Retrieved from http://www.thewhir.com/web-hosting-news/flexiscale-suffers-18-hour-outage

[24] Tim Ferguson, Salesforce.com outage hits thousands of businesses, January 8, 2009. Retrieved from http://news.cnet.com/8301-1001_3-10136540-92.html

[25] Mary Jo Foley, Microsoft apologizes for spate of recent Online Services outages, September 9, 2010. Retrieved from http://www.zdnet.com/blog/microsoft/microsoft-apologizes-for-spate-of-recent-online-services-outages/7337

[26] Fahmida Y. Rashid, Skype Outage Caused by Overloaded Servers, Outdated Desktop Client, Posted 2010-12-29, Retrieved from http://www.eweek.com/c/a/IT-Infrastructure/Skype-Outage-Caused-by-Overloaded-Servers-Outdated-Desktop-Client-749097/

[27] Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region, Retrieved from http://aws.amazon.com/message/65648/

[28] Beth Bacheldor, Lightning takes down a data center. Are solar explosions next?, August 10, 2011, Retrieved from http://www.itworld.com/data-centerservers/192207/lightning-takes-down-data-center-are-solar-explosions-next

[29] Anthony Wing Kosner, Amazon Cloud Goes Down Friday Night, Taking Netflix, Instagram And Pinterest With It, 6/30/2012 , Retrieved from http://www.forbes.com/sites/anthonykosner/2012/06/30/amazon-cloud-goes-down-friday-night-taking-netflix-instagram-and-pinterest-with-it/

[30] Jaspreet kaur, Comparison of load balancing algorithms in a Cloud, International Journal of Engineering Research and Applications, Vol. 2, Issue 3, May-Jun 2012, pp.1169-1173

[31] Kuyoro S. O., Ibikunle F., Awodele O., Cloud Computing Security Issues and Challenges, International Journal of Computer Networks (IJCN), Volume (3) : Issue (5) : 2011

[32] Jing Deng Scott C.-H. Huang Yunghsiang S. Han and Julia H. Deng, Fault-Tolerant and Reliable Computation in Cloud Computing, GLOBECOM Workshops (GC Wkshps), 2010 IEEE

[33] Ritambhara Agrawal, Legal issues in cloud computing, IndicThreads.com conference on cloud computing, June 2011

[34] Clive Longbottom, The Promise and Problems of Cloud Backup and Restore, Quocirca Ltd, 2011

[35] Robert H. Carpenter, Walking from cloud to cloud: the portability issue in cloud computing, Washington journal of law, Technology & Arts Volume 6, Issue 1 Summer 2010

[36] Raksha Nawal, Aashish Gupta, Pragya Nagar, Rajesh Gurjar, Emergence, Performance And Issues Challenging Cloud Computing, Proceedings of the NCNTE-2012, Third Biennial National Conference on Nascent Technologies, 2012

[37] Brian E. Mitchell, Cloud Computing: Intellectual Property Legal Issues, Strafford Publishing Webinar, 2011

[38] K. Ramana, A. Subramanyam and A. Ananda Rao, Comparative Analysis of Distributed Web Server System Load Balancing Algorithms Using Qualitative Parameters, VSRD-IJCSIT, Vol. 1 (8), 2011, 592-600

[39] Vlad Nae, Alexandru Iosup, Radu Prodan, Dynamic Resource Provisioning in Massively Multiplayer Online Games, Transactions on Parallel and Distributed Systems, 2010

[40] Toyoda Y. A simplified algorithm for obtaining approximate solutions to zero-one programming problems. Management Science, vol-21(12): 1417–27, 1975

[41] Aameek Singh, Madhukar Korupolu, Dushmanta Mohapatra, Server-Storage Virtualization: Integration and Load Balancing in Data Centers, Journal of Research and Development Vol-52, 2008

[42] Hao Liu, Shijun Liu, Xiangxu Meng, Chengwei Yang, Yong Zhang, LBVS:A Load Balancing Strategy for Virtual Storage, IEEE International Conference on Service Sciences, 2010

[43] Alan Massaru Nakai, Edmundo Madeira, and Luiz E. Buzato, Improving the QoS of Web Services via Client-Based Load Distribution, XXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (Aceito para apresentação). May, 2011

[44] Cao, J., Cleveland, W. Y. G., Jeffay, K., Smith, F., and Weigle, M. (2004). Stochastic models for generating synthetic http source traffic. In Proceedings of INFOCOM, Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, 2004

[45] Srinivas Sethi, Anupama Sahu, Suvendu Kumar Jena, Efficient load Balancing in Cloud Computing using Fuzzy Logic, IOSR Journal of Engineering (IOSRJEN) ISSN: 2250-3021 Volume 2, Issue 7(July 2012), PP 65-71

[46] Sandeep Tayal, Task Scheduling Optimization for the Cloud Computing Systems, IJAEST, 2011

[47] Zhangjun Wu, Zhiwei Ni, Lichuan Gu, Xiao Liu, A Revised Discrete Particle Swarm Optimization for Cloud Workflow Scheduling, IEEE 2010

Suriya Begum has completed her Bachelor of Engineering in Computer Science and Engg in 1995 from Bangalore University, India. She completed her Master of Technology in Computer Science in 2007 from Allahabad University, India and currently a research scholar in Visveswaraya Technical University, Belgaum, India. She is having almost 19 years of experience as an academician. Her research interest is cloud computing, networking and load balancing.

Dr. Prashanth C.S.R has completed Bachelor of Engineering in Computer Science and Engg from Bangalore University, India. M.S in Computer Science from University of Texas at Dallas and Completed his PhD in Computer Science, from Auburn University in 2006.His research interest is high performance computing, cloud computing, networks and Operating Systems. He is presently working as Professor and Head of Department of Computer Science and Engg, New Horizon College of Engineering, Bangalore. He has published many national and international papers.

A Comprehensive Review for Central Processing Unit Scheduling Algorithm

Ryan Richard Guadaña[1], Maria Rona Perez[2] and Larry Rutaquio Jr.[3]

**[1] Computer Studies and System Department, University of the East
Caloocan City, 1400, Philippines**

**[2] Computer Studies and System Department, University of the East
Caloocan City, 1400, Philippines**

**[3] Computer Studies and System Department, University of the East
Caloocan City, 1400, Philippines**

Abstract

This paper describe how does CPU facilitates tasks given by a user through a Scheduling Algorithm. CPU carries out each instruction of the program in sequence then performs the basic arithmetical, logical, and input/output operations of the system while a scheduling algorithm is used by the CPU to handle every process. The authors also tackled different scheduling disciplines and examples were provided in each algorithm in order to know which algorithm is appropriate for various CPU goals.

Keywords: *Kernel, Process State, Schedulers, Scheduling Algorithm, Utilization.*

1. Introduction

The central processing unit (CPU) is a component of a computer system that carries out the instructions of a computer program, and is the primary element carrying out the computer's functions. The central processing unit carries series of program instructions, executes both logical and arithmetical functions, and handles input/output operations of the system. The demand of activities to be performed by the CPU piqued the authors' interest on how CPU handles different tasks given by the user?

The question on how does a CPU handles different tasks given by the user is answered through scheduling. Scheduling is a key concept in computer multitasking, multiprocessing operating system and real-time operating system designs. Scheduling refers to the way processes are assigned to run on the available CPUs, since there are typically many more processes running than there are available CPUs, like shoppers sharing the checkout operators on their way out of the store. There are different types of Operating System schedulers that the authors focused on. First is the Long Term Scheduler also known as the admission scheduler that decides which jobs or processes are to be admitted to the ready queue; that is, when an attempt is made to execute a program, its admission to the set of currently executing processes is either authorized or delayed by the long-term scheduler. Second is the Mid-term Scheduler that temporarily removes processes from main memory and places them on secondary memory (such as a disk drive) or vice versa. Last is the Short Term Scheduler that decides which of the ready, in-memory processes are to be executed.

2. CPU Utilization

In order for a computer to be able to handle multiple applications simultaneously there must be an effective way of using the CPU. Several processes may be running at the same time, so there has to be some kind of order to allow each process to get its share of CPU time. One of the most important components of the operating system is the kernel, which controls low-level processes which is typically unknown to the average user. It controls how memory is read and written, the order in which processes are executed, how information is received and sent by devices like the monitor, keyboard and mouse, and decides how to interpret information received from networks. Kernel is also the central component of most computer operating systems that bridges applications and computer peripherals.

2.1 The CPU Process States

When a process is created, its state is set to new. Once the process is ready to use the CPU its state is set to ready. It is inserted into the ready queue waiting its turn to be assigned CPU time so that its instructions can be executed. Once the CPU is available the process next in line in the ready queue is set to running. This means that the process' instructions are being executed.

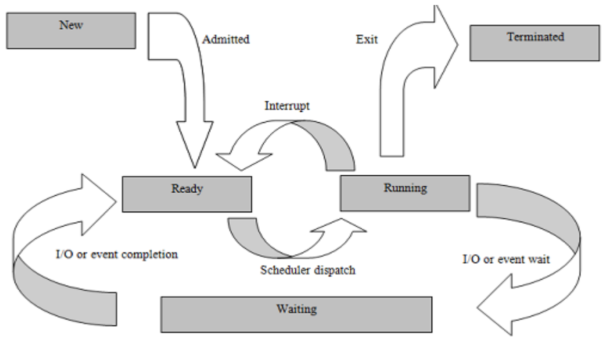


Fig. 1 CPU Process States.

Once the process is being executed two things can happen:

1) The process' instructions are all executed in which case its state will be set to terminate.

2) While the process is running an I/O interrupt or event wait is executed which stops the running program.

In the event the first case takes place, the program finishes executing and then terminates. This means all the instructions of the process have been executed and it has no more need for the CPU. However, this can also happen if there is some kind of error in the program that requires the process to be terminated prematurely. In the second case the procedures taken are much more complex. For example, let us say that there is a process that is currently occupying the CPU. As the instructions of this process are being executed the program needs to get input from the user at the keyboard. This causes the process to stop executing. In this situation the process will enter the waiting state. This means that the process will lose control of the CPU and be inserted into the waiting queue. Once the input is received from the user at the keyboard the process must go back to the ready state. The process cannot take hold of the processor; it must wait in the ready queue until it is assigned the CPU.

Once the process is assigned the CPU again, it will continue executing its instructions. Once again two things may happen. If there is need for more I/O then the process will once again enter into the waiting state. If not, then the process will complete and will become terminated once the final instructions are executed. As stated earlier a process may enter several states in its lifetime. However, where is this information stored? It is stored in the process control block (PCB). The process control block is a representative of each process. It contains information about the process, which it is associated with. The information it contains is the process state, program counter, CPU registers, CPU-scheduling information, memory management information, accounting information, and I/O status information.

CPU scheduling information is information that includes process priority, pointers to scheduling queues, and any other scheduling parameters. This is the basis of multi-programmed operating systems because the CPU is able to switch from process to process while the operating system is able to make the running programs seem as if they are being executed simultaneously. Whenever the CPU has to wait for I/O operations to occur, there are CPU cycles that are being wasted. The idea behind CPU scheduling is to be able to switch from process to process when the CPU becomes idle. In this way, while a process is waiting for an I/O request to complete, the CPU does not have to sit idle. It can begin executing other processes that are in the waiting state.

There are two scheduling schemes that are available. There is the non-preemptive scheduling scheme and there is the preemptive scheduling scheme. Different CPU scheduling algorithms have different properties and may have one class of processes over another. Many criteria have been suggested for comparing CPU scheduling algorithms. The characteristics used for comparison can make a substantial difference in the determination of the best algorithm. The criteria should include the following:

- CPU Utilization: This measures how busy the CPU is. CPU utilization may range from 0 to 100 percent. In a real system, it should range from 40% (for a lightly loaded system) to 90% (for heavily loaded system).
- Throughput: This is a measure of work (number of processes completed per time unit). For long processes, this rate may be one process per hour; for short processes, throughput might be 10 processes per second.
- Turnaround Time (TT): This measures how long it takes to execute a process. Turnaround time is the interval from the time of submission to the time of completion. It is the sum of the periods spent waiting to get into memory, waiting in the ready queue, executing in the CPU, and doing I/O.
- Waiting Time (WT): CPU scheduling algorithm does not affect the amount of time during which process executes or does I/O; it affects only the amount of time a process spends waiting in the ready queue. Waiting time is the total amount of time a process spends waiting in the ready queue.
- Response Time: The time from submission of a request until the system makes the first response. It is the amount of time takes to start responding but not the time that it takes to output that response. The turnaround time is generally limited by the speed of the output device.

3. Scheduling Algorithm

3.1 Non-preemptive scheduling algorithm

Non-preemptive or also known as the cooperative scheduling is the first scheme where once a process has control of the CPU no other processes can preemptively take the CPU away. The process retains the CPU until either it terminates or enters the waiting state. There are two algorithms that can be used for non-preemptive scheduling. There are different algorithms under non-preemptive scheduling scheme and these are the following:

3.1.1 First-Come, First-Served (FCFS) scheduling algorithm

In this scheduling algorithm the first process to request the CPU is the one that is allocated the CPU first. The First-Come, First-Served algorithm is very simple to implement. It can be managed using a First-In, First-Out (FIFO) queue. When the CPU is free, it is allocated to the first process waiting in the FIFO queue. Once that process is finished, the CPU goes back to the queue and selects the first job in the queue. An analogy for this is students waiting in line to pay for their lunch. When one student is ready to pay for their meal, they must go to the back of the line and wait for their turn. This is the idea behind the First-Come, First-Served algorithm.

Consider the following set of processes that arrive at time 0, with the length of the CPU burst given in milliseconds:

Table1. Given example of processes for FCFS

Process	Burst Time
P1	24
P2	3
P3	3

If the process arrives in the order P1, P2, P3, and are served in FCFS order, the gets the result shown in the following Gantt chart:

Gantt Chart			
	P_1	P_2	P_3
0	24	27	30

Fig. 2 Gantt Chart illustration of FCFS.

Therefore, the waiting time for each process is:

WT for P1 = 0 – 0 = 0
WT for P2 = 24 – 0 = 24
WT for P3 = 27 – 0 = 27

Average WT = (0 + 24 + 27) / 3
= 17 ms

The turnaround time for each process would be:

TT for P1 = 24 – 0 = 24
TT for P2 = 27 – 0 = 27
TT for P3 = 30 – 0 = 30

Average TT = (24 + 27 + 30) / 3
= 27 ms

3.1.2 Shortest Job First (SJF) scheduling algorithm

In this scheduling scheme the process with the shortest next CPU burst will get the CPU first. The movement of all the short jobs ahead of the longer jobs will decrease the average waiting time. If two processes have the same length of CPU burst, FCFS scheduling is used to break the tie by considering which job arrived first.

Consider the following set of processes that arrive at time 0, with the length of the CPU burst given in milliseconds:

Table2. Given example of processes for SJF

Process	Burst Time
P1	6
P2	8
P3	7
P4	3

Using SJF, the system would schedule these processes according to the following Gantt chart:

Gantt Chart				
	P_4	P_1	P_3	P_2
0	3	9	16	24

Fig. 3 Gantt Chart illustration of SJF.

Therefore, the waiting time for each process is:

WT for P1 = 3 – 0 = 3
WT for P2 = 16 – 0 = 16
WT for P3 = 9 – 0 = 9
WT for P4 = 0 – 0 = 0

Average WT = (3 + 16 + 9 + 0) / 4
= 7 ms

The turnaround time for each process would be:

TT for P1 = 9 – 0 = 9
TT for P2 = 24 – 0 = 24
TT for P3 = 16 – 0 = 16
TT for P4 = 0 – 0 = 0

Average TT = (9 + 24 + 16 + 0) / 4
= 12.25 ms

3.1.3 Priority (Prio) scheduling algorithm

A priority is associated with each process, and the CPU is allocated to the process with the highest priority. Equal priority processes are scheduled in FCFS order. An SJF is simply a priority algorithm where the priority (p) is the inverse of the next CPU burst (τ). The larger the CPU burst, the lower the priority, and vice versa.

Consider the following set of processes that arrive at time 0, with the length of the CPU burst given in milliseconds:

Table3. Given example of processes for Prio

Process	Priority	Burst Time
P1	3	10
P2	1	1
P3	4	2
P4	5	1
P5	2	5

Using priority algorithm, the schedule will follow the Gantt chart below:

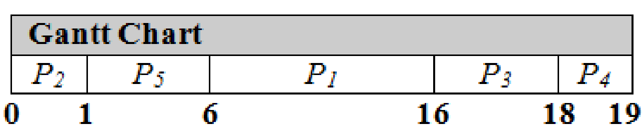


Fig. 4 Gantt Chart illustration of Prio.

Therefore, the waiting time for each process is:

WT for P1 = 6 – 0 = 6
WT for P2 = 0 – 0 = 0
WT for P3 = 16 – 0 = 16
WT for P4 = 18 – 0 = 18
WT for P5 = 1 – 0 = 1

Average WT = (6 + 0 + 16 + 18 + 1)/5
= 8.2 ms

The turnaround time for each process would be:

TT for P1 = 16 – 0 = 16
TT for P2 = 1 – 0 = 1
TT for P3 = 18 – 0 = 18
TT for P4 = 19 – 0 = 19
TT for P5 = 6 – 0 = 6

Average TT = (16 + 1 + 18 + 19 + 6)/5
= 12.25 ms

3.2 Preemptive scheduling algorithm

Preemptive scheduling is the second scheduling scheme. In preemptive scheduling there is no guarantee that the process using the CPU will continually run until it is finished. This is because the running task may be interrupted and rescheduled by the arrival of a higher priority process.

3.2.1 Shortest Remaining Time First (SRTF) scheduling algorithm

The SJF has a preemptive adaptation commonly referred to as shortest remaining time first; the process that is running is compared to the processes in the ready queue. If a process in the ready queue is shorter than the process running, then the running task is preempted and the CPU is given to the shorter process until it is finished.

Consider the following set of processes with the length of the CPU burst given in milliseconds:

Table4. Given example of processes for SRTF

Process	Arrival Time	Burst Time
P1	0	8
P2	1	4
P3	2	1
P4	3	5

If the processes arrive at the ready queue at the times shown and need the indicated burst times, then the resulting preemptive SJF schedule is as depicted in the following Gantt chart:

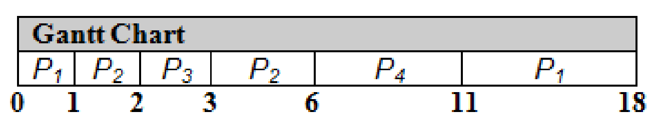


Fig. 5 Gantt Chart illustration of SRTF.

Therefore, the waiting time for each process is:

WT for P1 = 11 – 0– (1) = 10
WT for P2 = 3 – 1– (1) = 1
WT for P3 = 2 – 2 = 0
WT for P4 = 6 – 3 = 3

Average WT = (10 + 1 + 0 + 3) / 4
= 3.5 ms

The turnaround time for each process would be:

TT for P1 = 18 – 0 = 18
TT for P2 = 6 – 1 = 5
TT for P3 = 8 – 2 = 6
TT for P4 = 13 – 3 = 10

Average TT = (18 + 5 + 6 + 10) / 4
= 9.75 ms

3.2.2 Preemptive Priority (P-Prio) scheduling algorithm

Priority scheduling can either be preemptive or non-preemptive. When a process arrives at the ready queue, its priority is compared with the priority of the process, which is currently executing at the CPU. A preemptive priority scheduling algorithm will preempt the CPU if the priority of the newly arrive process is higher than the currently running process. A major problem with the priority scheduling algorithms, whether preemptive or non-preemptive is indefinite blocking or starvation. In a heavily loaded computer system, a steady stream of higher-priority processes can prevent a low-priority process from ever getting the CPU. A solution to the problem of indefinite blocking is aging. Aging is the technique of gradually increasing the priority of process that wait in the system for a long time.

Consider the following set of processes that arrive at time 0, with the length of the CPU burst given in milliseconds:

Table5. Given example of processes for P-Prio

Process	Arrival Time	Burst Time	Priority
P1	1	5	5
P2	2	10	4
P3	3	18	3
P4	4	7	2
P5	5	3	1

Using preemptive priority algorithm, the schedule will result to the Gantt chart as follows:

Gantt Chart

I	P_1	P_2	P_3	P_4	P_5	P_4	P_3	P_2	P_1

0 1 2 3 4 5 8 14 31 40 44

Fig. 6 Gantt Chart illustration of P-Prio

Therefore, the waiting time for each process is:

WT for P1 = 40 – 0– (1) = 38
WT for P2 = 31 –2– (1) = 28
WT for P3 = 14 – 3– (1) = 10
WT for P4 = 8 – 4– (1) = 3
WT for P5 = 5 – 5 = 0

Average WT = (38 + 28 + 10 + 3 + 0)/5
= 15.8 ms

The turnaround time for each process would be:

TT for P1 = 44 – 1 = 43
TT for P2 = 40 – 2 = 38
TT for P3 = 31 – 3 = 28
TT for P4 = 14 – 4 = 10
TT for P5 = 8 – 5 = 3

Average TT = (43 + 38 + 28 + 10 + 3)/5
= 24.4 ms

3.2.2 Round – Robin (RR) scheduling algorithm

This algorithm is specifically for time – sharing systems. A small unit of time, called a time quantum or time slice, is defined. The ready queue is treated as a circular queue. The CPU scheduler goes around the ready queue, allocating the CPU to each process for a time interval of up to 1 time quantum. The RR algorithm is therefore preemptive.

Consider the following set of processes that arrive at time 0, with the length of the CPU burst given in milliseconds:

Table6. Given example of processes for RR

Process	Burst Time
P1	24
P2	3
P3	3

If the system uses a time quantum of 4ms, then the resulting RR Gantt chart is:

Gantt Chart

P_1	P_2	P_3	P_1	P_1	P_1	P_1	P_1

0 4 7 10 14 18 22 26 30

Fig. 7 Gantt Chart illustration of RR

Therefore, the waiting time for each process is:

WT for P1 = 26 – 0– (20) = 6
WT for P2 = 4 – 0 = 4
WT for P3 = 7 – 0 = 7

Average WT = (6 + 4 + 7) / 3
= 5.67 ms

4. Analysis

The authors looked into a number of different scheduling algorithms and the two different scheduling schemes that was discussed in this paper, the preemptive and non-preemptive scheduling scheme. In order to know which algorithm to use for which CPU scheduling goal, different examples were given in each algorithm. Therefore, based on performance, the shortest job first (SJF) algorithm is recommended for the CPU scheduling problems of minimizing either the average waiting time or average turnaround time but the addition of preemption to the SJF algorithm gives supplementary increase in waiting and turnaround time, without affecting the response time. Long jobs have an even higher tendency to cause delay at the back of the queue since they can be interrupted by short jobs so even when long jobs get a chance to execute, they can be interrupted.

Also, the first come first serve (FCFS) algorithm is recommended for the CPU scheduling problems of minimizing either the average CPU utilization or average throughput but the discrepancy about FCFS is it promotes starvation[1].

The performance of the RR algorithm depends heavily on the size of the time quantum. It is concluded that if the quantum is too large, the RR policy degenerates into the FCFS policy. If the time quantum is too small, on the other hand, then the effect of the context – switch time becomes a significant overhead. As general rule, 80 percent of the CPU burst should be shorter than the time quantum.

In general, task given by the user to OS will use Priority based, Round Robin and preemptive while Real Time OS will use Priority and non preemption scheme.

Acknowledgments

The authors wish to thank their chairman, Prof. Amelia Damian, for sharing her knowledge on how to write a thesis onto journal format and her encouragement to finish this article. A special gratitude is also extended to their fellow faculty and colleagues.

[1] Starvation means that a job with low priority would never get a chance to enter the processor if there is steady stream of jobs or processes.

References

[1] Gisela May A. Albano, and Angelito G. Pastrana, Fundamentals of Operating System, A & C Printers, 2009.
[2] Silberschatz ,Galvin and Gagne, Operating systems concepts, 8th edition, Wiley, 2009.
[3] Silberschatz, Abraham, Peter B. Galvin, and Greg Gagne, Operating System Principles, 6th Edition, John Wiley and Sons Inc., 2006.
[4] Flynn, Ida M., and Ann Mclver McHoes, Understanding Operating Systems, 4th Edition, Thomson Learning Inc., 2007.
[5] Mehdi Neshat, Mehdi Sargolzaei, and Adel Najaran, "The New Method of Adaptive CPU Scheduling Using Fonseca and Fleming's Genetic Algorithm", Journal of Theoretical and Applied Information Technology, Vol. 37, No. 1, 2012, pp. 1-16.
[6] Huda Salih, and Yousra Fadil, "CPU Scheduling Simulation", Diyala Journal of Engineering, Vol. 02, No. 7, 2009, pp.39-52
[7] Abbas Noon, Ali Kalakech, and Seifedine Kadry, "A New Round Robin Based Scheduling Algorithm for Operating Systems: Dynamic Quantum Using the Mean Average", IJCSI International Journal of Computer Science Issues, Vol. 8, No. 1, 2011, pp. 224-229.

Ryan Richard H. Guadaña is a University of Caloocan City BS Computer Science graduate (2004) presently writing his thesis in Polytechnic University of the Philippines for his Masters of Science in Information Technology degree. He is at this time (2012) a full time faculty of University of the East Philippines. He has gained experience in Informatics International College as an OIC-Academic Head from 2008-2010 and as OIC-Office of Students Affairs from 2007-2009. Currently, he is interested and engaged in researches about computer aided training especially for Non-IT training such as disaster response. Now, he is a member of PSITE Philippines and ProICT Philippines.

Maria Rona Perez is a full time faculty of the Department of Computer Studies and Systems, College of Engineering, University of the East, Philippines while pursuing her Thesis Writing towards M.S. in Information Technology from Polytechnic University of the Philippines. She obtained B.S Information Technology "Cum Laude" from AMA Computer College - Manila Campus. Her research interest includes social studies, security and ethical issues and information system. She is a member of PSITE Philippines.

Larry T. Rutaquio Jr. is currently finishing his Graduate Study at Polytechnic University of the Philippines with the degree of Master of Science in Information Technology with specialization in Management Information System. Presently, he is working as a full-time faculty at the University of the East – Caloocan campus. He is also a member of Philippine Society of Information Technology Educators (PSITE) and ProICT Philippines. His area of interest are cloud computing and network security.

Implementation of DB-Scan in Multi-Type Feature CoSelection for Clustering

K.Parimala [1], Dr. V.PalaniSamy[2]

[1] **Asst Professor, MCA Department, NMS S.Vellaichamy Nadar College, Madurai-625019, TamilNadu, INDIA**

[2] **Professor & Head of the Deparment, Department of Computer Science & Engineering, Alagappa University, Karaikudi, TamilNadu, INDIA**

Abstract

Feature Selection is a preprocessing technique in supervised learning for improving predictive accuracy while reducing dimension in clustering and categorization. Multitype Feature Coselection for Clustering (MFCC) with hard k-means is the algorithm which uses intermediate results in one type of feature space enhancing feature selection in other spaces, better feature set is co-selected by heterogeneous features to produce better cluster in each space. Db-Scan is a density-based clustering algorithm finding a number of clusters starting from the estimated density distribution of corresponding nodes. It is one of the most common clustering algorithms and also most cited in scientific literature, as a generalization of DBSCAN to multiple ranges, effectively replacing the ε parameter with a maximum search radius.This paper presents the empirical results of the MFCC algorithm with Db-scan and also gives the comparison results of MFCC with hard k-means and DB-Scan. DB-Scan clustering is proposed for getting the quality clustering against the outliers and time criteria is less than any other clustering in high density data set.

Keywords: *Feature Selection, MFCC, Db-Scan.*

1. Introduction

Information or knowledge can be conceptualized as data. It reflects in the data norm, the size and dimensions have improved high and more. The feature selection plays a vital role in machine learning, data mining, information retrieval, etc. the goal of feature selection is to identify those features relevant to achieve a predefined task. Many researchers have been to find how to search feature subset space and evaluate them.

In supervised methods [1], the correlation of each feature with the class label is computed by distance, information dependence or consistency measures [2]. In unsupervised method the feature selection does not need the class of information such as document frequency and term strength [3]. The newly proposed methods namely Entropy based feature ranking method (En) proposed by Dash and Liu [4] in which feature importance is measured by the contribution to an entropy index based on the data similarity; the individual 'feature saliency' is estimated and an Expectation Maximization (EM) algorithm using Minimum message length is derived to select the feature subset and the number of clusters [5].

While the methods above are not directly targeted to clustering text documents, [6] proposes two other feature selection methods for text clustering. One is Term Contribution (TC) which ranks the feature by it overall contribution to the document similarity in the data set. The other is Iterative feature selection (IF), which utilizes some successful feature selection methods such as Information Gain (IG) and CHI-Square (χ^2) text to iteratively select features and performs text clustering at the same time.

[7] Combines information about document contents and hyper link structures to cluster documents. The hypertext documents in a certain information space were clustered into a hierarchical form based on contents as well as link structure of each hyper text documents.

From the ideas of [8] & [9] co-training algorithms learn through classifiers over each of the feature set and combine their predictions to decrease classification error. Cot raining algorithm can learn from unlabelled data starting from a weak predictor.

Clustering helps users, tackle the information overload problem in several ways: explore the contents of a document collection; group duplicate and near duplicate documents. Unsupervised method can hardly achieve a good performance when evaluated using labeled data.

Data fusion [10] is well suited to-problems involving massive amounts of data where each subsystem may not have entire data set, problems with many possible approaches, allows for natural and flexible distribution of resources aim to provide better performance than best input system. Voting procedures are examples of data fusion – results from identical data sets are merged.

This paper is devised to show the results of MFCC algorithm using density based clustering. This paper is organized as follows: Next we describe prior rela describing MFCC and Db-scan. Section 3 describes the learning of MFCC with Db-scan. Then in section 4, the experiments and evaluation results are explained and discussed finally, section 6 describes the conclusion and future works.

2. Related work

2.1. Multitype Features Coselection for Clustering (MFCC):

In this section we briefly discuss about MFCC. It is made clear that the selection of each type feature and the clustering is an iterative one. After one iteration of clustering, each data object will be assigned to a cluster. In [6], Liu et al. assumed each cluster corresponded to a real class. Using such information, they did supervise feature selection, such as Information Gain (IG) and $\chi 2$ statistic (CHI) [2] during k-means clustering. MFCC tries to fully exploit heterogeneous features of a web page like URL, anchor text, hyperlink, etc., and to find more discriminative features for unsupervised learning. We first use different types of features to do clustering independently. Then, we get different sets of pseudoclass, which are all used to conduct iterative feature selection (IF) for each feature space.

After normal selection, some data fusion methods are used to conduct iterative feature selection (IF) for each feature space, i.e., feature coselection. In each iteration of clustering, the coselections in several spaces are conducted one by one after clustering results in different feature spaces have been achieved before any coselection. Thus, the sequence of coselection will not affect the final performance. The general idea of coselection for k-means clustering is described in fig-1.

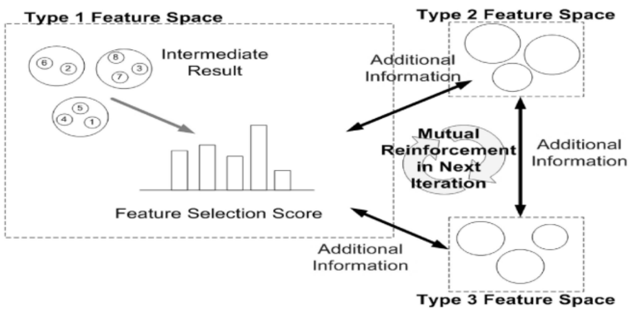


Fig – 1. The basic idea of Multitype feature coselection.

Suppose that we categorize data objects with M heterogeneous features into L clusters. Let fv_n be one dimension of the feature vector, icr_i be the intermediate clustering results in the i^{th} feature space, SF be the fusion function.

The pseudo algorithm is listed as follows:

```
Loop for N iterations of k-means clustering
  {
        Loop for m feature spaces
        {
          Do clustering in feature space m
        }
         Loop for M feature spaces
        {
```

For feature space m, do feature selection using results in all feature spaces.

For (fv_n) one dimension of the feature vector in space m, a feature selection score fss (fv_n, icr_i) is obtained by using intermediate clustering results icr_i in feature space i.

Then a combined score fss (fv_n) is achieved by fusing the scores based on different result sets.

$$fss(fv_n) = SF(fss(fv_n, icr_i)) \qquad (1)$$

```
        }
  }
```

In the equation (1), $fss(fv_n, icr_i)$ can be the value calculated by the selection function or rank among all features. The feature selection criteria, the six commonly used feature selection function mentioned in [2]:

Function	Mathematical form
$IG(t_k, c_i)$	$p(t_k,c_i)\cdot\log\frac{p(t_k,c_i)}{p(c_i)\cdot p(t_k)} + p(\bar{t}_k,c_i)\cdot\log\frac{p(\bar{t}_k,c_i)}{p(c_i)\cdot p(\bar{t}_k)}$
$\chi^2(t_k, c_i)$	$\frac{N\cdot(p(t_k,c_i)\cdot p(\bar{t}_k,\bar{c}_i) - p(t_k,\bar{c}_i)\cdot p(\bar{t}_k,c_i))^2}{p(t_k)\cdot p(\bar{t}_k)\cdot p(c_i)\cdot p(\bar{c}_i)}$
$CC(t_k, c_i)$	$\frac{\sqrt{N}\cdot(p(t_k,c_i)\cdot p(\bar{t}_k,\bar{c}_i) - p(t_k,\bar{c}_i)\cdot p(\bar{t}_k,c_i))}{\sqrt{p(t_k)\cdot p(\bar{t}_k)\cdot p(c_i)\cdot p(\bar{c}_i)}}$
$RS(t_k, c_i)$	$\log\frac{p(t_k \mid c_i)+d}{p(t_k \mid \bar{c}_i)+d}$
$OR(t_k, c_i)$	$\frac{p(t_k \mid c_i)\cdot(1-p(t_k \mid \bar{c}_i))}{(1-p(t_k \mid c_i))\cdot p(t_k \mid \bar{c}_i)}$
$GSS(t_k, c_i)$	$p(t_k,c_i)\cdot p(\bar{t}_k,\bar{c}_i) - p(t_k,\bar{c}_i)\cdot p(\bar{t}_k,c_i)$

Table-1 Feature Selection Functions.

Depending on the choices of fss and SF, we obtain five fusion models including voting, average value, max value, average rank, and max rank. The equations are listed as follows:

MaxRank(Rank(f_{vn})) = arg max(Rank(fvn,icr_i))
AverageRank(Rank(fvn))=(ΣRank(fvn,icr_i))/M
Voting(val(fvn)) = Σvote(fvn,icri)
Vote(fvn,icri)={0 val(fvn,icri) <st
1 val(fvn,icri) >=st
Average(val(fvn))= Σ val(fvn,icri)/M
Max(val(fvn)) = arg max(val(fvn,icri))

Table-2 Fusion Models

In the above equation, *val(fv_n,icr_i)* is the value calculated by selection function, *RANK(fv_n,icr_i)* is the rank of fv_n in the whole feature list ordered by *val(fv_n,icr_i)*, and st is the threshold of feature selection. After feature coselection, objects will be reassigned, features will be reselected, and the pseudoclass-based selection score will be recombined in the next iteration. Finally, the iterative clustering and feature coselection are well integrated.

In each of the iterations, the whole feature space should be reconsidered. The reason is that our method can help in finding more effective features through a mutual reinforcement process. Properly selected features will help clustering and vice-versa. That is to say, some discriminative features will not be found until late in the clustering phase. This can be proved by empirical results.

2.2 DBSCAN Clustering

DBSCAN regards clusters as dense regions of objects in the data space that are separated by regions of low density. A cluster is defined by this algorithm as a maximal set of density-connected objects. DBSCAN grows regions with sufficiently high density into clusters. Every object not contained in any cluster is considered to be noise.

In DBSCAN for each point of a cluster the neighborhood of a given radius (ε) has to contain at least a minimum number of points *(MinPts)* where ε and *MinPts* are input parameters [11].

The DBSCAN algorithm finds clusters as follows:

Let $X = \{x_1, x_2, x_3, ..., x_n\}$ be the set of data points. DBSCAN requires two parameters: ε (eps) and the minimum number of points required to form a cluster (MinPts).

1) Start with an arbitrary starting point that has not been visited.
2) Extract the neighborhood of this point using ε (All points which are within the ε distance are neighborhood).
3) If there is sufficient neighborhood around this point then clustering process starts and point is marked as visited else this point is labeled as noise (Later this point can become the part of the cluster).
4) If a point is found to be a part of the cluster then its ε neighborhood is also the part of the cluster and the above procedure from step 2 is repeated for all ε neighborhood points. This is repeated until all points in the cluster are determined.
5) A new unvisited point is retrieved and processed, leading to the discovery of a further cluster or noise.
6) This process continues until all points are marked as visited.

The computational complexity of DBSCAN is *O(NlogN)* if a spatial index is used. Otherwise, it is $O(N^2)$, where *N* is the number of objects. The main advantage of DBSCAN is that it is capable of discovering clusters of arbitrary shape. The main disadvantage of DBSCAN is that it leaves the user with the responsibility of selecting parameter values for e and *MinPts* that will lead to high quality clusters. The quality of the resulting clusters is sensitive to the user-defined parameters.

3. Proposed Work

3.1. Db Scan in MFCC

In this paper we present a method to build a clustering system that merges MFCC with density based clustering. The general idea for modification is based on the coselection and maximal set of *density-connected* objects. Db-scan grows regions with sufficiently high density in to clusters. Advantages of many of Db-scan algorithm include time efficiency and ability to find clusters of arbitrary shapes. MFCC reduces the noise feature effectively by and improved further performance. The modified MFCC got the idea from the arbitrary shapes, where we get intermediate membership to the noise features. So that the selection score for the modified MFCC will be as,

$$fss(\varepsilon, \min pts) = \mathop{sf}_{i=1}^{n} \left(\mathop{fss}_{c=1}^{npts} (p_{i,c}, \varepsilon, \min pts) \right) \qquad (2)$$

whereas,

SF – selection function to fuse the feature space selection (or, the intermediate clustering)

fss – feature selection score to select best center point (or, mean) from the specified feature space.

ε – max. distance between two samples for them to be considered as same in the neighborhood.

minpts – minimum number of points that must exist in the ε neighborhood.

npts – neighborhood points.

p – size of database.

c – cluster.

3.2 Experiments & Results:

The MFCC with Db-Scan algorithm proposed in the paper has been fully implemented and evaluated with extensive experimentation; this section presents the details of implementation, data set and text results.

3.2.1 .Evaluation metrics:

A number of metrics used in feature selection and clustering are evaluated and measures for categorization effectiveness. We use the best recall k precision metrics. Such measures are F-measure and time precision in each fss criteria.

F-measure is calculated by the harmonic mean of vocabulary terms (P) and total terms(R). Each fss criteria define the P & R terms.

We also use accuracy in this paper as a measure. Accuracy is computed as the ration of correctly classified testing documents to the total number of testing documents. Of course, all these performance metrics are computed for each category separately (i.e.) we apply all the testing documents to each fss criteria to compute P, R, f1, and accuracy for each fss criteria.

3.2.2 Experiment Results:

The experimental evaluation was performed on testdata data set. Here we can explain and results on testdata dataset. The testdata contains almost 255 articles, evenly distributed on 10 categories. Further each article can be assigned to one or more clusters. In our experiments run MFCC algorithm having db-scan is tested on above said database.

Density based algorithm – DB-SCAN with MFCC is verified with test data database (Table – 3). It contains feature classes of HTML, text files, word documents, jpeg files, user logs, etc.

Classes	No of documents	Related terms	Total term frequency
ASP	2	22	23
CSS	10	1439	6771
Gif	144	975	976
Html	25	14210	63392
Jpeg	18	3554	3659
Js	19	4935	38415
Pdf	5	249229	398036
Php	10	1670	4644
Png	9	245	245
Ppt	13	193505	208541

Table – 3. Feature Classes of test database.

MFCC algorithm clusters the dataset according to the query term. TF-IDF is calculated and the following result is got for CHI-square (Ψ^2), correlation coefficient (CC), GSS coefficient (GSS), and information gain (IG) for each feature class.

DBSCAN is that it is capable of discovering clusters of arbitrary shape. It leaves the user with the responsibility of selecting parameter values for ε and *MinPts* that will lead to high quality clusters. The quality of the resulting clusters is sensitive to the user-defined parameters. (refer fig – 2).

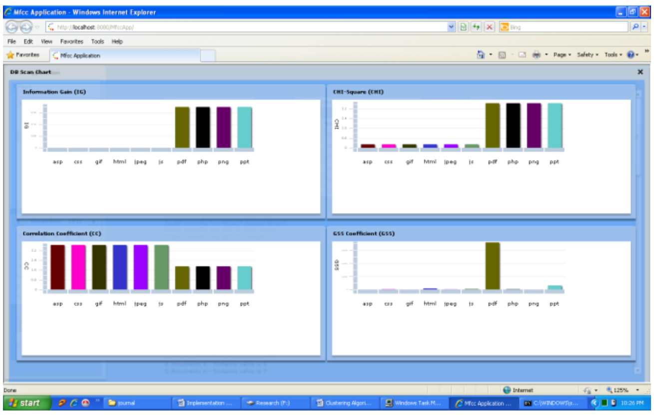


Fig – 2 DB-SCAN.

The testdata database is verified with hard k-means MFCC. The result is shown in Fig-3; the hard k-means clusters

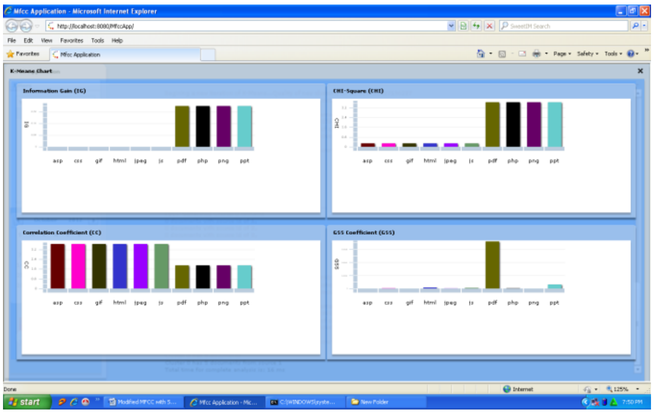


Fig – 3. Hard k-means

The hard k-means and Db-Scan shows the results more or less similar, they differ in time factor, Db-Scan works better in high density data sets. It shows the result in quality of cluster. Since ε and minpts are the two required user defined parameters, need not to specify the number of clusters as opposed to hard k-means. Db-Scan is notion of noise. If the ε neighborhood contains sufficient points then the particular point is marked as noise. Here in MFCC implementation the fss selects the feature space by intermediate clustering and then it fuses the fss score based on data sets. Thus the result is of db-scan and hard k-means is shown in table-4.

Feature selection functions	Cluster model	Number of clusters	Mean values	Time (ms)
IG	Db-scan	10	1,1,2,177,4,177,6,7,177, 177	109
	k-means	2	7.2707	94
Ψ^2	Db-scan	10	1,1,2,177,4,177,6,7,177, 177	15
	k-means	2	7.2707	32
CC	Db-scan	10	1,1,2,177,4,177,6,7,177, 177	15
	k-means	2	11.6937	31
GSS	Db-scan	10	1,1,2,3,4,5,6,7,8,9	0
	k-means	2	3.6349	16

Table – 4. Comparison of Db-Scan and hard k-means

The density based clustering shows better result than hard k-means clustering. Even though the two clusters show the same result, they differ in time factor (fig – 4)

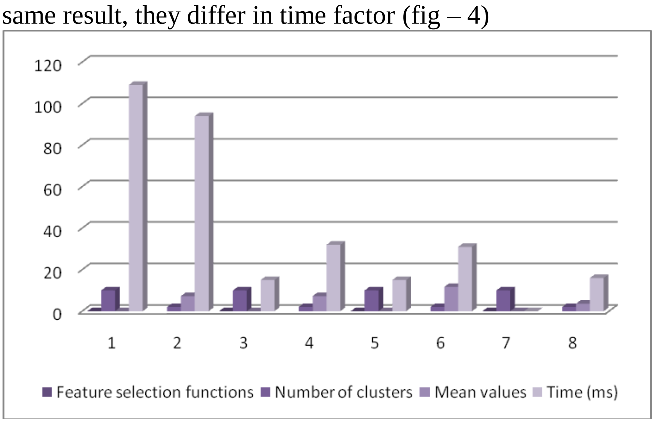


Fig – 4. Comparison of Db-scan & Hard k-me

4. Conclusion

The MFCC algorithm implemented with db-scan shows that it quality clusters and also maintains time criteria. So the higher dimension and high density data set can be clustered and features space can be framed by the fusion methods. The minpts and ε are the parameters determine the cluster shape and quality. The quality of the db-scan is implemented and shown. Db-scan is better in time factor and we need not to specify the number of clusters. Thus we have a chance of discovering further cluster or noise point to be revisited and processed. The MFCC algorithm can be implemented in other clustering algorithms or further extended to other data sets and applications.

References

[1] M. Dash and H. Liu, “Feature Selection for Classification,” Int’l J. Intelligent Data Analysis, vol. 1, no. 3, pp. 131-156, 1997.
[2] Y. Yang and J.O. Pedersen, “A Comparative Study on Feature Selection in Text Categorization,” Proc. Int’l Conf. Machine Learning (ICML ’97), pp. 412-420, 1997
[3] Shen Huang, Zheng Chen, Yong YU & WeiYing Ma, "Multi type Features Coselection for Web document Clustering", IEEE Transactions on Knowledge and Data Engineering; vol-18,no.4,April2006.
[4] M. Dash and H. Liu, "Feature Selection for Clustering," Proc. 2000 Pacific-Asia Conf. Knowledge Discovery and Data Mining, pp. 110-121,2000.
[5] H.C.L. Martin, A.T.F. Mario, and A.K. Jain, "Feature Saliency in Unsupervised Learning," Technical Report, Michigan State Univ.,2002
[6] T. Liu, S. Liu, Z. Chen, and W.-Y. Ma, "An Evaluation on Feature Selection for Text Clustering," Proc. Int'l Conf. Machine Learning(ICML'03), pp. 488-495, 2003.
[7]R. Weiss, B. Velez, M.A. Sheldon, C. Namprempre, P. Szilagyi, A.Duda, and D.K. Gifford, "HyPursuit: A Hierarchical Network Search Engine that Exploits Content-Link Hypertext Clustering, "Proc. Seventh ACM Conf. Hypertext, pp. 180-193, 1996.
[8] K. Nigam and R. Ghani, "Analyzing the Effectiveness and Applicability of Co-Training," Proc. Information and Knowledge Management, pp. 86-93, 2000.
[9] A. Blum and T. Mitchell, "Combining Labeled and Unlabeled Data with Co-Training," Proc. Conf. Computational Learning Theory, pp.92-100, 1998.
[10]M. Montague, "Metasearch: Data Fusion for Document Retrieval,"PhD Thesis, Dartmouth College, 2002.
[11]”DB SCAN algorithm” , Wikipedia, the free encyclopedia.

Mrs.K.Parimala.,MCA, Research Scholar in Computer Science in Alagappa University, Karaikudi, INDIA, under the guidance of **Dr.V.Palanisamy**, Professor & Head In-Charge, Department of Computer Science & Engineering, Alagappa University. Currently working as Assistant Professor, in NMS SVN College with a teaching experience of 14 years.

Dr. V.PalaniSamy, MCA, MTech (Adv.IT), Ph.D, Professor & Head In-Charge, Department of Computer Science & Engineering, Alagappa University, Karaikudi, TamilNadu, INDIA, specialized in Algorithms, Wireless Networks & Network Security. He has 20 years of teaching experience and 15 years of Research Experience.

Internet Geographic Information Security Supervision Platform Architecture Research

CHEN Wanzhi [1,2], LIU Jiping [2] and WANG Yong [2]

[1] Liaoning Technical University, FuXin, Liaoning Province,123000, P.R.China

[2] Chinese Academy of Surveying and Mapping, BeiJing,100830, P.R.China

Abstract

Nowadays every link of geographical information from products manufacture, management, and distribution to application would be faced with tremendous security risks, which urgent need to establish a geography information security management and monitoring system with unified and multilevel action manner by the new technology and information measures. This article presents the overall architecture of the platform and typical services node, based on analysis of construction requirements for the Internet geographic information security supervision platform, laid the theory ground work for further research of the final geographic information security management measures and technique adapt to the development of science technology level and the requirements of social widely, and realized the overall aim of geography information security supervision in the end.

Keywords: *Internet, Geographical Information, Security Supervision, Platform Architecture.*

1. Introduction

With the rapid development of our surveying and mapping cause, the geographical information industry developed rapidly and geographic information resources are continuously enriched, which has been playing an increasingly important role in the national economy, social development and national defense construction. The rapid popularization of Internet and cloud computing, Internet of things, etc. which new information and modern surveying and mapping technology has been developed rapidly, those making people could be efficiently to acquire and use a variety of the geographical information, at the same time, following a serious challenge of the geographical information safe at digital environment from production, management and approval to release and application. Users can easily obtained geographical information of high precision image and many sensitive areas through online service of the geographical information, also tagging on the map however their interest in the information by tag services features. The security problems for geographic information services have become increasingly prominent. Relevant departments of the country has always attached great importance to the geographical information safe of production, management and supervision of Internet web map, geographic information system services site. However, compare with higher maturity of Internet information security technologies and products, we still lack geographic information security management measures and technique, which adapt to the development of science technology level and the requirements of social widely. And foreign developed countries had set up special organizations to strengthen the geographical information resources collected and protected, such as "National Geospatial-Intelligence Agency" (NGA) for the US government, the same level with FBI, CIA and NSA, specialized responsible for the important geographical information collecting and supervision.

Geographical information is state infrastructure and strategic information resource, involved in national economic, political and military secrets and sensitive information. Thus aim to every link of geographical information products manufacture, management, distribution and application would be faced with tremendous security risks, which urgent need to establish a geography information security management and monitoring system with unified and multilevel action manner by the new technology and information measures, realized the overall process of supervision informationize for geographical information products manufacture, management, distribution and application, improving comprehensive ability of the map and geographic information credible, controlled and security in the course of management, distribution and services, and then the rapid respond ability of Internet (the Internet of things) geographic information security supervision come into being, set up more authorities cooperative mechanism of the geographical information security supervision, provide administrative law enforcement support for Internet maps

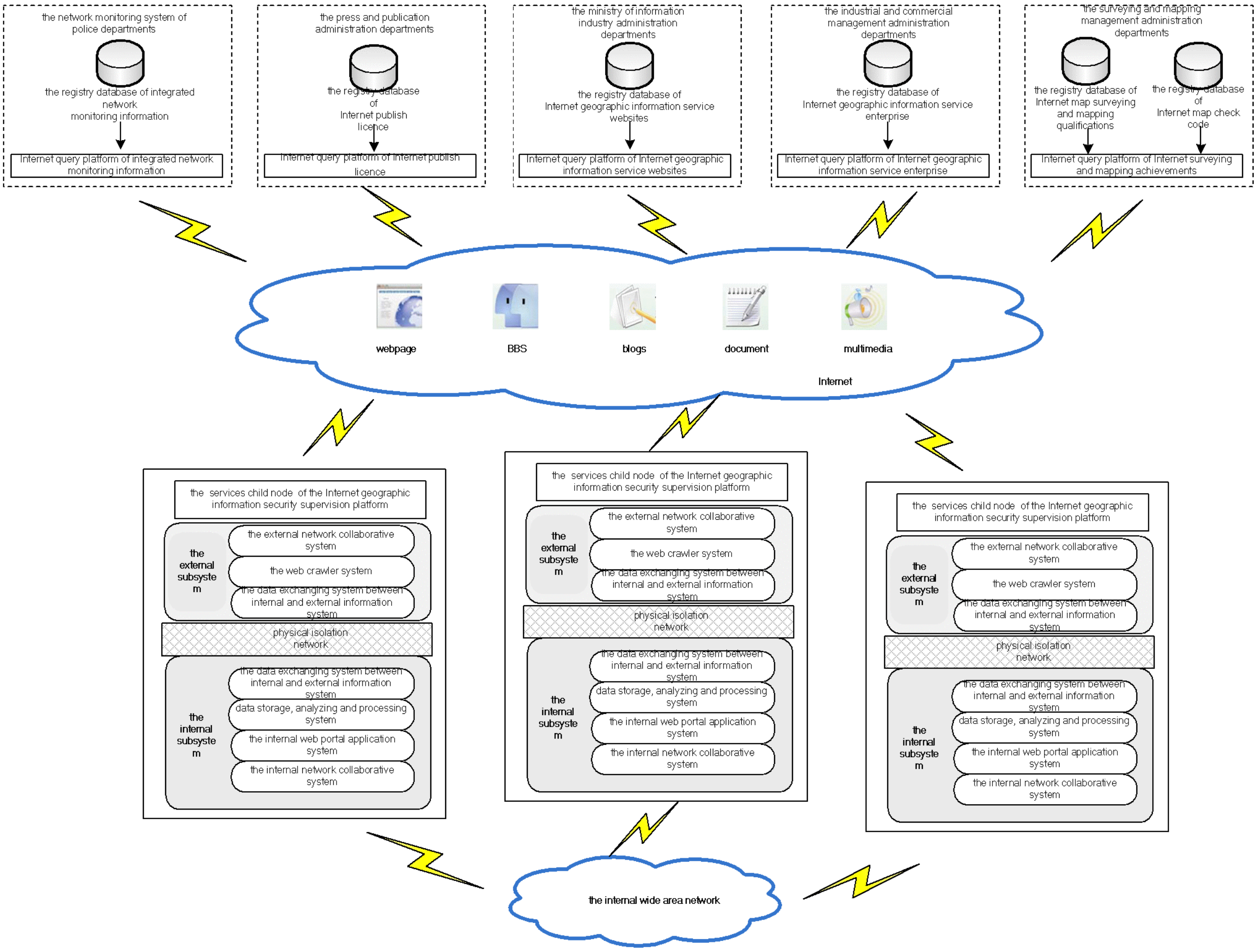


Fig. 1 The overall architecture of the geography information security supervision platform.

and geographical information lawbreaking behavior, ensure the national geographic information security. The geographical information security management project mainly including two kinds of development task, the secret-related geographical information security and the Internet (the Internet of things) geographical information security supervision. This article expatiates the research basis of the second field.

2. The Construction Requirements and Overall Architecture

The Internet geographical information security supervision is a collaboration work that need more administrative authorities and department in different regions to completed commonly, mainly responsibility for the administrative departments of surveying and mapping, the division of labour for relevant departments according to law, cooperate perfectly and set up a shortcut cooperative mechanism. In the end, supervision of Internet map and network geographical information service standardized and institutionalized, information shared and supervision profession streamlined, consist of to draw up the directory system of secret-related geographical information security, the security level of Internet geographical information security, the approved standard of networked map and geographical information; to establish the standard system of Internet geographical information security supervision perfectly; to give great impetus to construction of the map and geographical information approved network system; then to realized the approved procedure management networked for achievements in surveying and mapping, efficient identification and semi-automatization checked for important geographical element and secret-related geographical information by cooperative approved for map and intellectual processing technology, such as

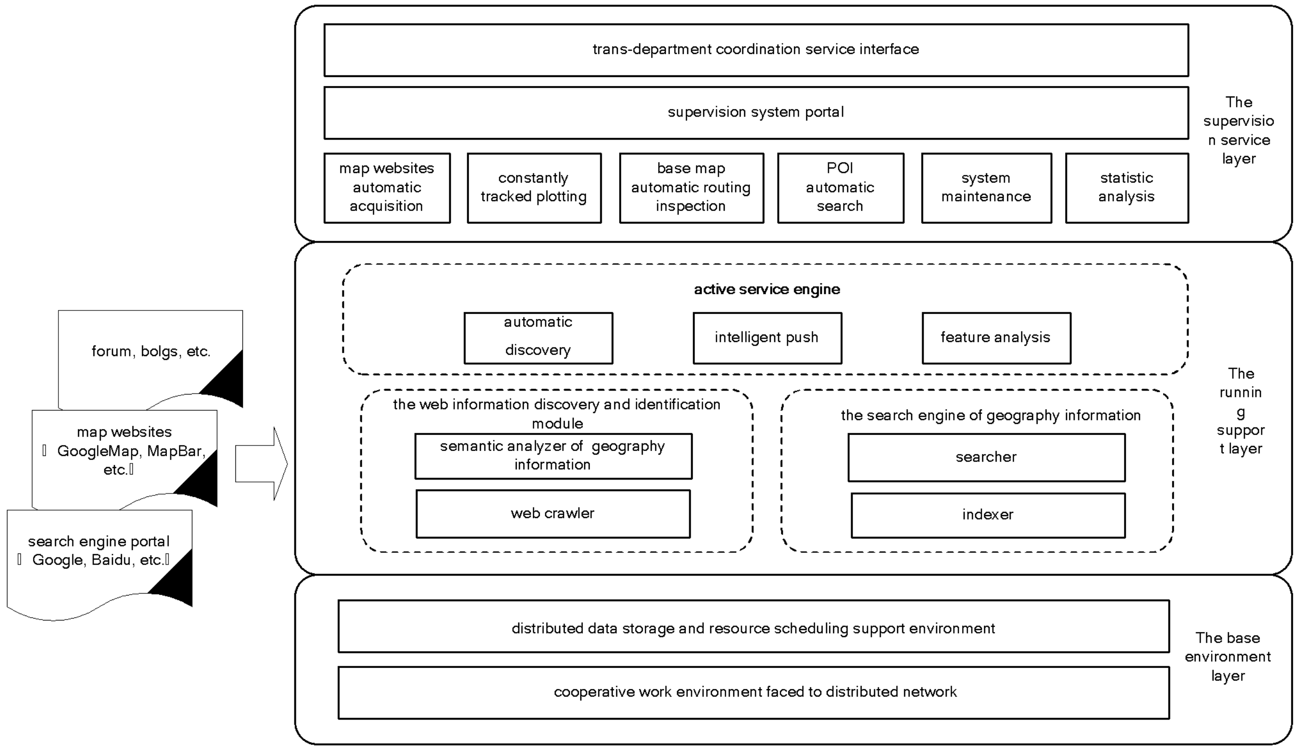


Fig. 2 The typical services node architecture of the Internet geographic information security supervision platform.

semantic analysis and pattern recognition; and in another hand, to accelerate the construction of Internet geographical information security supervision; to strengthen monitoring routinely of key point site for map and geographical information service, handle basic tasks: automatic discovery for Internet secret-related geographical information, intelligence identification, constantly tracked, investigate and collect evidence, analysis and evaluation, etc. So next to introduce primary service for the Internet geographic information security supervision platform in details.

2.1 The Capability of Discovery and Identification for Internet Geographical Information Service Site

The system through the Internet search engine scanning continuous, found them who provided geographical information service site, then remarked and archiving into database. Checking the civil sites' ICP/IP with the Internet recorded system of Ministry of Industry and Information Technology, at same time, making approval examine and verify of geographical information system more strictly, and recording all related items.

2.2 The Capability of Automated Acquisition and Analysis for Internet Geographical Information

The system analysis the content of the geographical information services site based on the acquired site records database, collect the effective information, and then place on formatted file.

2.3 The Capability of Intelligence Identification, Constantly Tracked and Take the Evidence, Signature Analysis for Secret-related Geographical Information

The system analysis the content of related to national security for politics, military, major objective, energy sources, etc., from acquired information by relevant analysis methods. Then relevance content would be recorded and marked, for subsequent information retrieval.

2.4 The Capability of Efficient Management, Sustainable Service and Quick Reaction

The system could acquire effective information effectively; the system maintainer could manage the whole system maintenance and setting easily, and control system application through short-term training. And the system could respond to operations of administrator rapidly, master the system current running condition in time. The system's redundant configuration could recovery from accidental interrupted situation.

2.5 The Capability of Linked Up and Down, Cooperate with Other Departments, Support Administrative Law Enforcement

The system record database could distribute shared for multi-level interactions, different levels and different departments by networked and operational collaboration. Legitimate users can easily log in platform system to extract relevant information, to provide evidence of law enforcement.

As suggested above, according to the national geographic information public services platform three levels construction requirement - shared, distributed and online; the basic principles – "makes overall plans, cent pace is carried out; overall planning and guidance, decentralized management"; the construction requirements for the Internet geographic information security supervision platform. The system platform's general architecture support by the newest computer network integration technology, which face to each department supervision on web maps and geographic information service websites for Internet with distributed B/S architecture, as shown in Fig.1. The information resources of platform derived from the each department network retrieval service, which use local distributed database, shared but isolated model between intranet and extranet. In fact, the local distributed database independent maintenance by each department except provided the information interface classes to the platform, not other sensitive data, each department could achieve the websites supervise object relevant details through retrieve function for the platform fast. This model more conducive to cooperation and distribution of responsibilities, sharing information between the multi-industry departments.

3. The Typical Services Node Architecture

The centre of Internet geographic information security supervision in the overall architecture is the services node with distributed, redundant and shared function for platform; consist of three-layers construction – base environment, running support and supervision service, as shown in Fig.2. The function of the typical services node as follows.

3.1 The Web Crawler System

To multiple sources and heterogeneous information for geographic information service from webpage, forum, Blogs, etc. of Internet, the system collect and store to related information with distributed multi-agent crawler network software, centralized logic and distributed physics search strategy. Which support to collect the spatial information automatically, classify and file them, including maps, vector data, POI(Point Of Interest) layer, geographical name and address, tag of user, map service, text, photograph, etc.

3.2 The Data Exchanging System between Internal and External Information System

The system realized data safety transmission with "ferry" operating mechanism, which must be internal and external network data security, dynamic and sustained in physical isolation network environment.

3.3 The Collaborative System between Internal and External Network

The system support the geographical information security exchange function between all WANs levels of nodes, dynamic and linkage; support the online or off-line mode data security exchange function between all supervision service node; support data exchange and collaborative function between all service node. At the same time, real-time monitoring all node system's services, security, running, etc.; recording all node system's running key information and log; alarm processing in the emergency; collecting popular feelings.

3.4 The Data Storage, Analyzing and Processing System

The system support efficient storage and fast scheduling for vast quantities of the Internet heterogeneous information by distributed storage mechanism, and data backup and storage management. Then the system analyzing and processing the Internet heterogeneous information on the basis of the structure characteristic for all classes of geographic information, support splitting and polymerizing plotting function for KML format file, and for point, line, plane object fast, rapid assessment function for secret-related content, support to identification of secret-related object from geographic information, collect the space-locating information, so that the geographical positions, objects, spatial attributes could be lot size automatic identification, collection and achievement management. In the end, the system could automatic discovery, automatic mark, and automatic alarm for secret-related content on the basis of semantic feature of secret-related geographic information.

3.5 The Internal Web Portal Application System

The system aim to routine geographic information security supervision, support to core application service function by web portal, mainly including: constantly tracked for Internet information; automatic alarm and collect evidence for secret-related geographic information; administrative supervision, and SMS presentation, mail alarm, Word/Excel import or export, automatic generation for official document or report forms, etc.; integrated with OA system; content statistic analysis by time, space, subject, etc. and user behavioral analysis, and forecast.

For the core function of supervision, the main parts are the web crawler system, data storage, analyzing and processing system, and internal web portal application system. The technicality key and difficult point are the automated acquisition and analysis for Internet geographical information, distributed storage and scheduling. Refer to the distributed cluster acquisition, multithreading concurrency scanning, real-time monitoring, metadata extraction, automatic classification, repetitive computation, automatic index, automatic keywords extraction, intelligent search and retrieval system in database, veridical search application, network popular feelings monitoring, special topic database, etc.

4. Conclusions

This article presents the overall architecture of the platform and typical services node, based on analysis of construction requirements for the Internet geographic information security supervision platform, laid the theory ground work for further research of the final geographic information security management measures and technique adapt to the development of science technology level and the requirements of social widely, and realized the overall aim of geography information security supervision in the end.

Acknowledgments

This work was supported in part by Hi-Tech Research and Development Program of China (863) under Grant Nos. SS2012AA120702.

References

[1] Surveying and Mapping Development Strategy Research Group for State Bureau of Surveying and Mapping, "Research report of Surveying and Mapping Development Strategy", unpublished.
[2] LIU Jiping, et al. "Research report of the geography information security supervision platform construction", unpublished.
[3] HAN Quanwei, "Thinking on Establishing National Internet and Web Geographic Information Services Supervision Platform", Geomatics & Spatial Information Technology, vol. 32(6), pp. 70-73, Dec 2009.
[4] Jia Lin, Alexander Halavais, Bin Zhang, "The Blog Network in America: Blogs as Indicators of Relationships among US Cities", CONNECTIONS, vol. 27(2), pp. 15-23, 2007

CHEN Wanzhi received his B.Sc. and M.Sc. degree in electrical engineering from Liaoning Technical University, FuXin, Liaoning Province, P.R.China. Now he is a PhD candidate of Chinese Academy of Surveying and Mapping.
His main research interests include GIS and information security. He have participated in several projects include two National Natural Science Foundations of China and one National Grand Fundamental Research 863 Program of China and published more than 10 research papers from 2007.

LIU Jiping received an M.S. degree from WuHan Technical University of Surveying and Mapping, WuHan, China, in 1992 and a Ph.D. degree from Information Engineering University of the People's Liberation Army, ZhengZhou, China, in 2004. Now he is a Research Prof and PhD supervisor at Chinese Academy of Surveying and Mapping. His research interests cover GIS of E-government and Spatial Data Mining and information security.

WANG Yong received an M.S. degree from Chinese Academy of Surveying and Mapping. Now he is a Assistant Prof at Chinese Academy of Surveying and Mapping. His research interests cover GIS of E-government and Spatial Data Mining and information security.

Study on Steering by Wire Controller Based on Improved H∞ Algorithm

Huipeng Chen [1,2], Qiqi Yu [1], Lingjun Zhu [1] and Guojin Chen [1]

[1] School of Mechanical Engineering, HangZhou DianZi University, Hangzhou 310018, Zhejiang, China

[2] Postdoc of Huazhong University of Science & Technology, Wuhan 430074, Hubei, China

Abstract

With the continuous development of vehicle and electronic industry, SBW (Steering by Wire) is replacing the traditional steering device of vehicle. Firstly this paper introduces the principle and structure of SBW, then discusses how to model the kinetic equation of SBW and how to simplify the model reasonably, and then obtains the control model of SBW. To meet the system requirement of robust performance, this paper selects the control strategy based on H∞ algorithm. According to the further observation and study on the curve of S/T singular value in H∞ algorithm, put forward a method which constructs a closed loop function and then back-step the system controller on the basis of key parameter in system frequency. The analysis of the time domain and the frequency domain of SBW, the simulation result under the interference and parameter perturbation of the model show that the controller designed by this method is simple and effective, and the controller owns perfect robust stability and robust performance.

Keywords: *Steering by Wire, Robust, H∞ algorithm.*

1. Introduction

With the continuous development of electronic and control technology recently, electronic power steering of the vehicle has a rapid development. The application of the various kinds of active and power steering system improves the transfer characteristic of steering angle, the steering response characteristic and the stability of steering for vehicle [1, 2]. However, nowadays the mechanical connect of the steering system leads that the steering performance has a strong non-linear time-varying characteristic with the change of speed, steering angle and the condition of the road adhesion. To make the vehicle drive along the expected path, drivers must adjust the characteristics of themselves frequently, so it would increase the mental and physical burden, and the nonprofessional driver would not adapt it especially [3]. So many researchers from different countries are studying the new technology for the steering system to solve the above problems, under this circumstance, the Steering-By-Wire system (SBW) emerge as the times requires. In recent decade, almost all the large scale automobile manufacturers have carry out deep studies on SBW. And the new technology has been applied in many new automobiles, such as the concept vehicle C-Grosser of CITROEN Corporation in France, the concept vehicle 8129 of the Daimler Chrysler Corporation in Germany, the concept vehicle Ray of KIA Corporation in South Korea and the test vehicle FASCarll designed by the German aerospace centre and traffic technology study institute in 2011. What's more, the manufacturers of the components, the vehicle design corporations and many universities have studied the SBW.

The main object of studying SBW control strategy is how to keep the stability, tracking and anti-interference under the complicated condition of work and road. At present, there are a large number of control strategies used in the SBW, and several typical cases acquire some good effect, for instance PID, LQG, H∞ etc [4, 5]. Aiming at the unreasonableness of design object and interference limit of LQG, the scholar Zames from Canada put forward the control idea of H∞ [6]. After more than twenty years development, H∞ control theory has been one of theoretical systems that solve the robust control problems successfully and perfectly. The selection for the weighting function is crucial and it will determine the performance of SBW. However, to select a appropriate weighting function should rely on the good engineer experience of designing SBW, and should take iterative and a lot of calculus, these problems bring many difficulties in designing the H∞ controller of SBW.

This paper avoids selecting the weight function with the method of reverse which is based on improving the H∞ control algorithm. This method constructs the closed-loop transfer function of system based on the H∞ algorithm directly with the further observation and study on the S/T singular curve by the author, and then reserves a closed-loop system with an expected S/T curve which meets the robust performance requirements of SBW. This is a kind of simplified H∞ circuit shaping algorithm based on the engineer significance. Its physical conception is distinct and the solving process is very easy, the orders of final controller are low. The S/T curve and the result of the simulation on the step response show that the robust stability and the robust performance of improved H∞ algorithm are better than the traditional H∞ algorithm.

2. KINETIC MODEL OF SBW AND ITS ANALYSIS

SBW adopts the X-By-Wire technology (XBW), and transfers the signal to the Electronic Control Unit (ECU), and then the commands transmitted by ECU will control the steering execute assembly to finish the steering commands, and finally the driver realize the driving intention. SBW gets rid of the limit of traditional mechanical connect completely. It can design transfer characteristics of the angle and force freely in theory and it offers tremendous space to design the steering characteristics. It has a extensive application market and vast development potential.

2.1 THEORY OF SBW STRUCTURE

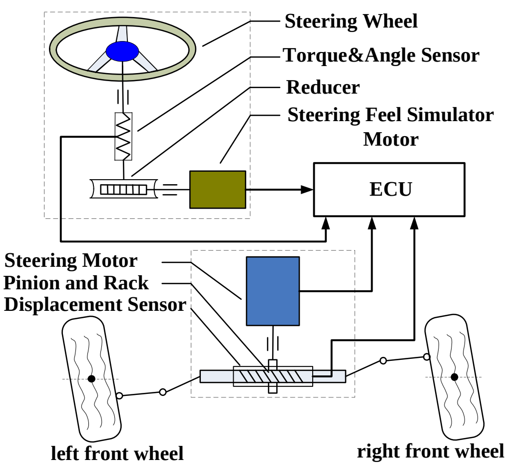


Fig.1 theory schematic diagram of SBW structure

The structure and theory schematic diagram of SBW studied in this paper is shown in figure 1, mechanical part contains the hand wheel part and the steering wheel part. The hand wheel part is used to simulate the transfer of driver's intention and the torque of the hand wheel; it contains the following main hardware: motor for simulating the steering feel, decelerating mechanism, rotation sensor, current sensor, torque sensor, general hand wheel and steering column. Steering part's function is to make the steering wheel turn; it contains the motor for executing turning, decelerating mechanism, rotation sensor, current sensor, rack displacement sensor, general steering gear-rack. Apart from the mechanical hardware, the distinct difference between SBW and traditional steering system is that SBW is equipped with a controller which has three functions: control the motor for simulating the steering feel, control the motor for executing turning and the fault tolerant control for the whole system's main components.

2.2. KINETIC MODEL OF SBW

As the figure 2 shown, SBW is divided into two parts according to the motor position, and the figure 2(a) shows the schematic diagram force of the simulation steering feeling assembly and the hand wheel assembly, and figure 2(b) shows the schematic diagram force of executing turning assembly.

In figure 2(a), I_h , D_h and D_m represent the moment of inertia of the hand wheel, damping coefficient of steering column and motor of simulating steering feel respectively. T_h is the steering torque and T_{hr} is resistance torque of the hand wheel. T_{ts} is resistance torque measured by torque sensor. T_m and T_{mr} represent output torque and resistance torque of motor. δ_h and δ_m represent the angle of the hand wheel and the motor of simulating steering feel. $f(\delta_h,\dot{\delta}_h)$, $f(\delta_m,\dot{\delta}_m)$ represent the nonlinear dry friction resistance torque of the hand wheel assembly and the motor for simulating the steering feel. and its kinetic equation as following:

$$I_h\ddot{\delta}_h + D_h\dot{\delta}_h + f(\delta_h,\dot{\delta}_h) = T_h - T_{hr} \quad (1)$$

$$I_m\ddot{\delta}_m + D_m\dot{\delta}_m + f(\delta_m,\dot{\delta}_m) = T_m - T_{mr} \quad (2)$$

$$T_{ts} = k_{ts}(\delta_h - i_{fw}x_r) + D_{ts}(\dot{\delta}_h - i_{fw}\dot{x}_r) \quad (3)$$

In figure 2(b), I_f , D_f is the moment of inertia of the execute motor, damping coefficient, and its angle is δ_m , the output torque is T_f and the resistance torque is T_{fr} . $f(\delta_f,\dot{\delta}_f)$ is the resistance of nonlinear dry friction of motor for executing steering. And its kinetic equation as following:

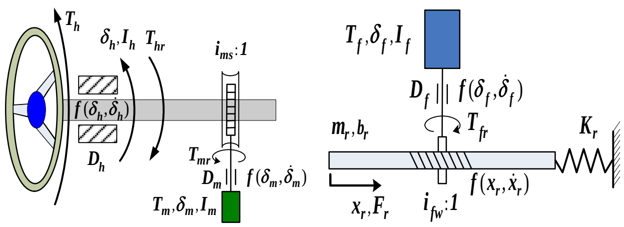
$$I_f\ddot{\delta}_f + D_f\dot{\delta}_f + k_f\delta_f + f(\delta_f,\dot{\delta}_f) = T_f - T_{fr} \quad (4)$$

The model of small gear-rack is similar to the traditional steering system and the model can be expressed as following:

$$m_r\ddot{x}_r + b_r\dot{x}_r + f(x_r,\dot{x}_r) = T_r / r_p - F_r + F_d \quad (5)$$

m_r and b_r represents the quality of rack and its damping coefficient, x_r is the displacement of rack. T_r is the reaction torque which acts on the small gear-rack. F_r is steering resistance and F_d is random interference resistance. i_{fw} is deceleration ratio of the motor for executing steering.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org



(a) (b)

Fig. 2 Force schematic diagram of SBW

According to the structure and force condition of SBW, the following relation can be concluded:

$$T_{mr} = T_{hr} / i_{ms}, T_w = i_{fw} T_{fr}, T_{hr} = T_{ts}, T_r = \rho T_{ts}$$

In the above relation, ρ is the coefficient of steering and i_{ms} is the deceleration ratio of the motor for simulating steering feel. Based on the above equation from (1)-(5) and the middle relation, the following kinetic equation of SBW could be obtaining:

$$\begin{aligned} & i_{fw}(D_{ts}\dot{x}_r + k_{ts}x_r) + T_h = \\ & I_h\ddot{\delta}_h + (D_h + D_{ts})\dot{\delta}_h + k_{ts}\delta_h + f(\delta_h, \dot{\delta}_h) \end{aligned} \tag{6}$$

$$\begin{aligned} & r_p m_r \ddot{x}_r + (r_p b_r + \rho_0 D_{ts} i_{fw})\dot{x}_r + \rho_0 k_{ts} i_{fw} x_r + f(x_r, \dot{x}_r) \\ & = \rho_0 (D_{ts}\dot{\delta}_h + k_{ts}\delta_h) - F_r + r_p F_d \end{aligned} \tag{7}$$

2.3. SIMPLIFY THE MODEL AND ANALYSIS

The resource of the steering resistance mainly comes from the effect between the road and tires, the inner steering system friction. The effect between the tire and the road is very complicated with the influence of the tire material, structure, pressure, vertical load and the working condition, and the inner friction of the system is quite complex, so the above factors lead that the steering resistance has the obvious nonlinear characteristic. In this paper, the overall analyze and accuracy model is not necessary for the steering resistance of the steering system. What we need is to grasp the relation between the front angle and the road resistance torque. So we can suppose the hypothesis that the relation between the front angle and road resistance torque is linear is tenable when we model based on the SBW.

Equivalent the aligning torque of front wheel as the linear spring and its stiffness is K_r, and it has the following expression:

$$F_r = K_r x_r \tag{8}$$

$f(x_r, \dot{x}_r)$ is the nonlinear dry friction torque of SBW, and its value is always small, so we can neglect it. So, the kinetic equation of SBW after simplifying:

$$T_h + k_s i_{fw} x_r = I_h \ddot{\delta}_h + D_h \dot{\delta}_h + k_s \delta_h \tag{9}$$

$$r_p m_r \ddot{x}_r + r_p b_r \dot{x}_r + (\rho_0 k_{ts} i_{fw} + r_p K_r) x_r = \rho_0 k_s \delta_h \tag{10}$$

We can get the following relation via Laplace transform and arrange them:

$X_r(s) = \rho k_{ts} \delta_h(s) / P(s)$,

$X_r(s) = \rho k_s T_h(s) / Q(s)$,

$\delta_h(s) = P(s) T_h(s) / Q(s)$

$X_r(s)$, $\delta_h(s)$, $T_h(s)$ is the Laplace transformation of x_r, δ_h, T_h, and the expressions of $P(s)$, $Q(s)$ as following:

$Q(s) = u_4 s^4 + u_3 s^3 + u_2 s^2 + u_1 s + u_0$,

$P(s) = r_p m_r s^2 + r_p b_r s + u_5$, $u_0 = k_s r_p b_r$,

$u_1 = D_h(\rho k_s i_{fw} + r_p K_r) + k_s r_p K_r$,

$u_2 = k_s r_p m_r + D_h r_p b_r + I_h u_5$, $u_3 = r_p(b_r I_h + m_r D_h)$,

$u_4 = r_p m_r I_h$, $u_5 = \rho k_s i_{fw} + r_p K_r$

3. ANALYSIS AND DESIGN OF ROBUST CONTROL SYSTEM

3.1. MODELING THE CONTROL SYSTEM OF SBW

According to the analysis of the kinetic model, we can conclude the following system block diagram as shown in figure 3.

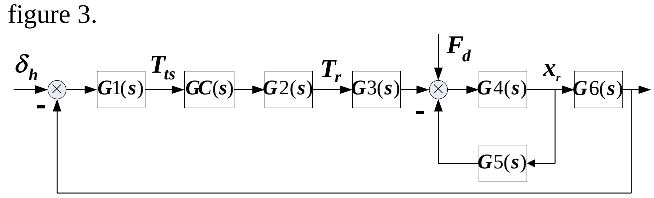


Fig. 3 System block diagram of SBW

$G1(s) = k_{ts}$, $G2(s) = \rho$, $G3(s) = 1/r_p$, $G6(s) = i_{fw}$, $G4(s) = 1/(m_r s^2 + b_r s)$, $G5(s) = K_r$ .Based on the work theory in the first chapter, SBW is an angle serve system actually that the controller control the power motor to offer the power according to the difference angle between the input and output that is the torque sensor signal. However, the road and work condition will affect the output signal from the sensor, and there is noise in the torque sensor, so the torque signal should be adjusted before transferring into the motor controller to make sure that the signal can be reflect the steering intention from the driver and then the power motor could be offer the accuracy power. Suppose the adjust controller $C(s)$, the improved mixed sensitivity H∞ circuit shaping in the robust control theory is used to design a controller which owns the robust stability and robust performance. Firstly, we supposed the input hand wheel angle $\delta_h = 0$, and transform the original system block diagram into standard H∞ feedback structure as shown in figure 4. In the figure 4, T_d is the random interference from the road, Δ is the model perturbation which contains uncertain parameters and non-structural uncertainty. WS , WR , WT is three extra weighting function to inhibit interference, limit control variable and improve the performance. w is the virtual input to meet the standard control structure. And the controller of this whole system is $K(s) = G2(s)G3(s)GC(s)$, and the control object is $G(s) = G1(s)G4(s)G6(s)/(1+G4(s)G5(s))$.

According to the SBW simulation parameters: $\rho = 4$, $k_{ts} = 121Nm/rad$, $r_p = 0.0088m$, $i_{fw} = 20$, $m_r = 5.28kg$, $b_r = 326.6N/(m/s)$, $K_r = 39951.6N/m$, we can obtain the specific transfer function expression:

$G1(s) = 121$, $G2(s) = 4$, $G4(s) = 1/(5.28s^2 + 326.6s)$, $G3(s) = 1/0.0088$, $G5(s) = 39951.6$, $G6(s) = 20$.

So, we can finally calculate the control object $G(s) = 2420/(5.28s^2 + 326.6s + 39951.6)$.

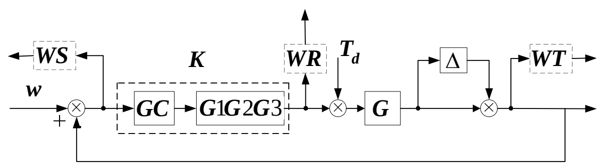


Fig. 4 Block diagram of controller

3.2.IMPROVED H ∞ ROBUST CONTROL ALGORITHM

H∞ robust control is a theory that we can obtain a controller with the robust performance by optimizing the infinite norms of some performance index in the H∞ space (that is Hardy space, the following brief write down for H)[7]. The H∞ robust control theory provides methods to solve the system robust control problems, for example the model may have uncertainties in a certain range and exist outside interference signal.

Uncertainty model includes two parts that one is uncertainty of non-structural uncertainty namely the dynamic characteristics of high frequency without modeling, another one is uncertainties of model parameters. The general means of modeling the uncertainty object uses an assembly to represent the object model; this assembly could be structured or unstructured. In this paper, the following unstructured multiplicative uncertainty is the unified expression of the above two uncertainty:

$$G^{'}(s) = G(s)(1+\Delta) \tag{11}$$

Generally speaking, multiplicative perturbation Δ has the high pass characteristic. What's more, we don't demand the display expression about Δ based on the standard frame of H∞, what we need is the knowledge of corresponding limit value $\|\Delta_{\max}\|_\infty$. H∞ mixed sensitivity control strategy shapes the closed transfer function such as sensitivity function and complementary sensitivity by gain shaping algorithm directly, so we can eliminate high peak value problem which may appear in the open loop gain shaping. Replace the effects of Δ with weighting function, and ensure the weighting function $WT(s)$ through $\|\Delta_{\max}\|_\infty$, and they offer great flexibility for the optimizing problem. Here we take the upper limit of Δ as the weighting function, and guarantee the system's robust stability with $WT(s)$ when the model is under perturbation, then can get rid of the Δ which is expressed in the original perturbation model, which means Δ =0.Moreover the abilities to anti-interference and signal tracking can be guaranteed through weighting function $WS(s)$. But it is a key to choose the weighting function in this process; it determines the performance of control system directly. And to obtain the expected weighting function, designers should perform large calculus in repeated iteration to accumulate the practical experience, and there is no shortcut[8].

Based on the observe and study the curve of S/T mixed sensitivity singular value in the H∞ circuit shaping control, from the practical engineer significance, constructing the complementary sensitivity function T according to the bandwidth frequency, high frequency asymptote slope, the largest singular value and then back stepping the controller K. Ascertain the shape of sensitivity function S through the correlation of S and T indirectly, and then guarantee the robust stability and robust performance of system .

Suppose that the bandwidth frequency closed loop demanded is $1/T_1$, to make the construction of T easier, here the corner frequency is approximate to bandwidth, the frequency of high frequency asymptote is -20ndB, n is an integer which range from 1～3, when n is too large and the control order will rise, this phenomenon is adverse to the controller and the control effect is not improved obviously. To guarantee the system to track the target value with non static difference, choose the largest singular value equals to 1 and then construct complementary sensitivity function T as following:

$$T = 1/(T_1 s + 1)^n = GK/(1+GK)$$

The controller with the ideal S/T curve:

$$K = 1/[(T_1 s + 1)^n - 1]G$$

The solving process is very easy which avoids a lot of iterative calculus on weighting function, and it is a simplified H∞ circuit shaping algorithm based on the engineer significance.

3.3. design the control system of SBW

To guarantee system to track the reference signal w with non static difference, choose the largest singular value of complementary sensitivity function equals to 1. The bandwidth of system decides the respond speed and the SBW system demands great quality respond speed, so the value of bandwidth here is not less than 100rad/s. To restrain influence on the control error produced by the uncertain signal such as the noise of sensor, and guarantee the system's robust performance, choose the slope of high frequency asymptote equals to -60dB/dec. Thus the three order inertial system spectrum curve with the largest singular value 1 is constructed by the singular value curve of T. To make the calculation convenient, the corner frequency is approximate to the bandwidth frequency, and then we obtain:

$$\frac{1}{((1/100)s+1)^3} = \frac{G(s)K(s)}{1+G(s)K(s)}$$

That is the controller of SBW:

$$K(s) = \frac{5.28s^2+326.6s+39951.6}{0.00242s^3+0.726s^2+72.6s}$$

And then regulate the controller as following

$$GC(s) = \frac{K(s)}{G2(s)G3(s)} = \frac{5.28s^2+326.6s+39951.6}{1.1s^3+330s^2+33000s}$$

To compare with the ordinary H∞ mixed sensitivity control strategy, referring to the method to choose weighting function in the literature[6], the this paper choose the three parameters of H∞ mixed sensitivity: $WS = 15/(s+0.5)$, $WR = 0.01$, $WT = 58(s+30)/(s+6000)$. By working out the function mixsyn(G,WS,WR,WT) with the aid of mixed sensitivity problem of the new robust control toolbox in Matlab, the process of designing the H∞ mixed sensitivity controller for the object $G(s)$ becomes more easier. And then we can conclude as following:

$$K'(s) = \frac{657039.8671(s+6000)(s^2+61.86s+7567)}{(s+0.5)\ (s+481.2)\ (s+5984)\ (s+1.404e004)}$$

And then regulate the controller as following:

$$C'(s) = \frac{K'(s)}{G2(s)G3(s)}$$
$$= \frac{1445.4877(s+6000)(s^2+75s+3616)}{(s+561.1)(s+1618)(s+5980)(s+0.5)}$$

4. SIMULATION AND ANALYSIS ON THE CONTROL SYSTEM

The controller designed by the improved H∞ algorithm is a three order controller, and the traditional H∞ controller designed by the choosing the weighting function is a four order controller.

By simulating and analyzing the unit step response of the rack displacement under the influence of two kinds controller and without controller, we can get the unit step response curve of front wheel angle as shown in figure 5. According the curve, when without controller works, the settling time for the unit step response is 0.138 seconds and the overshoot is 63.8%; when the improved H∞ controller works, the settling time is 0.075 seconds and the settling time is 0.082 seconds when the traditional H∞ mixed sensitivity controller works. This shows that the improved H∞ controller has the better response performance when we guarantee the robust stability.

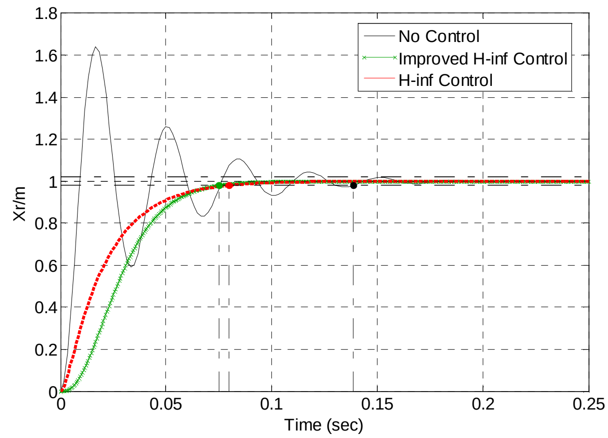


Fig. 5 step response curve of rack displacement

Analyzing the closed loop frequency spectrum of the controllers designed by the above two methods respectively and we can obtain the frequency spectrum as the figure 6. According to figure, two different kinds controllers all can obtain the expected curve shape of S/T curve; when the improved H∞ controller works, the gain of system sensitivity function is 0.316% in low frequency, the closed slope of closed loop is -60dB/dec; when the traditional H∞ controller works, the gain of system sensitivity function is 1.13% in low frequency, the closed slope of closed loop is -40dB/dec. So, the performance of robust and stability is not as perfect as the improved H∞ controller.

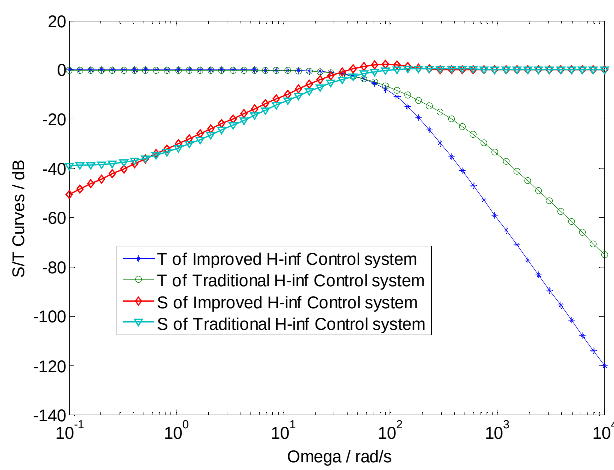


Fig. 6 frequency spectrum curve of SBW

5. CONCLUSIONS

This paper mainly discusses the dynamic modeling, analyzing and simplifying reasonably. Design the controller with the method based on the H∞ circuit shaping by deep analysis on the system. To avoid a large number of iterate calculus in selecting the weighting function of the classical H∞ arithmetic, put forward a new method to design an improved H∞ controller and design a controller for the SBW system. Simulate the system with the improved controller designed in this paper and traditional controller, then capture the results. The curves of closed loop frequency spectrum and the simulation results show that the method in this paper is simple and effective, and the improved controller has the better robust steady than the traditional controller. The interference step response shows that the new controller with the improved algorithm has a better robust performance. The results of the simulation taken under the condition of the parameter perturbation show that the new controller still can keep the system steady. These results declare that the controller design in this paper meet the control demands completely.

Acknowledgments

This work was supported in part by NSFC under Grant Nos. 51105125 and Open fund of Zhejiang Key Discipline of Marine Technology and Systems.

References

[1] Haggag, S., Alstrom, D., Cetinkunt, S., Egelja, "A Modeling, control, and validation of an electro-hydraulic steer-by-wire system for articulated vehicle applications", IEEE /ASME Transactions on Mechatronics,No.6, 2005, pp.688-692.

[2] Neville A.Stanton,Phipip Marsden. "From fly-by-wire to drive-by-wire:Safety implications of automation in Vehicles", Automotive engineering, Vol.109, No.9, 2001, pp.102-106.

[3] Choi S., Alonso J.J. and Kroo H.M. "Two-level multi-fidelity design optimization studies for supersonic jets", Journal of Aircraft, Vol.46, No.3, 2009, pp.776-790.

[4] Se-Wook OH, Ho-Chol CHAE, Seok-Chan YUN, et al. "The Design of a Controller for the Steer-by-Wire System", JSME international Journal Series C, Vol.47, No3, 2004, pp.38-44.

[5] M. Amrllein, P.T. Krein. "Dynamic simulation for analysis of hybrid electric vehicle systems and subsystems interactions, including power electronics", IEEE Trans. Veh. Technol. Vol. 54, No.3, 2005, pp. 825-836.

[6] Xiang Chen, Tiebao Yang, Xiaoqun Chen, "A Generic Model-Based Advanced Control of Electric Power-Assisted Steering Systems", IEEE Transactions on Control Systems Technology. Vol. 16, 2008, pp. 1289-1300.

[7] Paolo F., Francesco B., Jahan A., "Predictive active steering control for autonomous vehicle systems", IEEE Transactions on Control Systems Technology, Vol. 15, No. 3, 2007, pp. 566-580.

[8] Selamat H., Yusof R., Goodall R. M., "Self-tuning control for active steering of a railway vehicle with solid-axle wheel sets". IET Control Theory Appl., Vol. 5, No. 2, 2008, pp. 374-383.

Chen Huipeng, male, was born in 1978. He is both a doctor and a lecturer. From 2002 to 2005, he gained the postgraduate degree in the CNC center of School of Mechanical Science & Engineer of HUST. And from 2005 to 2009, he was studying in the CAD center of School of Mechanical Science & Engineer of HUST and gained the doctor degree in 2009. Chen has been engaged in the design based on Model-Based Design, hardware in the loop simulation technology, multidisciplinary design optimization, multi-domain unified modeling simulation and analysis, the electronic control system of automobile and the robust control research for a long time. He is the main host or project principal of several important projects and accumulates rich experience in the related research. More than 10 papers have been published in the academic journal and conference both in China and abroad.

Yu Qiqi, female, born in 1992, students, regular college course in reading, main research direction: Modeling and simulation, mechanical design and control.

Lingjun Zhu, male, born in 1992-06, undergraduate, major in the steer-by-wire and hardware in the loop.

Chen Guojin, male, was born in 1961 and he is both a doctor and professor. his major research contains the intelligence detection control and design of electromechanical system, multi-domain modeling and simulation.

Function Optimization Based on Quantum Genetic Algorithm

Ying Sun[1], Yuesheng Gu[2] and Hegen Xiong[1]

[1]College of Machinery and Automation, B.O.X 242, Wuhan University of Science and Technology, Wuhan, 430081, China

2Department of Computer Science, Henan Institute of Science and Technology, Xinxiang 453003, Henan, China

Abstract

Quantum genetic algorithm has the characteristics of good population diversity, rapid convergence and good global search capability and so on.It combines quantum algorithm with genetic algorithm. A novel quantum genetic algorithm is proposed ,which is called variable-boundary-coded quantum genetic algorithm (vbQGA) in which qubit chromosomes are collapsed into variable-boundary-coded chromosomes instead of binary-coded chromosomes. Therefore much shorter chromosome strings can be gained.The method of encoding and decoding of chromosome is first described before a new adaptive selection scheme for angle parameters used for rotation gate is put forward based on the core ideas and principles of quantum computation. Eight typical functions are selected to optimize to evaluate the effectiveness and performance of vbQGA against standard genetic algorithm (sGA) and genetic quantum algorithm (GQA). The simulation results show that vbQGA is significantly superior to sGA in all aspects and outperforms GQA in robustness and solving velocity, especially for multidimensional and complicated functions.

Keywords: function optimization; quantum genetic algorithm; variable-boundary coding; optimization algorithm.

1. Introduction

Quantum computation is a new and developing interdiscipline integrating information science and quantum mechanics. In the early of 1980's, Benioff[1] and Feynman[2] proposed the concepts of quantum computing. In 1994, Shor[3] presented a quantum algorithm used for factoring very large numbers, Grover[4] developed a quantum mechanical algorithm to search unsorted database in 1996. Since then, quantum computing has attracted serious attention and been widely investigated by researches. Nareyanan, Moore[5],and Han[6] proposed respectively quantum inspired genetic algorithm and genetic quantum algorithm in 1996 and 2000. These algorithms are inspired by certain concept and principles of quantum computing such as qubits and superposition of states. Chromosomes in these algorithms are probabilistically represented by qubits and so can represent a linear superposition of solutions. Many researches have found that these algorithms have excellent performance such as population diversity, rapid convergence and global search capability. Effective applications have been found in many domains such as shop scheduling [7,10], signal analysis[8], reactive power and voltage control[9], etc.

In classical quantum genetic algorithms, chromosomes are generally represented by two types, qubits and binary, during the algorithm procedure. Binary chromosomes are generated by observing (equating quantum collapsing in quantum mechanics) qubit chromosomes. The two types of chromosomes have the same length. As the more of dimension of optimization problems, the bigger of range of variables and the higher of precision of variables, the chromosome strings will become longer and then result in big memory requirement and long run time for a computer. In order improve this condition, this paper presents a novel quantum genetic algorithm, in which chromosomes are encoded by qubit and variable-boundary, to expect to short the length of chromosome strings and then cut down the memory requirement and speed up the run velocity of algorithm.

The organization of the remaining of this paper is as follows: In Section 2 the variable-boundary-coded quantum genetic algorithm is described in detail. Section 3 carries out the evaluation of effectiveness and performance of vbQGA by adopting eight typical functions and comparing with sGA and GQA. The result of this paper is summarized in the last Section.

2. Variable-boundary-coded Quantum Genetic Algorithm

Han[6] proposed a novel evolutionary computing method called a genetic quantum algorithm (GQA) and applied it to a well-known combinatorial optimization problem, knapsack problem. His research shows that GQA is superior to other genetic algorithm. Based on the GQA, we propose a novel quantum genetic algorithm called variable-boundary-coded quantum genetic algorithm, vbQGA, which we will introduce in this Section.

2.1 Representation in vbQGA

In GQA, the smallest unit of information is qubit. A qubit may be in the '0' state, in the '1' state, or in any superposition of the two. Based on the idea, in vbQGA, we represent the state of a qubit as follow:

$$|\Psi> \;=\; \alpha\cdot|x^l> \;+\; \beta\cdot|x^u> \tag{1}$$

where x^l and x^u are respectively the lower bound and the upper bound of some variable x, α and β are complex numbers that specify the probability amplitudes of the corresponding states. Obviously, a qubit may be in the 'x^l' state, in the 'x^u' state, or in any superposition of the two. The $|\alpha|^2$ and $|\beta|^2$ give respectively the probability that the qubit will be found in 'x^l' state and in 'x^u' state. Normalization of the state to unity guarantees

$$|\alpha|^2 + |\beta|^2 = 1 \tag{2}$$

Now suppose we have an *N*-dimension function optimization problem described as

$$\text{min: } f(X) = f(x_1, x_2, \cdots, x_i, \cdots, x_N)$$

$$\text{s.t.: } x_i^l \le x_i \le x_i^u, i = 1,2,\cdots,N \tag{3}$$

With respect to the chromosome k in generation t, the substring of variable x_i can be represented by qubit as follow:

$$q_{k,i}^t = \begin{bmatrix} \alpha_{k,i1}^t & \alpha_{k,i2}^t \\ \beta_{k,i1}^t & \beta_{k,i2}^t \end{bmatrix} \tag{4}$$

then, a whole qubit chromosome string for the *N*-dimension function optimization problem can be defined as

$$q_k^t = \left[\begin{array}{cc|cc|c|cc|c|cc} \alpha_{k,11}^t & \alpha_{k,12}^t & \alpha_{k,21}^t & \alpha_{k,22}^t & \cdots\ \cdots & \alpha_{k,i1}^t & \alpha_{k,i2}^t & \cdots\ \cdots & \alpha_{k,N1}^t & \alpha_{k,N2}^t \\ \beta_{k,11}^t & \beta_{k,12}^t & \beta_{k,21}^t & \beta_{k,22}^t & \cdots\ \cdots & \beta_{k,i1}^t & \beta_{k,i2}^t & \cdots\ \cdots & \beta_{k,N1}^t & \beta_{k,N2}^t \end{array}\right] \tag{5}$$

Apparently, the length of a qubit chromosome is $L = 2N$. Let s be the population size, then chromosome population in generation t can be described as

$$Q^t = \{q_k^t \mid k = 1,2,\cdots,s\} \tag{6}$$

2.2 Observation of qubit chromosomes in vbQGA

Observation in quantum genetic algorithm is similar to quantum collapse in quantum mechanics. In GQA, a probabilistic qubit chromosome will "collapse" into a binary chromosome through observation. However, in vbQGA, a qubit chromosome will "collapse" into a variable-boundary coded chromosome. For any a qubit $[\alpha_{k,ij}^t, \beta_{k,ij}^t]^T$ ($j = 1,2$), we generate a random number between 0 and 1, $r_{k,ij}^t$, if $r_{k,ij}^t \le |\alpha_{k,ij}^t|^2$, the qubit will be found in the 'x^l' state, otherwise, the qubit will be found in the 'x^u' state. With the substring of variable x_i of chromosome k in (4), we can "collapse" it into a substring of a variable-boundary coded chromosome, which we denote as $vb_{k,i}^t$. A $vb_{k,i}^t$ can be one of the four conditions defined as

$$vb_{k,i}^t \in \{[x_i^l \quad x_i^l], [x_i^l \quad x_i^u], [x_i^u \quad x_i^l], [x_i^u \quad x_i^u]\} \tag{7}$$

So, a whole variable-boundary coded chromosome may be, for example, is in follow form:

$$vb_k^t = [x_1^l \quad x_1^l | x_2^l \quad x_2^u | \cdots \quad \cdots | x_i^u \quad x_i^l | \cdots \quad \cdots | x_N^u \quad x_N^u] \tag{8}$$

and then the variable-boundary coded chromosome population in generation t can be described as

$$VB^t = \{vb_k^t \mid k = 1,2,\cdots,s\} \tag{9}$$

2.3 Rules of decoding

As described in (7), the substring of a variable-boundary coded chromosome with respect to variable x_i can be one of the four conditions. The four conditions correspond to four value regions (namely I, II, III and IV) of x_i, which are gotten by equally dividing the first quadrant, illustrated in Fig.1. Let $\Delta x_i = x_i^u - x_i^l$, then every region represents a value span of $\Delta x_i / 4$. The decoding rules of variable-boundary coded chromosome are given in Table 1.

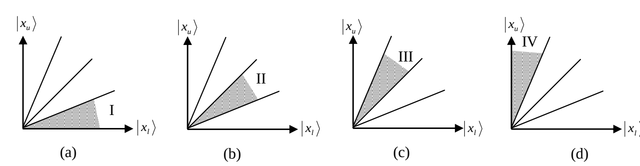


Fig.1. Four value regions of x_i corresponding to $vb_{k,i}^t$

In Table 1, r is a random number between 0 and 1. If $vb_{k,i}^t = [x_i^l \quad x_i^l]$, x_i will take a small value inclining to the lower bound, the corresponding value region is Region I. If $vb_{k,i}^t = [x_i^l \quad x_i^u]$ and $vb_{k,i}^t = [x_i^u \quad x_i^l]$, x_i will take a intermediate value, the corresponding value region are Region II and Region III respectively. Then if

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

$vb^t_{k,i} = [x^u_i \quad x^u_i]$, x_i will take a big value inclining to the upper bound, the corresponding value region is Region IV.

Table 1: Decoding rules of variable-boundary coded chromosome

$vb^t_{k,i}$	x_i	Illustrations
$[x^l_i \quad x^l_i]$	$x_i = x^l_i + r \cdot \frac{\Delta x_i}{4}$	Region I in Fig. 1. (a)
$[x^l_i \quad x^u_i]$	$x_i = x^l_i + (1+r) \cdot \frac{\Delta x_i}{4}$	Region II in Fig. 1. (b)
$[x^u_i \quad x^l_i]$	$x_i = x^u_i - (1+r) \cdot \frac{\Delta x_i}{4}$	Region III in Fig. 1. (c)
$[x^u_i \quad x^u_i]$	$x_i = x^u_i - r \cdot \frac{\Delta x_i}{4}$	Region IV in Fig. 1. (d)

2.4 Adaptive quantum rotation gate strategy

In many kinds of quantum-inspired algorithms, a primary updating operator for chromosomes is quantum rotation gate, which is defined as follow [6]:

$$U(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

where θ is rotation angle, which is generally looked up from a table. In our algorithm, quantum rotation gate for the substring of a qubit chromosome with respect to variable x_i is represented as

$$U^t_{k,ij}(\theta) = \begin{bmatrix} \cos(\theta^t_{k,ij}) & -\sin(\theta^t_{k,ij}) \\ \sin(\theta^t_{k,ij}) & \cos(\theta^t_{k,ij}) \end{bmatrix}, \; j = 1,2 \qquad (10)$$

In order to make the qubit chromosomes effectively converge to the fitter states, we put forward an adaptive rotation angles computing method, which is defined as

$$\theta^t_{k,ij} = sign\left\{\alpha^t_{k,ij} \cdot \beta^t_{k,ij} \cdot \left[f(b^{t-1}) - f(X^t_k) \right]\right\} \cdot \frac{b^{t-1}_i - x^t_{k,i}}{x^u_i - x^l_i} \times 0.05\pi \qquad (11)$$

where X^t_k is the solution k in generation t and b^{t-1} is best solution in generation $t-1$, $x^t_{k,i}$ and b^{t-1}_i are the value of their i th variable respectively, $f(X^t_k)$ and $f(b^{t-1})$ are their fitness respectively. $sign(\bullet)$ is a sign function which described as

$$sign(\bullet) = \begin{cases} +1 & if: \alpha^t_{k,ij} \cdot \beta^t_{k,ij} \cdot \left[f(b^{t-1}) - f(X^t_k) \right] > 0 \\ -1 & if: \alpha^t_{k,ij} \cdot \beta^t_{k,ij} \cdot \left[f(b^{t-1}) - f(X^t_k) \right] > 0 \\ \pm 1 & if: \alpha^t_{k,ij} = 0 \;\; and \;\; (b^{t-1}_i - x^t_{k,i}) \cdot \left[f(b^{t-1}) - f(X^t_k) \right] < 0 \\ \pm 1 & f: \beta^t_{k,ij} = 0 \;\; and \;\; (b^{t-1}_i - x^t_{k,i}) \cdot \left[f(b^{t-1}) - f(X^t_k) \right] > 0 \\ 0 & otherwise \end{cases} \qquad (12)$$

By analyzing formula (11) and (12) we can get:

Rotation angles are adaptive selected according to the difference values between $x^t_{k,i}$ and b^{t-1}_i. The bigger the difference values are, the bigger the absolute value of rotation angles are also.

The rotation directions, which can be gotten by (12), of quantum gate can make the solution converge to the fitter states. For example, if $f(b^{t-1}) - f(X^t_k) > 0$ (i.e., solution b^{t-1} is better than X^t_k) and $b^{t-1}_i > x^t_{k,i}$, then we should increase the $\beta^t_{k,ij}$ so as to augment the probability of ‘x^u’ state in the variable-boundary-coded chromosome. Hence, if $\alpha^t_{k,ij} \cdot \beta^t_{k,ij} > 0$ (i.e., in the first quadrant), the quantum gate should rotate in anticlockwise direction and the rotation angle should be positive. This just agrees with the result we can get from (11) and (12). Other conditions can be analyzed in the same method.

2.5 Procedure of vbQGA

The algorithm of vbQGA can be implemented as follows:

procedure vbQGA
begin
$t \leftarrow 0$
initialize Q^t
make VB^t by observing Q^t states
decode VB^t into $\boldsymbol{X}^t$ and evaluate them
store the best solution, $\boldsymbol{b}^t$, among $\boldsymbol{X}^t$
while (**not** termination-condition) **do**
begin
$t \leftarrow t+1$
make VB^t by observing Q^t states
decode VB^t into $\boldsymbol{X}^t$ and evaluate them
compare with $\boldsymbol{X}^t$ and $\boldsymbol{b}^{t-1}$, and update Q^t
using quantum gates
store the best solution, $\boldsymbol{b}^t$, among $\boldsymbol{X}^t$
end

end

3. Experimental Evaluation of VbQGA

3.1 Test functions

For the experimental evaluation of the algorithm presented in Section 2 eight typical test functions is chosen [11, 12].

De Jong function: De Jong function is defined as

$$F_1 = 100(x_1^2 - x_2)^2 + (1 - x_1)^2, \quad -2.048 \le x_i \le 2.048, \quad i = 1,2 \tag{13}$$

Although being mono-peak, DeJong function is ill-conditioned and intractable to search the global minimal solution: $f(1,1) = 0$.

Coldstein Price function: Coldstein Price function is described as

$$F_2 = [1 + (x_1 + x_2 + 1)^2 \cdot (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \cdot [30 + (2x_1 - 3x_2)^2 \cdot (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)], \quad -2 \le x_i \le 2, \quad i = 1,2 \tag{14}$$

This function has only one global minimal solution: $f(0,-1) = 3$.

Schaffer function: Schaffer function is given by

$$F_3 = 0.5 + \frac{\sin^2\sqrt{x_1^2 + x_2^2} - 0.5}{[1.0 + 0.001(x_1^2 + x_2^2)]^2}, \quad -100 \le x_i \le 100, \quad i = 1,2 \tag{15}$$

This function has only one global minimal solution: $f(0,0) = 0$.

Mono-pole and six-peak camelback function: Mono-pole and six-peak camelback function is formulated as

$$F_4 = 10 + \frac{\sin(1/x)}{0.1 + (x - 0.16)^2}, \quad 0 \le x \le 1 \tag{16}$$

The only one global maximal solution is $f(0.1275) = 19.8949$.

Dual-pole and six-peak camelback function: Dual-pole and six-peak camelback function is defined as

$$F_5 = (4 - 2.1x_1^2 + \frac{1}{3}x_1^4)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2, \quad -3 \le x_i \le 3, \quad i = 1,2 \tag{17}$$

This function has two global minimal solutions, i,e., $f(-0.0898,0.7126) = f(0.0898,-0.7126) = -1.031628$.

Multi-peak positive function: Multi-peak positive function is described as

$$F_6 = e^{-0.001x}\cos^2(0.8x), \quad x \ge 0 \tag{18}$$

This function has two local optimal solutions and one global maximal solution: $f(0) = 1$.

Ackley function: Ackley function is given by

$$F_7 = -20e^{-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_j^2}} - e^{\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_j)} + 22.71282, \quad -5 \le x_i \le 5 \quad i = 1,2,\cdots,n \tag{19}$$

This function has only one global minimal solution: $f(0.0,0.0,\cdots,0.0) = 0$. In the experimental evaluation we will take into account two conditions, $n = 2$ and $n = 10$.

Rastrigin function: Rastrigin function is formulated as

$$F_8 = 10n + \sum_{i=1}^{n}[x_i^2 - 10\cos(2\pi x_i)], \quad -5.12 \le x_i \le 5.12, \quad i = 1,2,\cdots,n \tag{20}$$

The only one global minimal solution is $f(-420.9687,-420.9687,\cdots,-420.9687) = 0$. We will take $n = 6$ for the experimental evaluation.

3.2 Optimization and results

In order to test and evaluate the effectiveness and performance of vbQGA, we will optimize the aforementioned eight functions with sGA, GQA and vbQGA.

In sGA, binary code，roulette wheel selection, one-point crossover and 0-1 mutation is adopted. The controlling parameters are: variable precision p=0.000001, population size s=50, crossover probability pc=0.8, mutation probability pm=0.01 and total generations of iteration t=500. The algorithm of GQA we used here is the same as that mentioned in [6]. With GQA we will take controlling parameters as: p=0.000001, s=10 and t=500, which are the same as those taken in vbQGA.

All the algorithms are integrated in a test system programmed by Java language. The test system is operated under the following environments: Microsoft windows XP 2002, Intel Pentium 1600*MHz* and 504*M* memory. For each algorithm 20 runs are performed with respect to the eight functions. The results are presented in Table 2.

In Table2, f_{opt} denotes the function value of optimum, $\bar{f}$ and sd are respectively average and standard deviation of function value over 20 runs, $\bar{t}$ (*sec/run*) represents the average elapsed time per one run. F_1 ~F_8 represent the corresponding functions described in the fore-subsection, e.g., F_1 represents De Jong function, F_2 represents Coldstein Price function etc.. We should notice that F_7 represents two conditions' Ackley function, so there are two lines of results, the upper one corresponding to $n=2$ and the lower one corresponding to $n=10$.

For giving a much clearer view of the results, the data in Table2 are illustrated by Fig.2, Fig.3, Fig.4 and Fig.5.

In the above four figures, the numbers of x-coordinate represent the index of the corresponding functions, e.g., '1' represents F1(i.e., De Jong function), '2' represent F2 (i.e., Coldstein Price function) etc.. We should also notice that '7'' and '7''' represent respectively Ackley function under condition $n=2$ and $n=10$.

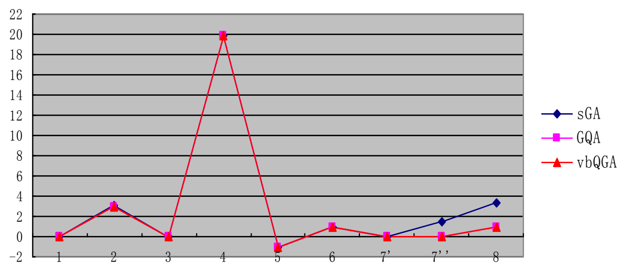


Fig.2. f_{opt} of the eight functions

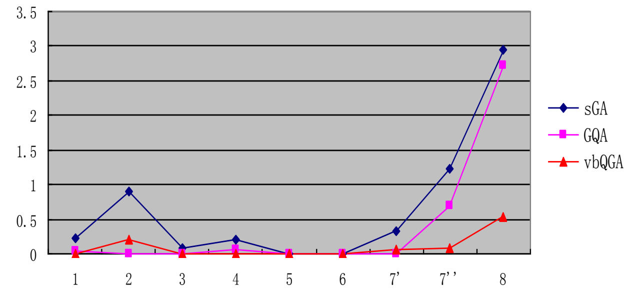


Fig.3. $\overline{f}$ of the eight functions over 20 runs

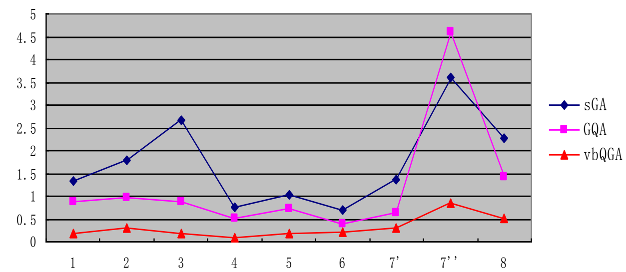


Fig.4. sd of the eight functions over 20 runs

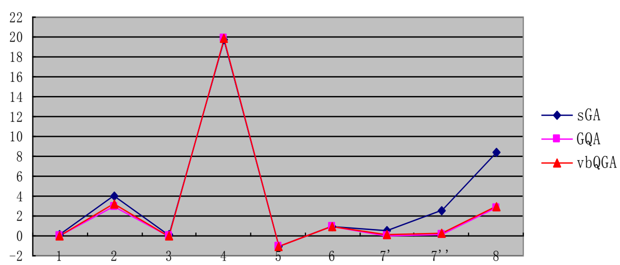


Fig.5. $\overline{t}$ of the eight functions over 20 runs

TABLE I. TABLE 2 RESULTS OF EXPERIMENTAL EVALUATION

Functions	sGA				GQA				vbQGA			
	f_{opt}	$\bar{f}$	sd	$\bar{t}$	f_{opt}	$\bar{f}$	sd	$\bar{t}$	f_{opt}	$\bar{f}$	sd	$\bar{t}$
F_1	0. 006	0. 145	0. 233	1. 326	0. 000	0. 015	0. 035	0. 883	0. 000	0. 002	0. 003	0. 187
F_2	3. 027	3. 951	0. 910	1. 801	3. 000	3. 000	0. 000	0. 983	3. 000	3. 228	0. 204	0. 299
F_3	0. 009	0. 121	0. 074	2. 668	0. 000	0. 006	0. 008	0. 886	0. 009	0. 013	0. 007	0. 193
F_4	19. 894	19. 790	0. 211	0. 756	19. 894	19. 810	0. 055	0. 530	19. 894	19. 894	0. 000	0. 096
F_5	−1. 032	−1. 024	0. 005	1. 042	−1. 032	−1. 030	0. 003	0. 730	−1. 032	−1. 031	0. 000	0. 188
F_6	1. 000	0. 999	0. 002	0. 683	1. 000	1. 000	0. 000	0. 395	1. 000	1. 000	0. 000	0. 210
F_7	0. 036	0. 534	0. 317	1. 350	0. 005	0. 005	0. 000	0. 625	0. 005	0. 087	0. 059	0. 306

	1.437	2.468	1.220	3.602	0.016	0.149	0.704	4.602	0.040	0.206	0.087	0.862
F_8	3.231	8.417	2.956	2.278	0.995	2.773	2.721	1.434	0.995	2.912	0.536	0.522

It can be known from Fig.2 that three algorithms under discussion can all get optimums with respect to F1~ F6. However, for F7 and F8, No one of the three algorithms can get optimums under the given controlling parameters. Though, the solutions of GQA and vbQGA are still obviously better than that of sGA. From Fig.6 and Fig.7 we can find that F7 and F8 are very complicated and intractable for there exiting many local optimums. By taking population size as s=50 and remaining other parameters unchanged, we carried out some test runs and the results show that GQA and vbQGA can exactly find out the optimums of F7 and F8.

Fig.3 tells us that sGA, GQA and vbQGA gain closely approximate averages of the function value of F1~ F6 over 20 runs. In contrast with this, for F7 and F8, the averages obtained by GQA and vbQGA are very approximate and evidently superior to those by sGA.

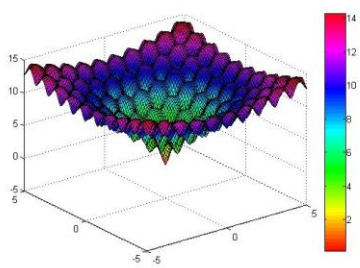


Fig.6. Figure of function Ackley(n=2)

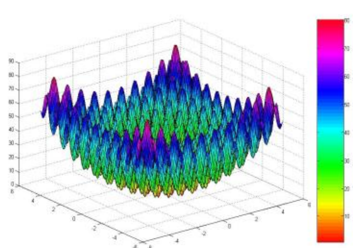


Fig.7. Figure of function Rastrigin (n=2)

Fig.4 shows that vbQGA gets the smallest standard deviations among the three algorithms and sGA gets the largest ones. It reveals that vbQGA is more robust than the two algorithms.

Fig.5 illustrates the comparison of average of elapsed time per one run among the three algorithms. It can be seen that vbQGA takes the least run time. Let us sum up all the $\bar{t}$ of eight functions for sGA, GQA and vbQGA and we can get 15.5062 seconds, 11.0675 seconds and 2.8631 seconds respectively. Obviously, vbQGA takes much less run time than the other two algorithms. In addition, it can be also seen that as the dimension and complexity of a function increase, this advantage will get more distinct.

To sum up, vbQGA is superior to sGA in all respects. Comparing with GQA, vbQGA can get very approximate quality of solutions. However, the standard deviation and the average elapsed time per one run of vbQGA are, especially for the multidimensional and complicated functions, less than GQA. This indicates that vbQGA has better robustness and solving velocity.

4. Conclusions

In this paper, we proposed a variable-boundary-coded quantum genetic algorithm, vbQGA, based on the core idea and principle of quantum computation. In this algorithm, qubit chromosomes are collapsed into variable-boundary-coded chromosomes instead of binary-coded chromosomes, a new adaptive selection strategy for angle parameters used for rotation gate is adopted. An experimental evaluation, in which eight typical functions are selected to optimize and sQA and GQA are selected as contrasts, has been conducted. Four statistical values have been used as measurements of performance to evaluate vbQGA. The results reveal that vbQGA is significantly superior to sGA in all aspects and outperforms GQA in robustness and solving velocity, especially for multidimensional and complicated functions. These demonstrate effectiveness and good performance of vbQGA.

Acknowledgment

This research wok has been supported by Hubei Province Department of Education in China (Grant No. D20121102). The author is grateful to the anonymous reviewers for their valuable suggestions and comments.

References

[1] P. Benioff, “The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machine,” J. Stat. Phys. 22:563-591 (1980).

[2] R.P. Feyman, “Simulating physics with computers,” Int. J. Theor. Phys., vol.21, July 1982, pp.467-488

[3] P. W. Shor, “Algorithm for quantum computation: Discrete logarithms and factoring,” Proc. of the 35th Annual

Symposium on the Foundation of Computer Sciences, IEEE Computer Society Press, Nov. 1994, pp.124-134,doi: 10.1109/SFCS.1994.365700.

[4] L.K. Grover, "A fast quantum mechanical algorithm for database search," Proc. of the 28th Annual ACM Symposium on the Theory of computing, Philadelphia, PA, May 1996,pp.212~221.

[5] A. Narayanan, M. Moore, "Quantum inspired genetic algorithm," Proc. of the Third IEEE International Conference on Evolutionary Computation, Piscataway, IEEE Press, NJ, May 1996, pp.61 – 66,doi: 10.1109/ICEC.1996.542334.

[6] K H. Han, "Genetic quantum algorithm and its application to combinatorial optimization problem," Proc. 2000 Congress on Evolutionary Computation. Piscataway, IEEE Press, NJ, July 2000, pp.1354~1360,doi: 10.1109/CEC.2000.870809.

[7] L. Wang, H. Wu, F. Tang, D. Zheng, "A Hybrid Quantum-Inspired Genetic Algorithm for Flow Shop Scheduling," in Advances in Intelligent Computing, Lecture Notes in Computer Science, Vol. 3645, Springer-Verlag, Berlin, Sep.2005, pp. 636-644,doi: 10.1007/11538356.

[8] G. Zhang, H. Rong, "Improved Quantum-Inspired Genetic Algorithm Based Time-Frequency Analysis of Radar Emitter Signals," in Rough Sets and Knowledge Technology, Lecture Notes in Computer Science, Vol. 4481, Springer-Verlag, Berlin, June 2007, pp. 484-491,doi: 10.1007/978-3-540-72458-2.

[9] J.G. Vlachogiannis, J. stergaard, "Reactive power and voltage control based on general quantum genetic algorithms," Expert Syst. Appl., vol.36, Apr.2009,pp.6118-6126, doi:10.1016/j.eswa.2008.07.070

[10] J. Gu, X. Gu, M. Gu, "A novel parallel quantum genetic algorithm for stochastic job shop scheduling," J. Math. Anal. Appl., vol. 355, Jan. 2009,pp.63-81, doi: 10.1016/j.jmaa.2008.12.065.

[11] S. Zhou, W. Pan, B.Luo, W. Zhang, Y. Ding, "A novel quantum genetic algorithm based on particle swarm optimization method and its application," Acta Electron. Sinica., Vol.34, May 2006, pp.897-901, doi: cnki:ISSN:0372-2112.0.2006-05-024.

[12] C. Zhou, F. Qian, "Improvement of quantum genetic algorithm and its application," Comput. Appl., Vol.28, Feb.2008, pp.286-288, doi: CNKI:SUN:JSJY.0.2008-02-007.

Ying Sun is an associate professor of College of Machinery and Automation, Wuhan University of Science and Technology. Currently, his research interests are manufacturing information, optimization design and computer aided engineering.

Yuesheng Gu birth 1973, male, associate professor of Henan institute of science and technology. His current research area is computer network technology and artificial intelligence.

Hegen Xiong received the Ph.D. degree in materials processing engineering from Huazhong University of Science and Technology, China, in 2005.He is a Professor of College of Machinery and Automation, Wuhan University of Science and Technology. Currently, his research interests are manufacturing information, job shop scheduling, flow shop scheduling and computer aided engineering.

Enhanced Intrusion Detection System for Input Validation Attacks in Web Application

Puspendra Kumar[1], R. K. Pateriya[2]

**[1] M Tech Scholar ,Computer Science & Engineering Department,
Maulana Azad National Institute of Technology
Bhopal, 462051, India**

**[2] Associate Professor, Computer Science & Engineering Department,
Maulana Azad National Institute of Technology
Bhopal, 462051, India**

Abstract

Internet continues to expand exponentially and access to the Internet become more prevalent in our daily life but at the same time web application are becoming most attractive targets for hacker and cyber criminals. This paper presents an enhanced intrusion detection system approach for detecting input validation attacks in the web application. The existing IDS for Input validation attacks are language dependent. The proposed IDS is language independent i.e. it works for any web application developed with the aid of java, php, dot net etc. In addition the proposed system detects directory traversal attacks, command injection attacks, cross site scripting attacks and SQL injection attacks, those were not detected in the existing IDS. This is an automatic technique for detection vulnerabilities over the internet. Our technique is based on the web application parameter which is in form of POST and GET which has generalized structure and values. This technique reduces analysis time of input validation attacks.

Keywords: SQL Injection attacks, XSS attacks, directory traversal attacks, GET and POST data, Detection.

1. Introduction

Web Application are most widely used for providing service to the user like online shopping, online reservation, and many more application which is designed in perspective of user. So the web application is popular attacks target due to time and financial constraints, limited knowledge of the programming, limited knowledge security awareness, misconfiguration that is meant lack of awareness of the security configuration deployment on the part of the programmer. With the aid of the input validation attacks attacker can steal the confidential data which decrease the market values of the organization. Web applications generally use TCP port for the communication with server. This communication is not protected by the IVAs [1].

Current Intrusion detection systems are designed in such a way that detects SQL injection attacks, XSS attacks but they do not detect the directory traversal attacks, and command injection attacks. Such IDS are designed language specific for example IDS for PHP based web application, IDS for JAVA based web application [2, 3, 4]. The proposed enhanced IDS approach detect SQL Injection attacks, XSS attacks, Directory Traversal attacks and command injection, and is not programming language specific. This IDS approach require only window environment for detecting IVA over the internet. It analyze web request data to detect that if any type of IVA exists. Then the detection system automatically generate the report against input validation attacks and send it to the server owner.

The enhanced IDS approach can be more effective for finding out any type of Input validation attacks and with the aid of this approach server administrator can take effective action against these attacks. So in this way this approach reduces the analysis time and also increase the efficiency of the system.

Rest of the paper is organized as follows: Section 2 describes the input validation attacks and section 3 describes the related work. Proposed enhanced IDS approach is given in section 4 this section also discusses how to analyze the raw data. Section 5 describes the comparison with existing IDS. Finally conclusion is given in section 6.

2. Input Validation Attacks

The Input Validation Attacks (IVAs) attempt to submit data which the web application does not expect to receive, that causes very serious consequences like session hijack, SQL poisoning, source code disclosure, directory browsing etc. Input validation is a security issue if an attacker discovers that the application makes unfounded assumptions about the type, length, format, or range of input data. The attacker can then supply carefully crafted input that compromises the application. When network and host level entry points are fully secured; the public interfaces exposed by the application become the only source of attack. The input to the application is a means to both test the system and a way to execute code on an attacker's behalf. If the applications blindly trust input. It may be susceptible to the following:

2.1 Buffer Overflows Attacks

Buffer overflow vulnerabilities can lead to denial of service attacks or code injection. A denial of service attack causes a process crash. Code injection alters the program execution address to run an attacker's injected code.

2.2 Cross-Site Scripting Attacks

An XSS attack can cause arbitrary code to run in a user's browser while the browser is connected to a trusted Web site. The attack targets the application's users and not the application itself, but it uses the application as the vehicle for the attack. Because the script code is downloaded by the browser from a trusted site, the browser has no way of knowing that the code is not legitimate. Internet Explorer security zones provide no defense. Since the attacker's code has access to the cookies associated with the trusted site and are stored on the user's local computer, a user's authentication cookies are typically the target of attack.
Example of Cross-Site Scripting:
To initiate the attack, the attacker must convince the user to click on a carefully crafted hyperlink, for example, by embedding a link in an email sent to the user or by adding a malicious link to a newsgroup posting. The link points to a vulnerable page in the application that echoes the invalidated input back to the browser in the HTML output stream. For example, consider the following two links.
Here is a legitimate link:
www.webapplication.com/logon.aspx?username=puspendra
Here is a malicious link:
www.webapplication.com/logon.aspx?username=<script>alert ('hacker code') </script>
If the Web application takes the query string, fails to properly validate it, and then returns it to the browser, the script code executes in the browser. The preceding example displays a harmless pop-up message. With the appropriate script, the attacker can easily extract the user's authentication cookie, post it to his site, and subsequently make a request to the target Web site as the authenticated user.

2.3 SQL Injection Attacks

A SQL injection attack exploits vulnerabilities in input validation to run arbitrary commands in the database. It can occur when the application uses input to construct dynamic SQL statements to access the database. It can also occur if code uses stored procedures that are passed strings that contain unfiltered user input. Using the SQL injection attack, the attacker can execute arbitrary commands in the database. The issue is enhance if the application uses an over-privileged account to connect to the database. In this instance it is possible to use the database server to run operating system commands and potentially compromise other servers, in addition to being able to retrieve, manipulate, and destroy data.

2.4 Canonicalization

Different forms of input that resolve to the same standard name (the canonical name), is referred to as canonicalization. Code is particularly susceptible to canonicalization issues if it makes security decisions based on the name of a resource that is passed to the program as input. Files, paths, and URLs are resource types that are vulnerable to canonicalization because in each case there are many different ways to represent the same name. File names are also problematic.

3. Related Work

3.1 ARDILLA Tool[2]

This tool is developed by Adam Kie˙zun, Philip J. Guo, Karthick Jayaraman, Michael D. Ernst. This tool is useful for identifying the SQL Injection attacks and XSS vulnerabilities. This technique works on unmodified existing code, generate concrete input that expose vulnerabilities and operate before software is deployed. ARDILLA is an automated tool for creating attacks. It is white box testing tool means that it requires source code of the application. It is based on the input generation, taint propagation and input mutation to find variants of an execution that exploit vulnerability.

3.2 Protect Web Application using Positive Tainting and Syntax-Aware Evaluation[3]

This approach is proposed by William G.J. Halfond, Alessandro Orso and Panagiotis Manolios. This approach uses four term to detect SQLIA Positive tainting, Accurate and efficient taint propagation, Syntax-aware evaluation of queries string and Minimal deployment requirement.

3.3 VIPER for Detecting SQL Injection Attacks[4]

In this techniques SQL Injection attacks detected by using heuristic based approach. It basically performs panetration testing of the web application .This approach analyzes the web application for detrmininng hyperlinks structure and input supplied from the user and gives error message, if any type of SQL injection occurres.

3.4 Runtime Monitoring Technique[5]

AMNESIA technique is used for detecting SQL Injection attacks over the web application. These technique workes on both static approach and runtime monitoring. It detects injected query before executed on the database server using model based approach.This approach have two part static part which automatically builds a legel queries using program analysis on the other hand in dynamic part it dynamically generates the queries against statically build queries using runtime monitoring. If queries violate the approach then this approach perevents the execution of the queries on the database server. This technique has four steps for preventing injection Identify the hotspot, Build SQL-query models, Instrument application, Runtime monitoring.

3.5 Detection by Feature of Single Character[6]

This techniques used to sigmoid function for detecting SQL injection attacks. They proposed detection algorithm of SQL Injection Attack based on single character. When the SQL character string is the SQL Injection, it call an attack character string. With the aid of this approach minimizes the predictive error in SQL Injection attack detection.

3.6 Obfuscation-based Analysis of SQL Injection Attacks[7]

This approach is proposed by Raju Halder and Agostino Cortesi. They implemented combined structure of static and dynamic analysis which is based on the obfuscation and de-obfuscation of SQLcommands SQL Injection attacks can be easily detected because dynamic verification is carried out on the obfuscated queries , at atomic formula level only those atomic formulas are tagged as vulnerable. And this approach finds the root cause of the SQL Injection attacks in dynamic query generation.

4. Enhanced IDS Approach

This section describe the enhanced IDS for Input Validation Attacks in web application. This proposed approach will perform on the Application layer of the OSI model. So that with the aid of application layer we can obtain POST and GET data over the network communication with server. C# language using socket programming [8] will be used to develop the proposed IDS. This approach will work in six steps for performing detection of input validation attacks (IVAs) as shown in following block diagram.

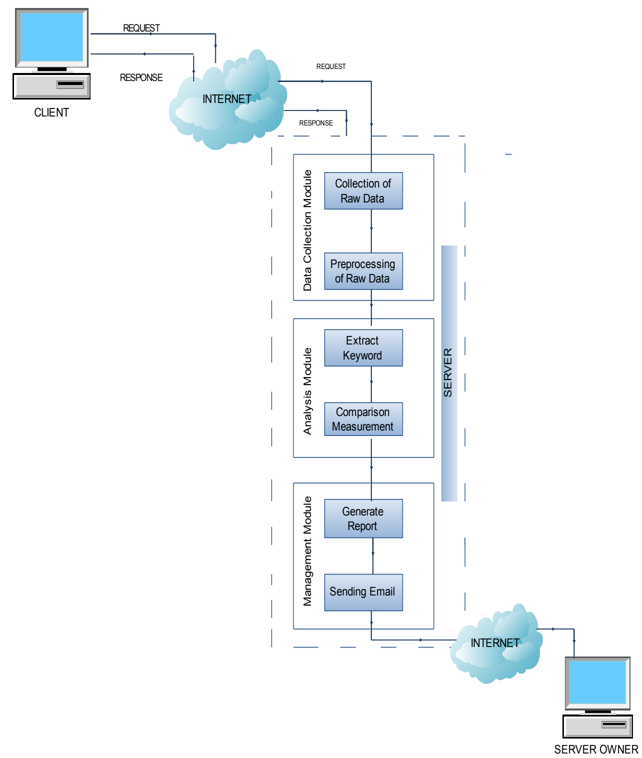


Fig. 1 Block Diagram of Proposed Enhanced IDS

4.1 Collection of Raw Data

In this step we collect the HTTP header data which contains GET and POST data which are used to pass the parameter value to the web server over the internet. The HTTP header data works as raw data for the proposed IDS. Raw data will be stored in text file. These file contains both request data and response data which is analyzed by proposed approach for finding out IVA attacks. The process of raw data collection is given stepwise in following DFD.

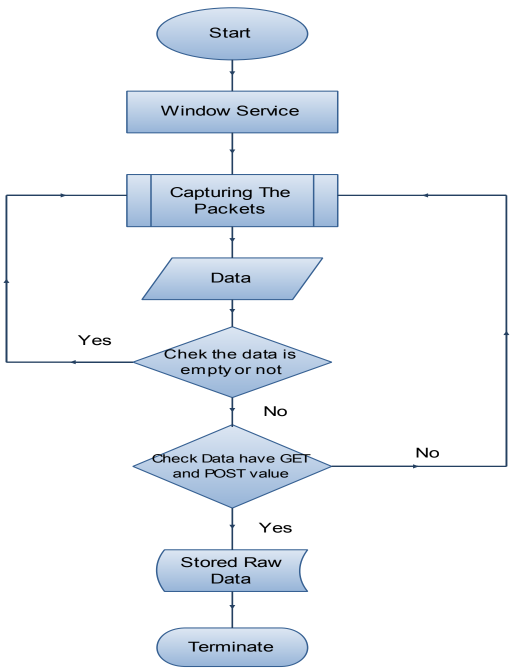
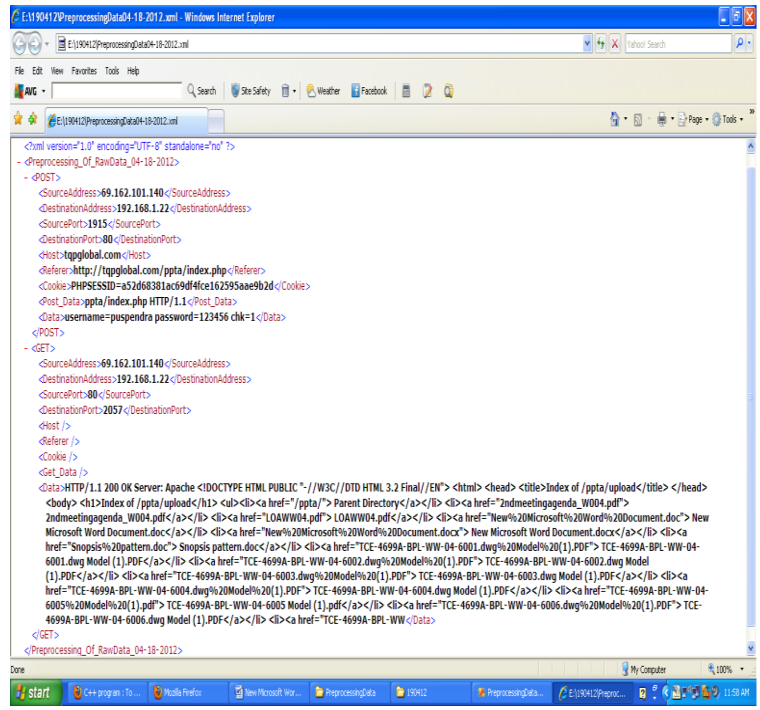


Fig. 2 DFD of Collection of Raw Data



4.2 Preprocessing of the Raw Data

The raw data collected in step 4.1 is preprocessed in this step the raw data is filtered to keep GET and POST data and to remove unnecessary information like http header details. These preprocessing data are generated in XML format. In this step we analyze the raw data and fetch useful data from raw text file. Following steps are used for preprocessing of the data.

Step 1: Input text file (.Text) containing raw data which is obtained in step1.
Step 2: Read file by line by line and match with HTTP header information like referrer, Host, and cookie etc. If match found then corresponding value will be stored in DataItem class. This task is performed until the entire HTTP header is not matched.
Step 3: After data stored in DataItem class then these add in the ArrayList collection.
Step 4: Repeat step 2 and step 3 until all of the data is read from text file.
Step 5: Add DataItem class in ArrayList collection. Now we need to fetch DataItem properties values like Host, Referrer and Cookie etc and assigned to the predefined the XML document.
Step 6: The generated output is in the form of XML format which contained the root node, child node.
Following screen shot shows the preprocessing of data in XML format.

4.3 Keyword Extraction

After preprocessing of the raw data we extract keywords from the xml file of the preprocessed data. In this step we extract GET and POST data from xml file or preprocessed file and perform analysis on extracted data. Following steps are used to Extract Keywords:
Step 1: Input the XML file which contains preprocessed data.
Step 2: Read the file according to Root node like POST and GET.
Step 3: If root node is POST then we extracts data of child node which will be source address, destination address, source port, destination port and its data, which is stored in Extracted DataItem collection class.
Step 4: If root node is GET then we extracts data of child node which will be source address, destination address, source port, destination port and its data, Get data to store in Extracted DataItem collection class. Otherwise go to step 6.
Step 5: Repeat step 3rd and step 4th until read operation in XML or preprocessing file is complete.
Step 6: End process.
The process of keyword extraction is given in following DFD.

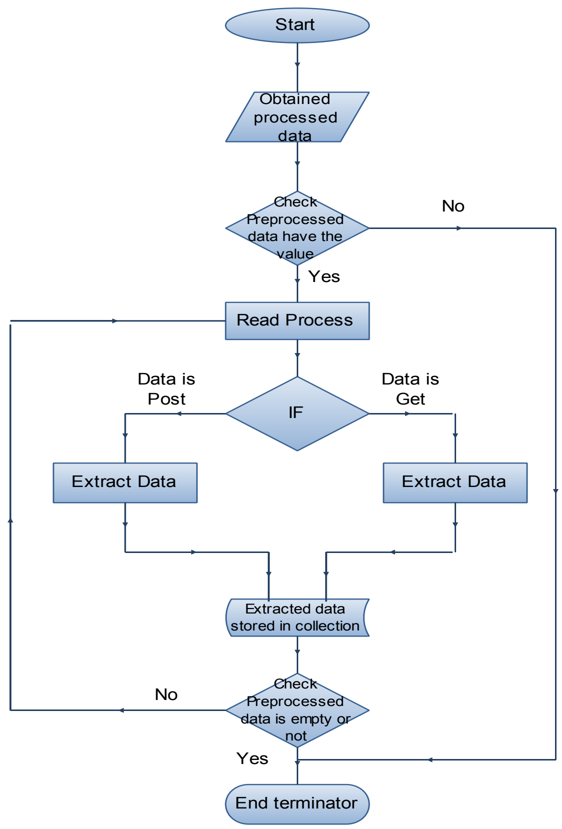


Fig. 3 DFD of Extracted Keyword

4.4 Comparison Measurement

In this step templates for SQL Injection attacks, cross site scripting attacks, command injection attacks directory traversal attacks will be created these templates will contain values corresponding to these attacks.
The data value of generated templates are compared with the data value of templates obtained in the step 4.3. If a match is found then these are marked as attacks like SQL Injection attacks, Cross site scripting attacks, command injection attacks and directory traversal attacks, and stored in the hard drive.
Following steps are used for the Comparison Measurement:
Step 1: Get the Extracted data which is found in step3.
Step 2: Perform matching process.
Step 3: Match data value with SQL Injection attacks template if match found then go to Step 7 otherwise go to next step.
Step 4: Match data value with Cross site scripting attacks template if match found then go to Step 7 otherwise go to next step.
Step 5: Match data value with Command Injection attacks template if match found then go to Step 7 otherwise go to next step.
Step 6: Match data value with Directory traversal attacks template if match found then go to Step 7 otherwise go to step 9.
Step 7: In this step data value is marked as an attack corresponding attacks template.
Step 8: In this step marked data value will stored in stored file.
Step 9: End process.
These steps clearly understood by data flow diagram of Comparison Measurement as shown in figure 5.

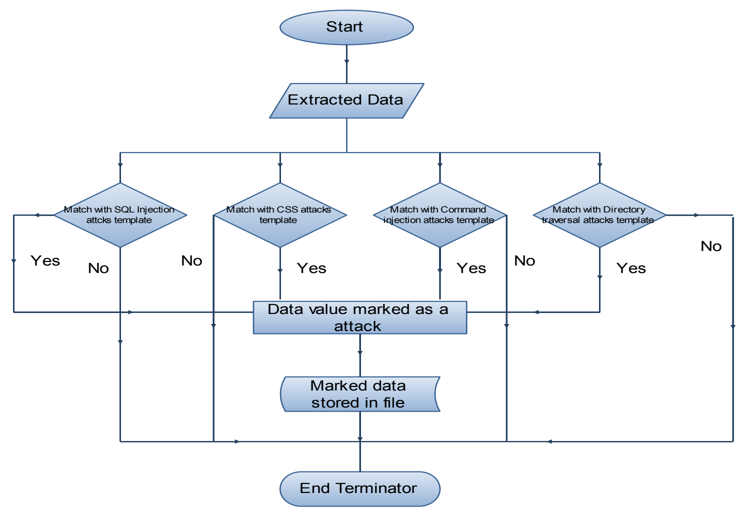
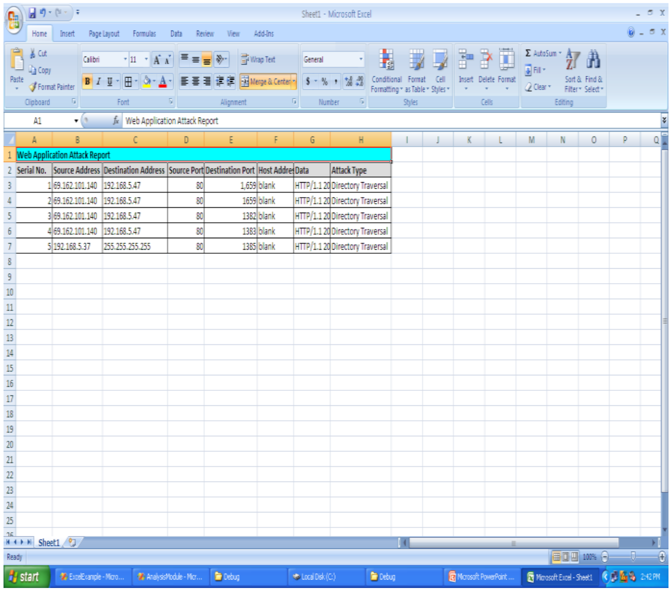


Fig. 4 DFD of Comparison Measurement

4.5 Report Generation

This step is for automatic report generation whenever any attacks is detected than automatic report is generated immediately in this step. This report has information about attacks like SQL Injection attacks, Command Injection Attacks, and Directory browsing attacks. The generated report also have the information about source address, destination address, source port, destination port, host, data and type of attacks performed on the data.

Following screen shots shows the attack report.



4.6 Sending Email

The generated report in step 4.5 is sent to server owner by email. With the aid of this report, server owner becomes aware with the online attacks on the server and take appropriate action to protect the server from web input validation attacks. This email sending is also automatic process.

5. Comparison with Existing IDS

This section shows that why proposed approach is better than previous IDS to detect input validation attacks in web application. Table 1 gives the comparative view of the existing IDS with proposed IDS.

Table 1: Comparison with existing IDS

Techniques	***Language***	**Attacks**
ARDILLA[2]	PHP based Web Application	Sql Injection , XSS attacks
WASP[3]	JAVA based Web Application	SQL Injection
VIPER[4]	ANY	SQL Injection
Our Proposed IDS	ANY	SQL Injection, XSS attacks, Command Injection, Directory Traversal Attacks

It is clear from the comparison that the proposed IDS will support to all web applications which are developed using any language like java, PHP, Dot Net etc. The proposed IDS will also detect to the SQL Injection attacks, XSS attacks, Directory Traversal attacks and Command injection attacks. Previous IDS system ARDILLA [2] do not detect command injection attacks, directory traversal attacks and they perform detection on only on PHP based web application. WASP [3] do not detect XSS attacks; command injection attacks and directory traversal attacks and they perform detection only on java based web application. And another tool is VIPER [4] do not detect XSS attacks, command injection attacks and directory traversal attacks.

6. Conclusion

In this paper, an enhanced intrusion detection system for input validation attacks on web application is proposed. The proposed intrusion detection system detect SQL injection attacks, XSS attacks, command injection attacks and also detect to directory traversal attacks. In addition with this the proposed IDS support all web applications which are developed using any language like java, php, Dot Net etc. Web application security can be improved with proposed IDS by making security provision active in advance .The analysis time to detect IVA is reduced in this IDS because whole process is operated without developer intervention.

References

[1] OWASPD-Open Web Application Security Project. “Top ten most critical Web Application Security Risks”, https://www.owasp.org/index.php/Top_10_2010-Main.

[2] Adam Kie˙zun,Philip J. Guo,Karthick Jayaraman,Michael D. Ernst:“Automatic Creation of SQL Injection and Cross-Site Scripting AttackS”, ICSE‟09, May 16-24, 2009, Vancouver, Canada.

[3] William G.J. Halfond, Alessandro Orso.” WASP: Protecting Web Applications Using Positive Tainting and Syntax-Aware Evaluation” IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 34, NO. 1, JANUARY/FEBRUARY 2008.

[4] Angelo Ciampa, Corrado Aaron Visaggio, Massimiliano Di Penta :”A heuristic-based approach for detecting SQL-injection vulnerabilities in Web applications”. *SESS'10* , May 2, 2010, Cape Town, South Africa Copyright 2010 ACM.

[5] William G.J. Halfond and Alessandro Orso,” Preventing SQL Injection Attacks Using AMNESIA” ICSE’06, May 20–28, 2006, Shanghai, China ACM 06/0005.

[6] Takeshi Matsuda,Daiki Koizumi,Michio Sonoda,Shigeichi Hirasawa, ”On predictive errors of SQL injection attack detection by the feature of the single character” Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on 9-12 Oct 2011, On Page 1722-1727.

[7] Raju haldar and Agostino Cortesi, “Obfuscation-based Analysis of SQL Injection Attacks” IEEE.

[8] C# Network Programming by Richard Blum ISBN: 0782141765, Sybex@2003.
[9] Y.Xie and A. Aiken." Static detection of Security vulnerabilities in scripting language", In USENIX.2006.
[10] A. Sabelfeld and A. Myers."Language-based information flow security". Selected Areas in Communications 2003.
[11] R.McClure and I. Kruger."SQL DOM: compile time checking of dynamic SQL statements". In ICSE, 2005.
[12] M. Martin and M. Lam."Automatic generation of XSS and SQL injection attacks with goal-directed model checking". In USENIX SEcurity,2008.
[13] S. Bandhakavi. P. Bisht, P. Madhusudan, and V.N. Venkatakrishan."CANDID: preventing SQL injection attacks using dynamic candidate evaluations". In CCS,2007.
[14] Z. Su and G. Wassermann,"The Essence of Command Injection Attacks in Web Applications". Proc.33rd Ann. Symp. Principles of Programming Language, p.372-382, Jan 2006.
[15] D. Scott and R. Sharp,"Avstracting Application-Level Web Security", Proc. 11th Int'l Conf. World Wide Web, p.396-407, May 2002.
[16] W. R. Cook and S. Rai, "Safe Query Objects: Statically Typed Objects as Remotely Executable Queries" Proc. 27th Int'l Conf. Software Eng. p.97-106, May2005.

First Author Puspendra Kumar perusing Master of technology in Information security from Maulana Azad national Institute of Technology Bhopal.

Second Author Dr. R.K. Pateriya working as Associate Professor in computer science and engineering department of Maulana Azad National Institute of Technology, Bhopal.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

A Comparative study of Analog and digital Controller On DC/DC Buck-Boost Converter Four Switch for Mobile Device Applications

Abdessamad Benlafkih[1],Salah-ddine Krit[2] and Mohamed Chafik Elidrissi[3]

[1]Laboratory of engineering physics and environment, Department of Physique, Faculty of Sciences, University Ibn Tofail BP 133, Kénitra 14000, Morocco

[2]Department of informatics, Polydisciplinary Faculty of Ouarzazate, University Ibn Zohr BP 638 Ouarzazate, Morocco

[3]Laboratory of engineering physics and environment, Department of Physique, Faculty of Sciences, University Ibn Tofail BP 133, Kénitra 14000, Morocco

Abstract

This paper presents comparative performance between Analog and digital controller on DC/DC buck-boost converter four switch. The design of power electronic converter circuit with the use of closed loop scheme needs modeling and then simulating the converter using the modeled equations. This can easily be done with the help of state equations and MATLAB/SIMULINK as a tool for simulation of those state equations. DC/DC Buck-boost converter in this study is operated in buck (step-down) and boost (step-up) modes.

Keywords: *Analog Controller, Digital Controller, system modeling, DC/DC Buck-boost converter, Matlab/ Simulink.*

1. Introduction

CURRENT trends in consumer electronics demand progressively lower supply voltages due to the unprecedented growth and use of wireless appliances. Portable devices, such as laptop computers and personal communication devices require ultra low-power circuitry to enable longer battery operation. The key to reducing power consumption while maintaining computational throughput and quality of service is to use such systems at the lowest possible supply voltage. The terminal voltage of the battery used in portable applications (e.g., NiMH, NiCd, and Li-ion) varies considerably depending on the state of their charging condition. For example, a single NiMH battery cell is fully charged to 1.8 V but it drops to 0.9 V before fully discharged [1]. Therefore, systems designed for a nominal supply voltage (say, 1.5 V with a single NiMH battery cell) require a converter capable of both stepping-up and stepping-down the battery voltage. While both buck (step-down) [2][3]and boost (step-up) [4]converters are widely used in power management circuits. The DC/DC Converter must provide a regulated DC output voltage even when varying load or the input voltage varies.

Therefore, the topologies for generating a voltage higher and lower than the supply is : non inverting buck-boost converter [5], [6],[7] which is essentially achieved by cascading a buck with a boost converter Fig.1, The trend in portable applications is to use the topologies that incorporate less number of external components and move closer to cost effective SOC designs [13].

Controller design for any system needs knowledge about system behavior. Usually this involves a mathematical description of the relation among inputs to the process, state variables, and output. This description in the form of mathematical equations which describe behavior of the system (process) is called model of the system [8][9]. This paper describes an efficient method to learn, analyze and simulation of DC/DC buck-boost converter four switch, with analog and digital Controller, The MATLAB/SIMULINK software package can be advantageously used to simulate power converters.

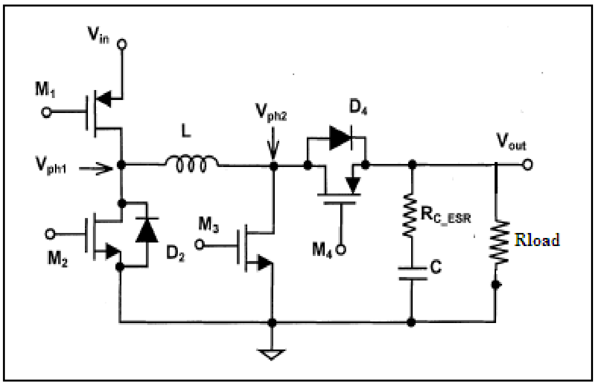


Fig.1 Noninverting synchronous DC/DC buck-boost converter.

2. buck-boost converter modeling

2.1 Open-loop synchronous buck-boost converter model.

In Figure1 a DC-DC buck-boost converter is shown. The switching period is T and the duty cycle is D. Assuming continuous conduction mode of operation, during T_{ON} the period of the cycle, switches M_1 and M_3 are ON and the input voltage is impressed across the inductor. Since the load current is instantaneously provided by the output capacitor during this interval, the capacitor voltage (output voltage) decreases, the state space equations are given by [9],

$$\begin{cases} \frac{di_L}{dt} = \frac{1}{L}\left[V_{in} - (R_L + R_{ON1} + R_{ON3}) \times i_L\right] \\ \frac{dv_c}{dt} = -\frac{1}{c} i_{out} \\ V_{out} = -R_{ESR} i_{out} + v_c \end{cases} \tag{1}$$

During the other interval of the switching period (T_{OFF}), switches M_2 and M_4 are turned ON and the inductor energy is transferred to the output, providing both the load current and also charging the output capacitor, the equations are given by [9],

$$\begin{cases} \frac{di_L}{dt} = -\frac{1}{L}\left[(R_L + R_{ON2} + R_{ON4}) \times i_L + V_{out}\right] \\ \frac{dv_c}{dt} = \frac{1}{c}(i_L - i_{out}) \\ V_{out} = -R_{ESR} i_{out} + v_c + R_{ESR} i_L \end{cases} \tag{2}$$

There is a time delay (known as dead-time) between turning OFF M_1, M_3 and turning ON M_2, M_4 to prevent shoot-through current. During this period, the inductor current flows through body diodes D_2 and D_4, from transistors M_2 and M_4, respectively.
The duty cycle (D) of the converter is given by

$$D = \frac{T_{ON}}{T_{ON} + T_{OFF}} = \frac{T_{ON}}{T} \tag{3}$$

Since the node V_{ph1} is connected to V_{in} for DT time over a period of T, the average voltage is $V_{ph1,avg} = DV_{in}$. Similarly, the average node voltage of V_{ph2} can be given by $V_{ph2,avg} = D'V_{out}$ $(D' = 1 - D)$.

Under steady-state operating condition, the inductor can be treated as short and the average voltage of V_{ph1} and V_{ph2} are equal.

$$DV_{in} = D'\,V_{out} \Rightarrow \frac{V_{out}}{V_{in}} = \frac{D}{1-D} \tag{4}$$

The Eq.(1) and Eq.(2) are implemented in Simulink as shown in Figure 2 to obtain the states, $i_L(t)$ and $V_{out}(t)$ [10][11][12].

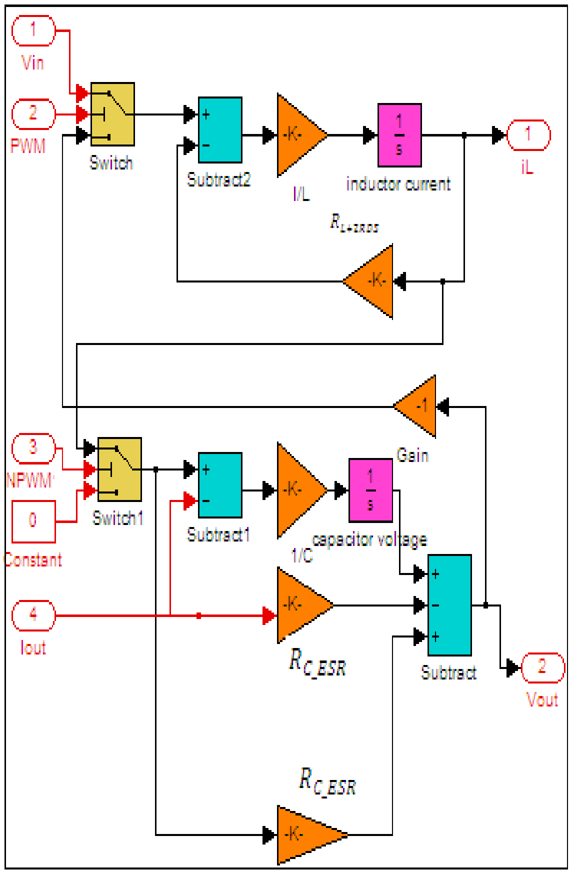


Fig.2 Open-loop of Buck-Boost Converter model

2.2 Close-loop synchronous buck-boost Converter model

a) Analog controller

The Figure 3 is presented the model by SIMULINK and MATLAB of buck-boost converter with analog controller, it uses the compensator of type III-A, and the model aims to regulate the output voltage in 3.24 (V) with variation of input voltage and load.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

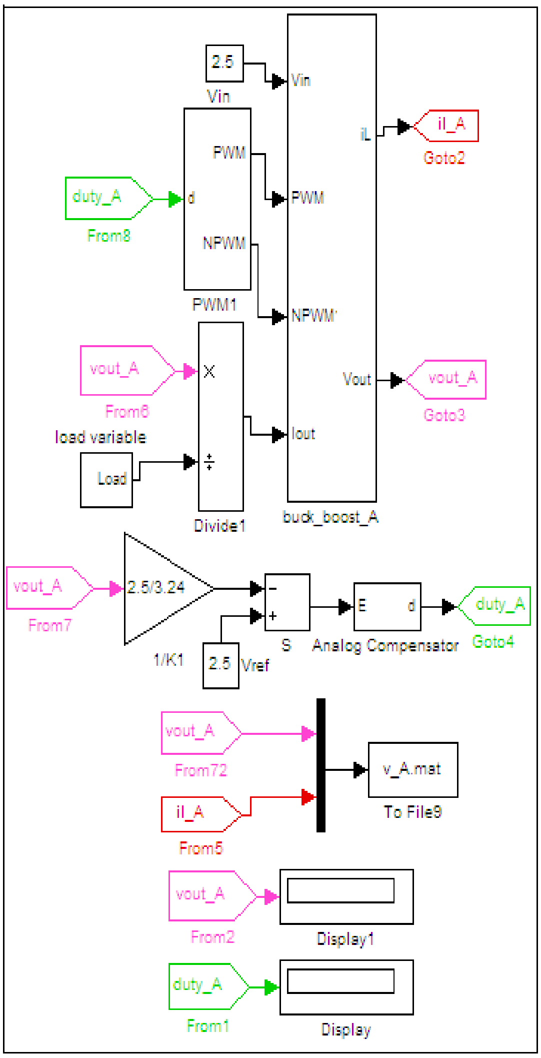


Fig.3 Close- loop buck-boost Converter model Analog controller

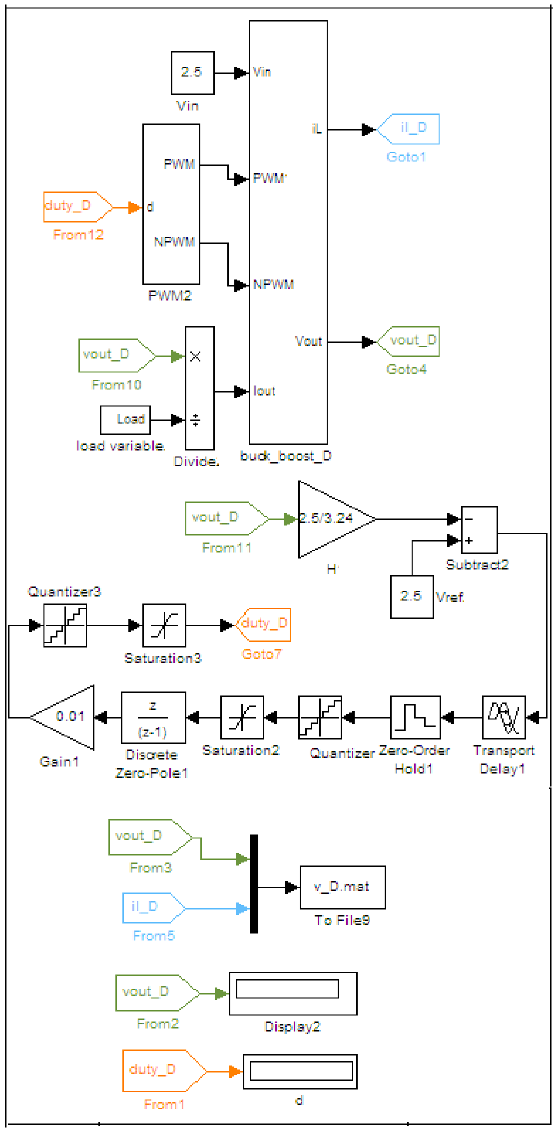


Fig.4 Close- loop buck-boost Converter model digital controller

b) Digital controller

The figure 4 is presented the model by SIMULINK and MATLAB of buck-boost converter with digital controller, the model aims to regulate the output voltage in 3.24 (V), with variation of input voltage and load.

3. Results and discussion

In this section, simulation results for Analog controller, digital controller and without feedback on buck-boost converter circuit.

3.1 Boost (step-up) mode

Table 1 shows the parameters of buck-boost converter on mode boost of three models, figure 5 and figure 6 show the inductor current and output voltage waveforms of three models on mode boost.

Table1: buck-boost converter parameters on mode boost

	Values without Feedback	*Values with Analog controller*	*Values with Digital Controller*
V_{in}*(V)*	2.5	2.5	2.5
V_{out} *(V)*	3.24	3.24	3.4
L (H)	280e-9	1e-6	280e-9
C (F)	250e-9	22e-6	250e-9
$R_{L+2RDS}(\Omega)$	0.5	8e-2	0.5
$R_{load}(\Omega)$	10	10	10
$R_{C\_ESR}(\Omega)$	1e-4	60e-3	1e-4
Duty cycle	D>0.5	D>0.5	D>0.5

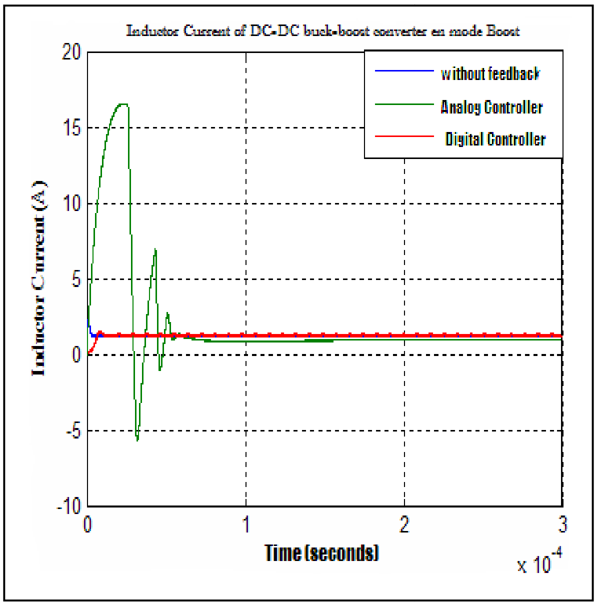


Fig.5 inductor current of buck-boost converter on mode boost

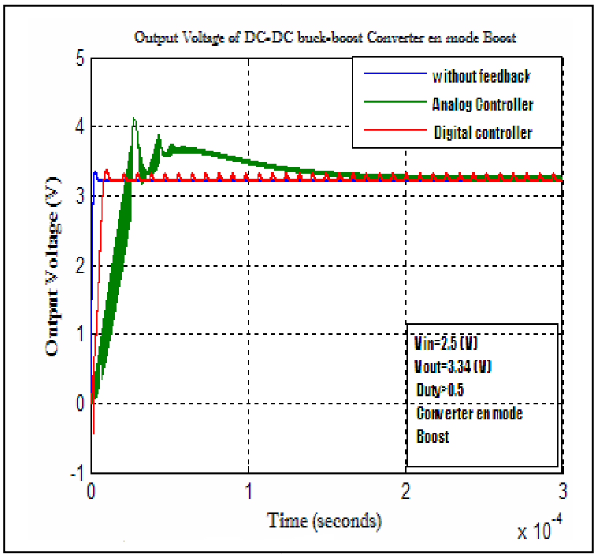


Fig.6 Output voltage of buck-boost converter on mode boost

3.2 Buck (step down) mode

Table 2 shows the parameters of buck-boost converter on mode buck of three models, figure 7 and figure 8 show the inductor current and output voltage waveforms of three models on mode buck.

Table2: buck-boost converter parameter on mode buck

	Values without Feedback	*Values with Analog controller*	*Values with Digital Controller*
V_{in}*(V)*	5	5	5
V_{out} *(V)*	3.24	3.24	3.4
L (H)	280e-9	1e-6	280e-9
C (F)	250e-9	22e-6	250e-9
$R_{L+2RDS}(\Omega)$	0.5	8e-2	0.5
$R_{load}(\Omega)$	10	10	10
$R_{C\_ESR}(\Omega)$	1e-4	60e-3	1e-4
Duty	D<0.5	D<0.5	D<0.5

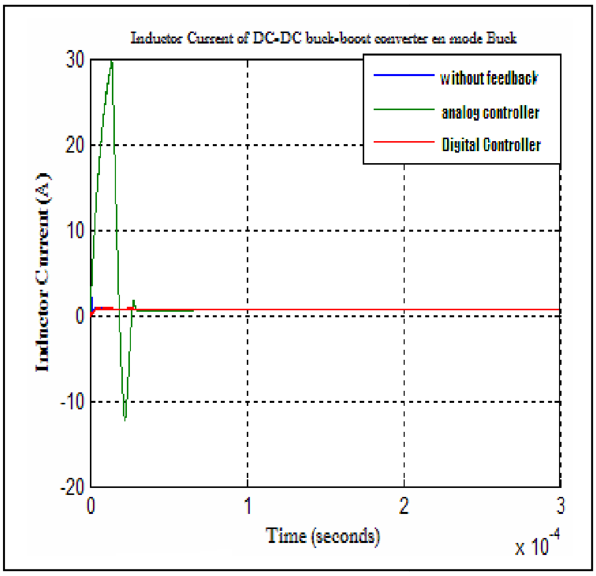


Fig.7 inductor current of buck-boost converter on mode buck

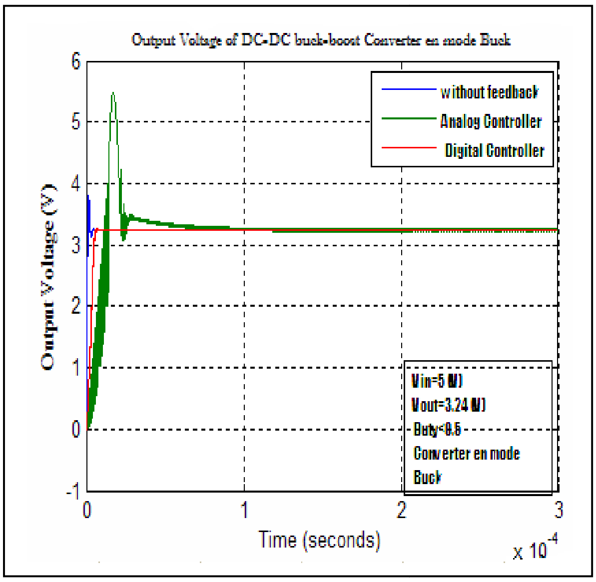


Fig.8 Output voltage of buck-boost converter on mode buck

3.3 Comparison between analog controller, digital Controller and without feedback.

The circuit has been modeled using Simulink/Matlab, the supply voltage ranges from 2.5V to 5V .the nominal switching frequency is 50 MHz Experimental results show that the output voltage is regulated in 3.24Vof three models and on two mode boost and buck independent of input voltage and load variation.

From figures 5 to 8 show that inductor current and output voltage of three buck-boost converter models in two mode boost and buck, Note that the transit waveforms of digital controller and without feedback are almost the same except for a slight difference, and the waveforms of analog controller model are deferent of previously waveform models.

From tables 1 and 2 we note that the values of the components digital controller and without feedback models are equal but they are less than the values component of analog controller model, and the duty cycle are almost equals in three models , in mode boost D>0.5 and in mode buck D<0.5 .

4. Conclusion

Matlab/Simulink provides an effective environment for modeling and simulation of DC/DC converters. As conclusion digital controller model gives very good dynamic respond compare with analog controller model in two mode buck and boost. And the digital controller model achieves our goal; it minimizes the values of components and conserves same results.

References

[1] C. Simpson. (2003) Characteristics of Rechargeable Batteries. National. com. [Online]. Available: http://www.national.com/appinfo/ power/files/fv.pdf.

[2] Benlafkih Abdessamad, Krit Salah-ddine and Chafik Elidrissi Mohamed " Designing a High Efficiency Pulse Width Modulation Step-Down DC/DC Converter for Mobile Phone Applications" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5,NO3 September 2012.

[3] Kaoutar ELBAKKAR and Khadija SLAOUI "Integrated Circuit of CMOS DC-DC Buck Converter with Differential Active Inductor"IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6,NO3 September 2011.

[4] P.Parvathy and Dr.N.Devarajan "Simulation and Implementation of Current-Fed FullBridge Boost Converter with Zero Current Switching for High Voltage Applications"IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3,NO3 May 2012.

[5] R. W. Erickson, Fundamentals of Power Electronics, 1st ed. New York: Chapman and Hall, 1997.

[6] “Design Applications Notes,” Tech. Rep., Linear Technology, Oct. 2001.

[7] B. Sahu and G. Rincon-Mora, “A low voltage, dynamic, noninvert-ing, synchronous buck-boost converter for portable applications,”IEEE Trans. Power Electron. , vol. 19, no. 2, pp. 443–452, Mar. 2004

[8] MOHAMED ASSAF, D. SESHSACHALAM, D. CHANDRA, R. K. TRIPATHI “ DC-DC CONVERTERS VIA MATLAB/SIMULINK” published in: proceeding ACMOS’05 Proceeding of the 7th WSEAS international conference on Automatic control ,modeling and simulation

[9] J.Mahdavi, A.Emadi, H.A.Toliyat,Application of State Space Averaging Method to Sliding Mode Control of PWM DC/DC Converters, IEEE Industry Applications Society October 1997..

[10] Vitor Femao Pires, Jose Fernando A.Silva, Teaching Nonlinear Modeling, Simulation, and Control of Electronic Power Converters Using MATLAB/SIMULINK, IEEE Transactions on Education, vol. 45, no. 3, August 2002.

[11] Juing-Huei Su, Jiann Shiuh Wu, Learning Feedback Controller Design of Switching Converters Via MATLAB/SIMULINK, IEEE Transactions on Education, vol. 45, November 2002.

[12] Daniel Logue, Philip. T. Krein, Simulation of Electric Machine Electronics Interfacing Using MATLAB/SIMULINK, in 7 Computer in Power Electronics, 2000,pp.34-39.

[13] Wang Langyuan Wu Xiaobo ; Lou Jiana "A Multi-Mode Four-Switch Buck-Boost DC/DC Converter" IEEE Power and Energy Engineering Conference, 2009. APPEEC 2009. Asia-Pacific

Abdessamad Benlafkih received the B.S. and M.S. degrees, from university sidi Mohamed Ben Abdellah Faculty of sciences Dhar El Mahraz Fez, In 2001 and 2003, Respectively. During 2011-2013, PhD researchers in electrical engineering and energy system , to study Analogue DC/DC Converter for mobile satellite communication systems, and wireless access network using CMOS Technology in Laboratory of engineering physics and environment, Department of Physique, Faculty of Sciences, University Ibn Tofail, Kénitra Morocco.

Salah-ddine Krit received the B.S. and Ph.D degrees in Software Engineering from Sidi Mohammed Ben Abdellah University, Fez, Morocco in 2004 and 2009, respectively. During 2002-2008, He worked as an engineer Team leader in audio and power management of Integrated Circuits (ICs) Research, Design, simulation and layout of analog and digital blocks dedicated for mobile phone and satellite communication systems using Cadence, Eldo, Orcad, VHDL-AMS technology. Currently, He is a professor of Informatics with the Polydisciplinary Faculty of Ouarzazate, Ibn Zohr university in Agadir, Morocco. My research interests include: Wireless Sensor Networks (Software and Hardware), Computer Engineering, Wireless Communications, Genetic Algorithms and Gender and ICT.

Mohamed Chafik Elidrissi. Professor (PES) of physics, Laboratory of engineering physics and environment, Department of Physique, Faculty of Sciences, University Ibn Tofail, Kénitra Morocco.

Structural Fatigue Reliability Based on Extension of Random Loads into Interval Variables

QIANGFENG WANG [1], JIANJUN LI [2] and SHENGLI ZHANG[2]

[1] Electromechanical, Northwestern Polytechnical University, Xi'an, China

[2] Luoyang Opto-Electro Technology Development Center, Luoyang, China

Abstract

According to the problem that for a structure under random loads, the structural fatigue life can't be directly calculated out by S-N curves and linear Miner cumulative damage rule. Owing to the uncertainty of loads, and the problem of the inaccuracy of calculated structural reliability index for the existence of deviation between measured data in projects and real data, the research method for structural fatigue reliability based on extension of random loads into interval variables is proposed. The innovation is that we can accurately calculate out the interval of the structural fatigue life and reliability index of a structure according to the probability density function of stress level of random loads and the coefficient of variation of measured loads. By practical calculation example, it is proved that this method is more suitable to practical engineering comparing to traditional methods. It will provide a perfect research approach for reliability analysis of the structure under random loads.

Keywords: *Random load, Structural reliability analysis, Fatigue life; Interval analysis, Rayleigh distribution*

1. Introduction

Study of structural fatigue lead by random loads, is an important aspect in the study of structural reliability. Fatigue failure is one of the main failure modes for a structure. According to the statistical data, 80% of structural failure belongs to fatigue fracture. For example, the shaft, bearings, springs and frames are mostly fractured under random loads. To accurately predict the fatigue life of a structure according to the random cyclic loads that the structure is bearing, is the foundation of limited fatigue lifetime design and reliability analysis for a structure. However, the current theoretical research mainly focuses on fatigue reliability of a structure under constant amplitude cyclic loads or multistage constant amplitude cyclic loads with ideal state, while fatigue reliability of a structure under random cyclic loads is relatively less. In fact, in practical engineering the structure under constant amplitude cyclic loads or multistage constant amplitude cyclic loads with ideal state is rare, and in the vast majority of cases, the structural fatigue failure is caused by random cyclic loads exerting repeatedly. In current theoretical analysis and practical engineering calculation, the conservative calculation method is commonly used, that is to choose a deterministic load with maximum peak value in all cyclic loads to substitute the actual loads, as well as to use the mathematical model of fatigue life of a structure under deterministic constant amplitude cyclic loads to predict the actual structural fatigue life. In fact, the results that calculated out are often far different from the practical condition of the structure.

2. Structural Fatigue Life Estimation Based on the Miner Rule

The calculation of the structural fatigue life is based on fatigue cumulative damage theory. When a structure is bearing constant amplitude cyclic loads with ideal state, according to S-N curves of the structural materials and structural characteristics, the fatigue life of the structure can be calculated out directly. When a structure is bearing multistage constant amplitude cyclic loads with ideal state, then the fatigue life of the structure can be calculated out according to linear Miner cumulative damage rule. Suppose that, a structure bears k stages stresses of fatigue loads in its fatigue life: $S_1, S_2, ..., S_k$, with the corresponding acting number for each stage of cyclic loads are respectively: $n_1, n_2, ..., n_k$. Then by Miner fatigue cumulative damage rule, and according to the S-N curves and structural characteristics of composing materials, we can get the fatigue cumulative damage of the structure produced by k stages of cyclic loads as follow:

$$\Delta D = \sum_{i=1}^{k} D(s_i) = \frac{n_1}{N_1} + \frac{n_2}{N_2} + ... + \frac{n_k}{N_k} = \sum_{i=1}^{k} \frac{n_i}{N_i} \quad (1)$$

During the above formula, n_i is the action number of cyclic loads of the i th stress level; N_i is the failure cycles number under the cyclic loads of the i th stress level. For a structure in service, when its fatigue cumulative damage produced by all levels of cyclic loads is up to the failure threshold (failure threshold is usually taken 1), it will fail.

The randomness of the cyclic loads will not only include the uncertainty of the stress level of loads, but also include the uncertainty of action times that the loads imposed to the structure. When taken the action times of cyclic loads as the lifetime measurement index, the uncertainty of loads mainly manifested as the uncertainty of stress level of loads, as shown in figure 1. Then, the uncertainty of stress level caused by the action of random cyclic loads can be described as the corresponding probability density function.

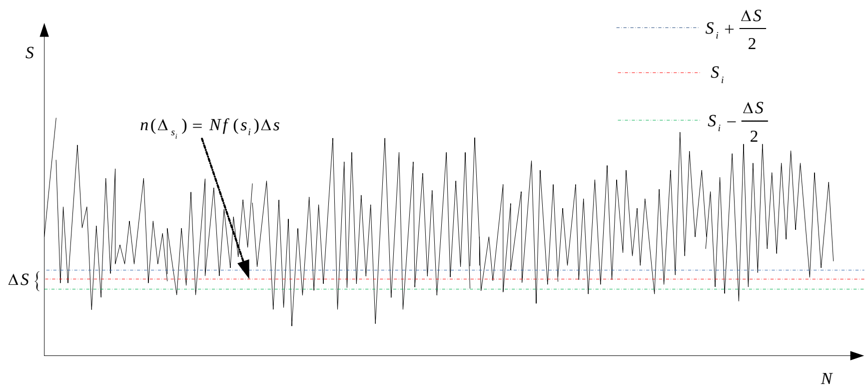


Fig. 1 Distribution of random cyclic loads

Suppose that, the stress levels of cyclic loads are random variables which obey probability density function $f(x)$. When the stochastic process is a narrow band process, the probability density of the peak values can be approximately considered as obeying Rayleigh distribution, that is:

$$f(S) = \frac{S}{\sigma^2}\exp\left(-\frac{S^2}{2\sigma^2}\right) \qquad (2)$$

During the above formula, σ is the shape parameter of state curve of Rayleigh distribution, and is a constant.

As shown in figure 2, according to the characteristics of probability density function, the probability $P(S_i)$ of stress values corresponding to the effect of per random cyclic loads acting, located in interval $\left[S_i - \frac{\Delta S}{2}, S_i + \frac{\Delta S}{2}\right]$ nearby S_i is

$$P(S_i) = f(S_i)\Delta S = \frac{S_i}{\sigma^2}\exp\left(-\frac{S_i^2}{2\sigma^2}\right)\Delta S \qquad (3)$$

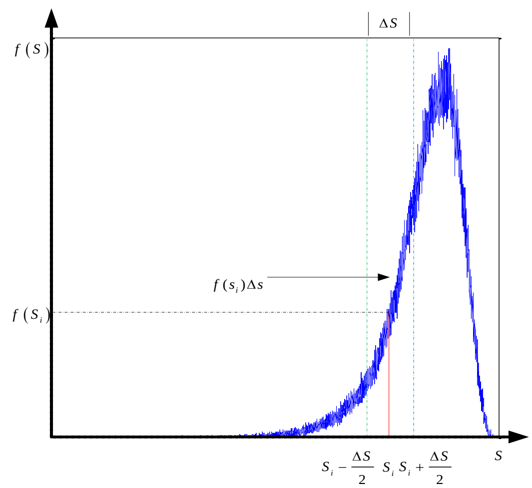


Fig. 2 Probability density distribution that the peaks of loads located in interval $\left[S_i - \frac{\Delta S}{2}, S_i + \frac{\Delta S}{2}\right]$ nearby S_i in Rayleigh distribution

If the action times of random cyclic loads is N, the times $n(\Delta_{S_i})$ that the stresses located in interval $\left[S_i - \frac{\Delta S}{2}, S_i + \frac{\Delta S}{2}\right]$ nearby S_i are

$$n(\Delta_{S_i}) = Nf(S_i)\Delta S = N\frac{S_i}{\sigma^2}\exp\left(-\frac{S_i^2}{2\sigma^2}\right)\Delta S \quad (4)$$

According to Miner fatigue cumulative damage rule, the structural fatigue cumulative damage $\Delta D_i(S_i)$ caused by the loads, whose stresses located in interval $\left[S_i - \frac{\Delta S}{2}, S_i + \frac{\Delta S}{2}\right]$, can be expressed as

$$\Delta D_i(S_i) = \frac{n(\Delta_{S_i})}{N_i} = \frac{N\dfrac{S_i}{\sigma^2}\exp\left(-\dfrac{S_i^2}{2\sigma^2}\right)\Delta S}{N_i} \quad (5)$$

For the reason of the structure are absorbing the acting of fatigue loads with k stress levels, let's divide the taking value of the whole interval of probability distribution of stress level into k subintervals correspondingly, and suppose that the midpoint of the i th subinterval is S_i. Form the above analysis, we can learn that, when the action times of the random cyclic loads is N, the stress acting times that located in the i th subinterval is $n(\Delta_{s_i})$, and the corresponding structural fatigue cumulative damage is $\Delta D_i(S_i)$.

According to formula (1), we can get the total structural fatigue cumulative damage under N times action of random cyclic loads is as follow

$$D = \Delta D_1(S_1) + \Delta D_2(S_2) + \ldots + \Delta D_k(S_k) = \sum_{k=1}^{k} \Delta D_i(S_i) \quad (6)$$

Substitute formula (5) into formula (6), we can get

$$D = \frac{Nf(S_1)\Delta S}{N_1} + \frac{Nf(S_2)\Delta S}{N_2} + \ldots + \frac{Nf(S_k)\Delta S}{N_k} = \sum_{k=1}^{k} \frac{N\dfrac{S_i}{\sigma^2}\exp\left(-\dfrac{S_i^2}{2\sigma^2}\right)\Delta S}{N_i} \quad (7)$$

Usually, the relationship between the fatigue life N of structural material and stress level S can be expressed as the form of power series, as follow:

$$S^m N = C \quad (8)$$

During the above formula, m and C are material constant.

According to the relationship between fatigue life and stress level shown in formula (8), the fatigue life N_i corresponding to stress level S_i can be expressed as:

$$N_i = \frac{C}{S_i^m} \quad (9)$$

Substitute formula (9) into formula (7), we can get

$$\begin{aligned} D &= \frac{Nf(S_1)S_1^m\Delta S}{C} + \frac{Nf(S_2)S_2^m\Delta S}{C} + \ldots + \frac{Nf(S_k)S_k^m\Delta S}{C} \\ &= \sum_{k=1}^{k} \frac{Nf(S_i)S_i^m\Delta S}{C} = \sum_{k=1}^{k} \frac{N\dfrac{S_i}{\sigma_S^2}\exp\left(-\dfrac{S_i^2}{2\sigma_S^2}\right)S_i^m\Delta S}{C} \end{aligned} \quad (10)$$

When $k \to \infty$, formula (10) is equivalent to the following formula:

$$\begin{aligned} D &= \lim_{k\to\infty}\sum_{k=1}^{k} \frac{Nf(S_i)S_i^m\Delta S}{C} = \int_0^{+\infty} \frac{N\dfrac{S}{\sigma^2}\exp\left(-\dfrac{S^2}{2\sigma^2}\right)S^m}{C} ds \\ &= N\int_0^{+\infty} \frac{\dfrac{S^{m+1}}{\sigma^2}\exp\left(-\dfrac{S^2}{2\sigma^2}\right)}{C} ds \end{aligned} \quad (11)$$

Furthermore, we can get the structural fatigue life N under the action of random cyclic loads as follow:

$$N = \frac{D}{\displaystyle\int_0^{+\infty} \frac{\dfrac{S^{m+1}}{\sigma^2}\exp\left(-\dfrac{S^2}{2\sigma^2}\right)}{C} ds} \quad (12)$$

When the fatigue failure of the structure happen, its fatigue cumulative damage is equivalent to the failure threshold, that is $D = \delta = 1$. Then formula (12) can be written as:

$$N = \frac{1}{\displaystyle\int_0^{+\infty} \frac{\dfrac{S^{m+1}}{\sigma^2}\exp\left(-\dfrac{S^2}{2\sigma^2}\right)}{C} ds} \quad (13)$$

During formula (12) and formula (13), owing to δ, C, σ and m are all for constant. We can calculate out the structural fatigue life on the basis of the mathematical model established above, according to the known probability density function of the random cyclic loads stress levels and the material parameters of the structure.

3. Mathematical Modeling of Structural Fatigue Reliability Based on Extension of Random Loads into Interval Variables

Suppose that the designed life of the structure is n_D, then let's first establish the structural fatigue state equation g_n (safety margin), measured by loading action times, that is:

$$g_n = \frac{1}{\int_0^{+\infty} \frac{\frac{S^{m+1}}{\sigma^2}\exp\left(-\frac{S^2}{2\sigma^2}\right)}{C} ds} - n_D \tag{14}$$

To taylor series expansion for the state equation g_n on the average point μ_g, and according to the linear feature of interval variables operation, the mean value and the deviation of the structural response are respectively:

$$\left.\begin{aligned} \mu_g &= g(x_1^c, x_2^c, \ldots, x_n^c) \\ \sigma_g &= \sum_{i=1}^{n}\left(\left|\frac{\partial g}{\partial x_i}\right|\right)_{\mu} \sigma_{x_i} \end{aligned}\right\} \tag{15}$$

Substitute formula (14) into formula (15), we can get the mean value and deviation of the fatigue state equation of the structure under the action of random cyclic loads, which are respectively:

$$\left.\begin{aligned} \mu_g &= \frac{1}{\left(\int_0^{+\infty} \frac{\frac{S^{m+1}}{\sigma^2}\exp\left(-\frac{S^2}{2\sigma^2}\right)}{C} ds\right)_{\mu_S}} - n_D \\ \sigma_g &= \left(\left|\frac{\exp\left(-\frac{S}{2\sigma^2}\right)\left[\left(\frac{S^{m+2}}{2\sigma^4}\right) - \frac{(m+1)S^m}{\sigma^2}\right]}{\frac{1}{C}\left[\int_0^{+\infty} \frac{S^{m+1}}{\sigma^2}\exp\left(-\frac{S^2}{2\sigma^2}\right) ds\right]^2}\right|\right)_{\mu_S} \sigma_s \end{aligned}\right\} \tag{16}$$

By the properties of Rayleigh distribution, we can get $\mu_S = \sigma\sqrt{\frac{\pi}{2}}$, $\sigma_S = \frac{4-\pi}{2}\sigma^2$.

Expand the random variables into interval variables, and the result is:

$$\left.\begin{aligned} \mu_g^I &= \left[\underline{\mu}_g, \overline{\mu}_g\right] = \left[\mu_g(1-\alpha_{\mu_g}), \mu_g(1+\alpha_{\mu_g})\right] \\ \sigma_g^I &= \left[\underline{\sigma}_g, \overline{\sigma}_g\right] = \left[\sigma_g(1-\alpha_{\sigma_g}), \sigma_g(1+\alpha_{\sigma_g})\right] \end{aligned}\right. \tag{17}$$

$$\beta^I = [\underline{\beta}, \overline{\beta}] = \left[\frac{\underline{\mu}_g}{\overline{\sigma}_g}, \frac{\overline{\mu}_g}{\underline{\sigma}_g}\right] \tag{18}$$

For the convenience of calculation, let be $\alpha_{\mu_g} = \alpha_{\sigma_g} = \alpha$, then substitute formula (17) into formulas (16) and (18), and the result is:

$$\underline{\beta} = \frac{\underline{\mu}_g}{\overline{\sigma}_g} = \frac{\dfrac{1}{\left(\int_0^{+\infty} \frac{\frac{S^{m+1}}{\sigma_S^2}\exp\left(-\frac{S^2}{2\sigma_S^2}\right)}{C} ds\right)_{\underline{\mu}_S = \mu_S(1+\alpha)}} - n_D}{\left(\left|\frac{\exp\left(-\frac{S}{2\sigma^2}\right)\left[\left(\frac{S^{m+2}}{2\sigma^4}\right) - \frac{(m+1)S^m}{\sigma^2}\right]}{\frac{1}{C}\left[\int_0^{+\infty} \frac{S^{m+1}}{\sigma^2}\exp\left(-\frac{S^2}{2\sigma^2}\right) ds\right]^2}\right|\right)_{\mu_S} \sigma_s(1+\alpha)} \tag{19}$$

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

$$\overline{\beta}=\frac{\overline{\mu}_g}{\underline{\sigma}_g}=\frac{\dfrac{1}{\left(\displaystyle\int_0^{+\infty}\frac{\frac{S^{m+1}}{\sigma_S^2}\exp\left(-\frac{S^2}{2\sigma_S^2}\right)}{C}ds\right)_{\underline{\mu}_S=\mu_S(1-\alpha)}}-n_D}{\left(\left|\dfrac{\exp\left(-\frac{S}{2\sigma^2}\right)\left[\left(\frac{S^{m+2}}{2\sigma^4}\right)-\frac{(m+1)S^m}{\sigma^2}\right]}{\frac{1}{C}\left[\int_0^{+\infty}\frac{S^{m+1}}{\sigma^2}\exp\left(-\frac{S^2}{2\sigma^2}\right)ds\right]^2}\right|\right)_{\mu_S}\sigma_s(1-\alpha)} \tag{20}$$

4. Example

Suppose that the fatigue loads that a structure bears are random cyclic loads and the random process is a narrow band one, then the probability density of the peak value of stress levels for cyclic loads can be approximately deemed to obey Rayleigh distribution. Let be $\sigma=1$, and the mean stress level for cyclic loads is $\mu_S=\sigma\sqrt{\frac{\pi}{2}}$, while the deviation is $\sigma_S=\frac{4-\pi}{2}\sigma^2$. Given the influence of environment and other factors in the actual engineering process, there will be some errors between the actual situation and the measured mean stress levels of cyclic loads and deviation. Let the coefficient of variation be $\alpha_{\mu_S}=\alpha_{\sigma_S}=0.1$, and the value of the parameter curve m is 3, and C value is 1×10^{10}. The designed life is $n_D=10^6$. Try to analyze and calculate out the reliability of the structure system.

Solution: (1) When the coefficient of variation of stress levels for cyclic loads during data acquisition is not taken into account, the stress levels for cyclic loads will be random variables. Substitute the known conditions into formula (16), and the result is:

$$\mu_g=1.313742\times10^6$$

$$\sigma_g=0.495317\times10^6$$

And then

$$\beta=\frac{\mu_g}{\sigma_g}=2.652326$$

(2) When the coefficient of variation of stress levels for cyclic loads during data acquisition is taken into account, the stress levels for cyclic loads will be interval variables. Substitute the known conditions into formulas (17), (19), and (20), and the result is:

$$\overline{\mu}_g=1.445116\times10^6$$

$$\underline{\mu}_g=1.182368\times10^6$$

$$\overline{\sigma}_g=0.544849\times10^6$$

$$\underline{\sigma}_g=0.445785\times10^6$$

And then

$$\underline{\beta}=\frac{\underline{\mu}_g}{\overline{\sigma}_g}=2.170084$$

$$\overline{\beta}=\frac{\overline{\mu}_g}{\underline{\sigma}_g}=3.241733$$

According to the above analysis, conclusions can be drawn that due to the existing of coefficient of variation of mean value and deviation of stress levels for cyclic loads, there will be some deviations between the calculated index of the structure reliability and the true value. And if the true index of structure reliability is substituted by this calculated value, the calculated structure reliability may be too conservative or too optimistic, and thus not in line with the actual engineering process. However, the use of the theory of expanding random variables into interval variables can help identify approximately the interval range of indexes of structure reliability.

5. Conclusions

（1） Due to the indeterminacy of loads, S-N Curves and linear Miner cumulative fatigue damage rules are not suitable for the study on fatigue life of structures under the action of random loads. This article builds a model for calculation of fatigue life of structures under the actions of random loads, and can obtain the corresponding fatigue life of structures relatively more accurate , according to the probability density function of stress distribution for random cyclic loads. Then the corresponding indexes of structure reliability can be determined with the relationship between the calculated structure fatigue life and designed life.

（2） As there are inevitable deviations between measured data in the actual engineering process and the authentic data, correspondingly, the calculated indexes of the structure reliability based on the measured data cannot be the exact values. If they are used for deduction of structure reliability, it will not be in line with the actual engineering process. This article expands the random variables of random cyclic loads into interval variables, builds a model for indexes of structure reliability with the method of interval analysis, and determines the fluctuation range for indexes of structure reliability, from which the lower limit can be taken in structure design.

References

[1] Songqing Shan, G. Gary Wang, “Reliable design space and complete single-loop reliability-based design optimization”, Reliability Engineering and System Safety, Vol. 93, No. 7, 2008, pp. 1218-1230.

[2] Philipp Limbourg, EtiennedeRocquigny, “Uncertainty analysis using evidence theory–confronting level-1 and level-2 approaches with data availability and computational constraints”, Reliability Engineering and System Safety, Vol. 95, No. 1, 2010, pp. 550-564.

[3] C. Jiang, G.R. Liu, X. Han, “A Novel Method for Uncertainty Inverse Problems and Application to Material Characterization of Composites”, Experimental Mechanics, Vol. 48, No. 7, 2008, pp. 539-548.

[4] D. Moens , D. Vandepitte, “Recent Advances in Non-probabilistic Approaches for Non-deterministic Dynamic Finite Element Analysis”, Archives of Computational Methods in Engineering, Vol. 13, No. 3, 2006, pp. 389-464..

[5] Juan Ma, Peter Wriggers, “Reliability-based optimization of trusses with random parameters under dynamic loads”, Comput Mech, Vol. 47, No. 1, 2011, pp. 627-640.

[6] Yoshihiro Kanno, “Worst scenario detection in limit analysis of trusses against deficiency of structural components”, Engineering Structures, Vol. 33, No. 04, 2012, pp. 33-42.

[7] Marco Lombardi, Raphael T. Haftkab, “Anti-optimization technique for structural design under load uncertainties”, Computer methods in applied mechanics and engineering, Vol. 157, No. 5, 1998, pp. 19-31.

[8] Wang Xiaojun, Wang Lei, “Response analysis based on smallest interval-set of parameters for structures with uncertainty”, Applied Mathematics and Mechanics, Vol. 33, No. 9, 2012, pp. 1153-1166.

[9] Marco Savoia, “Structural reliability analysis through fuzzy number approach with application to stability”, Computers and Structures, Vol. 80, No. 5, 2002, pp. 1087-1102.

[10] David Moens, Dirk Vandepitte, “A survey of non-probabilistic uncertainty treatment in finite element analysis”, Computer Methods in Applied mechanics and Engineering, Vol. 194, No. 3, 2005, pp. 1527-1545.

[11] A. S. Balu, B. N. Rao, “Multicut-High Dimensional Model Representation for Structural Reliability Bounds Estimation Under Mixed Uncertainties”, Computer-Aided Civil and Infrastructure Engineering, Vol. 27, No. 10, 2012, pp. 419-438.

[12] Ak. Ashakumar Singh, Momtaz Thingujam, “Fuzzy ID3 Decision Tree Approach for Network Reliability Estimation”, International Journal of Computer Science Issues, Vol. 9, No. 1, 2012, pp. 446-450.

[13] Ritika Wason, P.Ahmed, M.Qasim Rafiq, “Novel Software Reliability Estimation Model for Altering Paradigms of Software Engineering”, International Journal of Computer Science Issues, Vol. 9, No. 3, 2012, pp. 92-99.

QIANGFENG WANG is a doctorate student studying in machine-electronic engineering speciality, Northwestern Polytechnical University, Xi'an since 2009.He has done her graduation from Weifang University, Weifang and done her post graduation from Jiangsu Polytechnic University, Changzhou in the area of reliability analysis. He has contributed 2 articles in Conferences and 20 papers in research journals. His areas of interest are reliability analysis and failure diagnosis.

Clinical Relationships Extraction Techniques from Patient Narratives

Wafaa Tawfik Abdel-moneim[1], Mohamed Hashem Abdel-Aziz[2], and Mohamed Monier Hassan[3]

[1]Information System Department, Faculty of Computers and Informatics, Zagazig University
Banha, Egypt

[2]Information System Department, Faculty of Computers and Informatics, Ain-Shames University
Cairo, Egypt

[3]Information System Department, Faculty of Computers and Informatics, Zagazig University
Zagazig, Egypt

Abstract

The Clinical E-Science Framework (CLEF) project was used to extract important information from medical texts by building a system for the purpose of clinical research, evidence-based healthcare and genotype-meets-phenotype informatics. The system is divided into two parts, one part concerns with the identification of relationships between clinically important entities in the text. The full parses and domain-specific grammars had been used to apply many approaches to extract the relationship. In the second part of the system, statistical machine learning (ML) approaches are applied to extract relationship. A corpus of oncology narratives that hand annotated with clinical relationships can be used to train and test a system that has been designed and implemented by supervised machine learning (ML) approaches. Many features can be extracted from these texts that are used to build a model by the classifier. Multiple supervised machine learning algorithms can be applied for relationship extraction. Effects of adding the features, changing the size of the corpus, and changing the type of the algorithm on relationship extraction are examined.

Keywords: *Text mining; information extraction; NLP; entities; and relations.*

1. Introduction

Text mining can be defined as a knowledge-intensive process in which user deal with a document collection over time to extract useful and previously unknown information from data sources by using a suite of analysis tools. Text mining deals with the documents that are found in unstructured textual data. Text mining involves the application of techniques from areas such as Information Retrieval (IR), Natural Language Processing (NLP), Information Extraction (IE) and Data Mining (DM). NLP is commonly divided into several layers of processing: lexical, syntactic, and semantic level. The lexical level processing deals with words that can be recognized, analyzed, and identified to enable further processing. The syntactic level analysis deals with identification of structural relationships between groups of words in sentences, and the semantic level is concerned with the content-oriented perspective or the meaning attributed to the various entities identified within the syntactic level [1]. Natural Language Processing (NLP) has been widely applied in biomedicine, particularly to improve access to the ever-burgeoning research literature. Increasingly, biomedical researchers need to relate this literature to phenotypic data: both to populations, and to individual clinical subjects. The computer applications used in biomedical research therefore need to support genotype-meets-phenotype informatics and the move towards translational biology. This will undoubtedly include linkage to the information held in individual medical records: in both its structured and unstructured (textual) portions. Information extraction is the process of automatically obtaining structured data from an unstructured natural language document. Often this involves defining the general form of the information that are important in as one or more templates, which then are used to guide the extraction process. IE systems rely heavily on the data generated by NLP systems [2].

Information extraction system contains information such extract relations between entities from texts. Figure 1 contains example of relation mentions from the news data sets. The left side of the figure contains a pipeline representation of the RE task. The input consists of natural language documents containing e.g. unstructured text or speech. These documents are fed to the RE system, which identifies and characterizes the relations described in the text or speech data. The output of the RE system consists of relation mention triples which include the two entity mentions that take part in the relation and the relation type. The right side of Figure 1 contains example input document on the top and the relation mention triples from these sentence on the bottom. The document contains the sentence "George Bush traveled to France on Thursday for a summit". This contains relation mentions: a reference to a Physical.Located relation between "George Bush" and "France".

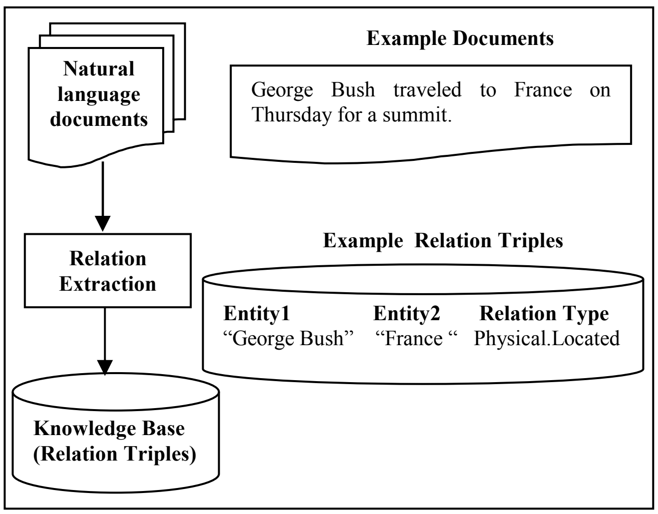


Fig. 1 Overview of relation extraction task with example input and output.

For processing the clinical information, a framework is defined which called the Clinical E-Science Framework (CLEF) project [3] for capture, integration and presentation of this information. The project's data resource is clinical narratives of the cancer patients from The University of Chicago medicine and Mayo Clinic. Information Extraction (IE) technology can be used by CLEF project to extract important information from clinical text. Entities, relationships and modifiers can be extracted from text by the CLEF IE system. The purpose of extracting these relationships is obtaining important information that is often not available in the structured record. What were the interventions for treating the problems? Where was the disease located? Which drugs were given for treating the problems? The extracted relationships are very important for clinical and research applications of IE. The supervised learning algorithm has two properties. The first property is that it does not learn any training example until an unseen example is given; it is called lazy based learning algorithm [4]. The second property is that it classified unseen objects based on target labels of their similar samples; it is called example based learning algorithm [4]. The project uses the guidelines of the CLEF IE system that concerned with extracting entities [5]. Gold standard – human annotated documents – is used to build models of patient narratives which can be applied to unseen patient files. This paper focuses on extracting relationships from patient narratives. Our approach uses different supervised learning algorithms such Support Vector Machine (SVM), Perceptron Algorithm with Uneven Margins (PAUM), NaiveBayes Weka, KNN Weka, and C4.5 Weka classifiers to extract these relationships. The classifiers use a gold standard corpus of oncology narratives which hand annotated with entities and relationships for system training and evaluation. A lot of experiments are applied to reach the much suitable algorithm that gives the high accuracy. The experiments are used different feature sets and see their effects on the system performance. These features sets derived from a linguistic analysis and syntactic analysis of sentence. Relationship is extracted in the same sentence which called inter-sentential relationships. Examine the influence of changing training corpus size for relationship extraction.

2. Previous Work

Now a day it is easy to store large amounts of data. Documents are available on the web, intranets, and news wires. However, amount of data available to us are still increasing, our ability is to extract the useful information from this data. Text mining is a good technique to extract useful information from texts. There are many forms of the useful information that can be extracted from the texts such as entities, events, attributes and facts. This information is helpful for researchers to understand the texts very easy. Information extraction (IE) framework has the practical goal of extracting structured information from natural language [6]. IE as a task was formalized largely in the context of the Message Understanding Conference (MUC) shared tasks (e.g., [7]; [8]). Message Understanding Conference (MUCs) [9] describes a method of classifying facts (information) into categories or levels. The researchers develop MUC project by adding new part that contains extracting relationships between entities that take place in MUC-7 [10] such employee_of, product_of, location_of.

In MUC-7, the training examples can be analyzed and hand annotated by the researchers to match contexts which expressed the relevant relation. The work in MUC can be classified into several dimensions: the text type (e.g. newswire, scientific papers, clinical reports); the relations addressed (e.g. part-of, located-in, protein-protein interaction); and the techniques used (e.g. rule-based engineering techniques, supervised learning techniques). In the case of rule-based engineering, writing extraction rules requires extensive effort from a rule engineering expert who is familiar with the target domain. In the case of supervised learning, annotation of training data and features/model parameters require extensive effort from at least one annotator (expert in the target domain) and from a natural language processing expert. In addition to MUC can be classified according to annotated corpora and evaluation software exist (e.g. the ACE relation extraction challenges [11], the LLL genic interaction extraction challenge [12], the BioCreative-II protein-protein interaction task [13]). Many systems use a syntactic parse with domain-specific grammar rules such Linguistic String project [14] to fill template data structures corresponding to medical statements. Other systems use a semantic lexicon and grammar of domain-specific semantic patterns such MedLEE [15] and BioMedLEE [16] to extract the relationships between entities. Other systems use a dependency parse of texts such MEDSYNDIKATE [17] to build model of entities and their relationships. MENELAS [18] also use a full parse. All these approaches are knowledge-engineering approaches. In addition to supervised machine learning has been applied to clinical text. There are many works on relation extraction from biomedical journal papers and abstracts. This work has been done within the hand-written rule base/knowledge engineering approaches.

Now a days the work on relation extraction using supervised ML techniques to train relation classifiers on human annotated texts. The annotated texts contain relation instances which contain relation type and their pair entities. There are many different approaches according to the ML algorithms and the features applied. There are several applications work on biomedicine such using maximum entropy approaches [19], conditional random fields [20] and rule learning methods such as boosted wrapper induction and RAPIER [21] and inductive logic programming [22]. SVMs also have been used for relation extraction [23] but not widely in biomedicine applications. Additional examples for relation extraction contains on [24], [25], and [26]. Currently researchers extract relations from clinical text (such patient narratives) using wide range of features by supervised ML approaches such as SVMs classifiers [27] and [28]. Relationships can be extracted from clinical texts as a part of clinical IE system. Different supervised Machine Learning approaches can be applied to show their affecting on the classification tasks specifically in relation extraction. Several features are used to extract relations such lexical, syntactic, and semantic features.

3. Techniques

3.1 Relationship schema

Firstly, this application focuses on extracting entities, relationships and modifiers from text. The real thing or event can be found in text documents called entity. Examples of entities in clinical text include diseases, location and drugs and so on. The words that describe an entity called modifiers such as the negation of a condition ("no sign of cancer"), the sub_location of an anatomical locus ("superior-vena-caval "). Relationships are entities that connected to each other and to modifiers e.g. linking an investigation to its result (CT scan shows no abnormality), linking the condition to an anatomical locus (back pain), and linking laterality to an anatomical locus (right groin). Entities, modifiers, and relationships can be determined in XML schema with hand- annotated. Table 1 shows relationship extraction and their argument types from clinical text with a description and examples of each type. Unified Medical Language System (UMLS) semantic network [29] was used to map each entity type into several UMLS types. Relationship shows the clinical dependencies between entities in patient narrative. The schema of relationship was described in [30]. Relationship is linking between pairs of specific entity types, e.g. the Has_finding relation link between investigation and result or between investigation and condition. The schema of relationship types and their argument types is shown graphically in Figure 2.

3.2 Gold standard corpus

Entities and relationships in oncology narratives can be hand-annotated by the schema and definitions to provide a gold standard for system training and evaluation. Narratives refer to notes, letters, and summaries written by the oncologist that describe the patient' care. Given the expense of human annotation, the gold standard portion of the corpus has to be a relatively small subset of the whole corpus. In order to avoid events that are either rare or outside of the main project requirements, it is restricted by diagnosis, and only considers documents from those patients with a primary diagnosis code in one of the top level sub-categories of ICD-10 Chapter II (neoplasms) [30]. In order to ensure even training and fair evaluation across the entire corpus, Narratives were selected by randomised and stratified sampling from a larger population of documents. The corpus contains 40 narratives, which were carefully selected and annotated according to a best approach, as described in [30]. This corpus is clinical narratives of the cancer patients from The

University of Chicago medicine [31] and Mayo Clinic hospital [32].

Table 1: Description of relationship types and their argument

Relationship type	***First argument type***	***Second argument type***	***Description***	***Example***
Has_target	Investigation, Intervention	Locus	Relates an intervention or an investigation to the bodily locus at which it is targeted.	• This patient has had a [arg1] bowel [arg2] ultrasound. • This patient has had a [ard2] chest [arg1] X-ray.
Has_finding	Investigation	Condition, Result	Relates a condition to an investigation that demonstrated its presence, or a result to the investigation that produced that result.	• This patient has had a [arg1] Ultrasound scanning which shows [arg2] hydronephrosis. • A chest [arg1] X-ray was [arg2] normal.
Has_indication	Drug or device, Intervention, Investigation	Condition	Relates a condition to a drug, intervention, or investigation that is targeted at that condition	• … [arg1] chemotherapy to treat the [arg2] cancer. • [arg1] remove brain [arg2] tumors.
Has_location	Condition	Locus	Relationship between a condition and a locus: describes the bodily location of a specific condition.	• This patient has had a [arg1] benign cyst on her [arg2] thyroid. • This patient has had a [arg1] lung [arg2] cancer.
Modifies	Negation signal	Condition	Relates a condition to its negation or uncertainty about it.	• There was [arg1] no signs of the [arg2] tumor. • There was [arg1] no-evidence of superior-vena-caval [arg2] obstruction.
Modifies	Laterality signal	Locus, Intervention	Relates locus or intervention to its sidedness: right, left, bilateral.	•…in her [arg1] left [arg2] breast. • [arg1] bilateral [arg2] mastectomies.
Modifies	Sub-location signal	Locus	Relates locus to other information about the location: upper, lower, extra- etc.	• [arg1] lower [arg2] lung. • [arg1] outside the [arg2] prostate. • This patient suffers from [arg1] upper [arg2] abdominal pain.

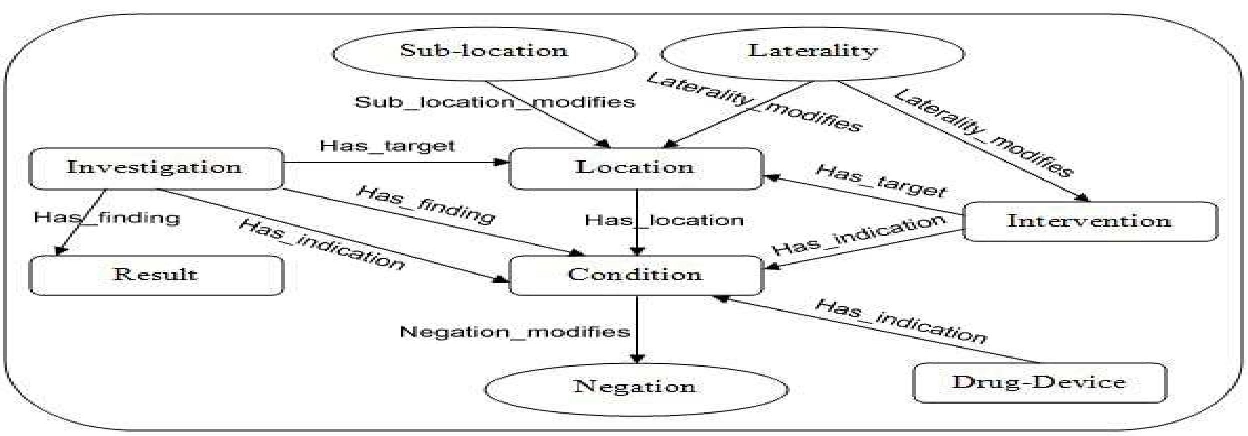


Fig. 2 The relationship schema, showing entities (Rounded rectangles), modifiers (ovals), and relationships (arrows).

3.3 Relationship extraction

The GATE NLP toolkit has built our system. GATE NLP toolkit is the tool that allows the applications to be constructed as a pipeline of processing resources [33]. Each resource in this pipeline analyzes the documents, the results of this analysis being available to later resources. The system is shown in Figure 3, and is described below [27]. The pre-processing technique of narratives carried out by using standard GATE modules. The processing resources that used to manipulate the narratives are tokeniser to split narratives into tokens, sentence splitter to split narratives into sentences, part-of-speech (POS) tagged for word tokens, and morphological analyser to find roots for word tokens. POS resource also provides each token with generic POS tag that contains of the first two characters of full POS tag, which called a "generalised" POS tag. After pre-processing technique, guidelines that described in [34] were used assuming that entities extraction is perfect recognition, as given by the entities in the human annotated gold standard described above. The relation extraction depends on the quality of entity extraction.

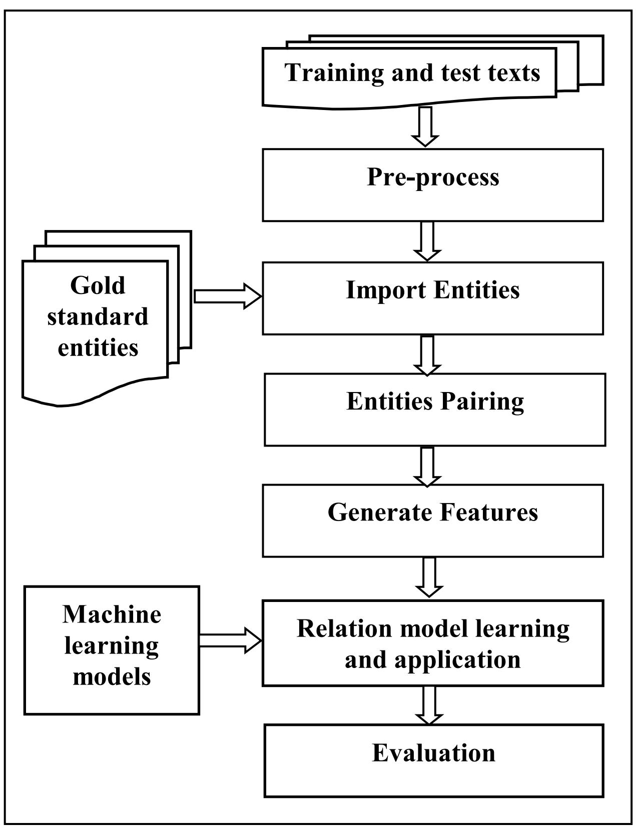


Fig. 3 Relationship extraction system as a GATE pipeline.

3.3.1 Classification

The clinical relationship extraction can be manipulated a classification task by assigning relationship type to an entity pair. Pairing the entities that may or may not be the arguments of a relation is called an entity pair. To apply an entity pairing task for each document, all entity pairs that are possible can be created under two constraints. The first constraint, entities pairs must be inside n sentences of each other. For all works in this paper, entities have paired in the same sentence, $n \leq 1$ (crossing 0 or 1 sentence boundaries). The second constraint, entity pairs must be constrained by argument type [35]. For example, there is no relationship between Drug or device entity and a Result entity as specified by the relationship schema. GATE resource developed specifically for extracting relationship from medical texts. This resource also assigns features that characterize lexical and syntactic qualities (described below) of each pair. Entity pairs are compatible with classifier training and test instances. In classifier training, there is two types of results are “class” if an entity pair is compatible with the arguments of a relationship present in the gold standard then there is a class of that relationship type and “class null” if an entity pair is not compatible with the arguments of a relationship. Features of entity pair training instances are used to build a model by the training classifier. In classifier application, unseen text can be used to create entity pairs, under the above constraints. In this classifier, each entity pair assigned class of relationship types or class null [27].

Because of the machine learning algorithms solve a binary-class problem thus to solve a multi-class problem, ML maps this problem to a number of binary classification problems. In multi-class problem the ML plugin implements two common methods are one-against-one and one-against-all. In one-against-one approach each pair of classes require one classifier. In one-against-all approach require a classifier for a binary decision of each pair of the n classes. One-against-all technique is used in our application to solve the multi-class problem.

4. Algorithms

There are different machine learning algorithms that can be used for classification to extract relations between entities such as Naïve Bayesian, Decision Tree, K Nearest Neighbour, Perceptron, and Support Vector Machine.

4.1 Naive Bayesian

Naive Bayesian classifier is a statistical classifier based on the Bayes’ Theorem and the maximum posteriori hypothesis [36]. Let T be a training set of instances. Considering that each data instance to be an n-dimensional vector of attribute values:

$$X = \{x_1, x_2, \ldots, x_n\} \qquad (1)$$

In a Bayesian classifier which assigns each data instance to one of k classes $C_1, C_2, \ldots, C_k$, a data instance X is assigned to the class for which it has the highest posterior probability conditioned on X. This means that, X is assigned to class Ci if and only if

$$P(C_i|X) > P(C_j|X) \text{ for all j such that } 1 \le j \le n, j \ne i. \qquad (2)$$

According to Bayes Theorem

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \qquad (3)$$

Since $P(X)$ is a normalizing factor which is equal for all classes, maximizing the numerator $P(X \mid C_i)P(C_i)$ is used to do the classification. Values of $P(X \mid C_i)$ and $P(C_i)$ can be estimated from the data that used to build the classifier.

4.2 Decision Tree

A decision tree is a tree data structure consisting of decision nodes and leaves. A leaf specifies a class value. A decision node specifies a test over one of the attributes, which is called the attribute selected at the node. The C4.5 algorithm constructs the decision tree with a divide and conquers strategy. In C4.5, each node in a tree is associated with a set of cases. Also, cases are assigned weights to take into account unknown attribute values. The C4.5 algorithm uses the concept of information gain or entropy reduction to select the optimal split [37]. Figure 4 shows the pseudo-code of the C4.5 Tree-Construction.

The information gain of an attribute a for a set of cases T is calculated as follow. If a is discrete, and $T_1, \ldots, T_S$ are the subsets of T consisting of cases with distinct known value for attribute a, then:

$$\text{gain} = \text{info(T)-}\sum_{i=1}^{s}\frac{|T_i|}{|T|}\times \text{info}(T_i). \qquad (1)$$

Where

$$\text{info}(T) = -\sum_{j=1}^{NClass}\frac{freq(C_j,T)}{|T|}\times \log 2(\frac{freq(C_j,T)}{|T|}) \qquad (2)$$

is the entropy function. While having an option to select information gain, by default, however, C4.5 considers the information gain ratio of the splitting $T_1, \ldots, T_S$, which is the ratio of information gain to its split information:

$$Split(T) = -\sum_{i=1}^{s}\frac{|T_i|}{|T|}\times \log 2(\frac{|T_i|}{|T|}). \qquad (3)$$

```
FormTree (T)
    (1) ComputeClassFrequency (T);
    (2) If OneClass or FewCases
            Return a leaf;
        Create a decision node N;
    (3) ForEach Attribute A
            ComputeGain (A);
    (4) N.test = AttributeWithBestGain;
    (5) If N.test is continuous
            Find Threshold;
    (6) ForEach T′ in the splitting of T
    (7) If T′ is Empty
            Child of N is a leaf
          Else
    (8) Child of N = FormTree (T′);
    (9) ComputeErrors of N;
       Return N
```

Fig. 4 Pseudo-code of C4.5 tree-construction.

4.3 K-Nearest-Neighbor Algorithm (KNN)

K-Nearest-Neighbor classifier is a statistical classifier. When a new sample arrives, *k*-NN finds the K neighbors nearest to the new sample from the training space based on some suitable similarity or distance metric. A common similarity function is based on the Euclidian distance between two data. There are three key elements [38]: a set of labeled objects (e.g., a set of stored records), a distance or similarity metric to compute distance between objects, and the value of *k*, the number of nearest neighbors. To classify an unlabeled object, the distance of this object to the labeled objects is computed, its *k*-nearest neighbors are identified, and the class labels of these nearest neighbors are then used to determine the class label of the object. Figure 5 shows the pseudo-code of the *k*-nearest neighbor classification algorithm. Where *v* is a class label, *yi* is the class label for the *i*th nearest neighbors, and *I* (·) is an indicator function that returns the value 1 if its argument is true and 0 otherwise. A majority vote can be a problem if the nearest neighbors vary widely in their distance and the closer neighbors more reliably indicate the class of the object. Another approach to solve this problem by weighting each object's vote by its distance, where the weight factor:

$$w_i = 1/d(x', x_i)^2 \qquad (1)$$

This amounts to replacing the last step of the KNN algorithm with the following: Distance-Weighted Voting:

$$y' = \arg\max_{\upsilon} \sum_{(X_i, y_i) \in D_z} w_i \times I(\upsilon = y_i). \qquad (2)$$

Input: D, the set of k training objects and test object $z = (x', y')$

Process:

Compute $d(x', x)$, the distance between and every object, $(x, y) \in D$

Select $D_z \subseteq D$, the set of closest training objects to z.

Output: $y' = \arg\max_{\upsilon} \sum_{(x_i, y_j) \in D_z} I(\upsilon = y_i)$

Fig. 5 *K*-Nearest Neighbor classification algorithm.

4.4 Perceptron with uneven margin (PAUM)

The advantages of Perceptron with margins are simple, effective, and on-line learning algorithm [39]. The training examples can be checked one by one by predicting their labels by Perceptron. The example is succeeded when the prediction is correct. The example is used to correct the model when the prediction is wrong. The algorithm stops when all training examples are classified by the model correctly. The margin Perceptron has better generalization performance than the standard Perceptron. Figure 6 describes the algorithm of Perceptron with uneven margin.

Require: A linearly separable training example
$z = (x, y) \in (X \times \{-1, +1\})^m$
Require: A learning rate $\eta \in R^+$
Require: Two margin parameters $\tau_{-1}, \tau_{+1} \in \square^+$
$w_0 = 0$; $b_0 = 0$; $t = 0$; $R = \max_{x_i \in x} \|X_i\|$
Repeat
  For $i = 1$ **to** m **do**
    If $y_i(\langle w_t, x_i \rangle + b_t) \le \tau_{y_i}$ **then**
      $w_{t+1} = w_t + \eta y_i x_i$
      $b_{t+1} = b_t + \eta y_i R^2$
      $t \leftarrow t + 1$
    End if
  End for
Until no updates made within the **for** loop
Return (w_t, b_t)

Fig. 6 Algorithm of PAUM (τ_{-1}, τ_{+1}).

4.5 Support vector machine (SVM)

One of the most successful machine learning methods for IE is Support Vector Machine (SVM), which is a general supervised machine learning algorithm. It has achieved state-of-the-art performance on many classification tasks, including named entity recognition [40]. The GATE-SVM system uses a variant of the SVM, the SVM with uneven margins, which has a better generalization performance than the original SVM on imbalanced dataset where the positive examples are much less than the negative ones. Formally, given a training set $Z = ((x_1, y_1), \ldots, (x_m, y_m))$, where X_i is the n-dimensional input vector and $y_i (= +1 \text{ or } -1)$ its label, the SVM with uneven margins is obtained by following the steps in figure 7. In these equations, τ is the uneven margins parameter which is the ratio of the negative margin to the positive margin in the classifier and is equal to 1 in the original SVM. The goal of the SVM learning is to find the *optimal separating linear hyper-plane* that has the maximum margin linear classifier to both sides.

The SVM problem can be extended to non-linear case using non-linear hyper-plane. Non-linear separation by mapping input data to a high-dimensional space which called kernel function. The new mapping is then linearly separable. Example of kernel function is Polynomial function:

$$K(x, y) = (x^T y + 1)^d \qquad (1)$$

Input: D, the set of m training objects $Z = ((x_1, y_1), \ldots, (x_m, y_m))$, and let $y_i \in \{1, -1\}^m$ be the class label of x_i.

Process: solve the quadratic optimization problem

$$\text{minimise}_{w,\, b,\, \xi} \langle W, W \rangle + C \sum_{i=1}^{l} \xi_i \qquad (1)$$

$$\text{subject to} \quad \langle W, X_i \rangle + \xi_i + b \ge 1 \quad \text{if } y_i = +1 \qquad (2)$$

$$\langle W, X_i \rangle - \xi_i + b \le -\tau \quad \text{if } y_i = -1 \qquad (3)$$

$$\xi_i \ge 0 \quad \text{for } i = 1, \ldots, m \qquad (4)$$

Output: The decision boundary should classify all points correctly

$$y_i (W^T X_i + b) \ge 1, \quad \forall i$$

Fig. 7 Support vector machine algorithm with uneven margin.

5. Features for classification

Lexical and syntactic features of tokens and entity pairs that created prior to classification are used to build the classification model. These features are a part of those described in [41] and [42]. These features are split into 14 sets as described in table 2.

TokN features are contained surface string and POS of the tokens that surrounding the entity pairs. This features are provide us with important information about the words surrounding entity pairs to decide if there is relationship between them. GentokN features are generalised tokN which containing morphological root and generalised POS. Str features are contained surface string features include all token features of both entity pairs, their heads, combine of their heads, first, last and other tokens between them, two tokens before the leftmost and after the rightmost entity pairs. POS features are created from POS tags of the entity pairs and the tokens that surrounding them. Root features are created from morphological analyzer of the entity pairs and the tokens that surrounding them. GenPOS features are created from generalised POS tags of the entity pairs and the tokens that surrounding them. Entities were divided into two categories are events and non-events entities. Event entities are Investigation and Intervention entities. Non-event entities are Condition, Location, Drug-device, Result, Negation, Laterality, and Sub-location. Inter features are contained intervening entities which mean types and numbers of entities between entity pairs. Event features are contained whether an entity pairs contain two events, two non-events, or one event and one non-event and if there are any intervening events or non-events between entity pairs. Allgen features are collection of all above features in root and generalised POS forms. Notok features are collection of all above features except for TokN.

Stanford Parser [43] can be applied to parse the corpus to generate a dependency analysis which contains syntactic relations between sentence tokens for the dep and syndist features sets. When the entities exist in the same sentence the dep feature set can be generated from the parse. This feature set consists of the dependency analysis of entity pairs, their heads, and combine of their heads, first, last and other tokens between them, two tokens before the leftmost and after the rightmost entity pairs. For the syndist feature set contains the number of links on the dependency path between the entity pairs and the number of tokens between two entities [28].

6. Results and Discussion

Evaluation of the system can be done by using the standard evaluation metrics of Recall and Precision. The terms of true positive (TP), false positive (FP) and false negative (FN) are used to determine Recall and Precision which matches between relations recorded in a system annotated response document and a gold standard key document. If the relation in the response exists in the key with the same arguments then the response relation is a true positive. If the relation in the response dose not exists in the key then the response relation is a false positive. If the relation in the key dose not exists in the response then the key relation is a false negative.

$$R=\frac{TP}{TP+FN} \qquad P=\frac{TP}{TP+FP} \qquad F1=\frac{2PR}{P+R}$$

Table 2: Feature sets for learning

Feature set	***Description***
TokN	Surface string and POS of tokens surrounding the arguments, windowed -*N* to +*N*, *N* = 6 by default.
GentokN	Root and generalised POS of tokens surrounding the argument entities, windowed *N* to +*N*, *N* = 6 by default.
Atype	Concatenated semantic type of arguments, in arg1-arg2 order.
Dir	Direction: linear text order of the arguments (is arg1 before arg2, or vice versa?).
Str	Surface string features based on Zhou et al [29], see text for full description.
POS	POS features, as above.
Root	Root features, as above.
GenPOS	Generalised POS features, as above.
Inter	Intervening mentions: numbers and types of intervening entity mentions between arguments.
Event	Events: are any of the arguments, or intervening entities, events?
Allgen	All above features in root and generalised POS forms, i.e. gen-tok6+atype+dir+root+genpos+inter+event.
Notok	All above except tokN features, others in string and POS forms, i.e. atype+dir+str+pos+inter+event
Dep	Features based on a syntactic dependency path.
Syndist	The distance between the two arguments, along a token path and along a syntactic dependency path.

Standard ten-fold cross validation methodology is used to split the corpus for evaluation in our experiments. There are scores for each type of relations and for relation overall. P, R and F1 scores are computed for each relation type on each fold and macro-averaging these values for individual relations.

6.1 Algorithm type

Multi algorithms are implemented on the training corpus of patient narratives to see which one is much suitable for relation extraction. Many algorithms of supervised machine learning are applied such as Naïve Bayes Weka, C4.5 Weka, KNN Weka, Perceptron algorithm with uneven margin (PAUM), and Support victor machine with uneven margin (SVM), the results of these algorithms are described in table 3.

Table 3: Relation extraction by different algorithms

Relationship type	*Metric (%)*	*Naive Bayes Weka*	*C4.5Weka*	*KNN Weka*	*PAUM*	*SVM UM*
Has_finding	P	48.48	0.42	70.76	0.67.5	76.66
	R	72.11	31.16	53.11	62.04	55.85
	F1	52.88	29.27	52.95	58.93	57.32
Has_indication	P	59.85	58.67	59.77	61.71	67.17
	R	80.39	94.79	84.76	85.59	72.85
	F1	67.57	71.43	68.75	70.6	68.89
Has_location	P	67.04	66.77	69.02	71.32	76.68
	R	94.98	95.63	91.73	92.28	85.67
	F1	78.12	78.2	78.28	79.7	80.11
Has_target	P	54.27	50.97	54.04	63.29	68.92
	R	95.39	96.65	86.16	86.98	76.45
	F1	68.71	66.36	66.01	72.88	71.59
Laterality_modifies	P	43.4	43.4	54.07	41.73	60
	R	58.57	58.57	68.57	58.57	51.9
	F1	47.85	47.85	58.57	47.52	54.23
Negation_modifies	P	62.44	72.38	71.11	70	80
	R	74.16	74.16	77.5	80	71.66
	F1	65.73	71.84	72.72	73.75	74.66
Sub-location_modifies	P	77.6	90.22	92.22	90.22	100
	R	98	98	98	98	93
	F1	85.24	93.14	94.03	93.14	95.55
Overall	P	60.95	60.81	63.33	67.34	73.99
	R	90.51	92.18	87.63	88.59	79.67
	F1	72.48	73.01	73.27	76.14	76.3
Run Time in seconds		28.563	29.999	25.148	18.736	27.487

Firstly; Naïve Bayes Weka algorithm is implemented. Different algorithm C4.5 decision tree is applied; overall F1value increases by around 0.5% than the value of NaiveBayesWeka algorithm. KNN Weka algorithm is used with the option '–k 2' to get the best results, there is small increase in the value of overall F1 around by 0.26% than the value of C4.5Weka. Another algorithm PAUM is implemented with the best options "–p 20 –n 5 –optB 0.0". Overall F1 value of PAUM improves the performance than the overall F1value of KNN Weka by around 3%. Finally; SVM with uneven margin algorithm is executed with the options "-c 0.7 -t 1 -d 2 -m 100 -tau 0.8" to get the best results. This means that polynomial kernel is used with degree 2 for quadratic kernel and parameter of uneven margin (τ) is 0.8. There is small change in the overall F1 value of SVM algorithm than overall F1 value of PAUM algorithm around by 0.16%.

From the results in table 3; SVM algorithm with uneven margin is the much suitable machine learning algorithm for relation extraction from medical texts. Figure 8 shows the graph of applying different algorithms for relation extraction from medical texts.

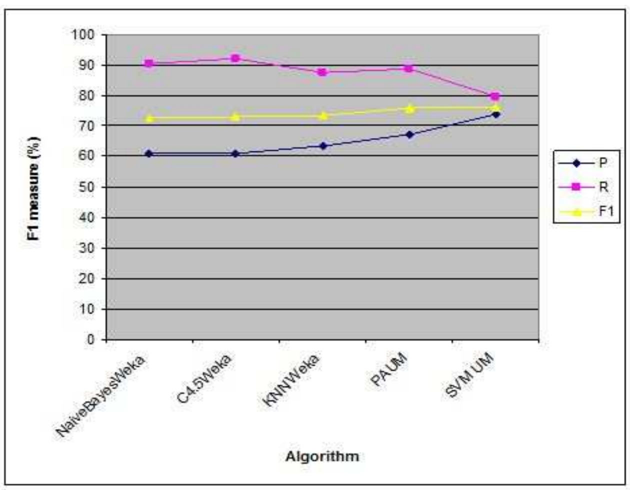


Fig. 8 Performance of different algorithms.

6.2 Run time

Different algorithms are implemented; the run time of each algorithm is the most important factor to known which one is much suitable with respect to time to run the application. The run times of each algorithm in seconds described in table 3. Each algorithm is applied on the same features which include the cumulative feature set +event which include different features are *TokN*, *Dir*, *Str*, *POS*, *Inter*, and *Event*.

The C4.5 weka algorithm spends more time to classify the data and extract the relation than other algorithms and the accuracy of the overall F1 measures not perfect very well comparing to other algorithms. The naive bayes weka algorithm needs small time compared to C4.5 weka but the overall F1 value is small than the overall F1 value of C4.5 weka. SVM algorithm is less in time than C4.5 and naïve bayes weka and the F1 measures is greater than these algorithms. This means that SVM is better than C4.5 and Naïve Bayes weka. KNN weka algorithm requires small run time comparing to C4.5, Naïve Bayes, and SVM algorithms. But the accuracy of overall F1 measure is less than SVM and better than C4.5 and Naïve Bayes. The PAUM algorithm considers the faster algorithm than other algorithms and their accuracy F1 measure is better than other algorithms except for SVM algorithm. PAUM algorithm is a faster on small data set than SVM algorithm and there is small difference on the accuracy of the overall F1 measures in between. Figure 9 shows the graph of the run time of each algorithm.

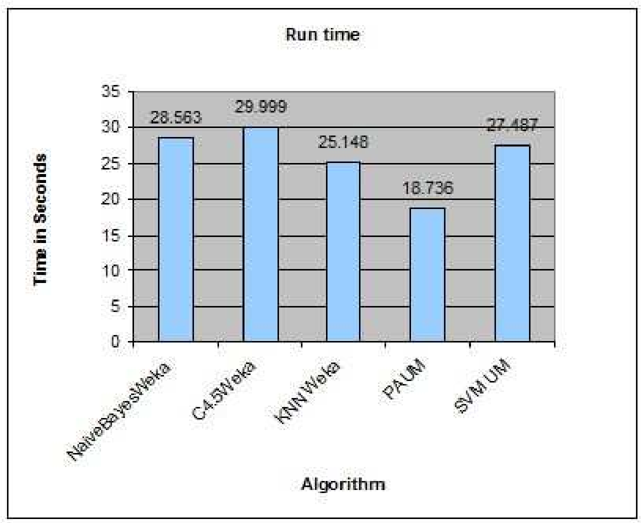


Fig. 9 Run times of different algorithms.

6.3 Uneven margin parameter

SVM with uneven margin is the better algorithm in the accuracy than other algorithm but not the faster one. The SVM algorithm is implemented with different uneven parameters to obtain the value of uneven margin (τ) that improves the performance of the system. Then SVM with this uneven margin value is applied with different features to see the effect of adding new feature to the model and also use different corpus size to known their effects on the performance of the system. Table 4 shows SVM with different uneven parameter values. The standard SVM use the uneven margin value 1, this gives bad results than SVM with uneven margin. When the value of uneven margin parameter decreases the results is improved. Figure 10 describes the graph of using different uneven margin parameters.

When the uneven margin value τ = 0.8 is applied, the performance is improved than τ =1 by around 2.6%. The value of τ is decreased than this level τ =0.8 the value of overall F1 measure decreased. Using τ = 0.6, there is small drop on the value of F1 by 0.1%. SVM with τ = 0.4 effects on the performance, this leads to a drop on the value of overall F1 than τ = 0.6 by around 0.36%. When the value of τ is changed to τ = 0.2, there is drop on the value of overall F1 than τ =0.4 by around 1.46%. This mean that while SVM is implemented with increasing the value of uneven margin τ, the performance of the system is improved until it is reached to the point that the performance is decreased with increasing the value of τ.

Table 4: SVM use different uneven parameter values

	Uneven margin (τ)					
	Metric (%)	*1.0*	*0.8*	*0.6*	*0.4*	*0.2*
Overall Relations	P	76.04	73.99	69.58	65.8	61.54
	R	72.16	79.67	85.29	90.78	94.91
	F1	73.7	76.3	76.2	75.84	74.36

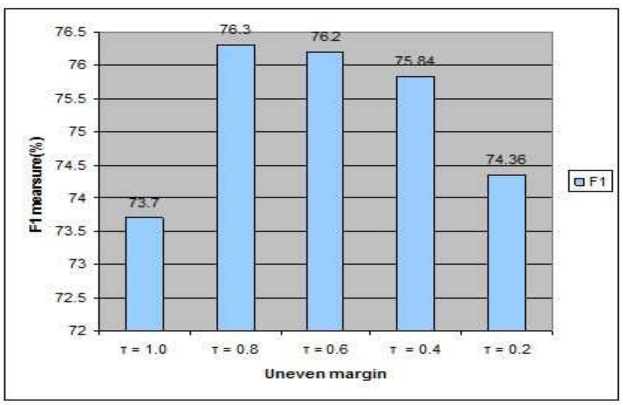


Fig. 10 Performance by different uneven margins.

6.4 Feature selection

The experiments searches for the performance of relation extraction with various feature sets, using the feature sets described in table 2. An additive strategy is used to select the feature. The experiments are divided into two cases, one case of feature sets that do not use syntactic parse information and the other case of feature sets that use syntactic parse information.

6.4.1 Non-syntactic features

Firstly, the experiments used the feature sets that do not use syntactic parse information for relation extraction. Starting with the basic features and then adding new feature set each time to measure the performance of the system. The results are described in table 5.

Starting with Tok6 and Atype features sets, the overall F1 value is 68.29%. Addition of Dir features leads to improve the performance in most metrics, there is improved in the overall F1 value by around 1.35%. Addition of Str features improves the performance in most metrics, there is improved in the overall F1 value by around 0.5%. Addition of the POS features leads to drop the performance in some metrics, overall F1 value drop by around 0.66%. Addition of the Inter features gives more improvements in all metrics, overall F1value increases by around 6.63%. Addition of the Event features gives more improvements in some metrics, overall F1 value increases by around 0.19%.

Generalizing features are used to see their effects on the performance of relation extraction. All Str features, POS features, and TokN features are replaced with their root features, generalized POS features, and generalized TokN features respectively. These results shown in the column Allgen, there is no change in overall F1 value. Notok features are implemented to see if it improves the performance. In this feature TokN features are removed from the full cumulative feature set, corresponding to column +event of table 5. These results are shown in the column Notok, this leads to drop the performance in some metrics, the overall F1value drop by around 0.71%. The graph of using non-syntactic feature sets is shown in figure 11.

6.4.2 Syntactic features

The second part of the feature selection experiments is using features that used syntactic parse information that derived from dependency parse analysis of the texts by using the Stanford parser [43]. The results of +event column in table 5 which corresponding to collection of all non-syntactic feature sets is copied to add in table 6 and then add the Dep features and Syndist features. Addition of the Dep features leads to drop the results that unclear. Addition of the Syndist features leads to a small drop in overall F1 that is unclear. Figure 12 shows the graph of the performance of syntactic feature sets. Addition of the Dep features leads to a drop the performance in some metrics, the overall F1value dropping by around 0.37%. Addition of the Syndist feature set leads to a drop the performance in some metrics, the overall F1 value dropping by 0.38%.

Table 5: Performance by non-syntactic feature sets

Relationship type	*Metric (%)*	*Tok6+ Atype*	*+Dir*	*+Str*	*+POS*	*+Inter*	*+Event*	*Allgen*	*NoTok*
Has_finding	P	5.25	55.16	66.83	45.83	61.66	76.66	76.66	74.76
	R	43.21	48.54	45.69	29.76	42.85	55.85	55.85	58.35
	F1	40.1	44.89	47.06	28.35	43.26	57.32	57.32	58.39
Has_indication	P	64.24	62.34	63.22	62.49	64.56	67.17	67.17	66.04
	R	68.48	69.5	70.16	71.1	69.86	72.85	72.85	71.43
	F1	65.26	64.76	65.37	65.47	66.16	68.89	68.89	67.52
Has_location	P	65.4	65.59	64.98	64.08	78.09	76.68	76.68	76.39
	R	85.5	85.24	87.54	90.3	85.68	85.67	85.67	85.67
	F1	73.38	73.39	73.91	74.4	80.82	80.11	80.11	79.94
Has_target	P	57.04	57.47	58.64	58.18	69.1	68.92	68.92	66.7
	R	67.52	76	76.96	73.68	79.9	76.45	76.45	75.74
	F1	60.08	64.31	65.52	63.77	73.46	71.59	71.59	69.99
Laterality_modifies	P	37.5	45.23	48.16	45	60	60	60	60
	R	37.14	53.57	57.14	47.14	58.57	51.9	51.9	51.9
	F1	36.9	48.47	50.61	45.23	59.23	54.23	54.23	54.23
Negation_modifies	P	70.71	70.71	70.71	70.71	75.71	80	80	80
	R	76.66	76.66	70.83	70.83	71.66	71.66	71.66	71.66
	F1	72.36	72.36	68.93	68.93	71.93	74.66	74.66	74.66
Sub-location_modifies	P	76.54	79.6	79.6	79.6	98.33	100	1.0	1.0
	R	85	93	93	93	93	93	93	93
	F1	78.1	83.02	83.02	83.02	94.64	95.55	95.55	95.55
Overall	P	63.45	63.2	63.45	62.81	73.85	73.99	73.99	73.05
	R	75.39	78.69	79.55	78.97	79.33	79.67	79.67	79.3
	F1	68.29	69.64	70.14	69.48	76.11	76.3	76.3	75.59

6.5 Size of training corpus

Changing the size of training corpus in the experiments is used to examine their effects on relationship extraction. Two subsets with size 20 and 30 documents is selected from 40 documents; referred to them as C20 and C30, respectively.

The collection feature set of *all* non-syntactic feature sets which represent in +event feature set is used in the experiments to show the effects of training corpus size on the performance, these results are shown in table 7.

Firstly, start the experiments with corpus size 20 documents. Increasing the corpus size to 30 documents leads to improve the performance in most metrics; overall F1 value improves by around 2.11%. Using corpus size 40 documents leads to improve the performance in most metrics; overall F1 value improves by around 2.09%. Increasing the size of the training corpus leads to improve the performance of relation extraction system. Figure 13 shows the effects of changing corpus size in the performance.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

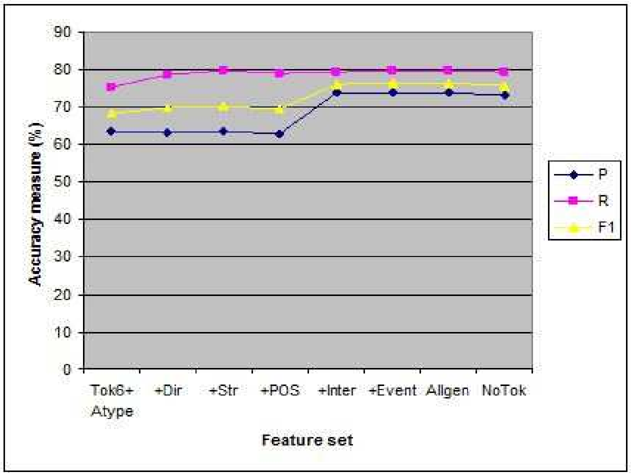


Fig. 11 Graph of non-syntactic feature sets performance.

Table 6: Performance by Syntactic Feature Sets

Relationship type	***Metric (%)***	***+Event***	***+Dep***	***+Syndist***
Has_finding	P	76.66	66.66	45
	R	55.85	44.85	17.02
	F1	57.32	46.49	24.44
Has_indication	P	67.17	66.38	65.37
	R	72.85	71.31	68.53
	F1	68.89	67.77	65.72
Has_location	P	76.68	77.15	79.16
	R	85.67	85.81	85.15
	F1	80.11	80.52	81.12
Has_target	P	68.92	69.23	71.56
	R	76.45	76.84	76.31
	F1	71.59	72.04	73.14
Laterality_modifies	P	60	60	50
	R	51.9	46.9	36.90
	F1	54.23	50.89	40.89
Negation_modifies	P	80	80	80
	R	71.66	71.66	71.66
	F1	74.66	74.66	74.66
Sub-location_modifies	P	100	100	100
	R	93	93	93
	F1	95.55	95.55	95.55
Overall	P	73.99	74	74.93
	R	79.67	78.75	76.98
	F1	76.3	75.93	75.55

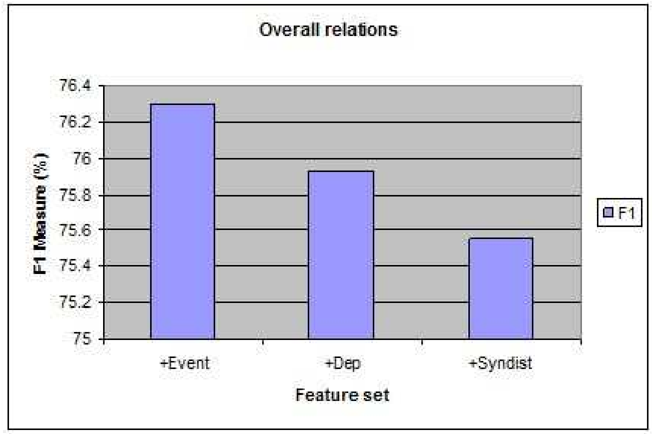


Fig. 12 Graph of syntactic feature sets performance.

7. Conclusion

From the results, the clinical relationships can be extracted from medical text using different supervised machine learning algorithm. SVM with uneven margin is much suitable algorithm which achieves high accuracy, but it takes more time in the run than Perceptron with uneven margin. Perceptron with uneven margin is very fast algorithm than others as well as the accuracy is relatively near to SVM, there is small change in between. SVM with uneven margin is implemented to show the effects of changing the values of *uneven margin* (τ) parameter, adding the feature sets, and changing the size of the training corpus for relationship extraction. Increasing the value of τ leads to improve the performance to reach the value that has high performance where $\tau = 0.8$ after that point the performance dropped. Adding new feature sets like *non-syntactic features* improves the performance. Adding the *syntactic features* leads to small drop in the performance that unclear. Changing the size of training corpus leads to improve the performance. Our future work on relationship extraction in CLEF includes the integration of a noun and a verb chunk tagger into the feature sets.

References

[1] Hahn U& Wermter J, Levels of Natural Language Processing for Text Mining. In: Ananiadou, S. & McNaught, J., ed. Text Mining for Biology and Biomedicine. Artech House: London, 2006.

[2] JISC, Text Mining: Briefing Paper. [Online] Available at: http://www.jisc.ac.uk/media/documents/publications/textminingbp.pdf [Accessed 16 April 2010], 2006.

[3] Rector A, Rogers J, Taweel A, Ingram D, Kalra D, Milan J, Singleton P, Gaizauskas R, Hepple M, Scott D, Power R: CLEF – joining up healthcare with clinical and post-genomic research. Proceedings of UK e-Science All Hands Meeting 2003, Nottingham, UK, 2003:264-267.

Table 7: Performance by corpus size

Relationship type	***Corpus size***			
	Metric (%)	***C20***	***C30***	***C40***
Has_finding	Count	16	23	36
	P	46.66	10	76.66
	R	21.66	3.33	55.85
	F1	29.04	5	57.32
Has_indication	Count	77	125	180
	P	58.69	65.39	67.17
	R	60.22	69.41	72.85
	F1	58.44	65.79	68.89
Has_location	Count	149	239	363
	P	75.05	79.79	76.68
	R	82.4	85.71	85.67
	F1	77.75	82.06	80.11
Has_target	Count	98	145	180
	P	69.4	69.14	68.92
	R	79.52	74.31	76.45
	F1	72.8	70.79	71.59
Laterality_modifies	Count	6	9	15
	P	20	40	0.6
	R	20	30	51.9
	F1	20	33.33	54.23
Negation_modifies	Count	9	11	20
	P	40	0.5	0.8
	R	40	0.45	71.66
	F1	40	46.66	74.66
Sub-location_modifies	Count	11	18	23
	P	60	80	100
	R	60	77.5	93
	F1	60	78.57	95.55
Overall	Count	366	570	817
	P	71.1	73.36	73.99
	R	74.2	75.46	79.67
	F1	72.1	74.21	76.3

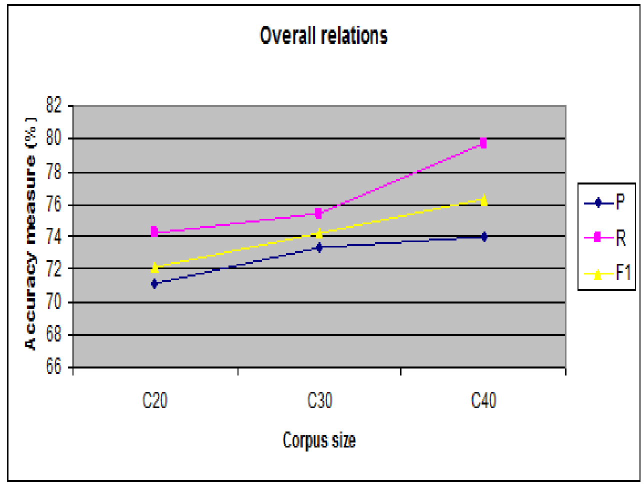


Fig.13. Graph of corpus size performance.

[4] Mitchell T, Machine Learning, McGraw-Hill, 1997.

[5] Roberts A, Gaizauskas R, Hepple M, Guo Y: Combining terminology resources and statistical methods for entity recognition: an evaluation. Proceedings of the Sixth International Conference on Language Resources and Evaluation, LREC 2008, Marrakech, Morocco, 2008.

[6] Turmo J., Ageno A., and Catal`a, N. Adaptive information extraction. ACM Computing Surveys, 2006:38(2).

[7] MUC-5. Proceedings of the 5th Message Understanding Conference, Baltimore, MD, USA. Morgan Kaufmann, San Mateo, CA, 1993.

[8] MUC-6. Proceedings of the 6th Message Understanding Conference, Columbia, MD, USA. Morgan Kaufmann, San Mateo, CA, 1995.

[9] Defense Advanced Research Projects Agency: Proceedings of the Seventh Message Understanding Conference (MUC-7) Fairfax, VA, USA: National Institute of Standards and Technology, 1998.

[10] Chinchor N: Overview of MUC-7/MET-2. Proceedings of the Seventh Message Understanding Conference (MUC-7), Fairfax, VA, USA, 1998.

[11] Doddington G, Mitchell A, Przybocki M, Ramshaw L, Strassel S, Weischedel R: The Automatic Content Extraction (ACE) Program – Tasks, Data, & Evaluation. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal,* 2004:837-840.

[12] Nédellec C: Learning language in logic – genic interaction extraction challenge. *Proceedings of the ICML05 workshop: Learning Language in Logic (LLL05), Bonn, Germany,* 2005:31-37.

[13] BioCreAtIvE II–Protein-Protein Interaction Task [http://biocreative.sourceforge.net/biocreative_2_ppi.html]

[14] Sager N, Lyman M, Bucknall C, Nhan N, Tick L: Natural language processing and the representation of clinical data. *J Am Med Inform Assoc,* 1994:1(2):142-160.

[15] Friedman C, Alderson P, Austin J, Cimino J, Johnson S: A General Natural-language Text Processor for Clinical Radiology. *J Am Med Inform Assoc,* 1994:1(2):161-174.

[16] Lussier Y, Borlawsky T, Rappaport D, Liu Y, Friedman C: PhenoGO: assigning phenotypic context to gene ontology annotations with natural language processing. *Pac Symp Biocomput,* 2006:64-75.

[17] Hahn U, Romacker M, Schulz S: medsynDikate – a natural language system for the extraction of medical information from findings reports. *Int J Med Inform,* 2002:67(1–3):63-74.

[18] Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux JF: A multi-lingual architecture for building a normalised conceptual representation from medical language. *Proc Annu Symp Comput Appl Med Care,* 1995:357-361.

[19] Grover C, Haddow B, Klein E, Matthews M, Nielsen L, Tobin R, Wang X: Adapting a Relation Extraction Pipeline for the BioCreAtIvE II Task. *Proceedings of the BioCreAtIvE II Workshop Madrid, Spain,* 2007.

[20] Bundschus M, Dejori M, Stetter M, Tresp V, Kriegel HP: Extraction of semantic biomedical relations from text using conditional random fields. *BMC Bioinformatics,* 2008:9(207).

[21] Bunescu R, Ge R, Kate R, Marcotte E, Mooney R, Ramani A, Wong Y: Comparative experiments on learning information extractors for proteins and their interactions. *Artif Intell Med,* 2005:33:139-155.

[22] Goadrich M, Oliphant L, Shavlik J: Learning to Extract Genic Interactions Using Gleaner. *Proceedings of the ICML05 workshop: Learning Language in Logic (LLL05), Bonn, Germany,* 2005:62-68.

[23] Giuliano C, Lavelli A, Romano L: Exploiting Shallow Linguistic Information for Relation Extraction from Biomedical Literature. *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL06), Trento, Italy,* 2006:401-408.

[24] Zelenko D, Aone C, Richardella A: Kernel Methods for Relation Extraction. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, PA, USA,* 2002:71-78.

[25] Zhou G, Su J, Zhang J, Zhang M: Exploring Various Knowledge in Relation Extraction. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor, MI, USA,* 2005:427-434.

[26] Bunescu RC, Mooney RJ: A Shortest Path Dependency Kernel for Relation Extraction. *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing,* 2005:724-731.

[27] Roberts A, Gaizauskas R, Hepple M: Extracting Clinical Relationships from Patient Narratives. In Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing Columbus, OH, USA: Association for Computational Linguistics, 2008:10-18.

[28] Roberts A, Gaizauskas R, Hepple M, Guo Y: Mining Clinical Relationships from Patient Narratives. In *Proceedings of the BioNLP 08 ACL Workshop: Themes in biomedical language processing*, 2008: 9(Suppl 11):S3.

[29] Lindberg D, Humphreys B, McCray A: The Unified Medical Language System. Methods Inf Med, 1993: 32(4):281-291.

[30] Roberts A, Gaizauskas R, Hepple M, Demetriou G, Guo Y, Setzer A, Roberts I: Semantic Annotation of Clinical Text: The CLEF Corpus. Proceedings of Building and evaluating resources for biomedical text mining: workshop at LREC 2008, Marrakech, Morocco, 2008.

[31] The University of Chicago medicine, URL http://www.uchospitals.edu/index.shtml.

[32] Mayo Clinic hospital, URL http://www.mayoclinic.com.

[33] Cunningham H, Maynard D, Bontcheva K, Tablan V: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 2002: 168-175.

[34] URL http://nlp.shef.ac.uk/clef/TheGuidelines/TheGuidelines.html.

[35] Rindflesch T, Fiszman M: The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform, 2003:36(6):462-477.

[36] K. Ming Leung: Naive Bayesian Classifier. POLYTECHNIC UNIVERSITY Department of Computer Science / Finance and Risk Engineering, 2007.

[37] Quinlan J. R., C4.5: Programs for Machine Learning, Morgan Kaufmann, Los Altos, 1993.

[38] Xindong Wu, et al. "Top 10 algorithms in data mining," Knowl Inf Syst, 2008:14:1–37.

[39] Li Y, Zaragoza H, Herbrich R, Shawe-Taylor J, and Kandola J. The Perceptron Algorithm with Uneven Margins. In *Proceedings of the 9th International Conference on Machine Learning, (ICML-2002)*:pages 379–386.

[40] Li Y, Bontcheva K, Cunningham H: SVM Based Learning System for Information Extraction. In Deterministic and statistical methods in machine learning: first international workshop no. 3635 in Lecture Notes in Computer Science, Springer, 2005:319-339.

[41] Zhou G, Su J, Zhang J, Zhang M: Exploring Various Knowledge in Relation Extraction. Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor, MI, USA, 2005:427-434.

[42] Wang T, Li Y, Bontcheva K, Cunningham H, Wang J: Automatic Extraction of Hierarchical Relations from Text. In The Semantic Web: Research and Applications. 3rd European Semantic Web Conference, ESWC 2006 no. 4011 in Lecture Notes in Computer Science, Springer, 2006:215-229.

[43] Klein D, Manning CD: Accurate unlexicalized parsing. In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics Morristown, NJ, USA: Association for Computational Linguistics, 2003:423-430.

Wafaa Tawfik Abdel-moneim is a teaching assistant in information system department in faculty of computers and informatics in zagazig university.

Mohamed Hashem Abdel-Aziz is a professor in information system department in faculty of computers and informatics in ain-shames university.

Mohamed Monier Hassan a professor assistant in information system department in faculty of computers and informatics in zagazig university.

E-Government Interoperability Framework based on a Real Time Architecture

Aris Puji Widodo[1], Jazi Eko Istiyanto[2], Retantyo Wardoyo[3], and Purwo Santoso[4]

**[1, 2, 3] Department of Computer Science and Electronics, Gadjah Mada University
Yogyakarta, Indonesia, 55287**

**[4] Department of Government Faculty Social and Political Sciences, Gadjah Mada University
Yogyakarta, Indonesia, 55287**

Abstract

One of E-Government (E-Gov) implementations is to produce a form of relationship namely Government to Government (G2G). The main factor of G2G concerning interoperability between central and local governments is the heterogeneity in developing E-Gov applications. The interoperability that has occurred are done by creating a brand new architecture that has a function to allow existing E-Gov applications to be able to communicate with each other. E-Govs in Indonesia are developed separately by the respective governments, i.e. central (national) and local (province, district/city). Therefore, interoperability and data integration have such a restriction.

This paper focused on creating an Enterprise Architecture Framework (EAF) for interoperability and data integration on E-Gov applications in Indonesia. This EAF combines the approach of Service Oriented Architecture (SOA) and Event Driven Architecture (EDA). The purpose of this combination is to produce EAF that is real-time, the relationship between the one-to-many services, and the message transmission models that are asynchronous. The services provided and the events that are used to trigger are defined by using Web Service Definition Language (WSDL), while the orchestration mechanism that occurs between services is defined by using the Business Process Execution Language (BPEL). BPEL is conducted by the Government Service Bus (GSB). To evaluate this EAF, the model of services which has been defined using a variety of data is measured based on execusion time.

Keywords: *Enterprise Architecture, E-Gov, SOA, EDA, Real Time Architecture.*

1. Introduction

E-Gov is used to improve the relationship between the central and the regional governments, between regions and other government parties, especially the citizens [1]. Implementation of e-Gov can produce a form of relationship such as: Government to Citizen (G2C), Government to Business (G2B) and Government to Government (G2G) [2].

As observed in Indonesia, the implementation of E-Gov in the level of preparation (web site design) is considered very good, meaning that almost every local government, central government and government agencies have a web site and the information is updated regularly [3]. The web site that is designed by the local government, central government or agency became a way to form a G2C relationship in terms of sharing information more openly to the citizens. In transaction and transformation level, the focus is still in separated development using a wide variety of development platforms [3][4]. The interoperability process that occurs between local government, central government and government institutions is still done by exchanging physical data storage media, such as floppy disks, compact disks and other storage media [5]. This condition means that the shape of the relationship of G2G on e-Gov in Indonesia has not worked well, and do not meet the spirit of the policy in INPRES No. 3 Year 2003 and the blue print of E-Gov applications systems [6] [7].

One of the main factors regarding the form of G2G relationships is the interoperability between local and central government, so the exchange, collecting and integration of data between local and central government can be achieved [8]. One of issues and challenges in conducting interoperability is the heterogeneity of the e-Gov applications that currently exist. The heterogeneity of existing e-Gov applications includes technology, architecture and platform which are used by each local government based on capability and knowledge. In order to create the interoperability of heterogeneous systems, the approach is to build an interoperability framework architecture that can support the existing previous architectures [9]. The approach of interoperability framework architecture aims to interconnect the applications of e-Gov that are separated and isolated so they are able to communicate with each other [10].

Interoperability framework architecture is developed with protocol-based approach and Extensible Markup Language (XML) scheme, which are distributed by a public network [11][12]. By using this protocol based framework architecture and XML schema, the interoperability process that occurs can only be done by exchanging XML scheme that is an open standard and does not provide services. To add services to the architecture of XML that are based on interoperability framework, an approach was used based on Service Oriented Architecture (SOA) using the web services technology [10][13][14][15][16][17][18]. The services provided in a SOA architecture approach are non-real-time to the business processes since the services provided are only based on request and response information, and are synchronous messaging [19].
In order to provide services that function in real time to the business process, this paper is focused on creating architectural framework for interoperability and data integration using a combination of SOA and EDA-based approach. This SOA approach that was used is based on the concept of making the communication architecture. This architecture can perform a wide range of heterogeneous platforms of existing e-Gov applications. The communication is represented among open standard services. The EDA approach that was used is based on the concept for provision of dynamic services. The combination of SOA and EDA approach aims to provide services that can be real time for business processes, because the services are available in a combination with the event as a trigger for asynchronous message sending.

Concerning the condition of e-Gov in Indonesian government which has heterogeneity ranging from technology, architecture and platform development, the architecture can be used as standard development of e-Gov in Indonesia in order to achieve interoperability between central and local governments, so later on it can lead to data integration as a whole.

2. SOA

SOA is a form of an architectural model that refers to the principle of technology-oriented service [20]. Figure 1 describes the scope of the service in SOA which consists of functions, procedures or a process that gives response if requested by a client or it can also be viewed as a logical encapsulation from an individual or a group of specific activities as the implementation of business processes. The concept of service-oriented allows to be approached by dividing the problem into a set of small service, and the solutions to these problems must be solved by allowing all to participate in a service orchestration. SOA is not associated with a particular technology, but more toward a modular approach.
To implement application with service-oriented approach to SOA, the implementation is in a layer that lies between business process layer and application layer. This layer is a part of the enterprise logic called the service interface layer, as shown in Figure 2. Service interface layer has a function to perform logical encapsulation in application logic, as well as the business processes that exist in business logic. Thus, the application can be more modular and be able to use a variety of technology platforms. Interface service layer is divided into several layers of abstraction: application service layer, business service layer, and the orchestration service layer [20].

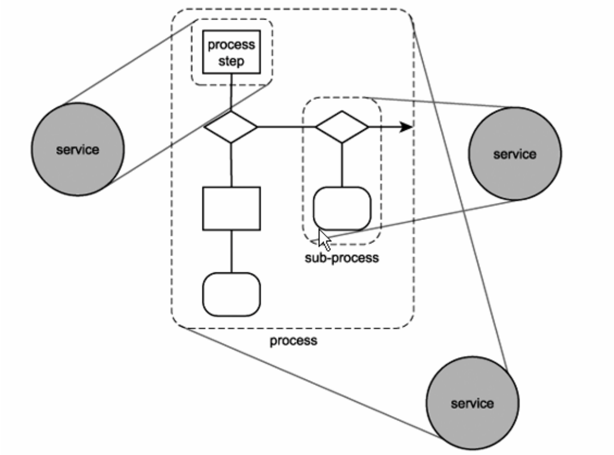


Fig. 1 Encapsulation Business Process with Services [20]

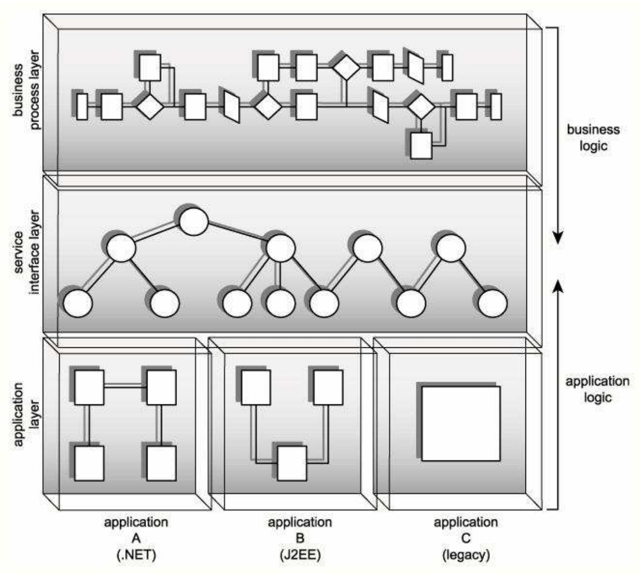


Fig. 2 The layers in Implementation on Enterprise [20]

Conceptually, SOA approach is based on the architecture depicting interaction among the main parts. These parts include Service Provider, Consumer Service, Service Registry, and Service Broker [21]. SOA uses find-bind-execute paradigm, where Service Providers register their service in the public registry. Then the registry is used by users to find services that match the desired criteria. If in the registry there is service that meets the needs, then the user will be given a contract and the addresses of the service [21].

3. EDA

An event is an occurance of something inside or outside the business process. It can be used to identify something that is happening or anything that will occur. Each occurrence of an event, this event has a header and a body [22]. Event header contains elements that describe the occurrence of such event, such as: identity, type, name, time-stamp, number, and event makers. The body contains the event that has occurred. In the concept of EDA approach, for each event that occurs inside or outside a business process, that event is immediately distributed to all parties having interest on it; then the parties will conduct the evaluation, and optionally take action in response to the event. EDA is extreme loosely coupled, and highly distributed. It means the event developer will know about the occurrence of the event, but does not have the knowledge about which parties will process the event and the people who has the interest [22]. In order to ease the search of an event, EDA will use an asynchronous flow model, where the flow of events begins at the time it is created and it ends at downstream execution of the action [22].

4. SOA and Web Services

IT architecture that was developed with heterogeneous platform has a challenge during application and data integration. The interoperability between the applications must ensure that the integration of business processes is more effective without affecting the availability of applications that exist previously. The distributed technology that already exists, such as CORBA, COM/DCOM, EJB, JAVA RMI and others can be used to implement interoperability between applications, but should be in the same development platform for each application [13]. In order to perform interoperability on heterogeneous platforms, SOA approach which is based on web services technology can be used. Web services technology has open standards, so it allows applications that are implemented by Web services platforms/different vendors to be able to communicate with each other [23]. Basically a service in the SOA is an application. This application represents the business logic (automation logic) of a large system that encapsulates the process, which must be able to stand alone and communicate with each other [23]. Figure 3 describes the communication between one of the service with other services that are defined using the Web Service Description Language (WSDL). WSDL describes the format of a message to be sent by web application services, so it can be understood by other web application services that receive the technology using Simple Object Access Protocol (SOAP) running on the Hyper Text Transfer Protocol (HTTP) [20].

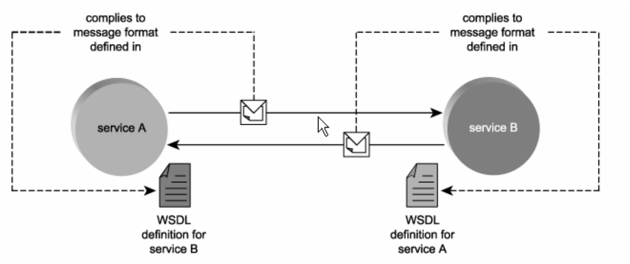


Fig. 3 Services Communication [20]

Web services are technology that is used to exchange information based on XML. It allows interoperability between multiple applications and platforms that are developed with different programming languages. SOA approach enables applications to communicate with other applications by requesting relevant service of respected application regardless of the platform and the programming language. Applications can adapt to the functions and services to suit the different processes in a flexible business processes [13].

5. Business Process Execution Language (BPEL)

BPEL is an XML-based language used to define business processes with web services. The main goal of BPEL is to standardize the business process flow that is defined to work with using the web services [24]. BPEL extends the interaction with web services model and enables to support the transaction. BPEL is based on web services, so it is assumed that each of the business processes involved is implemented by using web services, so that BPEL can regulate the interaction between web services using XML documents [25]. BPEL is used to describe a business process in two different ways, i.e. executable and abstract processes. The executable processes specification is to define the order of execution in a number of activities; the partners that are involved, exchanging messages between partners, and exception handling mechanisms. The abstract process specification is the behavior of exchanging messages with different parties without providing information about the internal behavior from the parties. The elements contained in the general structure of BPEL are composed of tag: Process, PartnerLinks, Variables, and Sequence [26]. Process is a main key element of the BPEL. The process name is defined on the tag having attribute name. In addition, the tag is also used to include information related to the process definition. PartnerLinks is an element to define the type of port from another service that participates in the execution of business processes. Variables are elements used to store the status that is used for workflow logic. Sequence is an element to organize a set of activities that can be executed consecutively. The elements that BPEL supports for sequence include receive, assign, invoke and reply.

6. Interoperability Architecture Conceptual

In this section is given a conceptual model of framework interoperability of E-Gov in Indonesia which can be used for interoperability of local governments to the central government or in the other way around. The case study used is the application of Information Systems and Population Administration (ISPA) developed by the Directorate General of Population Administration Department of Home Affairs [5].

6.1 Existing E-Gov Interoperability

ISPA applications that run on the sub-district, district/city and province consist of two versions, namely online ISPA and offline ISPA [5]. Among offline ISPA versions that are installed at each district/city (an application that has function for recording data of population registration and data of civil registration and for printing the population data), and province (an application that has a function to recap on the report of demographic data), each of them has its own applications and database. In offline ISPA, the data interoperability process is done by exchanging physical data storage media, such as floppy disks, CDs, and other storage media. Conversely, online ISPA is developed by a centralized architecture application, so the data and the application are located in one place, namely the Directorate General of Population Administration Department of Home Affairs. ISPA is planned to be accessed online at the district level (the application that has a function for recording data of population registration and data of civil registration and for printing the population data) that will be directly connected to the Directorate General of Population Administration Department of Home Affairs. The communications infrastructure of online ISPA will use VPN Dial with synchronous connections to the data center in the Directorate General of Population Administration Department of Home Affairs. The interoperability process of online ISPA is done directly because it was developed with centralized architecture [5].

6.2 Developed e-Gov Interoperability Architecture

Offline and online ISPA have limitations in terms of interoperability in both the application and the data. Online ISPA who are with Dial VPN connection models should be able to provide assurance that the connection to the data center in the Directorate General of Population Administration Department of Home Affairs will not break up, because when the connection is lost, ISPA does not operate. The number of districts is 5263. If the districts access at the same time, it will require a considerable amount of bandwidth. The IT infrastructure in the district level has not been perfect throughout Indonesia. Offline and online ISPA are developed by using client-server architecture.

Based on ISPA that is currently in operation, there should be an improvement pertaining to the architecture for better interoperability for applications and data. Interoperability architecture was developed using SOA approach to create services for business process, so interoperability can be performed by accessing these services. To create interoperability architecture that operates in real time, SOA approach needs to be combined with the EDA approach. The aim for this combination is to enable the service works with the event as a trigger, where the events in the services are defined using WSDL. To set the service orchestration which has occurred, a bus service called the Government Service Bus (GSB) is defined using BPEL. The development of interoperability architecture consists of the components shown in Figure 4:

1. Client (Services Consumer)
   This component is used to access the services provided by the service provider.
2. Services Provider
   This component is used to define business processes in a service using WSDL.
3. Event
   This component is used to trigger the services that are defined using WSDL.
4. BPEL
   This component is used to set the trigger based on orchestration between the services provided by the event.
5. GSB
   This component is used to perform detection, to trigger, and to distribute event (event services). Besides, this component also provides assurance that different protocols of heterogeneous platforms are

able to communicate. It is also responsible for transforming the message (to replace or to add a message) throughout the participating services (mediation service).

Interoperability scenarios are developed using an approach where each district/city, province, and national levels are assumed as elements involved in the mechanism of interoperability. Each element can act as a service provider as well as a service consumer. Architecture interoperability scenarios were developed using the model given in Figure 4 as follows:

1. National and province governments as a service provider for the functions of adding, deleting, and updating that will be used by the consumer service in districts/cities. This scenario is used to create a distributed database contained in each district/city, province, and national levels.
2. District/city as a service provider for the function of read data (data on population, birth data, mortality data, recap of the population, recap of the number of births, and recap the number of deaths) that will be used by the service consumer, i.e. province and national level in accordance with the hierarchical structure.
3. Each province and nationally has GSB that is equipped with BPEL to set orchestration between the services that have occurred. GSB is also defined as an event (the read data) that is used to trigger between the services performed by the BPEL orchestration. This scenario is based on the conditions in Indonesia that has 33 provinces, and each province has some districts/cities, thus it requires the orchestration mechanism for interoperability between services owned by each of the elements involved.

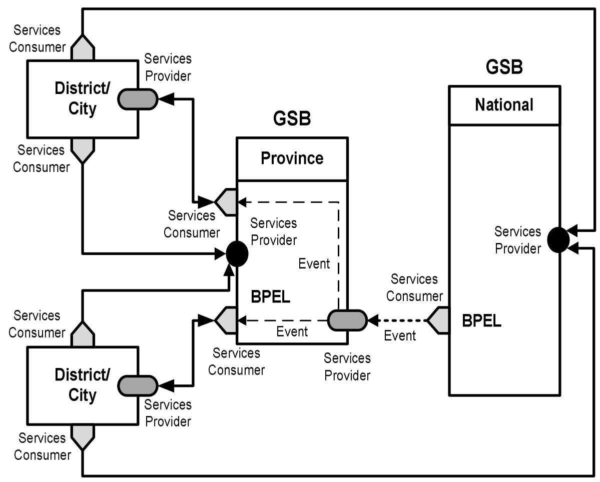


Fig. 4. Conceptual Interoperability Architecture

7. Implementation of Interoperability Architecture

To measure the validity of the conceptual interoperability architecture model, the conceptual model was then implementated as a simulation. To perform the simulation, there were used a variety of development platforms at the district/city, province, and national levels as given in Table 1.

Table 1: Development Platforms

Element	Programming Language	Database System
District/City	Php	MySQL
Province	Java	Postgre SQL
National	Java	SQL Server

The services that are used to define the base of consumer services and service providers will be placed on each of the elements, as given in Table 2.

Table 2 : Services Definition

Element	Services Consumer	Services Provider
District/City	Function add, delete, update	Read functions which includes: population data, birth data, mortality data, recap of the population, recap of the number of births, and recap the number of deaths
Province	Read functions which includes: population data, birth data, mortality data, recap of the population, recap of the number of births, and recap the number of deaths	Function add, delete, update
National	Read functions which includes: population data, birth data, mortality data, recap of the population, recap of the number of births, and recap the number of deaths	Function add, delete, update

The services used to add, delete, and update are placed in the province and national use one-to-one scenario without using BPEL, but using WSDL to be directly accessed by the district/city. This scenario aims to create a distributed database for each province and national level through a mechanism that is done together. That is when the district/city adds the population to the data, then the province and national levels will also be added to the population data. Then, the read services are placed at the district/city and province using one-to-many scenario where the event acts as a trigger while BPEL sets orchestration with other services as a partner. To be able to detect events and orchestration between services, it is necessary to use GSB. GSB which is used is GlassFish 3.1.2 server as an open ESB. The mechanism of message exchange uses In-Out model, which is a two-way messaging model to satisfy the asynchronous message exchanges. This architecture has a consumer service and a service provider, as well as other services unit including different orchestration that has occurred. It needs to be in a combined package called the service assembly as shown in Figure 5.

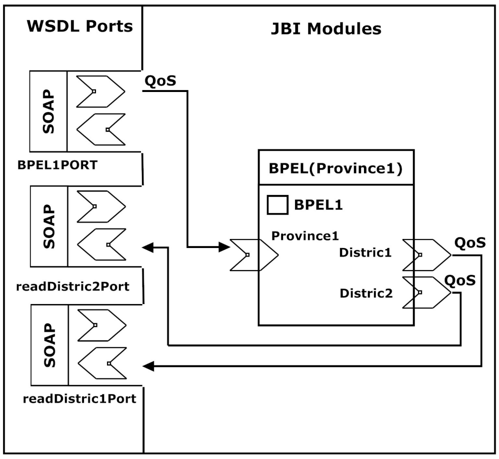


Fig. 5 Service Assemble in Province level

In Figure 5, Java Business Integration (JBI) module consists of several components that interact with the services model. The components that are used as a supply or consumption of services that are performed by the BPEL Service Engine will provide services in order to execute business processes. The executable business processes is defined by using BPEL which is an XML-based language. Execution includes the exchange of business process or business process orchestration messages between the web services with other services as partner. The contract between business processes with service is defined by the WSDL partner. Messages that are exchanged between business processes with service partners are wrapped into WSDL according to the JBI specification; then JBI will be redirected to the Normalized Message Router (NMR). NMR interacts with external web service through binding components. Binding component is responsible for wrapping the detailed protocol message exchange.

In Figure 6 the national does a request that consists of events and parameters of the data. Then the assignment is done by copying an event and the parameter data to be used in invokation as input parameter. Event and data parameters provide notification to the province, where the notification is then used as a request to the district/city based on events that are requested by the national. In BPEL of provinces as seen in Figure 7, event and parameters request are accepted as notification from the national, then re-assign and re-invoke are performed. Furthermore, based on the corresponding event the sequence province BPEL service requests are made to the district/city. District/city will give response to assign the appropriate parameters in response to invoke the BPEL province, in a way of parameter data concat from a number of districts/cities. Then the province also assigns the appropriate parameters response in the response from the national BPEL invokes, by using the parameter data concat from a number of provinces. Results that are given as a response to the national is in the form of parameter data that are obtained from the data in districts/cities in accordance with the national event given as requested.

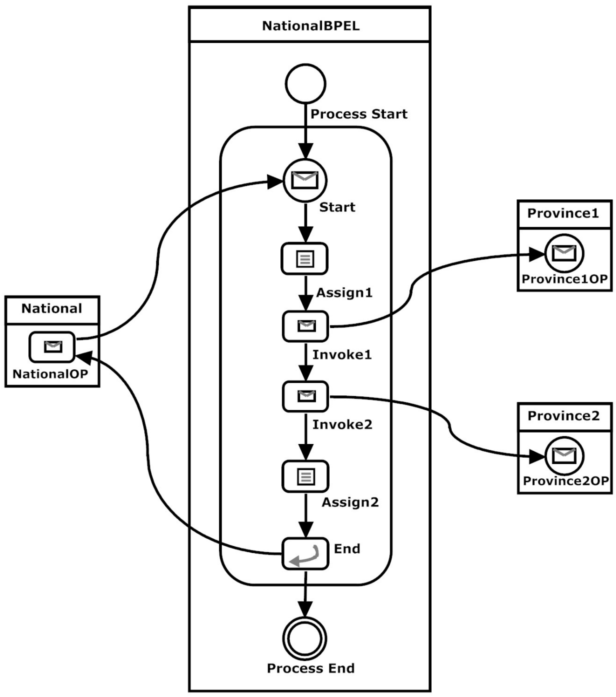


Fig. 6 National BPEL

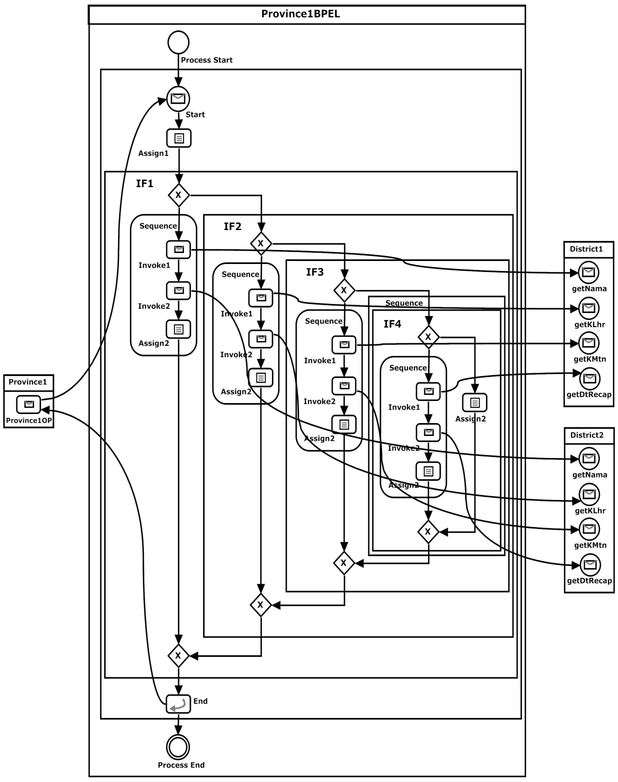


Fig. 7 Province BPEL

8. Performance Measure

In order to evaluate the interoperability architecture that has been generated, a measurement was done by calculating execution time using a number of mixed data on all services models that were defined. The environment that was used to evaluate the amount of time uses Intel Core 2 Duo 2.40 GHz processor with 2 megabyte-memory and 100 GB sized hard disk. The evaluation was done by measuring the length of the execution time for an access using BPEL (connections was not indirect to the database of the districts/cities, but using the services performed by orchestration), access without BPEL (direct connection to the districts/cities database), and LOCAL (direct national connection to the local database obtained from add, delete, and update service). This evaluation aims to determine the feasibility of the model services defined in the EAF by comparing the length of time between the execution of BPEL, without BPEL, and LOCAL. Table 3 presents the results of measurements of the length of execution time to show the recap amount of data on population by provinces, districts, sub-districts and villages. Table 4 presents the results of measurements of the length of execution time to show the whole population data.

Table 3 : Execution Time Recap of National Population

Number of Data (record)	Per Recap	Without BPEL (sec)	BPEL (sec)	LOCAL (sec)
6500	Province	7.833	0.492	0.201
	District	5.975	0.112	0.067
	Sub District	6.349	0.227	0.079
	Village	5.842	0.218	0.082
13000	Province	13.571	0.765	0.281
	District	10.662	0.187	0.124
	Sub District	11.992	0.374	0.202
	Village	11.721	0.312	0.202
19500	Province	20.443	1.153	0.369
	District	17.876	0.211	0.164
	Sub District	18.781	0.399	0.292
	Village	17.692	0.335	0.287
26000	Province	25.993	1.683	0.499
	District	23.723	0.322	0.183
	Sub District	22.918	0.413	0.399
	Village	21.769	0.467	0.401
32500	Province	33.842	2.118	0.572
	District	31.384	0.488	0.231
	Sub District	31.542	0.516	0.421
	Village	32.738	0.534	0.437
39000	Province	41.883	2.668	0.722
	District	38.667	0.589	0.301
	Sub District	38.994	0.692	0.481
	Village	38.478	0.671	0.492
45500	Province	53.881	3.276	0.871
	District	50.773	0.608	0.316
	Sub District	51.692	0.721	0.502
	Village	51.462	0.716	0.512

Execution time for using BPEL in Tables 3 and 4 were not used when the execution was first performed, because it was needed to compile the module and class firstly so it took a long time. The next acceses did not require module and class compilation, so the time was relatively shorter and stable for subsequent accesses.

Tabel 4 : Execution Time to Display National Population Data

Number of Data (record)	Without BPEL (sec)	BPEL (sec)	LOCAL (sec)
6500	35.233	17.354	1.742
13000	68.344	32.479	2.344
19500	110.276	47.199	3.988
26000	138.553	58.831	5.736
32500	167.439	70.773	7.271
39000	194.761	86.359	8.709
45500	222.992	98.794	10.851

In Figure 8 the time difference between that of BPEL and that of LOCAL execution was not significant, because the response given by the model of service was in the form of a single data output parameter. In the comparison between using BPEL and without BPEL, there was a significant difference because the connection was done in direct access to the database on each district/city. Therefore, more resource consumption was required.

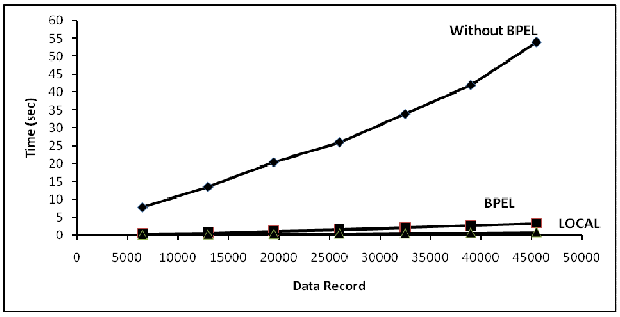


Fig. 8 Recap of National Population Execution Time

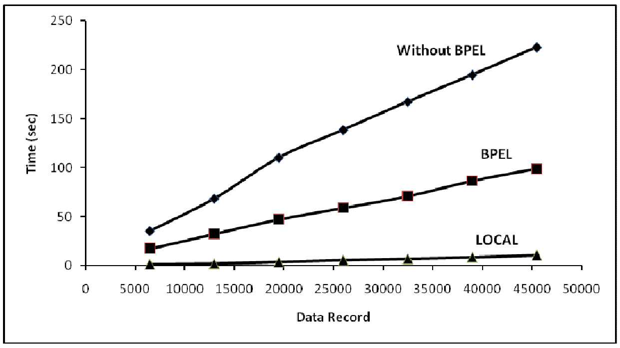


Fig. 9 Execution Time to Display National Population Data

In Figure 9 the execution time of BPEL compared with that of LOCAL result in a very significant difference, because the response given by the model of service was in the form of non-single output parameters meaning the form was a single array as a collection of the entire record of data and therefore more resources were required. The comparison between using BPEL and without BPEL was that they both require more resources. BPEL was faster because it did not connect directly to the database, but it accessed using a services-based text documents with XML format.

9. Conclusions

This paper aims to create EAF for interoperability and data integration for E-Gov applications in Indonesia based on real time architecture developed by the layer services. The aspects of real time are done by providing a trigger event to the other services in asynchronous message delivery model. The orchestration services are performed by GSB which is defined by BPEL. Based on the results of the evaluation by measuring the execution time of the services model in Table 3 and 4, the services model should be defined in EAF which is in the form of a model that has a response services with a single output parameter, such as the service recap of population, births, and deaths. The services model that has response to non-single output parameter should not be used on EAF. Rather, it uses local database (from each province and national) that records data obtained from services add, delete, and update which were requested by the district/city to produce a distributed database. The distributed database scheme aims to provide flexibility in each district/city, province, and national levels to meet different needs. The resources needed for interoperability by EAF is relatively less compared to the E-Gov application architecture where connections are made directly to the database, because EAF was built using a model services based on text documents with XML format.

Acknowledgments

The authors of this paper would like to thank Department of Computer Science and Electronics at Gadjah Mada University which provides guideline and technical assistance for the research.

References

[1] Patricia, J., Pascual, e-Government, e-Asean Task Force UNDP- APDIP, 2003.
[2] Hazlett, S., A., and Hill, F., E-government: the realities of using IT to transform the public sector, Managing Service Quality, EJEG, 2003, Vol. 13, No. 6, pp. 445-452.
[3] Sosiawan, E.,A., “Tantangan Dan Hambatan Dalam Implementasi E-Government Di Indonesia (Challenges and Obstacles in Implementing E-Government in Indonesia)”, National Proceeding e-Indonesia, ITB, Bandung, 2008, pp. 46-59.
[4] Hasibuan, Z. A., “Langkah-langkah Strategis dan Taktis Pengembangan E-Government untuk Pemda (Strategy Steps and Tactical Development of E-Government for Local Government)”, Information System Journal, MTI UI, Vol. 3 – No. 1, 2007, pp. 61-70.
[5] Setiadi, H., Hasibuan, Z. A., and Fahmi, H., “Perubahan Arsitektur Database dan Aplikasi Administrasi Kependudukan Yang Sejalan Dengan Otonomi Daerah (Changing Database Architecture and Population Administration Application toward Local Autonomy)”, Information System Journal, MTI UI, Vol. 3 – No. 1, 2007, pp. 42-51.
[6] Instruksi Presiden Nomor 3 Tahun 2003 tentang Kebijakan dan Strategi Nasional Pengembangan e-Government (Presidential Instruction No. 3 of 2003 on National Policy and Strategy Development of e-Government), Jakarta, 2003.
[7] Blue Print Sistem Aplikasi E-Government, Departemen Komunikasi dan Informatika Republik Indonesia (Blue Print System Applications E-Government, Ministry of Communications and Information of the Republic of Indonesia), Jakarta, 2004.

[8] Rabaiah, Abdelbaset, "Federation of E-government: A Model and Framework", Information Systems Audit and Control Association Journal, Vol. 4, 2007, pp. 1-4.
[9] Allen, B.A., Juillet, L. Paquet, P. and Roy, J., "E-Governance & government on-line in Canada:Partnerships, people & prospects", Government Information Quarterly, 2001, Vol. 18, pp. 93–104.
[10] Janssen, Marijn, and Cresswell, Anthony, "Enterprise Architecture Integration in E-government", Journal of Enterprise Information Management, Vol. 18, Issue 5, 2005, pp.531 - 547.
[11] Ioannidis, Anastasios, Spanoudakis, Manos, and Priggouris, Giannis, "An Xml-Based Platform For E-Government Services Deployment", Journal of Web Engineering, Vol. 1, No.2, 2003, pp. 147-162.
[12] Lee, Thomas, Hon , C., T., and Cheung, David, "XML Schema Design and Management for e-Government Data Interoperability", Electronic Journal of e-Government, Vol. 7 Issue 4, 2009, pp. 381 – 390.
[13] Phu, Phung Huu, and Yi, Myeongjae, "A Service Management Framework for SOA-based Interoperability Transactions", Proceedings The 9th Russian-Korean International Symposium on, Korean, 2005, pp. 680-684.
[14] Contenti, Mariangela, "A Distributed Architecture for Supporting e-Government Cooperative Processes", TCGOV, 2005, pp. 181–192.
[15] Weerakkody, V., Janssen, M., and Hjort-Madsen, Kristian, "Integration and Enterprise architecture challenges in E-Government: a European perspective", International Journal of Cases on Electronic Commerce, Vol. 3, Issue 2, 2007, pp. 13-36 , April-June.
[16] Müller, Viktor, Simon, Balázs, László, Zoltán, and Goldschmidt, Balázs, "Business Monitoring Solution for an e-Government Framework based on SOA Architecture", Proceeding of the 13th IASTED International Conference : Software Engineering and Applications (SEA), Cambridge, MA, USA, 2009, November 2-4.
[17] Peristeras, Vassilios, Tarabanis, Konstantinos, and Goudos, Sotirios K., "Model-driven eGovernment interoperability: A review of the state of the art", Computer Standards & Interfaces Journal, Science Direct, Vol 31, 2009, pp. 613–628.
[18] Stefanova, Kamelia, "Design Principles of Identity Management Architecture Development for Cross-Border eGovernment Services", Journal of e- Government, Vol. 8, Issue 2, 2010, pp. 189-202.
[19] Braun, Christian, and Winter, Robert, "Integration of IT Service Management into Enterprise Architecture", Proceedings of the 2007 ACM symposium on Applied computing, ACM New York, NY, USA, 2007, pp. 1215 – 1219.
[20] Erl, Thomas, Service-Oriented Architecture: Concepts, Technology, and Design, Prentice Hall PTR, New Jersey, 2005.
[21] Open SOA, Service Component Architecture (SCA), http://osoa.org/display/Main/Service+Component+Architecture+Home, 2007. (Last access: Juli 3, 2009).
[22] Michelson, B.,M., Event-Driven Architecture Overview, Patricia Seybold Group, OMG, February 2, pp. 1-8, 2006.
[23] Sahin, K., "Service Oriented Architecture (SOA) Based Web Services For Geographic Information Systems", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing, 2008, Vol. XXXVII, Part B2, pp. 625-630.
[24] Zhong D., et al., "Reliability Prediction for BPEL-Based Composite Web Service", Proceedings of The First International Conference on Research Challenges in Information Science, Ouarzazate, Morocco, 2007, pp. 265-270.
[25] Hafid, B., Balouki, Y., Laassiri, J., Benaini, R., Bouhadadi, M., and Hajji, S. E., "Using BPEL for Behavioral Concepts in ODP Computational Language", Proceedings of the World Congress on Engineering, London, U.K., 2009, Vol. I, pp. 123-129.
[26] Juric, M., B., and Pant, K., "WSDL and UDDI Extensions for Version Support in Web Services", Journal of Systems and Software, Vol. 82, 2009, pp. 1326-1343.

Aris Puji Widodo is a PhD student at Department of Computer Science and Electronics, Gadjah Mada University. His subject is on aspect software engineering architecture.

Jazi Eko Istiyanto is a professor at Department of Computer Science and Electronics, Gadjah Mada University. He holds a PhD in Electronics Systems Engineering, University of Essex, UK since 1995. He is a member of IEEE Robotics and Automation Socienty, IEEE Computer Socienty and Institution of Engineering and Technology.

Retantyo Wardoyo is an associate professor at Department of Computer Science and Electronics, Gadjah Mada University. He holds a PhD in Computation Technology, UMIST, UK since 1996.

Purwo Santoso is a professor at Department of Government Faculty Social and Political Sciences, Gadjah Mada University. He holds a PhD in Political Sciences, London School and Political Sciences, UK since 1998.

Design and Implementation of Geographic Information System on Tourism Guide Using Web-Based Google Maps

Agus-Sudarma[1], Nyoman-Piarsa[2] and Wira-Buana[3]

**[1] Department of Information Technology, Udayana University
Bali, 80361, Indonesia**

**[2] Department of Information Technology, Udayana University
Bali, 80361, Indonesia**

**[3] Department of Information Technology, Udayana University
Bali, 80361, Indonesia**

Abstract

Tourism is one of the world appeal to come to the State of Indonesia, especially Bali. The government was aware of the potential of the tourism so that the government continues to campaign to introduce tourism in Bali to the public. But the promotion which is done by government is still using conventional method. That promotion is promoted by distributing brochures, pamphlet, billboards, etc. The weakness of that method is the tourists cannot access information in real time. Therefore, it need a system that contains information on tourism as a tour guide for travelers who want to visit to Bali so that the information obtained about the tourist attractions are complete. This system will be able to help travelers to determine their location and tourist destination. In addition, this system can also be accessed anywhere with an Internet connection. Sources of data in this system is informant, event or activity, and the document or archive. While the research methods used are as follows : literature review, interviews, and collect the necessary data. Implementation of system will be created use the PHP programming language, while the DBMS used is My SQL.

Keywords: *Tourism, Google Maps, Google Maps API.*

1. Introduction

There are many advantages increase in gained by the state from tourism activities, such as: national income and foreign exchange, a lot of available jobs, and cultural preservation can work together to build Bali. Therefore, the government continues to engage in many activities that support the promotion of tourism.

The way to support and promote tourism activities is to introduce tourism products directly to tourists, both domestic and foreign tourists. Promotion may be in a form of tourism product info, history, location, maps, accommodation information, transportation, etc.

Nevertheless the promotion is still done traditionally. So to solve this problem, this paper will be design geographic information system on tourism guide using a web-based google maps. Moreover, Bali is a tourist destination of the world, therefore geographic information system on tourism guide using a web-based google maps system is needed to support the tourism activities.

2. Methodology

This section, will explain about the methodology used in research that's titled Design and Implementation of Geographic Information System on Tourism Guide Using Web-Based Google Maps.

2.1 Collaborative GIS

The development of GIS dates back to 60's, with the apparition of GIS as computational tools for capturing, managing and transforming data of spatial references to be used in planning and decision-taking processes. Other work that followed after these, was trying to provide a support for the integration between groups of people and for the decision-taking based on geo-referential data, studying also the impact of them on its users. The development of these systems continued and took advantages of the Internet era, in which the Web maps has a fundamental situation. [4]

2.2 Definition of Google Maps

Google Maps is an application made by Google group. Google Maps associated with digital maps. Digital maps displayed by Google Map are various such as map and satellite. Currently, the number of users of Google Maps is growing rapidly because the information provided by Google Maps is very complete. In addition, when there is changing of data in a location, Google Maps will quickly make corrections and update data.

Google Maps can be accessed using a stable Internet connection. The user interface of Google Maps is also very user friendly. So, for those who are new in using Google Maps surely will be interested in this digital map system. Therefore, no wonder if many people in the world use Google Maps to locate and specify a location. [8]

2.3 Definition of Google Maps API

So many features that could be developed by the Google Maps that make the Google Maps is growing very rapidly. Based on these ideas, the founder of Google Maps was made an API (Application Programming Interface) for Google Maps which later was named Google Maps API.

Google Maps API is very useful in the creation of a geographic information system because it does not need to make a digital map for these systems since the digital maps that will be used is provided by Google that can save time and effort. Thus, we can focus on the data and information that will be displayed by the geographic information system. The understanding of the Google Maps API requires a basic knowledge of HTML and javascript. [9]

Using Google Map to work on a geographic information system provides several benefits including the following :

1. Simplify storage system map.
2. Can be accessed anywhere with an internet connection.
3. No need to create a new map.
4. Change of map data faster.
5. The system is going to be lighter.

2.4 System Overview

The overview diagram of this research is shown in Fig 1.

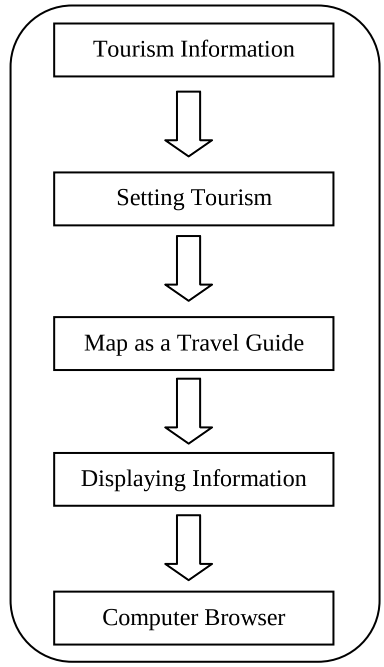


Fig. 1 General Overview System

This system has two types of user, admin and user (public user). Two types of users have different access rights depending on the user level. Admin has the right to access more than the user (public user).

The tourism information is added by admin. Then the system will bring up the map information from Google Maps, and the tourism information will appears as digital map by category. For example, if the tourists want to search for information based on tourism attractions category, it just need to select attraction category. When the category has been selected then the next step is to find information or the location of the selected category. Locations can display the desired information as a tour guide.

This system not only has advantage on the selection of category, but also has other advantages, as the map can become digital maps and normal maps (street maps). The travelers are highly facilitated by the selection of maps so the system can be used as a tourist guide for tourists. The following is an explain of the process that occurs in the geographic information system on tourism guide using a web-based google maps and illustrated by the context diagram described in Fig. 2.

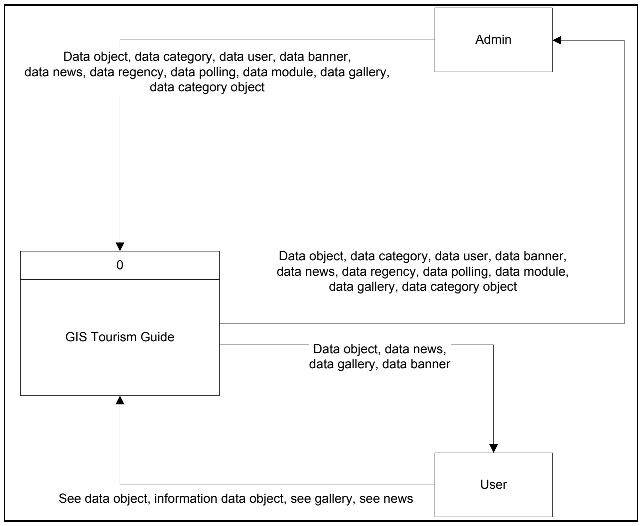


Fig. 2 Context Diagram

In Fig. 3 will be explain about the business processes that occur in the geographic information system on tourism guide using a web-based google maps.

Business process is demonstrated through entities called data flow diagrams. Entities on data flow diagram level 0 is login, view object, view news, create object, and management master data.

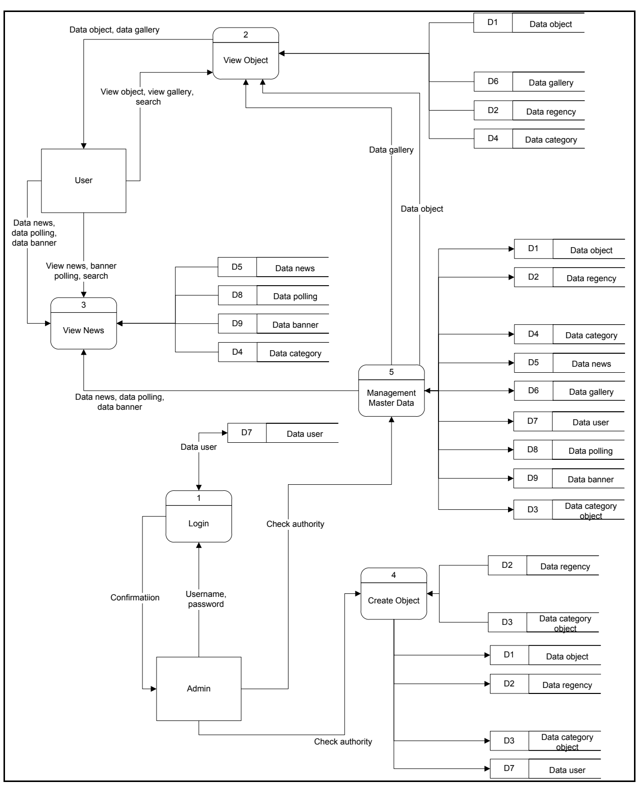


Fig. 3 Data Flow Diagram Level 0

Figure 4. illustrating table design geographic information system on tourism guide using a web-based google maps. Table formed of the design were 10 tables : tb_gallery, tb_object, tb_category_object, tb_polling, tb_banner tb_module, tb_regency, tb_category, tb_user, and tb_news.

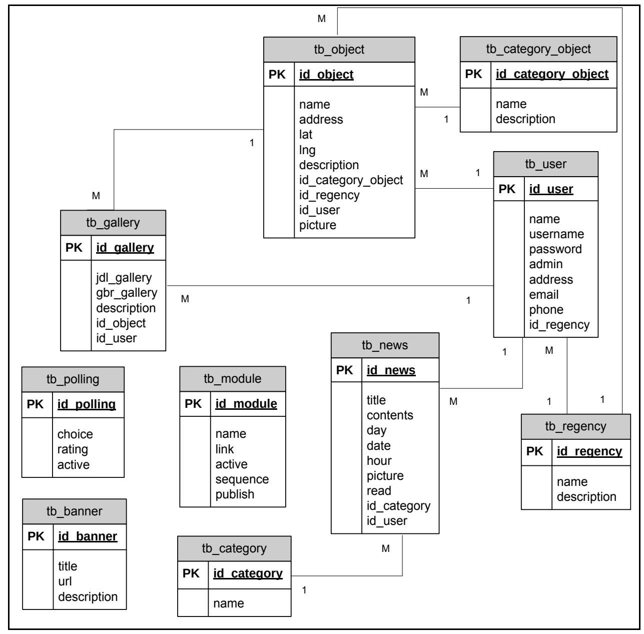


Fig. 4 Desgin Table

The system consist of 10 tables :

1. tb_gallery used to store data about picture.
2. tb_polling used to store data and information about polling such as rating, choice, etc.
3. tb_banner used to store data and information about banner.
4. tb_category used to store data and information about category of news.
5. tb_module used to store data and information about module such as name, link, publish, etc.
6. tb_news used to store data and information about news such as title, contents, picture, etc.
7. tb_object used to store data of attractions such as data or information a beach, park, etc.
8. tb_category_object used to store data about the attractions of category.
9. tb_user used to store data and information about user such as name user, address user, email, etc.
10. tb_regency used to store data and information about regency such as name and description.

The degree of relation indicates the maximum number of entities that can be filled with other entities in the entity set. The degree of relationship that occur in sets of entities on the system :

1. tb_category with tb_news
   The degree of relation that occurs between tb_category and tb_news is 1: M (one to many) because a *category* can have a lot of *news*.
2. tb_object with tb_gallery
   The degree of relation that occurs between tb_object and tb_gallery is 1: M (one to many) because a *object* can have a lot of *gallery*.
3. tb_regency with tb_object
   The degree of relation that occurs between tb_regency and tb_object is 1: M (one to many) because *regency* can have a lot of *object*.
4. tb_category_object with tb_object
   The degree of relation that occurs between tb_category_object and t_object is 1: M (one to many) as a *category_object* can have many *object*.
5. tb_user with tb_object
   The degree of relation that occurs between tb_user and tb_object is 1: M (one to many) because *user* can send or have a lot of *object*.
6. tb_regency with tb_user
   The degree of relation that occurs between tb_regency and tb_user is 1: M (one to many) because *regency* can have a lot of *user*.
7. tb_user with tb_gallery
   The degree of relation that occurs between tb_user and tb_gallery is 1: M (one to many) because user can have or send a lot of *gallery*.
8. tb_user with tb_news
   The degree of relation that occurs between tb_user and tb_news is 1: M (one to many) because user can have or send a lot of *news*.

3. Experiments and Results

Once the system design is completed, the next will be the trial process. This step holds an important role in programming. It lets the programmer do trial and error fixing before it launched officially for the viewers.

Fig. 8 Homepage

Menu on Fig. 8 contains recent news about tourism in Bali as well as the homepage menu search button, polling, and banner.

Fig. 9 Gallery Album

Gallery is a page that displays album contains image galleries of attractions in the island of Bali. Initially the gallery menu will display gallery album if the gallery album is clicked to it will display a gallery of images.

Fig. 10 Gallery Images

On Fig. 10, there are two album pictures related to Bali Bird Park, labeled as “Bali Bird Park” and “Koleksi Bali Bird Park”.

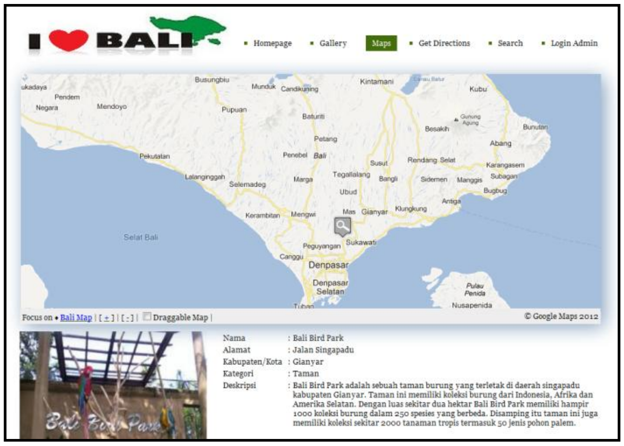


Fig. 11 Gallery Map

Besides being able to display images in the selected album, gallery menu can also display maps of locations corresponding to the album. On Fig. 11 is corresponding map of the location of Bali Bird Park.

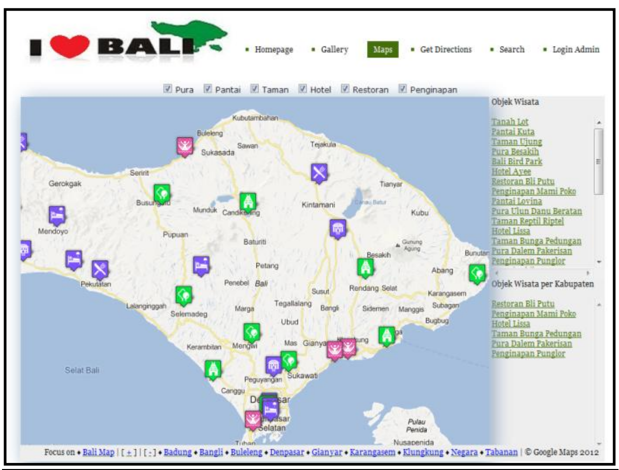


Fig. 12 Map

Fig.12 Map is a page that displays a map of attractions in the island of Bali. Attractions are displayed on maps, visualized with marker according to the category object. Maps using a check box that can be used to select categories of objects based on user desires such as *pura (temple)*, *pantai (beach)*, *taman (park)*, *hotel*, *restoran (restaurant)*, *penginapan (lodging)*.

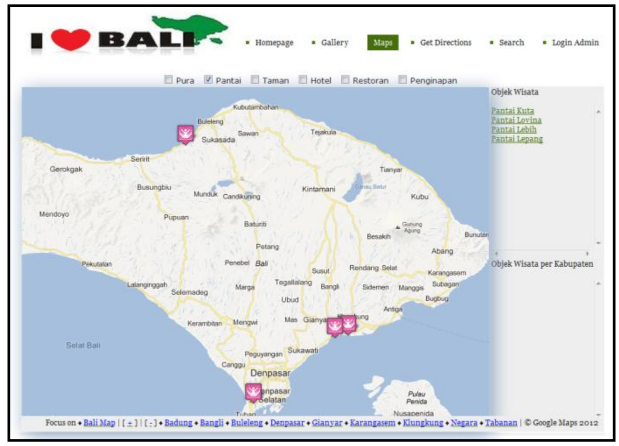


Fig. 13 Map Object Categories *Pantai (Beach)*

Fig. 13 is an example of the user who just wants to see the object categorized *pantai (beach).* There is only one marker is shown on the map.

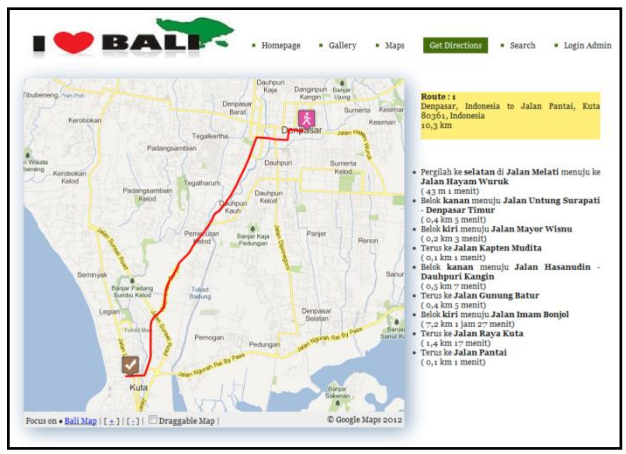


Fig. 14 Get Directions

Get directions is a page that will show you the road to the location of the existing tourist attraction on the island of Bali with a drag marker starting location to a destination location.

Fig. 14 is a view to using the get directions feature. Pink marker indicating the starting location with an image of people walking (walking) while a brown marker with pictures lawn signs are markers showing the location of the destination.

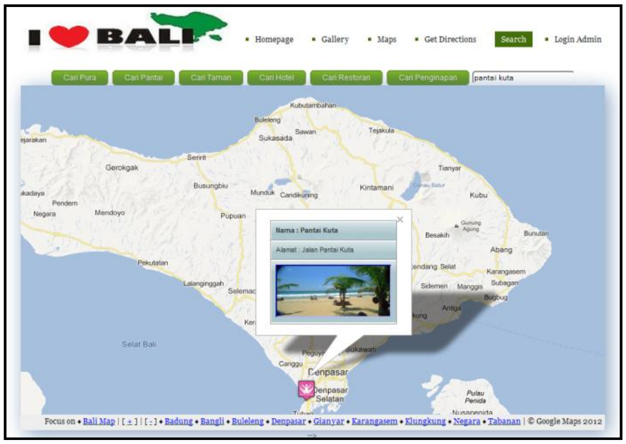


Fig. 15 Search

Fig. 15 Search is a page that will display a searchable user attraction. Users need to type the area or location to search. After that, the system will bring up the location of the object by name or location of the area that has been input by the user.

Fig. 16 Login Admin

Fig. 16 Login Admin is special page for the user who has full access rights to the system admin. This page serves as the admin login page.

4. Conclusions

The result on this paper is a design data flow diagrams and tables. Data flow diagrams and tables have a strong relationship. Data flow diagrams and tables can be used as a basis and foundation to create a geographic information system.

The system will be created is a geographic information system on tourism guide using a web-based google maps. The system will help travelers to choose tourist destinations in Bali. The system can be accessed anywhere in real-time by using an Internet connection. Basically, making this system aims to help the government to promote tourism.

Acknowledgments

Our thank goes to Department of Information Technology Udayana University, Bali, who has helped organize this research.

References

[1] Purvis, Michael, Jeffrey Sambells, Cameron Turner. Beginning Google Maps Applications with PHP and Ajax From Novice to Professional. Apress, 2006
[2] Google Code. Map V3 Tutorial. http://code.google.com/apis/maps/documentation/
[3] Svennerberg, Gabriel. Beginning Google Maps API 3. Apress, 2010
[4] P. M. Rafael, et al. Technology Integration around the Geographic Information: A State of the Art. IJCSI International Journal of Computer Science Issues, Vol. 5, 2009
[5] Kadir, Abdul. 2008. *Belajar* Database *menggunakan* MySQL. Yogyakarta : Andi.
[6] http://en.wikipedia.org/wiki/Latitude
[7] http://en.wikipedia.org/wiki/Longitude
[8] http://en.wikipedia.org/wiki/Google_Maps
[9] http://en.wikipedia.org/wiki/Google_Maps#Google_Maps_API
[10] http://repository.upnyk.ac.id/905/1/SKRIPSI.pdf
[11] https://developers.google.com/maps/documentation/javascript/
[12] http://maps.google.co.id/
[13] http://themeforest.net/

I Putu Agus Sudarma studied Information Technology in Departement of Information Technology Udayana University since August 2008, and now working her research fo S.Ti. degree in Information Technology.

I Nyoman Piarsa, ST., MT received his ST degree in Electrical Engineering from Institut Teknologi Sepuluh Nopember University in 1998, and his MT received in Electrical Engineering from Gadjah Mada University in 2001. He is lecturer at Electrical Engineering Departement (major in Computer System and Informatics) and at Information Technology Departement of Udayana University.

Putu Wira Buana, S.Kom., MT received his ST degree in Informatics Engineering from Institut Teknologi Sepuluh Nopember University in 2005, and his MT received in Electrical Engineering from Brawijaya University in 2008. He is lecturer at Electrical Engineering Departement (major in Computer System and Informatics) and at Information Technology Departement of Udayana University.

Combined Architecture for Early Test Case Generation and Test suit Reduction

Mr. Saurabh Karsoliya, Prof.Amit Sinhal, Er.Amit Kanungo

Computer Science and Engineering, MANIT
Bhopal, M.P., India

Information Technology Department, TIT
Bhopal, M.P., India

Information Technology Department, TIT
Bhopal, M.P., India

Abstract

Model based combinatorial testing is the most effective and efficient method of systematic interaction testing. In this multiple variables can interact with test model in the form of combination and each set is considered as a test pairs. It is used mainly for reducing the test pair's size, complexity and time of test generation. Accomplishment of this reduction task is the most vibrant and tedious activity because this consist of test data extraction, boundary analysis, path measurement analysis, condition coverage etc.
The problem with Model based combinatorial testing is optimization of cost and effort which can only be achieved by early test case generation methods. For this we are using UML diagrams from design phase.
For reducing the test size and complexity we have proposed a new MBTGA framework and an MTGIPO. In this paper we focus our research on empirical study and result obtained by developed framework. We are also proposing a new UML design data extraction (DDE) algorithm for getting the useful information from a given UML diagram as input.
Keywords: *Combinatorial test, Pairwise test, UML, NP-complete, Genetic algorithm, MBT, UML, MBTGA, MTGIPO, DDE*

1. Introduction:

Combinatorial Testing As Research Domain

Systematic testing of highly-configurable software systems, e.g. systems with many optional features, can be challenging and expensive due the exponential growth of the number of configurations to be tested with respect to the number of features. It is estimated that 30% of an enterprise's IT budget is devoted to the original development and 70% is for enhancements and fixing bugs not discovered during original development [3]. Thus the usage of Combinatorial Interaction Testing (CIT) technique can improve the effectiveness of the testing activity for these kinds of systems, at the only cost of modeling the system's configurations space.

Combinatorial testing can help detect problems like interaction failures of combinations this early in the testing life cycle. The key insight underlying *t*-way combinatorial testing is that not every parameter contributes to every failure and most failures are triggered by a single parameter value or
interactions between a relatively small numbers of parameters [4]. In fact, CIT consist in systematically testing all possible partial configurations (that is, involving up to a fixed number of parameters only) of the system under test. The most commonly applied combinatorial testing techniques pairwise testing, which consists in applying the smallest possible test suite covering all
the pairs of input values (each pair in at least one test case). It has been experimentally shown that a test suite covering just all pairs of input values
can already detect a large part (typically 50% to 75%) of the faults in a program [3, 4]. Moreover, incorrect behaviors or failures due to unintended feature interaction, detected by CIT, may not be detectable by other more traditional approaches to systematic testing [1, 5].

2.Purpose of Study (Problem Statement):

The immediate purpose of this research is to Design & Develop a MBTGA Framework for Programmed Combinatorial Test (PCT) that can help to provide assurances of reduced overhead of testing . Some techniques for providing such assurances have been developed in the past, but no single technique has provided a complete solution to the problem. Thus, this thesis will explore the effectiveness of combining two such techniques (Unified Modeling & Combinatorial Testing) into a single tool. The more general purpose of this research is to improve the available methods of software testing.
There are several major challenges that completely resolved by our tool with testing modern software.

Some are as follows

- Automation of the test case generation and their execution.
- Development of domain and software engineering expertiseneeded for adequate testing.
- Formalization and modeling of the software specifications and implementations, and software testing process and effects. The reduction in growing complexity of the modern software-based systems.
- Generating Test Cases Criteria at the Time of Design and Requirement Analysis (Early Test Case Creation saves time & cost).
- Extracting the data from UML diagram to generate test cases.
- Generating the reduced number of test cases (Test Suite Size).
- Controlling the Test Case Creation by Constraints & Relations.
- Providing the maximum test coverage (100% for 2-way pairwise testing).
- Reducing the execution time for testing.

Thus, by considering the entire above research objective we have developed a Automated Tool which follow the proposed design architecture of MBTGA & Implemented an algorithm for this interaction testing based on Enhanced MBTGA based IPOG (MTGIPO). We also proposed a UML based data extraction algorithm (DDE). Our tool is capable of generating the minimum test cases in comparison with other publicly available tool.

2.1 Literature Survey

Proposed Methodology: - Complexity of software needs to identify better techniques for different functionalities in the software development life cycle. This complexity is truly reduced by deterministic decision environment of parameters in quality assurance of software. It is mainly done by ways of testing, which is an activity that faces constraints of both time & resources. Testing the outbound of any software is been judged by means of special type of test in black box testing known as combinatorial testing. It is a well known dynamic approach for quality improvements because it provides effective error detection at very low cost. Creation of optimal set of test will effectively decreases complexity of the software system by pairing the input parameters through pairwise testing using orthogonal arrays & Latin squares. Hence an efficient strategy is required to reduce the number of test cases formed by above mentioned method.

MBTGA with enhanced IPOG has contributed to enhance many known CA and MCA that exist in the literature. In our research & implementation we proves that the given algorithm & design architecture of MBTGA (Model Based Test Generation Architecture) is well defined for improving efficiency thorough multiple parameters (Size, Time, Complexity, Cost etc.).We have implemented our tool based on two proposed approaches. First is MBTGA Algorithm & Design Architecture for 2-way (Pairwise Testing) & a Modified Test Generation IPO (MTGIPO) for 3,4 way interaction testing. The MTGIPO uses algorithmic approach based on IPOG (In Parameter Order General) strategy for test case reduction with improved parameters. While for pairwise test suite creation phase of MBTGA requires searching the best pair combination for pairtest (2 way Test).

Design Architecture

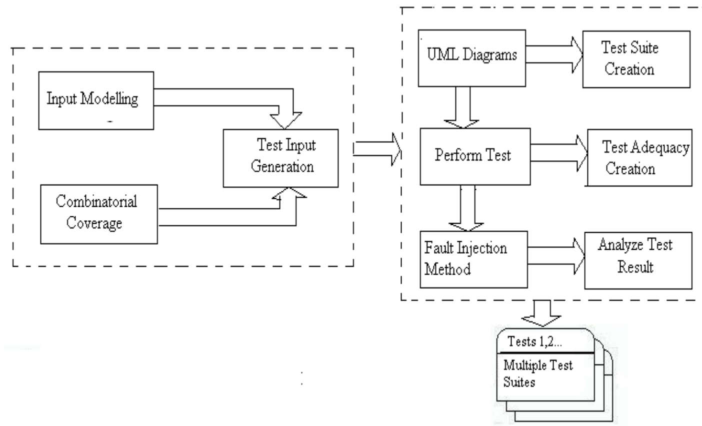


Fig:- A New Model Based Test Design Architecture (MBTDA) For Test suite creation

3. PROPOSED SOLUTION:

During our research & implementation we had used a result which is been proven that the problem of combinatorial testing is NP complete & whose solution may be achieved by heuristic or genetic based search algorithm. We are calculating our result of reduced pairtest through genetic algorithm. It has two basic operator’s mutation & crossover. We first construct the activity diagram for the given problem and then randomly generate initial test cases, for a program under testing. Then, by running the program with the generated test cases, we obtain the corresponding program execution traces. Next, we compare these traces with the constructed activity diagram according to the specific coverage criteria. We use path coverage as test adequacy criteria. Next, we propose a novel approach to generate test cases from test scenarios using UML activity, sequence and class diagram. First we generate test scenarios from the activity diagram and then for each scenario the corresponding sequence and class diagrams are generated. After that we analyze the sequence diagram to find the interaction categories and then use the class diagrams to find the settings categories. After analyzing each category using category partitioning method, its significant values and constraints are generated and respective test cases are derived. Finally, we propose a technique to optimize the generated test cases automatically. We define an error minimization technique in our approach, which works as the basic principle for optimized test case generation. Transition coverage is used as test adequacy criteria in this approach.

3.1 Developing a Common Framework (MBTGA)

Several approaches to design test cases and application of software testing have been proposed by researchers. We also know that testing is a very important phase of software development and always comes at the last. No such technique is been developed to generate test cases before the implementation of code. So we are focusing our research on development of a framework on the basis of which it is been possible of developers and stakeholders to generate different test scenarios previously. We are also contributing our work towards the optimization of test case generation strategies via combinatorial testing. Thus by enriching the above two main goals of early generation of test case by UML through Model based testing & improving the test case efficiency by enhancing the combinatorial testing methodologies, we had developed a UML based combinatorial approach(MBTGA) framework. It will fulfill all the requirements and serve best to accomplish our research objectives. Also the prototype tool being developed to prove efficiency and effectiveness of the proposed methodology.

We separate our research in two domain in which first one focuses on reducing the test case size, complexity and efforts. Second one is early generation of test cases, at the time of requirement analysis and design which saves cost, time and efforts. Our project also makes it possible to identify the bugs which tend to come at the time of development and installation. To overcome our first research domain objective we developed the solution in various phases. For this initially we proposes a new flow structure to inculpate both the concept in one. Then we develop an algorithmic approach based on Unified Model Architecture of MBTGA for pairwise testing (Interaction 2-way) & MTGIPO (Enhanced Unified In Parameter Order General for Interaction > 2 way) strategy for test case reduction with improved parameters. In this test suite creation phase of MBTGA requires searching the best pair combination for pairtest (2 way Test).We also applying test suite prioritization methodology to enhance the performance of testing & also reduces the test suite size & complexity. To accomplish our second research domain goal we use model based testing (MBT). From this we uses UML, which supports object-oriented technology, is widely used to describe the analysis and design specifications of software development. UML models are an important source of information for test case design. UML activity diagrams describe the realization of the operation in design phase and also support description of parallel activities and synchronization aspects involved in different activities perfectly. Now this part must extract data from UML diagram so to overcome this challenge we had also proposed a new UML test data extraction (DDE) algorithm.

3.2 Taking UML as Model in Model Based Testing

Model based testing (MBT) refers to the type of testing process that focuses on deriving a test model using different types of formal ones, then converting this test model into a concrete set of test cases [5]. Models are the intermediate artifacts between requirement specification and final code. Models preserve the essential information from the requirement, and are the basis for implementation. Instrumentation of models into testing process is the prime subject of concern of our thesis. Testing methodologies which uses model is called *model based testing (MBT)*. Development of *unified modeling language (UML)* has helped a lot to visualize/realize the software development process. At the earliest stage of *software development life cycle (SDLC)*, no one including user and developer can see the software; only at the final stage of the product development it is possible. Any errors/problems found out at the final stage, it incurs a lot of cost and time to rectify, which is very much crucial in IT industry.

UML is the modeling language, which supports object-oriented features at the core. In the last few years, *object-oriented analysis and design (OOAD)* has come into existence, it has found widespread acceptance in the industry as well as in academics. We concentrate here on widely accepted practices based on the use of the Unified Modeling Language (UML) to support an object-oriented development process [6]. The main reason for the popularity of OOAD is that it holds the following promises:

• Code and design reuse

• Increased productivity

• Ease of testing and maintenance

• Better code and design understandability

UML accomplish the visualization of software at early stage of SDLC, which helps in many ways like confidence of both developer and the end user on the system, earlier error detection through proper analysis of design and etc. UML also helps in making the proper documentation of the software and so maintains the consistency in between the specification and design document. The key advantage of this technique is that the test generation can systematically derive all combination of tests associated with the requirements represented in the model to automate both the test design and test execution process.

3.3 Problem Statement

To develop early test case generation strategy Our one of the main research objective is to make it possible for designer and developer to generate the test case at the time of requirement gathering & design phase. From this early information of test case failure & success reports the designer can easily remove the bugs & error at very early stages of project development. It improves quality of product in well defined standards. For this we are using the UML diagram which is part of design phase & flowchart which is a part of requirement gathering phase to extract the sufficient amount of test data. Later on the proposed reduction algorithm is applied to get better results.

To reduce the Test Suite size, complexity & cost here our aim is to develop a strategy which can be able to generate combinatorial test cases for variable parameters value which reduces efforts and cost. It also extracts what information is necessary to test the integration of components in the process of system composition;

To extracts test data from different design diagrams (UML) Here we investigate which individual or combination of UML diagram types, offer sufficient information to generate test cases; Also shows how the proposed strategy reports on the amount of testable information contained in a model.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

3.4 UML Test Data Extraction Algorithm

Extracting the data from UML diagram is most difficult task. For this there is no such software developed which will extract the things from a png or jpeg format. It should be considered as new project. So we focus our main aim towards extracting all the useful information from a fixed textual format which is automatically generated from an well recognized tool PlantUML. It is a UML Based tool which takes the diagram as input and gives the respective textual notation for UML. To accomplish this reading we use pattern matching functions and operators or regular expression and then transform the values to a text file.

In our proposed strategy we had used activity diagram as a test model or a formal model. Here activity diagram is used to parse the information to generate test scenarios for various path available in it. Covering all the path will ensure that the maximum coverage is been achieved in it.

First of all our approach parses the activity diagram and generates the test scenarios which satisfy the path coverage criteria. As activity diagrams represent the implementation of an operation like the flow chart of code implementation and an executing path is a possible execution trace of a thread of a program, the executing paths are derived directly from the activity diagrams. We have considered path coverage in our approach, since it has the highest priority among all the coverage criteria for testing. Our approach also handles the complicacy of nested fork joins using a criterion that checks whether the target activity state of a transition is a fork or an activity state. If the target of the transition is a fork, then the fork has higher priority over the activity state. So it should be considered first and then only the other path is considered. As a result of this priority criterion the complicated nested fork-join pair is handled properly in our approach. After all the possible test scenarios are generated we generate the corresponding sequence diagram, and class diagrams for each scenario. Then test cases are then derived by finding significant values of environment conditions and parameters.

The Proposed Algorithm for UML test data extraction (DDE):-

Algorithm: DDE (Parameter, values)

1. Initialize the diagram as @startuml and assign it parameter 0, initial value is "Start"
2. Scan the file completely
3. Check for →" "
   If (pattern==found) consider it as next parameter and value in " " as its parameter value

   Repeat above till all identified

4. Now again scan for →< >
   If found read for next two [ ]

   For [Yes] add value in →" " as its value in above parameter

   For [No] add default value *

   Else for no value found consider both as parameter value.

5. Now check for York condition pattern
   If (===_===→" "equals to ===_===→" ")

   The above condition is consider as values of same parameter

6. Repeat till 5 until find →(*)
7. If(found== →(*))
   Then add next parameter and its value as "End"

8. Scan till all pattern matches
9. Exit

Mathematical Expression

Result Achieved: - We demonstrate our result by showing the improved performance MBTGA based on MTGIPO over other existing strategies. For this evaluation certain parameters is been identified and the result is been compared. These parameters are Reduced Test Size, Coverage, Time required, Complexity, Don't Care Conditions and last one is most important term possibility of early generation of test case.

To compare against other existing strategies we found that about 40% of the conditions of the program were usually covered by random test data generation, genetic approach covered 60% of the conditions and pairwise testing outperform former two by a considerable margin in most of our experiments. Genetic search achieved about 85% condition-decision coverage on average, while the random test-data generator consistently achieved just over 55%. So the pairwise testing strategy is proved to be an efficient test generating strategy. The following table shows the size of generated test set obtained by our technique as well as two other methods. Note that the size of test suite in case of pairwise strategy using MTGIPO is less than other tools and pairwise testing result obtained is in the form of Comparison Table's, Graphs, Utilities Functions, Features, Parameter Covered tables.

Table: - Test Data 1:- 5 parameter, 15 parameter values
Method:-Variable value

Tool Name	Number of Pairs Covered in all Combinations	Test Case Generate	Coverage Achieved	Time Required	Don't care Conditions
ACTS	270	40	0.254	0.015 sec	Not Count
PICT	NA	NA	NA	NA	NA
TCG	270	43	NA	NA	Not Count
MBTGA	**270**	**40**	**0.467**	**0.019 ml sec**	**Not Present**

Table 1:-Comparison of MBTGA with other tools available for test data 1 for 3 way

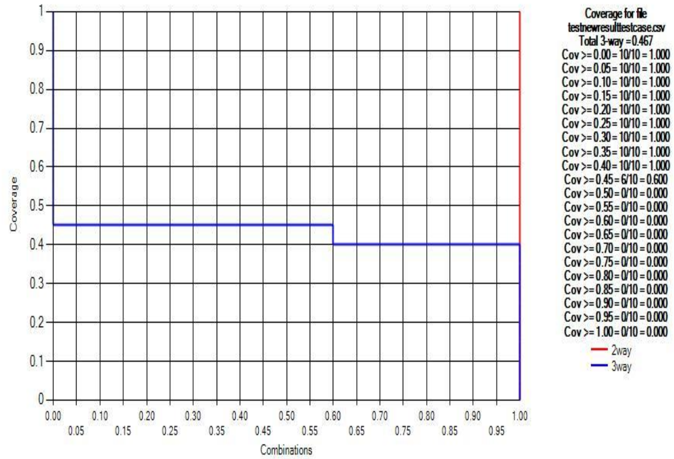
Table Conclusion:-The above table shows that the tool is capable of achieving reduction up to 85% and covers all pairs in just 40 test cases with 3 way interaction probability in just 0.019 mili second. It also shows that how the different tools perform under same conditions.

Table 11:-Empirical study of MBTGA on the basis of various parameters

Data Extraction(DDE)	Pairs Covered	Reduced Test Cases	Tool Name MBTGA	Coverage Achieved	Time	Don't care	Comap-atibility
6 Param, 9 values	32	6	Activity Diagram 1	1	2.609 ml sec	0	N/A In Any tool
5 Param, 7 Values	19	4	Activity Diagram 2	1	0.683 ml sec	0	N/A
7 Param, 9 Values	34	4	Activity Diagram 3	1	0.8066 ml sec	0	N/A
6 Param, 9 Values	32	6	Activity Diagram 4	1	0.849 ml sec	0	N/A
11 Param, 15 Values	101	6	Activity Diagram 5	1	1.5089 ml sec	0	N/A

Table Conclusion: - The above table shows an effective early test case generation feature of our tool which implies on the data extraction from UML activity diagram. We used our developed DDE algorithm for getting this result. We shows the performance and result evaluation of our tool on the basic of 7 parameters. Firstly DDE is been capable of extracting the correct data from given UML textual notations. The table shows how our tool effectively reduces the test size in very less time and gives maximum coverage. The feature which we have proposed and implemented is not present in any combinatorial testing tool and serves as add on module for our research. In future its improved versions are likely to be developed.

Graph:



Graph Conclusion: - The above graph is been generated from combinatorial coverage measurement tool by NIST. It shows the coverage achieved by our MBTGA tool based on MTGIPO and gives 100 % for 2 way (Red) and 0.467 for 3 way (Blue) interaction test, which is higher than any other available tool.

4. Conclusion & Future Work:

- The empirical study & the above chart easily prove that the strategy which we are proposing is effective & efficient for pairwise coverage problem. The test cases will 70 % reduced in most of the cases depending upon the seed value selected as per the given test criteria based on requirement. Future work will more effectively enhance the above proven results in a systematic way so as to generalize the tool. Also some researchers were focusing on improving solution domain of this genetic theory by NP complete & hard relations. So while proposing a new strategy in combination with pair wise approach always kept in mind its practical implementation so as to make the tester's work easy. MBTGA Tool is a great deal in test suite reduction & in addition it also provided maximum coverage based on genetic theory.

5. References:

[1] Pairwise Testing concept & strategy available from ;http://www.pairwise.org/tools.asp.

[2]B. Selic, and J. Rumbaugh, "UML for Modeling Complex Real Time Systems", Available Online Via www.rational.com/Products/Whitepapers/100230.Jsp.

[3]E. Holz, "Application of UML within the Scope of New Telecommunication Architectures", GROOM Workshop on UML, Mannheim:Physicaverlag, 1997.

[4]G. Booch, J. Rumbaugh and I. Jacobson, "The Unified Modeling Language User Guide", Addision Wesley, Reading, MA,1999.

[5]Clementine Nebut, Franck Fleurey, Yves Le Traon, "Automatic Test Generation: A Use Case Driven Approach", IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 32,NO. 3, MARCH 2000

[6]A. Abdurazik, J. Offutt, and A. Baldini. "A controlled experimental evaluation of test cases generated from UML diagrams". Technical report, George MasonUniversity, Department of Information and Software Engineering, 2004.

[7]Lei, Y and Tai, K. C., In-Parameter-Order: A Test Generating Strategy for Pairwise Testing, IEEE Trans. on Software Engineering, 2002, 28 (1), 1-3.

[8]J.Bach and P. Shroeder. Pairwise testing - a best practice thatisn't. In Proceedings of the 22nd Pacific Northwest Software Quality Conference, pages 180– 196, 2004.

[9]Mandl, R., Orthogonal Latin Squares: An Application of Experimental Design to Compiler Testing, Comm. ACM, 1985, 28 (10), 1054-1058.

[10]Cohen, D. M., Dalal, S. R., Fredman, M. L., and Patton,G. C., The AETG system: An Approach to Testing Based on Combinatorial Design, IEEE Transaction on Software Engineering, 1997, 23 (7), 437-443.

[11]Colbourn, C. and Dinitz, J. (Ed.) The CRC Handbook of Combinatorial Design, CRC Press, 1996.

[12]A. Andrews, R. France, S. Ghosh, and G. Craig. "Test adequacy criteria for UML design model. Software Test Verification and Reliability", 13:97–127, 2003.

[13]Stefania Gnesi, Diego Latella, and Mieke Massink. "Formal test- case generation for UML state charts". In ICECCS '04: Proceedings of the Ninth IEEE International Conference on Engineering Complex Computer Systems Navigating Complexity in the e-Engineering Age, pages 75–84, Washington, DC, USA, 2004. IEEE Computer Society.

[14]J. Hartmann, C. Imoberdorf, and M. Meisinger." UML-based integration testing". In ISSTA '00: Proceedings of the 2000 ACM SIGSOFT international symposium on Software testing and analysis, pages 60–70, New York, NY, USA, 2000.ACM.

[15]S. Helke, T. Neustupny, and T. Santen, "Automating Test Case Generation from Z Specifications with Isabelle," ZUM '97: The Z Formal Specification Notation, LNCS 1212, pp. 52-71.
J.P. Bowen, M.G. Hinchey and D. Till, eds. Springer-Verlag, 1997.

[16]Davis, Lawrence (1985). Job Shop Scheduling with Genetic Algorithms, *Proc. Int'l Conference on Genetic Algorithms* and their Applications.

[17]S. R. Dalal, A. Jain, N. Karunanithi,J. M. Leaton, C. M. Lott, G. C. Patton "**Model-Based Testing in Practice"** To appear in *Proceedings of ICSE'99, May 1999 (ACM Press).*

[18]P. E. Ammann and A. J. O_utt. "Using formal methods to derive test frames in category partition testing". In Ninth Annual *Conference on Computer Assurance (COMPASS'94),* Gaithersburg MD, pages 69–80,1994.

[19]P. E. Ammann and P. E. Black. "A specification-based coverage metric to evaluate test sets". In Proceedings of Fourth IEEE International High-Assurance Systems Engineering Symposium (HASE 99), pages 239–248. IEEE Computer Society, November 1999. Also NIST IR 6403.

[20]P. E. Ammann, P. E. Black, and W. Majurski." Using model checking to generate tests from specifications". In Proceedings of the Second IEEE International Conference on Formal *Engineering Methods (ICFEM'98),* pages 46–54. IEEE Computer Society, Dec. 1998

An Advanced Survey on Secure Energy-Efficient Hierarchical Routing Protocols in Wireless Sensor Networks

Abdoulaye Diop[1], Yue Qi[2], Qin Wang[3] and Shariq Hussain[4]

[1,2,3,4] School of Computer and Communication Engineering, University of Science and Technology Beijing Beijing, 100083, China

Abstract

Wireless Sensor Networks (WSNs) are often deployed in hostile environments, which make such networks highly vulnerable and increase the risk of attacks against this type of network. WSN comprise of large number of sensor nodes with different hardware abilities and functions. Due to the limited memory resources and energy constraints, complex security algorithms cannot be used in sensor networks. Therefore, it is necessary to balance between the security level and the associated energy consumption overhead to mitigate the security risks. Hierarchical routing protocol is more energy-efficient than other routing protocols in WSNs. Many secure cluster-based routing protocols have been proposed in the literature to overcome these constraints. In this paper, we discuss Secure Energy-Efficient Hierarchical Routing Protocols in WSNs and compare them in terms of security, performance and efficiency. Security issues for WSNs and their solutions are also discussed.

Keywords: *Wireless sensor network, Hierarchical routing protocol, Security.*

1. Introduction

The tremendous development in the electronics technology lead the way to development of micro-electronics thus enabling production of small chips and micro devices. The communication technology is being reformed due the design and development of micro devices and hence enabled the design and development of WSNs with low cost, low energy consumption and high utilization. WSNs have lot of applications in military, health and other industrial sectors. Because of the characteristics of WSNs, sensor nodes are usually characterized by limited power, low bandwidth, memory size and limited energy [1].

Due to the scalability and energy efficiency characteristics, researchers proposed many routing protocols for cluster-based WSNs [2]. In WSNs, routing protocols can be classified into two categories: Network Structure and Protocol Operation. Hierarchal routing protocol is one of the categories in their classification of WSN routing protocol based on the Network Structure. In cluster-based routing protocols, network is divided into cluster and each cluster has its own cluster head (CH). Further, CHs are responsible for relaying of messages from ordinary nodes to the Base Station (BS). CHs can communicate directly with the BS, can be anywhere in the network and change per interval, which also improves network's energy efficiency [2].

Several enhanced secure hierarchal routing protocols have been proposed in literature [19, 21, 23, 24-28], to attempt to achieve both security and efficiency for WSNs. Most routing protocols are vulnerable to a number of security threats [3]. Attacks involving CHs are the most damaging.

Due to the resource constraints of wireless sensors, public-key based cryptographic algorithms like RSA and Diffie-Hellman are too complicated and energy-consuming for WSNs. However the symmetric cryptographic technique has its own qualities that always make it more favourite as compared to public key cryptography for WSNs. Furthermore to provide security in WSN, encryption keys must be established among sensor nodes. Key distribution refers to the distribution of multiple keys among the sensor nodes. Key management also receives a great deal of attention in data encryption and authentication in WSNs security.

Hence, it is necessary to well balance security level and the associated energy consumption overhead, to mitigate the security risks. Keys that are necessary for security and efficiency requirements of WSNs are listed in Table1.

In this paper, we present an Advanced Survey on Secure Energy-Efficient Hierarchical Routing Protocol in WSNs. Security issues are discussed and their solutions presented.

The rest of the paper is organized as follows. Section II describes the Energy constraints in WSNs while Section III presents a review of energy-efficient hierarchical cluster routing protocols. Security requirements in WSNs are presented in Section IV. Section V discusses various attacks that can be launched on routing protocols in WSNs.

Table 1: Design requirement of energy-efficient security scheme

S. No	***Requirement Type***	Requirements
1.	Security Requirement	Authentication Secrecy Integrity Resilience against node capture Resistance against node replication Compromised node revocation Fresh node addition
2.	Efficiency Requirement	Energy efficiency Network connectivity Maximum supported network size Minimum memory storage Low computational overhead Low communication overhead

Basic security mechanisms in WSNs are presented in section VI. In Section VII, secure hierarchical routing protocols in WSNs are discussed. Security analysis is presented in Section VIII and finally Section IX concludes our work.

2. Energy constraints in WSNs

The biggest constraint among the rest of the major constraints of a WSN is energy. In most cases the battery replacement is impossible. This means that the lifetime of a sensor depends greatly on the life of the battery. Fig. 1 illustrates sensor node architecture with four major components and associated energy cost parameters [2]. Basically, energy consumption in sensor nodes can be classified into following three parts, as shown in Fig. 2.

In WSNs, communication is more costly than computation [4] and many operations are energy intensive. It is for this reason that the current research focuses primarily on ways to reduce energy consumption.

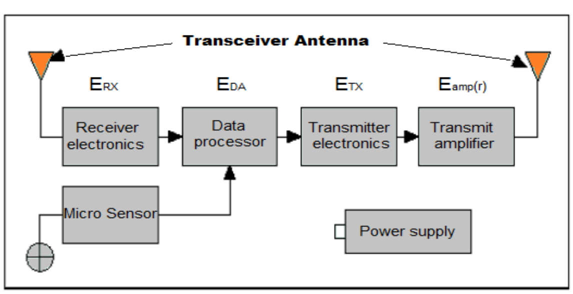


Fig. 1 Major components and associated energy cost parameters of a sensor node.

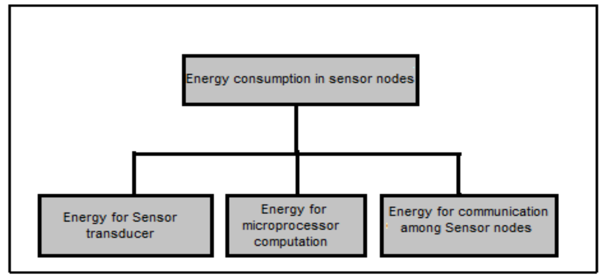


Fig. 2 Sensor nodes energy consumption.

In WSNs, the security mechanisms (e.g., encryption, decryption, signing data, verifying signatures) are the main factors that influence power consumption by sensor nodes. Hence, adding security to WSNs also impose overhead to power consumption, the energy required to store security parameters and the energy required to transmit the security parameters. Limited energy prohibits the use of complex security mechanisms for message expansion. Furthermore, in WSNs security mechanism usually use more energy consumption for higher security levels. Thus, WSNs could be classified into different security levels following energy cost [5].

3. Energy-Efficient Hierarchical routing Algorithms in WSNs

HWSNs is one of the main research areas in WSNs and behave the most energy-efficient among the rest of protocols for WSNs. Table 2 shows routing protocols classification in WSNs regarding to different categories.

Many research projects during the last few years have explored cluster based routing protocols in WSNs from different perspectives. Most of them have been proposed for routing the correct data to the BS and prolonging the life of sensors node. Hence, each protocol has advantages and disadvantages.

Table 2: Routing protocols classification in WSNs

Routing Protocols in WSNs	Network Structure	Flat network routing Hierarchical Location Based-routing
	Protocol Operation	Negotiation-based routing Multipath-based routing Query-based routing QoS-based routing Coherent-based routing

3.1 Low Energy Adaptive Clustering Hierarchy (LEACH).

LEACH [2], first energy-efficient hierarchical routing protocol, is proposed for WSN using homogenous stationary nodes. In LEACH, Sensors nodes choose their leader based on some parameters such as the strongest signal received from a CH. After certain interval, new nodes are selected as CH. LEACH reduces energy consumption by utilizing randomize rotation of CHs to evenly distribute the energy load in the network and turning off ordinary nodes when not required.

3.2 Power-Efficient Gathering in Sensor Information Systems (PEGASIS)

PEGASIS [6] is an extension of LEACH protocol. PEGASIS forms chains from sensors nodes. Sensors nodes transmit or receive data from a neighbor, in this way PEGASIS avoids cluster formation and uses only one node in a chain to transmit to the base-station. Therefore, increases the network lifetime.

3.3 Hybrid Energy-Efficient Distributed Clustering (HEED).

HEED [7] is an improvement of LEACH. HEED clustering randomly selects CHs and improves the lifetime of the network over LEACH clustering.

3.4 Energy-Efficient Homogeneous Clustering Algorithm (EEHCA) for Wireless Sensor Networks

In EEHCA [8], a new CH is selected based on the residual energy of existing CH, nearest hop distance of the node and holdback value. The uniform distribution of the cluster members extended the network lifetime.

4. Security Requirements in WSNs

To address the security issues in WSNs, we come across certain security requirements that must be addressed in WSNs environment. Here are some core security properties, implementation of which can contribute in development of more secure WSNs.

4.1 Authentication

It enables entities to cooperate within WSN without risk, by identifying and controlling participants in the network. It appears to be the cornerstone of a WSN. We cannot ensure confidentiality and the integrity of exchanged messages, if from the start we are not sure to communicate with the correct nodes. Therefore, it is essential for a receiver to have a mechanism to verify that the received packets have indeed come from the actual sender node. We can use Message Authentication Code (MAC) to ensure both the authentication of the origin of the message and integrity. An example of MAC is HMAC [9].

4.2 Data Integrity

It ensures that no message can be altered by an entity as it traverses from the sender to the recipient. It can be ensured by the use of cryptographic hash functions, which require obtaining a fingerprint for each digital message. MD5 function and Secure Hash Algorithm-1 (SHA-1) [10] are some examples of most used hash functions.

4.3 Data Confidentiality

Once the message parts are authenticated, confidentiality remains an important point. It is to keep the secrecy of exchanged messages. The confidentiality can be ensured by the use of cryptography keys (i.e. symmetric or asymmetric).

4.4 Availability

It ensures that the services of a WSN should be always available even in the presence of an internal or external attack. A central access control system is used to ensure successful delivery of every message to its recipient.

4.5 Data Freshness

This service ensures that the data is up-to-date and ensures that no adversary can replay old messages. Data freshness is important when the WSN nodes use shared keys for message communication. The risk is that a potential adversary can launch a replay attack using the old key, as the new key is being refreshed and propagated to all the nodes in the WSN. A nonce or time-specific counter may be added to each packet to check the freshness of the packet.

4.6 Self-organization

In a WSN, each node should be self-organizing. This requirement of WSN also poses a great challenge to security. The dynamic nature of a WSN makes it sometimes impossible to deploy any preinstalled shared key mechanism among the sensors nodes and the BS [11]. It is desirable that in WSNs, the nodes self-organize among themselves not only for multi-hop routing but also to carryout key management and developing trust relations.

5. Routing Attacks in WSNs

The network layer of WSNs suffers from different types of attacks such as: (i) Sybil, (ii) sinkhole, (iii) hello flood, (iv) wormhole, (v) selective packet forwarding, etc. These attacks are described briefly. Table 3 illustrates the routing attacks on WSNs and some solutions to defeat them.

5.1 Sybil Attack

The attacker presents multiple identities on one node in the network. In this way, the attacker mostly affects the routing mechanism. Generally Sybil attacks are prevented by validation techniques.

5.2 Sinkhole Attack

In this type of attack, attacker presents himself in the network with high capability resources, by which announces a short path to destination to attract packets and then may drop them [12]. In this way, sinkhole attack gives birth to some attacks like blackhole, selective forwarding, etc.

5.3 Hello Flood Attack

Strong hello message broadcasted by attacker with high transmission power is to be received by every node in the network [12]. Other nodes may think this message is nearest to them and sends packets by this node. In this way, attack congestion occurs in the network. Hello flood attacks are prevented using blocking techniques.

5.4 Wormhole Attack

An adversary launch wormhole with tunneling mechanism to establish him between entities by confusing the routing protocol. Using out-of-bound channel to route packets, makes this kind of attack very difficult to detect.

Table 3: Routing attacks on WSNs and countermeasures

Layer	***Attacks***	Solutions
Network	Spoofed routing information & selective forwarding	Egress filtering, authentication, monitoring
	Hello Flood	Authentication, packet leashes by using geographic and temporal info
	Wormhole	Authentication, probing
	Sinkhole	Redundancy checking
	Sybil	Authentication, monitoring, redundancy

5.5 Selective Forwarding

Generally two factors are important in this attack. The first is location of attacker as it will attract more traffic if the location of malicious node is close to base. The number of dropped messages is another factor, the more messages drops, the more energy it has in order to attack. An adversary can selectively forward some messages and drops others, therefore may compromise a node [12].

6. Basic Security Mechanisms in WSNs

Security in sensor networks poses different challenges than conventional network, due to inherent resources and computing constraints. However, secure communications in some WSNs are critical. Two security aspects such as the area of cryptography and key management received a great deal of attention in WSNs. Cryptography and key management mechanisms for WSN security are presented below.

6.1 Key Management

Key management is the process in which keys are created, stored, protected, transferred, used between authorized parties and destroyed when they do not need [13]. Key management establishes the keys that are necessary to provide confidentiality, integrity and authentication services. Due to the limited memory resources and energy constraints of sensor nodes, complex security algorithms cannot be used in sensor networks. The main goal of key management in WSNs is to ensure security requirements of WSN by encrypting messages and authenticates the communicating nodes. Key management is quite challenging issue in WSNs and researchers presented a large number of approaches in literature due to the importance of key management in WSNs. Some researchers have investigated the WSNs key management schemes and divided them into different categories.

From the work of Xiangqian and Makki [14], key management schemes in WSNs can be classified as following: key pre-distribution schemes, hybrid cryptography schemes, one-way hash schemes and key management in hierarchy networks.

– Key Pre-Distribution Schemes: refers to how many keys are needed and how should the keys be distributed before the nodes are deployed? Key pre-distribution schemes can be classified as probabilistic schemes and deterministic schemes.

In probabilistic scheme, the existence of a shared key between a particular pair of nodes is not certain and instead guaranteed only probabilistically. The basic idea of these schemes is to randomly preload each sensor with a subset of keys from a global key pool before deployment. So, these schemes can also be called Random Key Pre-distribution (RKP). The first probabilistic key pre-deployment scheme is introduced by Eschenauer and Gligor [11], which consists of three phases: key pre-distribution, shared-key discovery, and path-key establishment.

Contrary to probabilistic schemes, deterministic schemes guarantee that any two intermediate nodes can share one or more pre-distribution keys.

LEAP [15] (Localized Encryption and Authentication Protocol), is a basic example of deterministic key management scheme. The authors of LEAP establish four types of keys that must be stored in each sensor to ease the overhead of key management and to provide secure communications in WSNs.

– Hybrid Cryptography Schemes: use both asymmetric-key and symmetric-key cryptographs.

– One-way Hash Schemes: is used in many approaches that come from one-way hash function technique to ease key management.

– Key Management in Hierarchy Networks: many key management approaches are based on a normal flat structure. There are still some approaches that utilize a hierarchical structure in order to ease the difficulties by balancing the traffic among a BS, CHs, and sensors. These are the three parts of networks that have different resources.

Zhang and Varadharajan [1] also considered three important factors for classification of key management schemes in WSNs based on the encryption techniques. These include symmetric, asymmetric and hybrid. Based on the key establishment mechanism, Zhang and Varadharajan [1] divided the symmetric and asymmetric schemes into eight and three subcategories respectively. This classification is shown in Fig. 3.

– Most of the WSNs use the symmetric key schemes because these schemes requires less computation time than other schemes. Based on the key distribution, key discovery and key establishment in the schemes, symmetric schemes can be divided into six categories: entity based schemes, pure probabilistic-based schemes, polynomial-based key pre-distribution schemes, matrix-based key pre-distribution schemes, tree-based key

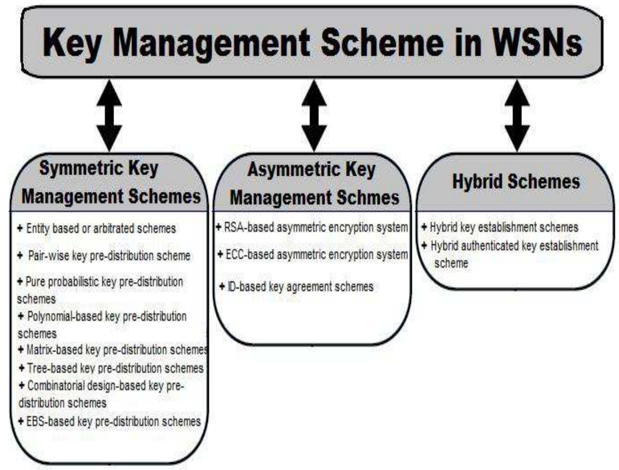


Fig. 3 Key management scheme in WSNs

pre-distribution schemes and EBS-based key pre-distribution schemes.

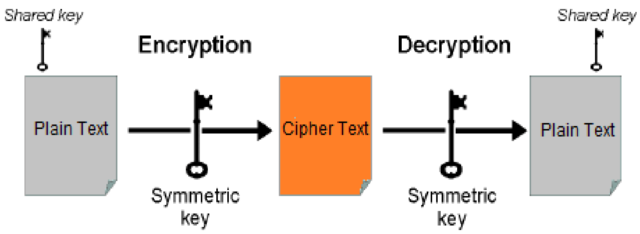
– In asymmetric key management schemes, RSA and Elliptic Curve Cryptography (ECC) are two major public key techniques. Public key technology is widely used in the security of Internet. On the other hand, some researchers believe that these techniques are too heavy-weight for sensor networks because of requirement constraints.

– In Hybrid schemes, several research groups [16] proposed the hybrid key establishment schemes for WSNs. The motivation is to exploit the difference among the BS, the CH and the sensor, and place the cryptographic burden on the BS or the sensors whose sources are less constrained. Sensors have limited computational power and energy resources, whereas BS has much more computational power and other resources. The hybrid key establishment schemes reduce the high computational cost on the sensors by placing them on the BS side.

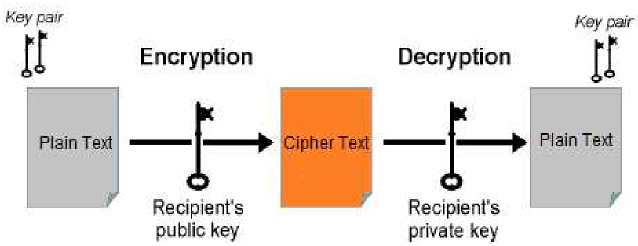
6.2 Cryptographic Mechanisms

There are two types of cryptography techniques depending on the key. First is symmetric key cryptography that uses the same key for encryption and decryption (e.g., AES). Second is asymmetric key cryptography that uses different keys to encrypt and decrypt (e.g., RSA), requires more computation resources than symmetric key cryptography. Symmetric and Asymmetric encryption are illustrated in Fig. 4 (a) and (b) respectively. Due to the limited resources, public key cryptographic algorithms are not suitable for WSNs. However, the symmetric cryptographic technique has its own qualities that make it more favourable as compared to public key cryptography for WSNs.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org



(a) Symmetric Encryption



(b) Asymmetric Encryption

Fig. 4 Types of cryptography

For secure communication in WSNs, it is necessary to choose the most efficient cryptographic algorithm. A benchmark on cryptographic algorithms was presented by Law et al. [17] for WSNs. They focus on storage, energy-efficiency and security properties of Skipjack, RC5, RC6, and Rijndael.

In [17], the result of analysis show that in an environment where security is important, memory efficient cryptographic algorithm is required. And in an environment where availability of network is important, energy-efficient cryptographic algorithm has to be used. TinySec [18] provides Skipjack and RC5 as recommended cryptographic algorithm. Each algorithm has its own property, memory, energy efficiency and security. Table 4 summarized the execution time of ECC, RSA, RC5 and Skipjack implemented on an Atmega128 processor.

Table 4: Execution time of some cryptographic algorithms

S. No	***Algorithms***	Operation Time
1.	RC5	0.38 ms
2.	Skipjack	0.26 ms
3.	ECC-160	810 ms
4.	RSA-1024	10990ms

7. Secure Hierarchical Routing Protocols in WSNs

Security in cluster based routing protocols is a particularly challenging task. Many works have been proposed to secure the hierarchical routing protocols. In this section, some approaches are reviewed and compared. Their advantages and disadvantages are also discussed.

RLEACH [19] is a secure routing protocol for cluster-based WSNs, using group key management, was proposed by Zhang et al. to solve the problem of secure LEACH. In this protocol, clusters are formed dynamically and periodically, can be thought as security extension of LEACH. RLEACH uses improve random pair-wise key management scheme (RPK), which use the one-way hash chain, symmetric and asymmetric cryptography to ensure security in LEACH. RLEACH resists to different attacks such as selective forwarding, sinkhole attacks, sybil attacks, and hello flood attack. Another advantage of RLEACH is that it is energy-efficient. In case, when a node transmit data to its CH, the member nodes among a cluster can close their wireless devices during the schedule creation phase or can sleep during the data transmission phase to save energy. Therefore, RLEACH balance the network security and the energy consumption in cluster-based WSNs.

EECBKM [20] Energy-Efficient Cluster Based Key Management is a cluster based technique for key management in WSNs. In this protocol, initially the clusters are formed in the network and the CHs are selected based on the energy cost, coverage and processing capacity. An EBS key set is assigned by the sink to every CH and cluster key to every cluster. The EBS key set contains the pair-wise keys for intra-cluster and inter-cluster communication. The data is made to pass through two phases of encryption during data transmission towards the sink. In this way security is ensured in the network. These keys are distributed to the nodes by the CH prior to communication. Secure channel is established between the nodes and the CH after the key distribution. Results have shown that this proposed technique reduces node-capture attacks and efficiently increases packet delivery ratio with reduced energy consumption.

SHEER [21] is a secure hierarchical energy-efficient routing protocol proposed by Ibriq and Mahgoub, which provides energy-efficient and secure communication on the network layer. For key distribution and authentication, securing the routing mechanism, SHEER uses HIKES (Hierarchical Key Establishment System) and also implements a probabilistic transmission mechanism to improve the network energy performance and lifetime.

SHEER defends the network against hello flood attack, sybil attack. The sinkhole attack will also fail because the attacker does not possess all keys, required for authentication. SHEER fail to protect the network from selective forwarding attacks.

SecLEACH [22] is a protocol for securing node-to-node communication in LEACH-based networks. Using random key pre-distribution, SecLEACH provides security in LEACH, introduced symmetric key and one-way hash chain to provide different performance numbers on efficiency and security depending on its various parameter values.

SecLEACH is an improvement of SLEACH [23], the first study in homogeneous WSNs focused on adding security to cluster-based communication protocols with resource constrained sensor nodes. SecLEACH provides authenticity, confidentiality, integrity and freshness for node-to-node communication. Otherwise SecLEACH is vulnerable to key collision attacks and do not provide full connectivity. The overheads in SecLEACH were computed using the number of CH value in the network which decrease the total energy consumption, and prolong the network's lifetime.

SS-LEACH [24] is a protocol based on LEACH protocol, considering routing security and network lifetime. Improving the method for electing CHs, the SS-LEACH protocol forms dynamic stochastic multipath CHs chains using nodes self-location technology and key pre-distribution strategy. So the SS-LEACH protocol strongly improves the energy-efficiency and hence prolongs the lifetime of the network. The SS-LEACH protocol can prevent compromised node and preserve the secrecy of the packet. It also can avoid sybil attack, selective forwarding and hello flooding.

NSKM [25], a Novel Secure Key Management module for Hierarchical Clustering WSNs provides an efficient scalable post-distribution key establishment that allows the hierarchical clustering topology platform to provide acceptable security services. In NSKM, there are three categories of keys; pre-deployed keys, network generated keys and the BS broadcasted keys. This module is the first implemented security module for WSNs that provides reasonable resistance against replay and node capture attacks. This work couples hierarchical clustering based routing with NSKM module. The selection of SCH among CHs is based on its location and its distance to BS. Most of communication types in WSNs have unique features of this work, using in-network keys generation and blending. The NSKM module is energy-efficient, has strong flexibility against susceptible attacks on WSNs, keeping the resource starved nature of sensor nodes. NSKM also ensures that the whole network is never compromised even if there has been an attack in the network. Furthermore, it is highly lightweight and scalable and is acceptable to be used in large WSNs.

AKM [26] is an Authenticated Key Management scheme for hierarchical networks based on the random key pre-distribution. Security is provided by using two kinds of keys, a pair-wise shared key between nodes, and a network key. To divide nodes into clusters, AKM scheme use an existing ring structure energy-efficient clustering architecture (RECA). Using more than one encryption key, AKM provides multiple level of encryption, secure cluster formation algorithm and avoid node captures. AKM provides confidentiality, global and continuous authentication of nodes in the network by periodically refreshing the network key. In general AKM scheme can be applied for different energy-efficient data dissemination techniques for sensors networks. However, if adversary re-enters the compromised node into the network before refreshing the current network key, the resiliency of AMK scheme will be same as given in Eschenauer et al. [11].

SRPSN [27] is a Secure Routing Protocol for Sensor Networks consists of a hierarchical network with CHs and cluster member nodes. CHs route the messages from sensor nodes. A preloaded symmetric key is shared between all CHs and the BS to protect data. SRPSN is also designed to safeguard the data packet transmission on the sensor networks under different types of attacks. A group key management scheme is proposed, which contains group communication policies, group membership requirements and an algorithm for generating a distributed group key for secure communication. Every sensor node contributes its partial key for computing the group key. One drawback associated with this protocol is that there is no authentication in the mechanism. Hence, SRPSN fail to protect against attacks like spoofing, altering, replaying. If the adversary uses the sybil attack, the problem will be more severe. The malicious node can also become a sinkhole. Another problem of this scheme is that children nodes will select a largest NBR node to relay data. However, energy consumption will be increased in this case.

SecRout [28], a Secure Routing Protocol for sensor networks is proposed by Yin and Madria to provide security against attack from compromised nodes in sensor networks. SecRout can detect if packets are dropped or modified by malicious nodes. In the SecRout protocol, only high efficient symmetric cryptography is used to secure messages, and the partial routing path is recorded in sensor nodes memory. Further, SecRout uses two types of keys: the master shared key used between the sink and CHs, and the cluster key among the clusters to encrypt the message. In SecRout all messages will be verified through MAC. It ensures that the messages received are not tampered, hence guarantees freshness. In SecRout, two-level architecture can greatly lower the message overhead. Therefore, SecRout can greatly save the energy, and decrease the usage of memory and bandwidth.

IKDM [29] is an Improved Key Distribution Mechanism, based on hierarchical network architecture and bivariate polynomial-key pre-distribution mechanism. In IKDM, each sensor has a unique id in the network. An offline Key Distribution Server (KDS) first initializes sensors before deployment by giving each sensor node a polynomial share. In order to setup a pair-wise key between two sensor nodes, they exchange their node ids first, and then nodes evaluate their stored polynomial. Since, sensors nodes can obtain the same value from the two distinct calculations, which can be used as their pair-wise communication key. Note that in IKDM, two communicating parties can establish a unique pair-wise key between them. IKDM scheme can achieve better network resilience against node capture attack, hence can provide efficient security and is not affected by the number of compromised sensors. IKDM scheme provides better scalability, network throughput, fixed key storage overhead, full network connectivity and is suitable for large-scale WSNs. Therefore IKDM scheme is more energy-efficient due to the lower communication overhead for sensor nodes during the pair-wise key establishment process.

8. Security Analysis

We describe some secure hierarchical routing protocols selected based on security mechanisms, security requirements, various routing attacks and performance metrics.

Security requirements for several routing protocols are summarized in Table 5. We observe that SecLEACH, SHEER, EECBKM, AKM, IKDM and NSKM address all the listed security requirements (authenticity, confidentiality, freshness and integrity) thus they are more secure than rest of the protocols if the security requirements is taken as criteria. According to the security requirements, selected protocols classification show that authentication and integrity are the most satisfied.

An overview of routing attacks in WSNs is shown in Table 6. From the table, it is clear that certain schemes defeats or mitigate the effect of various routing attacks. Considering the resistance against the routing attacks, Table 6 shows

Table 5: Security requirements for secure hierarchical routing protocols

Secure Hierarchical Routing Protocol	***Security Requirements***			
	Authenticity	*Confid-entiality*	*Integrity*	*Freshness*
RLEACH	Yes		Yes	
EECBKM	Yes	Yes	Yes	Yes
SHEER	Yes	Yes	Yes	Yes
SLEACH	Yes		Yes	
SecLEACH	Yes	Yes	Yes	Yes
SS-LEACH	Yes	Yes		
NSKM	Yes	Yes	Yes	Yes
AKM	Yes	Yes	Yes	Yes
SRPSN	Yes	Yes	Yes	
SecRout	Yes		Yes	
IKDM	Yes	Yes	Yes	Yes

'Yes' means that protocol can achieve that security requirement.

Table 6: Resistance of routing attacks for secure hierarchical routing protocols

Secure Hierarchical Routing Protocol	***Routing Attacks in WSNs***					
	Selective Forwarding	*Sink-hole*	*Worm-hole*	*Sybil*	*Hello Flood*	*Node Capture*
RLEACH	Yes	M		Yes	Yes	
EECBKM		M		Yes	Yes	Yes
SHEER	Yes	M		Yes	Yes	
SLEACH	Yes	M			Yes	
SecLEACH	Yes			Yes	Yes	
SS-LEACH	Yes			Yes	Yes	M
NSKM	Yes	M	M	Yes	Yes	Yes
AKM	Yes	Yes	M	Yes	Yes	Yes
SRPSN	Yes					
SecRout	Yes	M		M		Yes
IKDM	Yes	M	M	Yes	Yes	Yes

'Yes' means that protocol defeats the attack and 'M' means that protocol mitigates the effect of attack based on our pre-evaluation.

that RLEACH, NSKM, EECBKM, AKM, SecRout, and IKDM are more resistant to routing attacks than rest of the secure protocols.

The detailed comparison results are summarized in Table 7. We observe that energy efficiency depends strictly on the communication overhead. Therefore, all schemes with lower communication overhead achieve energy efficiency (e.g. IKDM, SHEER, SS-LEACH, etc.). This is due to the fact that communications consume much more energies than the code execution or computation in WSNs. We also remark that approaches based on probabilistic key distribution (e.g. SLEACH, Sec-LEACH, etc.) are less

Table 7: Comparison summary based on security mechanisms, performance and efficiency of some selected secure hierarchical routing protocols implemented for WSNs.

Protocol Name	***A Comparative Overview Of Representative Secure Hierarchical Routing Protocols For WSNs***									
	Ref	*Cryptography Scheme*	*Key distribution and Management Scheme*	*Authentication Scheme*	*Storage Load*	*Comm. Load*	*Scalab-ility*	*Robust-ness*	*Connectivity*	*Energy Efficiency*
RLEACH	[19]	Symmetric key cryptography	Improved Random pair-wise key management (IRPK)	Authentication is achieved via IRPK	High	Medium	Good	Good	Probabilistic	Medium
EECBKM	[20]		EBS-based key Management schemes	Via Key Management	Low	Low	Medium	Good	100%	Good
SHEER	[21]	Symmetric key cryptography	Hierarchical key management and authentication scheme	Authentication is achieved via HIKES	Medium	Low	Good	Good	100%	Good
SLEACH	[23]	Symmetric key cryptography		MAC	High	Medium	Medium	Limited	Probabilistic	Medium
Sec-LEACH	[22]	Symmetric key cryptography	Random key pre-distribution scheme	Don't provide broadcasts authentication	High	Medium	Medium	Limited	Probabilistic	Medium
SS-LEACH	[24]	Symmetric key cryptography	Keys pre-distribution strategy		Medium	Low	Medium	Limited	100%	Good
NSKM	[25]		Key management schemes based	MAC	Low	Low	Good	Good	100%	Good
AKM	[26]		Random Pre-distribution Key Management	Via Key Management and MAC	High	Medium	Good	Good	Probabilistic	Medium
SRPSN	[27]	Symmetric key cryptography	Group key management scheme	MAC	Medium	Low	Medium	Low	100%	Good
SecRout	[28]	Symmetric key cryptography	The Scheme introduced in LEAP [15]	MAC	Low	Low	Good	Limited	100%	Good
IKDM	[29]		Bivariate polynomial-key pre-distribution mechanism	Via polynomial key pre-distribution mechanism	Low	Low	High	Good	100%	Good

energy-efficient than other schemes (e.g. IKDM based on deterministic approaches). Due to the fact that approaches based on probabilistic key distribution generate a lot of messages, require much more memory space. In contrast, the deterministic key distribution requires more computation time for nodes. Note that in WSNs, computation consumes less energy compared to the exchange of messages between sensor nodes. Considering the scalability, Table 7 also shows that probabilistic approaches are less scalable than other schemes.

There exist some surveys on secure hierarchical WSNs [30]. However none of them address the energy consumption constraints following the security mechanisms.

Based on hierarchical topology, Sharma and Jena [30] consider that all selected works are energy efficiency. However, they do not pay much attention on energy constraints when different security mechanisms are used. This is very crucial because technique based probabilistic and deterministic don't have the same impact on energy consumption. In addition, Sharma and Jena [30] did not address the performance requirements study (e.g. memory overhead, computation overhead etc.), which is more important because it is strictly bound to the energy consumed.

In the paper, we presented an overview of well-known routing protocols for WSNs and a technical overview of each protocol. We also provide a comprehensive and informative comparison of them, which we believe is a significant improvement, when compared to other comparative studies in WSNs.

4. Conclusions and Future Research

The main goal of a routing protocol design is to provide energy efficiency and extend network lifetime. Sensor nodes are susceptible to a number of routing attacks depending on the nature of the WSNs, the limited memory resources and energy constraints. In order to provide security in WSNs and mitigate the security threats to routing protocols, secure routing protocols to be used. In this paper, we reviewed and analysed some secure cluster-based routing protocols. The comparative study show that some selected schemes can well balance between security level and the associated energy consumption overhead. An informative overview of protocols is given and their advantages and disadvantages listed. We also presented detailed comparison based upon various criteria in the analysis section. Further, research would be needed to address issues related to secure routing under the mobility for resource constrained WSN. The study may help to orient the development of future proposals well adapted in the area of security issues in routing protocols for WSNs.

Acknowledgments

This work was supported by the High-tech Research and Development Program of China (Grant No.2011AA040101-3).

References

[1] J. Zhang, V. Varadharajan, “Wireless sensor network key management survey and taxonomy,” Journal of Network and Computer Applications, Vol. 33, No. 2, 2010, pp. 63-75.
[2] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, “Energy-efficient communication protocol for WSNs,” In Proc. of the 33rd Hawaii International Conference on System Sciences, 2000.
[3] A. Modirkhazeni, N. Ithnin and O. Ibrahim, “Empirical Study on Secure Routing Protocols in Wireless Sensor Networks,” International Journal of Advancements in Computing Technology, Vol. 2, No. 5, 2010, pp. 25-41.
[4] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. E. Culler, and K. Pister, “System architecture directions for networked sensors”, In Proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems, 2000, pp. 93-104.
[5] L. Yuan and G. Qu, “Design space exploration for energy efficient secure sensor networks”, In Proceedings of IEEE International Conference on Application-Specific Systems, Architectures, and Processors, 2002, pp. 88-100.
[6] S. Lindsey and C. S. Raghavendra, “Pegasis: Power-efficient gathering in sensor information systems”, In IEEE Aerospace Conference Proceedings, 2002, pp. 1125–1130.
[7] O. Younis and S. Fahmy, “Heed: A Hybrid, energy, Distributed Clustering Approach for Ad-hoc Networks,” IEEE Transactions on Mobile Computing Issues in Wireless Networks and Mobile Computing, Vol. 3, No. 4, 2004, pp. 366-369.
[8] S. K. Singh, M. P. Singh and D. K. Singh, “Energy– efficient Homogeneous Clustering Algorithm for Wireless Sensor Network,” International Journal of Wireless & Mobile Networks (IJWMN), Vol. 2.No. 3, 2010, pp. 49-61.
[9] H. Krawczyk and M. Bellare and R. Canetti, “HMAC: Keyed-Hashing for Message Authentication,” RFC 2104, 1997.
[10] D. Eastlake and P. Jones, "US Secure Hash Algorithm 1 (SHA1)", RFC 3174, 2001.
[11] L. Eschenauer and V. D. Gligor, “A key management scheme for distributed sensor networks,” In Proc. of the 9th ACM conference on Computer and communications security (CCS’02), 2002, pp. 41-47.
[12] C. Karlof and D. Wagner, “Secure routing in wireless sensor networks: Attacks and countermeasures”, In Proceedings of First IEEE International Workshop on Sensor Network Protocols and Applications, 2003, pp. 113-127.
[13] A. Menezes, P. van Oorschot, and S. Vanstone. Handbook of Applied Cryptography. CRC Press, 1996.
[14] Xiangqian Chen, Kia Makki, Kang Yen, and Niki Pissinou, “Sensor network security: a survey,” IEEE Communications Surveys Tutorials, Vol. 11, No. 2, 2009, pp. 52–73.
[15] S. Zhu, S. Setia, and S. Jajodia, “Leap: efficient security mechanisms for large-scale distributed sensor networks,” In Proc. ACM CCS, 2003, pp. 62–72.
[16] J. Zhang, V. Varadharajan, “Group-based Wireless Sensor Network Security Scheme,” In The fourth international conference on wireless and mobile communications (ICWMC 2008), 2008.
[17] Y. W. Law, J. M. Doumen, and P. H. Hartel, “Benchmarking block ciphers for wireless sensor networks (extended

abstract)," In 1st IEEE Int. Conf. on Mobile Adhoc and Sensor Systems (MASS), 2004.

[18] A. C. Karlof, N. Sastry, and D. Wagner, "TinySec: a link layer security architecture for wireless sensor networks," In Proceedings of the 2nd international conference on Embedded networked sensor systems (SenSys 04), 2004, pp. 162.175.

[19] K. Zhang, C. Wang, and C. Wang, "A secure routing protocol for cluster-based wireless sensor networks using group key management," In Proc. 4th IEEE International conference on Wireless Communications, Networking and Mobile Computing (WiCOM'08), 2008, pp. 1–5.

[20] T. Lalitha and R. Umarani, "Energy efficient Cluster Based Key Management Technique for Wireless Sensor Network," International Journal of Advances in Engineering & Technology (IJAET), Vol. 3 No. 1, 2012, pp. 186-190.

[21] J. Ibriq and I. Mahgoub, "A secure hierarchical routing protocol for wireless sensor networks," In: Proc. 10th IEEE International Conference on Communication Systems, 2006, pp. 1-6.

[22] L. B. Oliveira, A. Ferreira, M. A. Vilaca, H. C. Wong, M. Bern, R. Dahab, and A. A. F. Loureiro, "Secleach-on the security of clustered sensor networks," Signal Processing, Vol. 87, No. 12, 2007, pp. 2882–2895.

[23] A. C. Ferreira, M. A. Vilaca, L. B. Oliveira, E. Habib, H. C. Wong, and A. A. Loureiro, "On the security of cluster-based communication protocols for wireless sensor networks," In Proc. 4th IEEE International Conference on Networking (ICNŠ05), 2005, pp. 449–458.

[24] D. Wu, G. Hu, and G. Ni, "Research and improve on secure routing protocols in wireless sensor networks," In 4th IEEE International Conference on Circuits and Systems for Communications (ICCSC 2008), 2008, pp. 853–856.

[25] I. S. Gawdan, C. O. Chow, T. A. Zia, Q. I. Sarhan, "A Novel Secure Key Management for Hierarchical Wireless Sensor Networks," In Proceeding of 2011 Third Conference on Computational Intelligence, Modeling and Simulation (CIMSiM), 2011 , pp. 312 - 316.

[26] F. Kausar, A. Masood and S. Hussain. "An Authenticated Key Management Scheme for Hierarchical Wireless Sensor Networks," In Advances in Communication Systems and Electrical Engineering, Lecture Notes in Electrical Engineering, Vol. 4, 2008, pp. 85-98.

[27] M. Tubaishat, J. Yin, B. Panja, and S. Madria, "A secure hierarchical model for sensor network," ACM SIGMOD Record, Vol. 33, No. 1, 2004, pp. 7–13.

[28] J. Yin and S. Madria, "SecRout: A Secure Routing Protocol for Sensor Network," IEEE 20th International Conference on Advanced information networking and applications, Vol. 1, 2006.

[29] Y. Cheng and D. Agrawal, "An improved key distribution mechanism for large-scale hierarchical wireless sensor networks," Ad Hoc Networks (Elsevier), Vol. 5, No. 1, 2007, pp. 35–48.

[30] S. Sharma and S. K. Jena, "A Survey on Secure Hierarchical Routing Protocols in Wireless Sensor Networks," In Proceedings of International Conference on Communication, Computing (ICCCS'11), 2011, pp. 146-151.

Intelligent Modeling and Decision Making for Product Quality of Manufacturing System Based on Fuzzy Cognitive Map

JIHONG PANG

College of Mechanical and Electronic Engineering, Wenzhou University
Wenzhou 325035, Zhejiang, China

Abstract

Recent research finds that consumers pay more and more attention to the high grade product. An intelligent decision making system is proposed in this paper, the purpose of which is to monitor product quality of manufacturing system and give warnings to the quality managers accordingly. Since the complex interaction among the multivariate quality characteristic (QC) and the intelligent model is hard to be manually managed, the well-performing machine learning algorithms need to be proposed to support an automatic control of product quality. And many decision making tools have been developed to monitor product quality with the intelligent control model. Fuzzy cognitive map (FCM) is a popular tool for intelligent modeling, and is usually applied to many fields. In this paper, FCM is used for intelligent modeling in the decision making process of manufacturing system. Based on the mechanism, we design an intelligent model to solve the decision making problem of product quality control. Then, the effect of varying the support scale of trained data is tested by using FCM in different environments. Finally, we get some useful conclusions and discuss several potential research developments in the future.

Keywords: *Product Quality, Intelligent Modeling, Decision Making, Manufacturing System, Fuzzy Cognitive Map*

1. Introduction

In recent decades, the intelligent modeling and decision making method has attracted considerable attention from different fields of scientific research and technological experimentation. The intelligent modeling approaches have become increasingly available for use in product quality decision making related research [1, 2]. On the other hand, as the global market becomes more and more competitive, consumers become more discriminating product quality [3]. So the fierce market competition has certainly made it easier for consumers to search for a high quality product [4, 5]. In fact, consumers often make a decision on product quality based on the QCs that the products have [6]. In order to improve product quality, a specific attention has to be paid to modeling intelligent system [7].

In a word, the intelligent modeling is one of the most important factors which can influence product quality [8]. And the conventional wisdom is that intelligent modeling has been a boon to decision makers. Furthermore, the intelligent modeling is a powerful technology and complete facility to support decision making process [9]. In more recent years, a lot of intelligent systems based on different machine learning models have been intensively applied in many areas [10]. With the development of the internet and computer technology, intelligent modeling techniques for their utilization to decision making problems have been investigated [11, 12]. Due to the rapid growth of computer network and intelligence technologies, the product quality is increasingly being controlled in decision making system [13]. Besides, the product quality monitoring of manufacturing system can provide a powerful tool for advanced warning for quality supervisors in decision making process [14].

Quality managers often rely on prior knowledge during decision making tasks due to information processing limitations. While saving time and costs have been mentioned as the primary drivers of decision making, saving time may be more important than saving costs for most quality managers of enterprises [15]. So the importance of product quality can be related through an artificial intelligence system. The field of intelligent system has grown significantly and is believed to be capable of monitoring product quality in modern manufacturing system [16]. So quality managers looking for the best product fit take longer to make a decision because they spend more time searching for scattered product information [17]. Once quality managers begin to focus more on benefits and less on costs, the potential for making better quality decisions of intelligent system can be realized.

The decision making process discussed in this paper strives to develop an intelligent model that can be used to monitor and control the product quality of smart manufacturing system. This study is divided into the following subsections. Firstly, the modeling framework of intelligent decision making system is created based on FCM. Second, we will probe deeper into the modeling process of product quality control, and discuss relevant applications that have been found in the open literatures.

Finally, we will conclude this paper with a summary of the advances of product quality intelligent modeling that is expected to be realized based on FCM.

2. Description of Intelligent Modeling System

In fact, the implementation of high performance quality processing of manufacturing system is a challenging task. A lot of affective factors may also influenced decision making in intelligent system, because it touched on the human's subjectivity and experience. In the next section, intelligent modeling is identified in this paper based on FCM.

2.1 Modeling Framework of Intelligent Decision Making System

The main goal of the modeling framework is that decision efficiency is likely to improve when quality managers focus both on cost reduction and quality improvement. Then, the previous section was firstly used to introduce to the engineering applications customers, the FCM method and the various kinds of decision-making tools that are used to build the intelligent model. Several efforts have tried to overcome the limitations of the previous research approach by using intelligent decision making system.

On the other hand, the databases are created for data storage and model validation. The databases should have a sufficient size and include the maximum combination of inputs and outputs with types of information. In this approach, a series of raw data information is used as inputs into the database. Traditionally, intelligent manufacturing system employed for product quality decision making from other easily available data has used multivariate QC as inputs. The decision making process is based on fuzzy logic and reinforcement learning, which allows the intelligent system to deal with the uncertainty and ambiguity of the data input in different settings.

Moreover, the main objective of the intelligent system is the aggregation of individual judgments into a global value function. The ability to detect abrupt changes in manufacturing system is one of the most important functions in intelligent quality controlling [18]. However, the differences between traditional decision and intelligent decision making can be attributed to the technology that is available to the decision maker. Different types of modeling data input can be provided to the fuzzy inference engine for the decision making process. In this framework, the intelligent modeling system has been effectively developed.

It is easy to implement with few parameters in the process of construction of growth intelligent models. However, the complexity and difficulty of decision-making process can become cumbersome as the number of parameters grows. In this study, the modeling framework can be expressed with an analytical expression in Fig. 1.

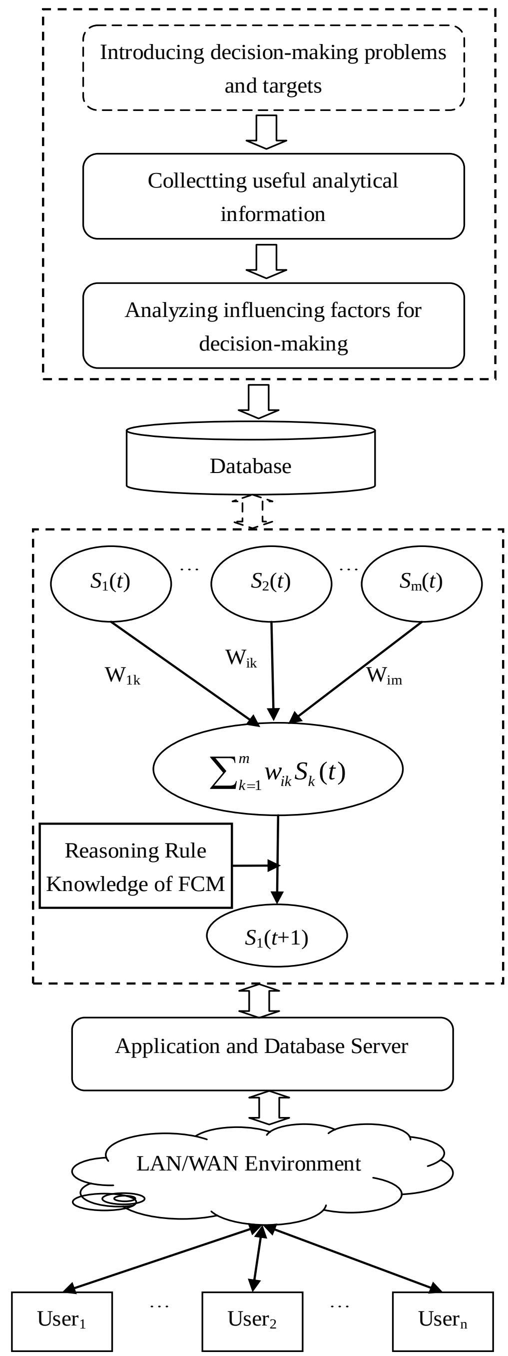


Fig. 1 Modeling framework of intelligent decision making system

2.2 Intelligent Modeling Based on FCM

The goal of this approach is to aid the intelligent reasoning technology to improve the knowledge about the decision situation. We confirm that the FCM method for intelligent modeling is precise and applicable to decision making of manufacturing system. In recent years, FCM has become a hot spot of research [19]. The analytical expression of FCM has been successfully implemented in lots of manufacturing systems [20, 21]. In this paper, We applied our policies and methodologies in a real example based on FCM.

However, the prior researches have only solved the traditional decision making problems. In the traditional decision-making process, the criteria aggregation model is known a priori, while the total preference is unknown. The approaches that have been identified in the literature both involve the use of FCM and can be summarized as follows. FCM has been a preferred tool for intelligent modeling, and many researchers are now paying more and more attention to decision-making logic with it [22]. It has been shown by previous studies that FCM method is being successfully used in decision process [23]. This is due to the large scale regulations encompassed in reasoning process with FCM [24, 25].

Besides, FCM is usually designed to express the correlation and influence between different nodes by obtaining arrange of weights. Many such combinations of input and layer weights can be determined by the theory of fuzzy mathematics [26]. And the weights of FCM are generally scalars, but the causal knowledge can be implemented with a fuzzy number [27]. More detailed discussions about FCM are made the following arrangements [28].

The topological structure standard of FCM is described in the following:

$$U = (S, E, W) \tag{1}$$

Where, $S=\{S_1, S_2, \ldots\ldots, S_n\}$ denotes the Concept node of FCM; $E=\{<C_a, C_b>; C_a, C_b \in C\}$ denotes the causal relationship between nodes to arc; $W=\{W_{11}, W_{12}, \ldots\ldots, W_{nn}\}$ denotes the data correlation between nodes or influence. For instance, we can confirm the weight question with fuzzy numbers in the case of the global value function.

On the basis of above work, the uncertainty inference model based on FCM is constructed. The inference model and data structure of intelligent modeling system is discussed below. In this paper, the inference model of FCM method is described in Fig. 2.

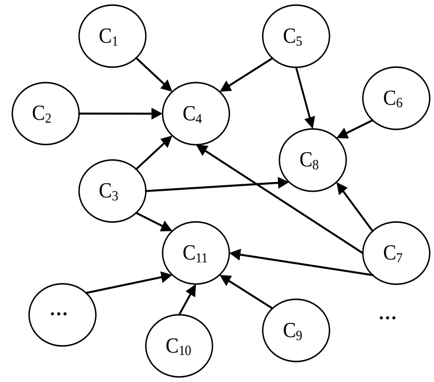


Fig. 2 The inference model of FCM

After all nodes were currently defined based on FCM theories and methods, the state matrix can be represented as shown below:

$$\alpha_U(t) = (S_1(t), S_2(t), \cdots, S_n(t)) \tag{2}$$

With the incidence matrix to express network topology, we can obtain the adjacency matrix from structure type synthesis in the following:

$$W_U(t) = \begin{bmatrix} w_{1,1} & w_{1,1} & \vdots & w_{1,m} \\ w_{2,1} & w_{2,2} & \vdots & w_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,1} & \vdots & w_{m,m} \end{bmatrix} \tag{3}$$

In order to improve the veracity of modeling, a comprehensive evaluation method based on fuzzy synthesis judgment was presented in this paper. Based on fuzzy number definition and properties of principle, the decision of the manufacturing system was made available in the product quality control. This can take a value in [0, 1]. And the qualitative judgments of subject degree and weight are shown in Fig. 3.

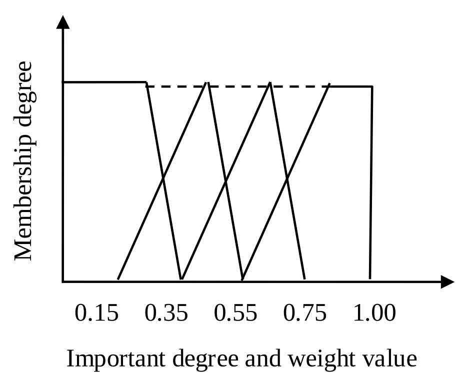


Fig. 3 The value of subject degree and weight

In order to reason the node in next state, we can apply the inference rules of FCM to make useful inferences about different message sequences in the following formula:

$$S_j(t+1) = f(\sum_{i=1}^{m} S_j(t) w_{ij})$$
$$= f(\begin{bmatrix} s_1(t) \\ s_1(t) \\ \vdots \\ s_m(t) \end{bmatrix} \begin{bmatrix} w_{1,1} & w_{1,1} & \vdots & w_{1,m} \\ w_{2,1} & w_{2,2} & \vdots & w_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,1} & \vdots & w_{m,m} \end{bmatrix}) \tag{4}$$

In addition, FCM refers to a system modeling approach that is patterned after a comprehensive understanding of the fuzzy inference mechanism.

3. An Intelligent Decision Making System For Product Quality

The goal of this study is to implement an intelligent modeling system for product quality monitoring based on FCM method. So this paper presents the final results of a feasibility study on intelligent modeling for product quality monitoring in smart manufacturing system. Generally speaking, intelligent modeling capabilities have already been incorporated in a modern manufacturing system for experimental purposes. On the other hand, the hybrid algorithm developed has been applied on an original database consisting of quality control processes with multivariate QCs in the smart manufacturing system.

In order to describe the decision making of in quality control in a manufacturing system, we introduced the mechanism for managing the quality assurance process. Here, we analyze experimental results on the accuracy decision of FCM algorithm. The fuzzy cognitive map of multivariate QCs is shown in Fig. 4.

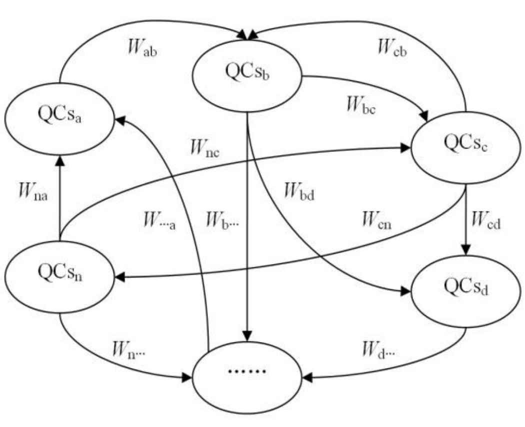


Fig. 4 The fuzzy cognitive map of multivariate QC

To assess the performance of the FCM algorithm, the cross-validation strategy has been used as a validation method. A brief description of intelligent modeling is given below for completeness in an example. The feasibility study, presented in this paper, is described below.

This paper mainly applies the means of decision analysis of product quality control in manufacturing system and chooses the case examples to carry on analyzing. Here is the example of standards and guidelines to consider. The manufacturing system consists of 8 QCs for product quality intelligent modeling process. These include quality awareness (QC_1), accuracy of the equipments (QC_2), stability of the equipments (QC_3), material strength (QC_4), material hardness (QC_5), system measurement (QC_6), environment characteristics (QC_7), feasibility of manufacturing methods (QC_8). Then, we can obtain the adjacency matrix from 8 QCs in the following:

$$W_8(t) = \begin{bmatrix} 0 & 0.15 & 0 & 0 & 0.50 & 0 & 0 & 0.75 \\ 0 & 0 & 0.65 & 0.25 & 0 & 0.20 & 0.25 & 0 \\ 0.35 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.80 & 0 & 0.15 & 0 & 0 & 0.80 \\ 0 & 0.40 & 0 & 0.35 & 0 & 0.65 & 0.35 & 1.00 \\ 0.60 & 0.25 & 0 & 0 & 0.15 & 0 & 0 & 0.20 \\ 0 & 0 & 0.25 & 0 & 0 & 0 & 0 & 0 \\ 0.15 & 0 & 0 & 0 & 0.35 & 0 & 0.65 & 0 \end{bmatrix}$$

In this paper, we use vertex w to denote the concept and for its state. Given the binary concept value set used in the FCM, we can apply to represent the situation of different QC. Moreover the idea of credibility weights for decision makers is introduced. According to the FCM algorithm, the proposed linguistic weights are checked in terms of their neighborhood. The proposed method has the advantage that experts do not have to assign numerical causality weights for each indicator.

The first step in constructing the FCM model of product quality controlling is the determination of the concepts with FCM method. Many different decision makers following the algorithm developed the FCM model that is comprised of concepts group FCM. When the decision makers have described the concepts of the FCM model for monitoring the product quality in manufacturing system, the causal interconnections between different QCs have to be determined with our major manufacturing partners. A transformation process with the corresponding mechanism is needed to register each type of data source in manufacturing system. And the decision-making process of group FCM is described Fig. 5.

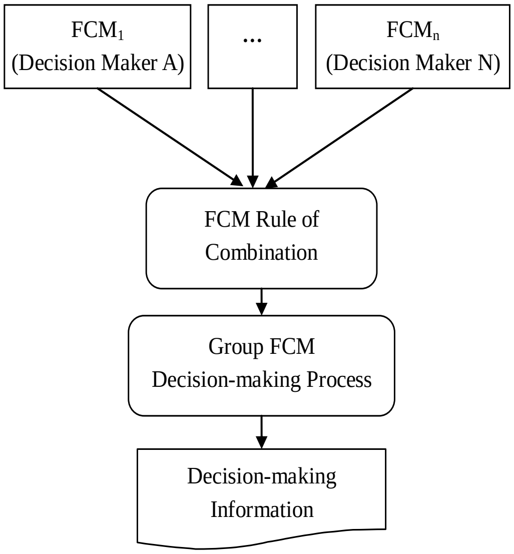


Fig. 5 The decision-making process of group FCM

FCM approach is an inference mechanism that allows the fuzzy causal relations among different factors. FCM models the decision making of a manufacturing system and offers a support to produce new knowledge based system applications. Here only a part of the intelligent decision making process has been developed to illustrate the role of the FCM model. Though the decision makers did not interact with each other during the modeling process, the degrees of influences from each expert's judgments individually can be obtained based on FCM model. In this paper, FCM can yield insights into indirect effects among all nodes under uncertain environment. The iterative process of FCM intelligent modeling is shown in Table 1.

Table 1: Iterative process of FCM intelligent modeling

	QC_1	QC_2	QC_3	QC_4	QC_5	QC_6	QC_7	QC_8
Iteration(1)	0.25	0.50	0.75	0.50	0.35	0.80	0.65	0.45
Iteration(2)	0.42	0.61	0.57	0.68	0.54	0.62	0.85	0.76
Iteration(3)	0.76	0.82	0.21	0.79	0.81	0.22	0.94	0.97
…	…	…	…	…	…	…	…	…
Hidden Pattern	0.95	0.97	0.07	0.93	0.98	0.05	0.97	1.00

It may be observed that the FCM was applied successfully in the study. From an engineering point of view, the decision making process can be thought a sort of statistics and artificial intelligence with database management. Extensive numerical experimentation is carried out in order to evaluate and validate different computational solutions for the implementation in decision making process. The level of such decisions for product quality is improving all the time. Besides, the statistical analysis assessing effectiveness of the intelligent system involved lots of comparison tests applied to the independent samples of computer runs.

4. Conclusions

In order to improve the usability of intelligent modeling for product quality decision making, a novel method has been proposed in this study based on FCM method. Different from a conventional product quality monitoring system, this approach aims to reduce the response time by processing the multivariate QC before they are transmitted to quality managers. Final results show good promise of the suitability of FCM approach for product quality intelligent modeling and decision making.
The intelligent system is implemented in product quality decision making process. It has been shown that the FCM could be used in intelligent modeling with effective results. The development of new decision making techniques based on intelligent modeling is necessary for increasing the correct ratio of quality control. Future research regarding intelligent modeling system with FCM aims to explore further the potentials in product quality decision making process.

References

[1] M. A. Al-Khedher, C. Pezeshki, J. L. McHale, and F. J. Knorr, "Empirical modeling of nanoindentation of vertically aligned carbon nanotube turfs using intelligent systems", Fullerenes Nanotubes and Carbon Nanostructures, Vol. 20, No. 3, 2012, pp. 200-215.
[2] K. N. Gowtham, M. Vasudevan, V. Maduraimuthu, and T. Jayakumar, "Intelligent modeling combining adaptive neuro fuzzy inference system and genetic algorithm for optimizing welding process parameters", Metallurgical and Materials Transactions B: Process Metallurgy and Materials Processing Science, Vol. 42, No. 2, 2011, pp. 385-392.
[3] A. Pariyani, W. D. Seider, U. G. Oktem, and M. Soroush, "Dynamic risk analysis using alarm databases to improve process safety and product quality: Part I-Data compaction", AIChE Journal, Vol. 58, No. 3, 2012, pp. 812-825.
[4] Y. K. Tse, and K. H. Tan, "Managing product quality risk and visibility in multi-layer supply chain", International Journal of Production Economics, Vol. 139, No. 1, 2012, pp. 49-57.
[5] V. Ravi Kumar, and K. Raghuveer, "An unsupervised approach to analyze users opinion on products using

customer reviews", International Journal of Computer Science Issues, Vol. 8, No. 4, 2011, pp. 380-385.

[6] D. P. McIntyre, "In a network industry, does product quality matter?", Journal of Product Innovation Management, Vol. 28, No. 1, 2011, pp. 99-108.

[7] H. A. Rakha, K. Ahn, W. Faris, and K. S. Moran, "Simple vehicle powertrain model for modeling intelligent vehicle applications", IEEE Transactions on Intelligent Transportation Systems, Vol. 13, No. 2, 2012, pp. 770-780.

[8] N. Park, and S. Lee, "Ambient Intelligent models for remote resource control and software maintenance in manufacturing globalisation", International Journal of Services Operations and Informatics, Vol. 5, No. 3, 2010, pp. 291-311.

[9] A. Martin, M. Manjula, and P. Venkatesan, "A business intelligence model to predict bankruptcy using financial domain ontology with association rule mining algorithm", International Journal of Computer Science Issues, Vol. 8, No. 3, 2011, pp. 211-218.

[10] S. T. Mueller, and G. Klein, "Improving users' mental models of intelligent software tools", IEEE Intelligent Systems, Vol. 26, No. 2, 2011, pp. 77-83.

[11] M. C. Panda, and V. Yadava, "Intelligent modeling and multiobjective optimization of die sinking electrochemical spark machining process", Materials and Manufacturing Processes, Vol. 27, No. 1, 2012, pp. 10-25.

[12] R. Prakash, and R. Anita, "Robust model reference adaptive intelligent control", International Journal of Control, Automation and Systems, Vol. 10, No. 2, 2012, pp. 396-406.

[13] A. Siltepavet, S. Sinthupinyo, and P. Chongstitvatana, "Improving quality of products in hard drive manufacturing by decision tree technique", International Journal of Computer Science Issues, Vol. 9, No. 3, 2012, pp. 29-34.

[14] H. J. Smadi, and A. K. Kamrani, "Product quality-based methodology for machine failure analysis and prediction", International Journal of Industrial Engineering : Theory Applications and Practice, Vol. 18, No. 11, 2011, pp. 568-581.

[15] A. Rahim, and M. Shakil, "A tabu search algorithm for determining the economic design parameters of an integrated production planning, quality control and preventive maintenance policy", International Journal of Industrial and Systems Engineering, Vol. 7, No. 4, 2011, pp. 477-497.

[16] K. Y. Chan, T. S. Dillon, and C. K. Kwong, "Handling uncertainties in modelling manufacturing processes with hybrid swarm intelligence", International Journal of Production Research, Vol. 50, No. 6, 2012, pp. 1714-1725.

[17] D. Pandey, M. S. Kulkarni, and P. Vrat, "A methodology for joint optimization for maintenance planning, process quality and production scheduling", Computers and Industrial Engineering, Vol. 61, No. 4, 2011, pp. 1098-1106.

[18] R. S. Wadhwa, "Towards modeling changeovers for flexible foundry manufacturing", International Journal of Computer Science Issues, Vol. 9, No. 4, 2012, pp. 85-90.

[19] G. Acampora, and V. Loia, "On the temporal granularity in fuzzy cognitive maps", IEEE Transactions on Fuzzy Systems, Vol. 19, No. 6, 2011, pp. 1040-1057.

[20] R. Furfaro, J. S. Kargel, J. I. Lunine, W. Fink, and M. P. Bishop, "Identification of cryovolcanism on Titan using fuzzy cognitive maps", Planetary and Space Science, Vol. 58, No. 5, 2010, pp. 761-779.

[21] D. K. Iakovidis, and E. Papageorgiou, "Intuitionistic fuzzy cognitive maps for medical decision making", IEEE Transactions on Information Technology in Biomedicine, Vol. 15, No. 1, 2011, pp. 100-107.

[22] S. Bueno, and J. L. Salmeron, "Benchmarking main activation functions in fuzzy cognitive maps", Expert Systems with Applications, Vol. 36, No. 3, 2009, pp. 5221-5229.

[23] B. Lazzerini, and L. Mkrtchyan, "Analyzing risk impact factors using extended fuzzy cognitive maps", IEEE Systems Journal, Vol. 5, No. 2, 2011, pp. 288-297.

[24] N. H. Mateou, and A. S. Andreou, "A framework for developing intelligent decision support systems using evolutionary fuzzy cognitive maps", Journal of Intelligent and Fuzzy Systems, Vol. 19, No. 2, 2008, pp. 151-170.

[25] Y. G. Petalas, K. E. Parsopoulos, and M. N. Vrahatis, "Improving fuzzy cognitive maps learning through memetic particle swarm optimization", Soft Computing, Vol. 13, No. 1, 2009, pp. 77-94.

[26] J. L. Salmeron, "Augmented fuzzy cognitive maps for modelling LMS critical success factors", Knowledge-Based Systems, Vol. 22, No. 4, 2009, pp. 275-278.

[27] W. Stach, L. Kurgan, and W. Pedrycz, "A divide and conquer method for learning large fuzzy cognitive maps", Fuzzy Sets and Systems, Vol. 161, No. 19, 2010, pp. 2515-2532.

[28] W. Stach, W. Pedrycz, and L. A. Kurgan, "Learning of fuzzy cognitive maps using density estimate", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 42, No. 3, 2012, pp. 900-912.

Jihong Pang, Dr. Jihong Pang received the Ph.D. degree in mechanical engineering from Chongqing University of China. Currently, he is a researcher at Wenzhou University, China. His major research interests include quality and reliability engineering, industrial engineering (IE), enterprise resource planning (ERP). He has published nearly thirty papers in related journals.

Analysis of Pipe Size Influence on Pipeline Displacement with Plain Dent Based on FE Calculation

Ying Wu[1], Peng Zhang[1] and Xia Han[2]
[1] School of Civil Engineering and Architecture, Southwest Petroleum University
Chengdu, Sichuan, China

[2] Materials Supply Divisions, Southwest Petroleum And Gas Branch, SINOPEC
Chengdu, Sichuan, China

Abstract
According to the report of the United States transportation department, mechanical damage is one of the most important reasons for the pipeline accident .The most typical form of mechanical damage is indentation. Dent defect is one of the important factors affecting pipeline fatigue life, and it will greatly reduce the fatigue life of the pipeline in service. Meanwhile the dent displacement will be changed with the operation pressure fluctuations of the in-service pipeline, resulting in a circular bending stress, which directly affect the pipeline fatigue life. For typical plain dent on pipeline, the finite element models were established under different circumstances. A large number of the calculation results were sorted and inducted. On this basis, the results were analyzed by univariate. Non-linear regression analysis was utilized to fit the results, some specific expressions of the relationship between the dented pipeline displacements and the diameter and wall thickness of pipeline are obtained after much calculation and analog.

Keywords: *Pipeline Diameter, Wall Thickness, Plain Dent, Displacement, Finite Element*

1. Introduction

According to the report of the United States transportation department, mechanical damage is one of the most important reasons for the pipeline accident. The most typical form of mechanical damage is indentation (or simply dent).The Figure 1 is dent cross-section shape.

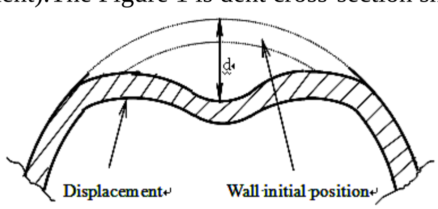


Fig. 1 The dent cross-section shape.

Under the effect of internal pressure, the dented pipe can have certain recovery, which will increase with the increase of internal pressure [1]. Dent recovery can be described by the recovery ratio, which is the ratio of the ultimate dent depth that undergoes the dent recovery and the original dent depth after the external disturbance is removed [2] .The dent recovery ratio is the very important parameter to influence the fatigue life of the dented pipeline [3,4,5].

There is a direct correlation between the dent recovery and the maximum displacement when the dented pipeline suffered the operating pressure. In order to accurately judge the fatigue life of the dented pipeline, based on finite element calculation results, the maximal displacement of the dented pipeline is studied in single-factor and multi-factor analysis by regression method [6,7].

2. Dent Defect Classification and Modeling

According to the analysis of the data accumulated at the scene, the dent on pipeline can be classified by three typical types: type I, type II and type III (see figure 2) [8,9]. The finite element analysis is carried out for type III (plain dent) in this paper.

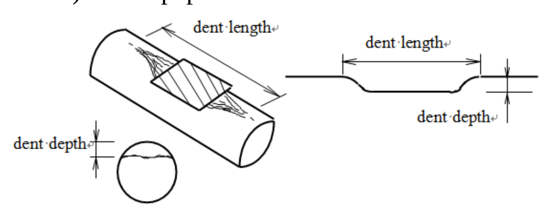


Fig. 2 The shape of plain dent.

The finite element software is utilized to establish the dented pipeline model under different circumstances. According to the real stress state and geometry shape of the pipe under internal pressure, a half of pipe is chosen for analysis by the axisymmetric principle, which includes all the model of dent defect.

It is apparent that different boundary conditions will have different impacts on the deformation even for the same load cases. However, compared to the dimension of pipelines, the dent in a pipe is so small that it could always be treated as a local deformation problem. Therefore, in order to not limit the generality of the research and eliminate the impact of the boundary conditions, the length of the pipe was taken as long as possible in the FE model (6 times the outer radius of pipe was used). In addition, according to the actual situation, fixed boundary conditions were applied on both ends of the pipe.

In fact, most of dent defects is caused in construction stage, and the external disturbance would be removed after dent defects was formed, so the FEM model only considered the formed dent. In addition, the influence of external load was not considered, and only the inner pressure was considered.

The 4-node shell element SHELL 63 was used to mesh the solid model. Different meshing density was used to simulate the dent, and according to some factors such as analytical process, calculation accuracy and calculation time and so on, a best meshing density was chose.

Pipeline material is used API X-60. The model of the original dent is result from a flat compressing pipeline along the longitudinal pipe. Pipe diameter is D. Pipe wall thickness is t. The longitudinal dent length of FE model is L. Dent length is the distance between intactness cross-section that is adjacent on both sides of the dent. The dent depth of FE model is d. Dent depth is the maximum displacement of the pipe wall thickness on the place of dent. The internal pressure of FE model is P. In addition, the finite element model conformed to some existing papers that study the similar problems.

The figure 3 and figure 4 are respectively the finite element mesh and displacement of the dented pipeline when P is 3.105Mpa, the ratio of pipe diameter and wall thickness D/t is 30 and the ratio of dent depth and pipe diameter is 0.08.

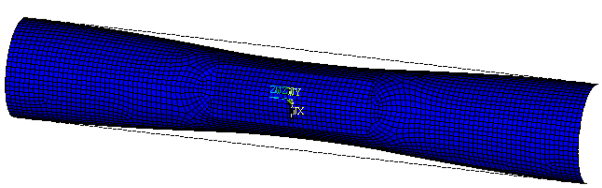


Fig. 3 The finite element meshing map of the part dented pipe.

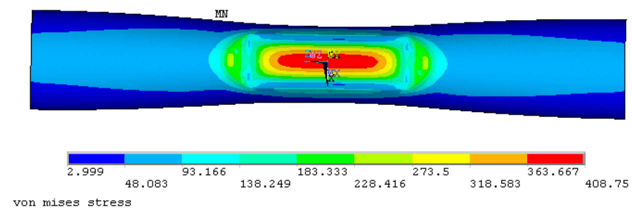


Fig. 4 The displacement of the part dented pipe (mm).

3. The Single Factor Analysis of Displacement

3.1 Pipe Wall Thickness

The figure 5 shows the relationship between the maximum displacement (DMX) and pipe thickness (t) under the different ratio of dent depth and pipe diameter (d/D) when the other parameters is unchanged.

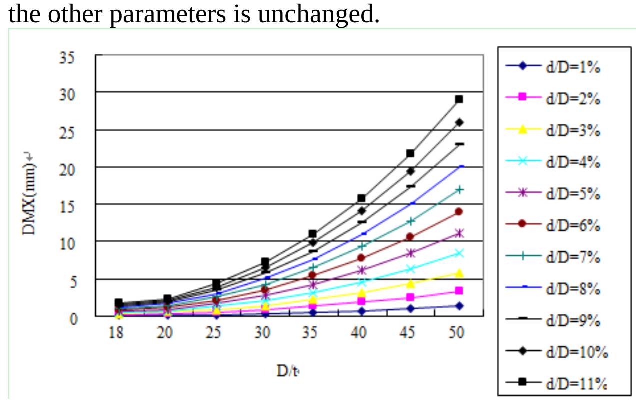


Fig. 5 The relationship of the maximum displacement and the ratio of pipe diameter and wall thickness

The figure 6 shows the relationship between the maximum displacement and pipe thickness when the pipe diameter is equal 1000mm.

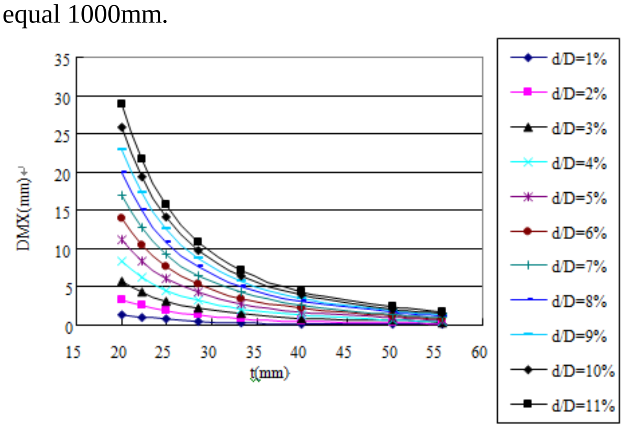


Fig. 6 The relationship of the maximum displacement and pipe wall thickness.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

As can be seen from figure 6, at different values of d/D, when pipe diameter is unchanged, the maximum displacement of the dented pipe decreased with the increase of the pipe wall thickness.

After a series of comparison and calculation, the relationship between the maximum displacement and pipe wall thickness can be expressed by power function, which fit it best. The common expression of the power function is as follows:

$$DMX = at^{b} \quad (1)$$

The table 1 and table 2 list out the coefficient a and b of the expression between the maximum displacement and pipe thickness under the different values of d/D.

Table 1: The values of *a, b* under different *d*

d(mm)	**a**	**b**
10	1388.2	-2.3186
20	6693.5	-2.5387
30	11303	-2.5377
40	23506	-2.6503
50	32999	-2.6676
60	42604	-2.6767
70	53464	-2.6887
80	65432	-2.7018
90	76681	-2.7084
100	89545	-2.7192
110	101772	-2.7261

The correlation coefficient square values of the table 3 are all equal to 1, so the fitting effect is quite good.

Table 2: The values of *a*, *b* and R^2 under different *d*

d(mm)	**a**	**b**	R^2
(10,20)	4802	-2.5411	0.67
(20,30)	11362	-2.6101	0.84
(30,40)	19566	-2.6452	0.92
(40,50)	28885	-2.6672	0.95
(50,60)	38160	-2.6757	0.97
(60,70)	48083	-2.6837	0.98
(70,80)	58806	-2.6926	0.98
(80,90)	70393	-2.7025	0.98
(90,100)	82616	-2.7123	0.99
(100,110)	95396	-2.722	0.99

As can be seen from table 2, when the dent depth is more than 20mm, the fitting effect is quite good.

3.2 Pipe Diameter

The figure 7 shows the relationship between the maximum displacement and pipe diameter under the different ratio of longitudinal dent length and pipe diameter (L/D) when the other parameters is unchanged.

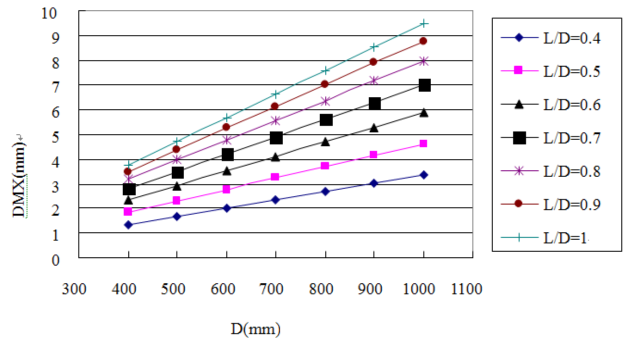


Fig. 7 The relationship of the maximum displacement and pipe diameter.

As can be seen from figure 7, at different ratio of longitudinal dent length and pipe diameter, the maximum pipe displacement increased with the increase of the pipe diameter. After a series of comparison and calculation, when the other parameters are unchanged, the relationship between and the maximum pipe displacement and pipe diameter can be expressed by linear function under the different values of L/D, which fit it best. The common expression as follows:

$$DMX = a \times D + b \quad (2)$$

The table 3 and table 4 list out the coefficient *a* and *b* of the expression between the maximum displacement and pipe diameter under the different values of L/D.

Table 3: The values of *a, b* under different *L/D*

L/D	**a**	**b**
0.4	0.0034	-0.0121
0.5	0.0046	-0.0125
0.6	0.0059	-0.0122
0.7	0.0007	-0.0101
0.8	0.008	-0.0105
0.9	0.0088	-0.0084
1	0.0095	-0.0085

The correlation coefficient square values of the table 3 are all equal to 1, so the fitting effect is quite good.

Table 4: The values of *a*, *b* and R^2 under different *L/D*

L/D	***a***	***b***	R^2
(0.4,0.5)	0.004	-0.0123	0.75
(0.5,0.6)	0.0053	-0.0123	0.84
(0.6,0.7)	0.0064	-0.0112	0.91
(0.7,0.8)	0.0075	-0.0103	0.95
(0.8,0.9)	0.0084	-0.0094	0.97
(0.9,1)	0.0087	-0.0091	0.94

As can be seen from table4, when the values of L/D are less than 0.5, the fitting effect is quite good.

4. Conclusions

(1) At different values of d/D, the maximum displacement of the dented pipe decreased with the increase of the pipe wall thickness. When the other parameters are unchanged, the relationship between the maximum displacement and pipe wall thickness of the dented pipeline can be expressed by power function model, as follows: $DMX = at^b$.

(2) At different values of L, the maximum displacement of the dented pipe increased with the increase of the pipe diameter. When the other parameters are unchanged, the relationship between the maximum displacement and pipe diameter of the dented pipeline can be expressed by linear function model, as follows: $DMX = a \times D + b$

Acknowledgments

This work described here is partially supported by the grants from the National Natural Science Foundation of China (No. 50974105), the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20105121110003), and the applied basic research project of Sichuan province (2011JY0138).

References

[1] 7th Report of the European Gas pipeline Incident Data Group, 2008.
[2] Y. Wu, P. Zhang, L. Q. Chen, "Study of the Factors Involved in the Strength of Dented Pipeline", in Proceedings of the International Conference on Pipelines and Trenchless Technology 2009, 2009, pp. 2092-2099.
[3] J.Adam, Rinehart and B.Pete, "Length effects on fatigue behavior of longitudinal pipeline dents", in Proceedings of the International Conference on Pipeline 2002, 2002, pp. 1849-1858.
[4] Y. Wu, P. Zhang, W. Liu, "Fatigue Life Calculation of the In-Service Dented Pipeline", Key Engineering Materials, Vol. 467, 2011, pp. 1327-1332
[5] Y. Wu, P. Zhang, W. Liu, “Analysis of the factors Involved in the Fatigue Life of Dented Pipeline”, J. Oil and Gas Technology, Vol. 30, No. 3, 2008, pp. 411-413.
[6] Y. Wu, P. Zhang, X. Y. Peng, “The Assessment of the Burst Strength of a Dent and Gouge on Pipeline”, pressure vessel. Vol. 26, No. 4, 2009, pp. 47-50.
[7] Y. Wu, P. Zhang, “Peak Cycle Stress Analysis of the II Type Dent on Pipeline Based on FE Calculation”, J. Applied Mechanics and Materials. Vol. 55-57, 2011, pp. 1659-1664.
[8] Fatma Ayari1, Fayza Ayari , Emin Bayraktar and Chokri Ben Amar, “Image processing and Finite Element modeling for analysis of a metal matrix composite”, International Journal of Computer Science Issues, Vol. 9,No. 5, 2012, pp. 448-457.
[9] Olufemi Ayinde Folorunso and Shahrizal Sunar Mohd, “Visualising Pipeline Sensor Datasets with Modified Incremental Orthogonal Centroid Algorithm”, International Journal of Computer Science Issues, Vol. 8,No. 5, 2011, pp. 268-278.

Ying Wu received a Master Degree in Engineering from Southwest Petroleum University, Chengdu, China, in 2007 and a Ph.D. also from Southwest Petroleum University in 2010. She has been a lecturer at Southwest Petroleum University, school of Civil Engineering and Architecture, China. Her main research interests in the safety and reliability of the oil and gas storage and transportation.
Peng Zhang has been a professor at Southwest Petroleum University, school of Civil Engineering and Architecture, China. Her main research interests in the safety and reliability of the oil and gas storage and transportation.
Xia Han received a Master Degree in Engineering from Southwest Petroleum University, Chengdu, China, in 2007 and a Ph.D. also from Southwest Petroleum University in 2010.

Study of the Video Monitoring System Image Recognition Solutions Based on Mathematic models

Peilong XU[1], Cheng ZHANG[2]

[1] **The Growing Base for State Key Laboratory, Qingdao University, No. 308, Ningxia Road, Qingdao 266071, China.**

[2] **Office of Facilities and Laboratory Management, Qingdao University, Qingdao, China.**

Abstract

objective: Through establishment a set of image recognition system based on mathematic models, to develop a auto alarm solution for the video monitoring system. Methods: compare the images the video monitoring system collected according to the time sequences. Then after binaryzation and wave filtering, the images were converted into numerical values using autocorrelation function, and the alarm threshold value was confirmed by experiences. Results: Through experiments, the change ratios of the two images before and after image processing were inversely proportional to the autocorrelation function. When the function value is less than 0.8, it indicates that there is an object volumes larger than 1m3 has invaded into 15m distances, and when the function value is less than 0.6, it indicates that there is an object volumes larger than 1m3 has invaded into 30m distances. Conclusion: Through calculation of autocorrelation functions, auto alarm for the images collected by video monitoring system could be effectively realized.

Keywords: *Video monitoring system, Autocorrelation function, Image processing Mathematic models.*

1. Introduction

Dynamic image recognition and self alarm is a key point in monitoring system studies for many countries, and the research fruits have been used in public safety areas widely. Now, public safety agencies in every countries are seeking an auto alarm system for video monitoring system which could not only save labor forces, but also decrease rates of false and leak alarms. But because of the imperfect of motive images recognition methods, the existing solutions are high in time complexity. The present automatic alarm systems mostly distinguish the object by perceiving the sound, light reflection signals or heat radiation, such as vehicle installation back-draft radar and so on. However, the object recognition rate for the objects with no reflection signal or far heat radiation is not high. Our study combines with the image processing technology and develops an alarm system based on the graphics change mathematical model. This is a good way to solve the above problems.

2. The Automatic Alarm Image Recognition Scheme and Method Of The Video Security Monitoring System

2.1 The network structure of the video security monitoring system with the function of automatic alarm

The safety monitoring system includes the following several parts: monitoring camera equipment, coaxial cable used for the transmission of analog signal, DVR hard disk video camera, graphics workstations equipped with a video acquisition card, video matrix distributor, TV wall. The system acquires the analog video signal using various forms of industrial camera. And the analog signal is transmitted to the DVR hard disk video cameras and video collection card through the double loop line, respectively. The image acquisition card acquires an image after a certain time according to the sampling rate. Though the binarization of the image, the computer stores the document in the bitmap file format to the hard disk for processing. The real-time images are displayed in the TV wall, and at the same time, the system warns about the alarm area and sends the discriminant result to the user terminal through the Internet. DVR hard disk video camera records all the images. Fig. 1 shows the network topological structure of the video safety monitoring system.

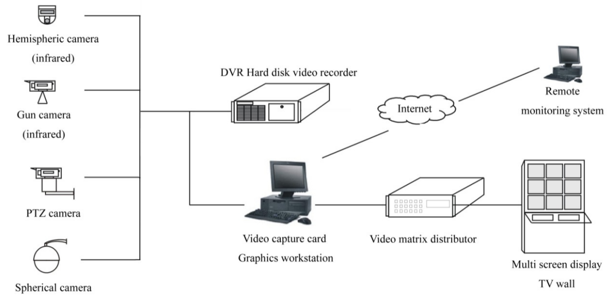


Fig. 1 the network topological structure of the video safety monitoring system

2.2 The basic scheme of the image recognition

The images stored in the hard disk are removed the color information, and make the gain processing. Then the two adjacent images are compared, and through the judgment of the autocorrelation function for the adjacent images, it can be told whether someone or other alien come into the monitoring area.

2.3 The method of image gain processing

Due to the difference of the image acquisition devices and the difference of the environment, some noise cannot be avoided in the direct image. So the images should experience the gain processing before the comparison and identification. Among many filtering processing methods, the system chooses the following two methods:

(1) Neighborhood average method: the basic thought of this method is to get the average value A' of the point A pixel and point B pixel's gray value in the image. Then replacing the point A's gray value uses the valve A'. It is assumed that the coordinate of the image pixel point A is (x, y) and its gray value is f (x, y). The size of neighborhood S is the size of L×W and after the neighborhood average processing, the gray value g(x, y) of the image is (1).

$$g(x,y)=\int \begin{matrix} \overline{f}(x,y) \left| f(x,y)-f(x,y) \right| > T \\ f(x,y) \quad other \end{matrix} \tag{1}$$

In this formula, T is the given threshold (2).

$$\overline{f}(x,y)=\frac{1}{LW}\sum_{(x,y)\varepsilon S} f(u,v) \tag{2}$$

This processing is effective to the light and dark point noise produced during the processing of image sensor, transmission channel, decoding. At the same time, it can also make a smooth transition for the images only having some small differences.

(2) Median filtering is based on the order statistic theory and is a nonlinear signal processing technology which can reduce the noise effectively. Its basic principle is replacing the value of a point in the digital image or in the digital sequence with the mid-value of the every point value in its adjacent domain. This method can get the pixel value closed to the true value and eliminate the isolated noise point. Its outstanding merit is that it can protect the boundary information in the processing of noise elimination. However, its efficiency becomes lower and lower with the increase of the noise. So the median filtering is not suitable when the number of the noise pixel is more than half of the pixel.

In the digital image system, the typed images are acquired firstly, and then transformed into the electrical signals by scanning. After this, the signal can be used to store and compare. Finally, a new restored image is shown to the user. In this process, the high Gaussian noise is usually produced. Moreover, due to the limit of the processing speed for the dynamic image acquisition system, the system is unfavorable to use complicated method. In this system, the neighborhood average method is employed due to its lower time complexity. This method can remove the Gaussian nose efficiently and keep the edge of the image well.

2.4 Gray level expansion

Because the contrast of the darker area is very poor for the infrared image in the night, it is difficult to distinguish without processing. The grays level should be stretched in the image contrast in order to improve the grayscale dynamic range in the image processing. The treatment process is as follows. The range of the gray level for the image f (x, y) is [a, b]. The dynamic range of the transformed image g(x, y) is [c, d]. This transformation can be realized as the following formula (3):

$$g(x,y)=\begin{cases} c & f(x,y)<a \\ \dfrac{(d-c)[f(x,y)-a]}{b-a}+c & a\le f(x,y)\le b \\ d & f(x,y)\ge b \end{cases} \tag{3}$$

This piecewise transformation will elongate the original grey interval [a, b], but compress [0, a] and [b, 255]. If there are only a few of points in the interval, this transformation is reasonable.

2.5 The autocorrelation function of the safety monitoring system image

The image collected by safety monitoring system is transformed into the black and white image after the gain and gray level expansion processing. The mathematical formula of the acquired image signal is (4)

$$X(t) = X_e(t) + X_0(t) \tag{4}$$

The Xe(t) is even field signal and its value is (5)

$$X_e(t) = \sum_{i=0}^{M-1}\sum_{k=0}^{(N-1)/2}\sum_{l=-\infty}^{\infty} x[iM_5, 2kN_5, lT_p] \bullet \sin c\{2w[t-(i+kM+lNM)T]\} \tag{5}$$

The X0(t) is odd field signal and its formula is(6):

$$X_e(t) = \sum_{i=0}^{M-1}\sum_{k=0}^{(N-1)/2}\sum_{l=-\infty}^{\infty} x[iM_5, 2k(2K+1)N_5, lT_p] \bullet \sin c\left\{2W[t-(i+kM+lNM+\frac{M(N+1)}{2})T]\right\} \tag{6}$$

In this formula, W is the highest frequency of the image signal and the sampling period is T=W/2.

So the autocorrelation function of the image acquired by the safety monitoring system is (7)(8)(9)(10)

$$R_\tau = R_{\tau 1} R_{\tau 2} R_{\tau 3} \tag{7}$$

$$R_{\tau 1} = \sum_{k=-\infty}^{\infty} \exp\left[-a_m\left|\tau - kT_L\right|\right] \tag{8}$$

$$R_{\tau 2} = \sum_{l=-\infty}^{\infty} \left\{ \begin{array}{l} \exp\left[-a_m\left|\tau - lT_P\right|\right] \\ +C \bullet \exp\left[-a_m\left|\tau - (l+\frac{1}{2})T_P\right|\right] \end{array} \right\} \tag{9}$$

$$R_{\tau 3} = \exp[-\beta\left|\tau\right|] \tag{10}$$

Through the analysis of the autocorrelation function, it can be concluded as follows:

(1) The real space image can be scanned into one dimension function, so all of the autocorrelation function can be transformed into τ function.

(2) Rτ1 reflects the correlation of the inline. The other two directions are regarded as the constants and the direction of m is considered as stochastic. And then find out the Rτ1. When α equals to kTL, the Rτ1 get the extremum value, namely, Rτ1=1. In other position, Rτ1 declines as the e index law to the sides. The decrease speed depends on the parameter αn. Because the one dimension scanning signal is periodic, Rτ1 shows the trait of periodicity.

(3) Rτ2 reflects the correlation of the interval. According to the formula of Rτ2, there will be a relative peak every other frame. But because the TV signal is scanned interlacedly, there will be another related peak in the even field and odd field besides in the even field and even field, odd field and odd field. So they have great relationships.

(4) Rτ3 reflects the correlation of the interframe and has no periodicity.

(5) Rτ is the product of the three items. Product can be considered as the modulation of every item in mathematics. Because the periodicity of Rτ1 is smaller than that of Rτ2, it can be regarded that Rτ2 modulates the wave of Rτ1 and then gets modulated by Rτ3.

2.6 Activation Mechanism of the Alarm System

The alarm of this system is completed depending on the autocorrelation function calculation. The value got by the function is in an inverse ration of image date change. Therefore, the alarm activation can be achieved by setting a threshold value. But if the monitor camera is sheltered by something, the pictures will drastically change, such as the leaves falling down. It can also active the alarm system. In our research, the physical method is temporarily used to solve this false, such as the necessary protection of the camera.

3.Experimental Method and Results

3.1 System Information

The image processing uses the WV-CW860B/CH all-weather integration camera produced by Toshiba Company. And the image collection uses the AP-6816-4 acquisition card produced by APYLJ Company. The camera can work in the infrared and visible light modes. Because the automatic alarm function of video monitoring system is usually open in the night, the infrared acquisition mode is used to test. The AP-6816-4 acquisition card uses the 10 Bit 6816 chip, and the resolution of the image collection is set to PAL mode, which is 704×576.

The “Symbolic Math Toolbox” in Matlab software produced by the Mathworks Company USA is employed to transform the formula into the program of Matlab. Then through the Matcom software, the program of Matlab is transformed into the C++ program codes .

3.2 Noise Reduction of Image Filtering

During the process, the image gets gain process first, and then experiences gray level expansion. In order to ensure the image processing speed, Dell Precision T7500 graphics workstation is employed to process the image. This graphic workstation can ensure the processing speed to 20 FPS (frame per second). The image IMG001.jpg is processed by neighborhood average method using Matlab 7.0 software. The 5×5 neighborhood average and neighborhood standard deviation is calculated by nlfilter function. The program written is as follows:

```
figure;
a=imread('e:\test3_1.jpg');
a=im2double(a);
f1 = inline('max(x(:))');
f2 = inline('std2(x)*ones(size(x))');
A1 = nlfilter(a,[5 5],f1);
subplot(1,3,1);
imshow(a);title('input image');
subplot(1,3,2);
imshow(A1);title ( The 5×5 image after the neighborhood average) ;
A4 = blkproc(a,[ 5 5 ],f2);
subplot(1,3,3);imview(A4,[]);% (neighborhood standard deviation image)
```

In the test, the fuzzy figure A becomes to a clear figure B after the gain and gray level expansion processing which is shown in Fig. 2.

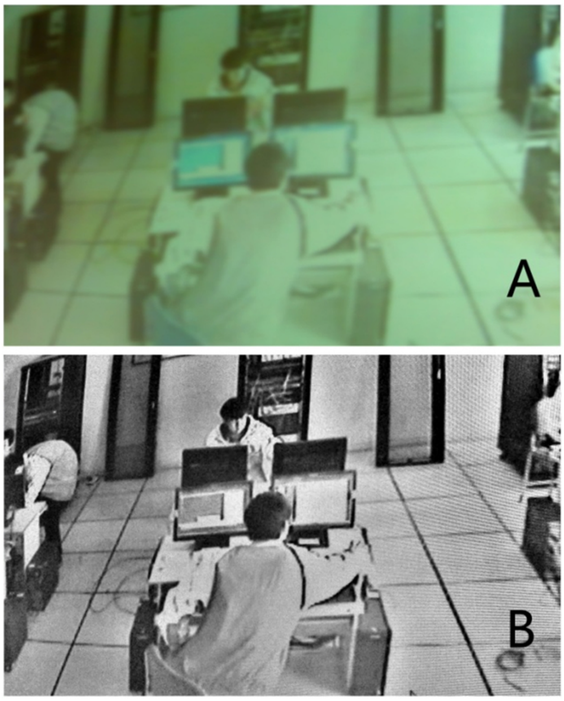


Fig.2. The Effect Image After The gain and Gray Level Expansion Processing.

3.3 The application of the autocorrelation function in alarm system

According to autocorrelation function calculation formula of the safety monitoring system image, the formula is programed by Matlab as follows:

```
R1 = 0;
R2 = 0;
for k = -inf:inf
R1 = R1 + exp(-am*abs(t-tl));
end
for l = -inf:inf
R2 = R2 + exp(-am*abs(t-tp)) + C*exp(-am*abs(t-(l+1/2)*tp));
end
R3 = exp(-b*abs(t));
R = R1 * R2 * R can 3;
```

After the operation and correction, the program can be changed into C++ program code using Matcom and applied into the system. The human body and the 1 m^3 carton are used to intrusion test for the video monitoring system. Table 1 shows the results of the intrusion test.

Table1 the autocorrelation function change result of the adjacent images

Invasion object	***Distance***	Rτ Value
person	10m	0.310
person	20m	0.461
Person	30m	0.562
Carton	5m	0.610
Carton	10m	0.671
Carton	15m	0.735
None		1.158

The experimental results show that the value Rτ and the image difference are inverse ratio, the smaller the value Rτ and the larger the image difference. Through the test, among the images taken by the camera, when the $R\tau \leqslant 0.6$, the human body size object invades in the range of 30 m; when the $R\tau \leqslant 0.8$, 1 m^3 size cube invades in the range of 15 m.

4. Conclusions

This study established a set of mathematic model for collection, compare and recognition of dynamic images. The experiments indicate that through calculation of autocorrelation functions, auto alarm for the images collected by video monitoring system could be effectively realized.

References

[1] Eric Olmedo, Jorge De La Calleja, Antonio Benitez and Ma. Auxilio Medina, " Point to point processing of digital images using parallel computing", International Journal of Computer Science Issues, Vol. 9, No.3, 2012, pp.1-10.

[2] Ziguan CUI, Zongliang GAN, Xuefang ZHAN, Xiuchang ZHU, "Error Concealment Techniques for Video Transmission over Error-prone Channels: A Survey", Journal of Computational Information Systems , Vol.8 No.21, 2012, pp.8807-8818.

[3] Wei Hui, Liu Zupeng, "Study on the Removal Method of Image Depth Impulsive Noise", Journal of Convergence Information Technology, Vol. 7, No. 18, 2012, pp. 316- 323.

[4] Wei Yan-jun, Zhao Xin, "An Novel Grayscale Image Segmentation Method based on 2D Entropy and Particle Swarm Optimization", International Journal of Digital Content Technology and its Applications, Vol. 6, No. 14, 2012, pp. 68-76.

[5] Wang Xin-cheng, "Multimedia technology （Image section）". Chengdu: Chengdu University of Electronic Science and technology Publishing, 2000.

[6] Abdur Shahid, Shahriar Badsha, Md. Rethwan Kabeer, Junaid Ahsan and Mufti Mahmud, " A Non-Blind Watermarking Scheme for Gray Scale Images in Discrete Wavelet Transform Domain using Two Subbands", International Journal of Computer Science Issues, Vol. 9, No.5, 2012, pp.101-109.

[7] Jong-Bae Jeon, Jung-Hyun Kim, Jun-Ho Yoon. "Performance evaluation of teeth image recognition system based on Difference Image Entropy", Proceedings - 3rd International Conference on Convergence and Hybrid Information Technology, ICCIT 2008, Busan, Korea,2008.pp.967-972.

[8] Banerjee Satyajit, Chowdhury Atish Datta, Ghosh Subhas Kumar, "Video surveillance with PTZ cameras: The problem of maximizing effective monitoring time", Lecture Notes in Computer Science: Distributed Computing and Networking - 11th International Conference, ICDCN 2010, Proceedings, Kolkata, India,2010, pp,341-352.

[9] Chohra Amine, Kanaoui Nadia, Madani Kurosh. "Image recognition based analysis and comparison of hybrid intelligent approaches for computer aided diagnosis", 2007 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS, Dortmund, Germany,2007 pp,367~372.

First Author Peilong XU, born in 1977. The author has achieved Master degree from Tongji University of Shanghai in 2007. He is currently an engineer of computer science in the State Key Laboratory, Qingdao University. The author's research interests include software engineering, image processing, and spectral analysis. Recently, he have published paper named Design and Implementation of Landscape System for East and West Huashi Street in Beijing Based on Virtual Reality Technology on journal Applied Mechanics and Materials（ Trans Tech Publications Inc, Switzerland）etc.

Second Author Cheng ZHANG, Office of Facilities and Laboratory Management, Qingdao University, Qingdao, China.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

BP Neural Network Algorithm Optimized by Genetic Algorithm and Its Simulation

Junsheng Jiang

School of Mechatronics and Vehicle Engineering, Weifang University
Weifang, 261061, China

Abstract

Aiming at the drawbacks of slowly converging and easily getting into the local minimum appearing in the BP neural network, this paper combines the general optimization of genetic algorithms together with the local optimization of BP neural networks to improve the performance of BP neural networks. The numerical experiment shows that, compared with the original BP neural network, the optimized BP neural network can effectively reduce the average error of the model calculation and prediction, greatly cut the times of iteration, and raise the calculation accuracy and convergence speed. This paper also demonstrates the ability of genetic algorithms to improve the performance of BP neural networks.

Keywords: *BP neural network, genetic algorithm, optimization*

1. Introduction

In recent years, artificial neural networks have been successfully applied to system identification, pattern identify and control fields. As the neural network theory is still in progress, there is not a set of ways to guide the design process. Now the design of neural network mainly focuses on two sides: one is the optimum structural design of neural networks, the other the algorithm of neural network's weight. At homes and abroad, many researches have been conducted on these two sides and proposed a lot of optimization methods.

The gradient descent method is usually used for training BP networks, whose advantages are fast in local optimization because of only one direction. However, it easily converges to local optimization and then has poor performance on global optimization. The genetic algorithm was put forward by professor Holland in 1975, which is based on biological evolution and genetic variation. The genetic algorithm makes population evolution ceaseless by means of choosing copy and genetic operator and can obtain optimal solutions at last. It has good performance of global optimization because of its searching the optimal solution in the whole solution space, however, at high cost. Some researches indicate that the genetic algorithms can reach the 90 percent of the global optimization quickly, but the convergent speed becomes slow clearly in the last period.

The paper introduces a BP neural network optimized by genetic algorithms and the BP neural network takes advantages of the gradient descent method and genetic algorithms. The simulations demonstrate that the optimized algorithm has faster convergent speed than the original algorithm.

2. Principle of Genetic Algorithm

The genetic algorithm is a random search algorithm that simulates natural selection and evolution. It searches the total solution space and can find the optimal solution globally over a domain. The basic operation of a genetic algorithm is as follows:

(1) Choosing Operation

The choosing operation means to choose the individuals from the old group and add them to the new group in a certain probability. The probability of the individual selected is relevant to the fitness value: the better fitness value of the individual, the greater probability of being chosen.

(2) Crossover Operation

The crossover operation means to select two individuals from the existing individuals and generate a new better individual through the exchange and combination of two chromosomes. The crossover process is to choose two chromosomes from the groups and select randomly one point or more points to be exchanged. The crossover operation is shown in Figure 1.

$$\begin{array}{l} A:1100\vdots 01011111 \\ B:1111\vdots 01010000 \end{array} \rightarrow \begin{array}{l} A:1100\vdots 01010000 \\ B:1111\vdots 01011111 \end{array}$$

Fig.1. Crossing Operation

(3) Mutation Operation

The mutation operation is to choose one individual from the group and then select a point of chromosomes for variation operation to generate better individuals. The mutation operation is shown in Figure 2.

$$\mathrm{A}:1100\ \ 0101\ \ 1111 \rightarrow \mathrm{A}:1100\ \ 0101\ \ 1101$$

Fig.2. Variation Operation

3. BP Neural Network Optimized by Genetic Algorithm

3.1 Flowchart of BP Neural Network Optimized by Genetic Algorithm

The BP neural network optimized by the genetic algorithm consists of three parts: network structure determination, genetic algorithm optimization and network prediction. The structure of a BP neural network is determined by the input and output parameters of the fitting function, by which the length of individuals of the genetic algorithm can be determined. The genetic algorithm is exploited to optimize the weights and thresholds of the BP neural network. Each individual of the population consists of all weights and thresholds of a network and the individual fitness value is calculated by the individual fitness function. Subsequently, the genetic algorithm finds the best fitness value corresponding to the individual by means of the selection, crossover, and mutation. Next, the BP neural network obtains the optimal initial weights and threshold provided by the genetic algorithm and predicts the function output after the network has been trained.

3.2 The Procedure of the Genetic Algorithm

The BP Neural network optimized by the genetic algorithm uses the genetic algorithm to optimize the initial weights and thresholds of the BP neural network and makes the optimized BP neural network better predict the output. The BP neural network optimized by genetic algorithm is as follows:

(1) Population Initialization

The individual encoding is real-coded and each individual is a real string which consists of connection weights of the input layer and the hidden layer, the threshold of hidden layer, connection weights of hidden layer and output layer, and the threshold of output layer. The individual contains all weights and thresholds of the neural network. We can form a neural network with the determined structure, weights and thresholds under the conditions of the network structure being known.

(2) Fitness Function Calculation

To begin with, the initial weights and thresholds of the BP neural network are obtained by the individual; then, the output is predicted after the BP neural network has been trained with the training data. Use the sum of the absolute value of error between the prediction output and desired output as the fitness value of individual and the formula is shown as

$$F = k(\sum_{i=1}^{n} abs(y_i - o_i)) \tag{1}$$

In the formula, n is the number of output nodes of the network, y_i is the desired output of the ith node of the neural network, o_i is the prediction output of the ith node and k is the coefficients.

(3) Selection Operation

The genetic algorithm operation has the roulette selection method, the tournament method and so on. This paper adopts the roulette method, which is based on the proportional fitness. The choice probability p_i of each individual i is

$$f_i = k / F_i \tag{2}$$

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \tag{3}$$

where, F_i is the fitness value of the individual i, k is the coefficients and N is the number of individual species.

(4) Crossover Operation

The crossover method uses the real crossover method because the individuals use real-coded. The crossover between the kth chromosome a_k and the lth chromosome a_i is as follows:

$$\left.\begin{aligned} a_{kj} &= a_{kj}(1-b) + a_{lj}b \\ a_{lj} &= a_{lj}(1-b) + a_{kj}b \end{aligned}\right\} \tag{4}$$

where, b is a random number between zero and one.

(5) Mutation Operation

The jth gene a_{ij} of the ith individual is selected to mutate. The mutation operation is as follows:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max}) * f(g) & r \geq 0.5 \\ a_{ij} + (a_{\min} - a_{ij}) * f(g) & r < 0.5 \end{cases} \tag{5}$$

$$f(g) = r_2(1 - g / G_{\max})$$

where, a_{max} and a_{min} are the upper and the lower bounds of the gene a_{ij}, respectively, r_2 is a random number, g is the current iteration number, G_{max} is the maximum number of evolution and r is a random number between zero and one.

4. Simulation example

The approximation function which has been used widely in the neural network was here used to verify the feasibility and effectiveness of the optimized algorithm. The function is as follows:

$$f(x) = \frac{\sin(8x)}{(x-0.25)^2 + 0.1} \tag{6}$$

Using each of the gradient descent algorithm and the generic method, only low approximation accuracy can be derived. If we further raise the approximation accuracy, then the gradient descent method is easy to fall into local minima. In addition, the genetic algorithms can quickly converge to the field of minimal errors, but it is difficult to achieve the final convergence. In this paper, the BP neural network and the genetic algorithm were together used to approximate the function. The best individual fitness value obtained by the optimized BP neural network is shown in Figure 3. Comparisons between the network prediction errors of the original and the optimized networks are shown in Figure 4. As seen in Figure 4, the BP network optimized by the genetic algorithm has more prediction accurate than the original network, since the root-mean-square of the optimized network is 5.2284×10^{-5} and that of the original BP network is 2.2356×10^{-4}. Thus, compared with the original network, the optimized network delivers better performance.

5. Conclusions

The original BP neural network is easy to fall into local minimum. To overcome the shortcoming, the paper combined the genetic algorithms with the gradient descent. The proposed algorithm in this paper can fully take the advantages of two algorithms, effectively avoid the local minimum and be able to achieve the desired weights of the BP network quickly. The simulation shows that the proposed algorithm can effectively solve the local minimum problem in the neural network training, and the mean square error of the prediction can greatly reduced.

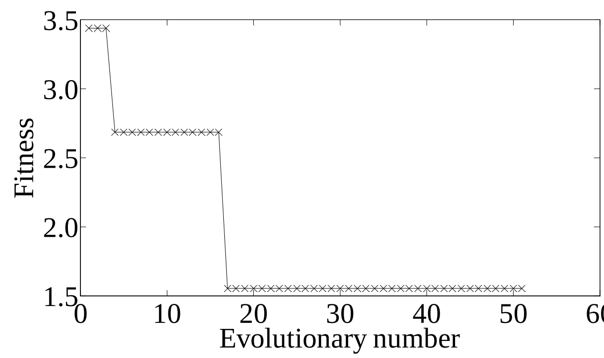


Fig. 3 The best individual fitness value

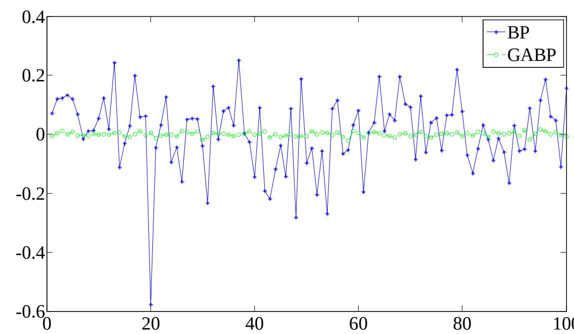


Fig. 4 Comparisons between forecasting errors of the un-optimized and optimized networks

References

[1] Wang Xiaoping, Cao Liming, Genetic Algorithm-Theory, Application and Realization. XI'AN: Xi'An JiaoTong University Press, 2002.
[2] LI Mu, HE Yi-gang, ZHOU Shao-wu, TAN Wen, “Hybrid Genetic Neural Network Method for Predicting Chaotic Time Series”, Journal of System Simulation, vol.20, no.21, 2008,pp.5824-5828.
[3] WU Ling, JIANG Zhou-hua, GONG Wei, LI Yang, “GA-NN-Based Predicting Model of Activity of Multiple Slag System”, Journal of Northeastern Universty (Natural Science), vol.29, no.12, 2008, pp.1725-1728.
[4] Xie Hongmei, Liao Xiaoping, Lu Yuhai, “GA-BP Algorithm and Its Application in the Forecasting of Part Quality” , China Mechanical Engineering, vol. 19, no. 22, 2008, pp.2711-2714.
[5] Seyedyaser Bozorgirad, Mohammad Ishak Desa and Antoni Wibowo ,”Genetic Algorithm Enhancement to Solve Multi Source Multi Product Flexible MultistageLogistics Network”, International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012 pp.157-164.

[6] Sayedmohammadreza Vaghefinezhad and Kuan Yew Wong", A Genetic Algorithm Approach for Solving a Flexible Job ShopScheduling Problem" International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012 pp.85-90.
[7] V. Sai Shanmuga Raja and Dr. S.P. Rajagopalan ,"A Neuro-Genetic System for Face Recognition", International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012 pp.263-267.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Energy Efficient Protocols in Wireless Sensor Networks: A Survey

Dr. Kavita Taneja[1], Amandeep kaur[2]

**[1] MMICT&BM, Maharishi Markandeshwar University
Ambala, Haryana, India**

**[2] Computer Science & Engineering Department, M.M.E.C, Maharishi Markandeshwar University
Ambala, Haryana, India**

Abstract

Wireless Sensor Networks (WSNs) consists of huge number of Sensor Nodes (SNs) with sensing, communication and processing capabilities. SNs have limited energy supply, storage and computational capacity. In recent years energy efficient computation is a major concern in WSN. The critical aspects include reduction in the energy consumption of SNs so that the network lifetime can be extended to reasonable times. For this purpose many novel innovative techniques based on energy efficient computation have been proposed. In this paper, we present a brief analysis on energy efficient computation protocols. We have also presented a comparison of these protocols.

Keywords: *WSN, Energy Efficiency, Aggregation.*

1. Introduction

WSNs have gained worldwide attention in recent years; particularly with the proliferation in technology which has facilitated the development of smart SNs. WSN consist of small SNs with sensing, computation, and wireless communications capabilities. These SNs measure ambient conditions in the environment surrounding them and then transform these measurements into signals that can be processed to reveal some characteristics about phenomena located in the area around these SNs. WSN contain hundreds or thousands of these SNs, and these SNs have the ability to communicate either among each other or directly to an external Base Station (BS). Basically, each SN comprises sensing, processing, transmission, mobilizer, position finding system, and power units [1]. The use of pervasive networking technology gives WSN a new kind of scope that can be applied to a wide range of uses. These can be roughly differentiated into monitoring space, monitoring things, and monitoring the interactions of things with each other and the encompassing space. The first category includes environmental and habitat monitoring, precision agriculture, indoor climate control, surveillance, and intelligent alarms. The second includes structural monitoring, condition-based equipment maintenance, medical diagnostics, and urban terrain mapping. The most dramatic applications involve monitoring complex interactions, including wildlife habitats, disaster management, emergency response, ubiquitous computing environments, asset tracking, healthcare, and manufacturing process flow [2]. Energy efficient computing is necessary for at least two reasons as shown in Fig 1. First, sending all the raw data to a BS for centralized processing is very costly in terms of energy consumption. Second, merely communicating user's requirement need to be compromise lot of QoS Parameters in light of energy constraints.

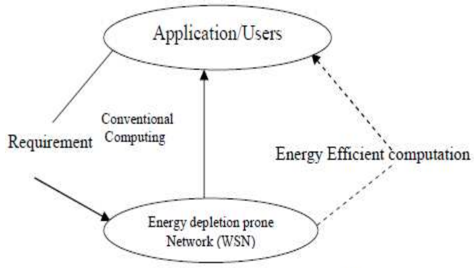


Figure 1: Computing in WSN

2. Energy Efficient Computing in WSN: Challenges

WSN offer a powerful combination of distributed sensing, computing and communication. The emerging WSN technology promises to fundamentally change the way humans observe and interact with the physical world. To realize such a vision, efficient computing is necessary for following reasons.

Resource constraints: The resources constraints in WSN involve energy, memory, bandwidth, processing capability, buffer size, and limited transmission power. Among them, energy is a primary issue since energy is severely constrained in SN and it is not feasible to replace or recharge the battery for SNs that are often expected to work in a remote or unattended environment. Computation intensive algorithms, expensive signaling protocols, or overwhelming network states maintained at SNs are not feasible.

Data redundancy: WSNs are characterized by high redundancy in the sensor data. However, the redundancy in the data helps loosen the robustness/reliability requirement of data delivery, but it unnecessarily spends more precious energy. Data fusion or data aggregation is a solution to reduce redundancy in the data.

Network dynamics: Network dynamics causes are SN failures, wireless adaptive link failures, SN mobility, and SN state transitions. Such a highly dynamic network greatly increases the need of distributed computation.

Load balance: In order to prolong network lifetime, energy load must be evenly distributed among all SNs so that the energy at a single SN or a set of SNs is not be drained out.

Scalability: The number of SN deployed in a region may be in the order of hundreds or thousands. Based on the application, the number may reach value of millions. WSN must be scalable enough to operate over long period of time.

Coverage: In WSNs, each SN obtains a certain view of the environment. A given SNs view of the environment is limited both in range and in accuracy; it can only cover a limited physical area of the environment. Hence, area coverage is also an important design parameter in WSN.

3. Infrastructure Support for Computing in WSN

In order for WSN to effectively perform computation, some necessary infrastructure needs to be established. This includes neighbor discovery and management, synchronization, localization, clustering and grouping, and data collection infrastructure.

Neighbor discovery and management refers to the process in which SNs discover their neighbors, learn their properties, and control which neighbors to communicate with expensive signaling protocols, or overwhelming network states maintained at SNs are not feasible.

Synchronization refers to the process in which SNs synchronize their clocks. It is necessary because sensory data is often not useful in absence of proper temporal reference frame.

Localization refers to the process in which SNs obtain their position or coordinate information. Similar to synchronization, localization is necessary because sensory data needs to be put in a spatial reference frame.

Clustering and grouping refers to the process in which SNs organize themselves into clusters or groups for some specific function.

Data collecting infrastructure ensures that sensory data is transported correctly and efficiently to one or a few collection points, sometimes called data sinks. Data aggregation deals with this distributed computing of data within network. These techniques are tightly coupled with gathering of data at the SNs as well as how packets are routed, and have a significant impact on energy consumption.

4. Energy Efficient Techniques in WSN

Based on the above issue, several techniques have to be exploited, even simultaneously, to reduce the energy consumption in WSN. At a very general level, we identify some enabling approaches as explained below:

Distance-Energy Cluster Structure Algorithm (DECSA) [3] is a distributed competitive unequal clustering algorithm based on clustering routing algorithm LEACH. The algorithm continues in round and each round consists of initialization and stable working stages. In initialization phase Cluster Head (CH) and base station CH is elected based on distance and residual energy and TDMA time slots are assigned to member nodes by the appropriate CH. In working stage after receiving message from Base Station (BS), according to different values of TBCH, BCH select maximum TBCH CH as a next hop and rest hop can be selected in same manner to form complete communication path.

Energy Balancing Cluster Routing Protocol Based on Mobile Agent (EBMA) [4] is energy balanced protocol based on mobile agent. In EBMA, the complete network is divided into clusters and each cluster consists of many hexagon cells and obtains cellular topology. The ordinary nodes connect to the closest mobile agent which acts as CH and is responsible for managing all SNs in that cluster and needs to be active, while the ordinary SNs can go to a sleep state when they have no tasks to perform. The number of mobile agents should be limited to reduce the cost but must ensure efficient gathering.

Energy Efficient and Balanced Cluster-Based Data Aggregation Algorithm (EEBCDA) [5] is unequal cluster based algorithm. The operation of EEBCDA is also divided into rounds and every round consists of a set-up phase and a steady-state phase, especially, there is a network-division phase before the first round. The network is divided into rectangular regions called swim lanes, then, each swim lane is further partitioned into smaller rectangular regions, called grids. The node with maximal residual energy of each grid is selected as CH. The grids away from BS are bigger and have more nodes to participate in CHs rotation to balance energy consumption. Then, each ordinary node chooses the closest CH to join a cluster. And TDMA time slots are assigned to member nodes by the appropriate CH. In steady state phase every member sends data to CH during its allocated transmission slot according to the TDMA schedule. Afterward, every CH aggregates the collected data and sends the fused data to BS. For the sake of CHs selection in next round, each member transmits its residual energy along with its data to CH at the last time of data gathering in every round.

dYnamic and scalable tree Aware of Spatial correlaTion (YEAST)[6] is a spatial correlation aware algorithm to perform efficient data collection in WSN. Based on the correlated region, which is the region SNs collect similar readings, the algorithm maximizes data aggregation along the communication route, and decreases the costs in the route discovery. A hybrid routing approach composed of proactive part is responsible for discovering the positions of both the neighbors and the sink used later to construct the routing infrastructure and the routing itself and reactive part is responsible for clustering nodes aware of spatial correlation and constructing data aggregation aware routes to transmit the collected data. Finally, the YEAST algorithm uses dynamic routes to ensure load balancing in the delivery of data. The YEAST algorithm is performed in four phases. In phase 1, SNs store the sink's and neighbor's position. In phase 2, consists of cluster formation, the election of a coordinator among the nodes that detected the occurrence of a new event, and the division of the event area into cells. In Phase 3, when an event happened, the coordinator sends a packet to the sink node informing its position. The sink node sends notifications to all other coordinators about the new coordinator Position and to the new coordinator the positions of coordinators that already exist. Finally, Phase 4 is responsible for creating the routing tree connecting all coordinators to the sink node and sending the collected data to the sink node. Each variation defines a new approach for a coordinator node to create its straight lines. In YEAST-CF (Closest First) the closest coordinators to the sink node are the first to create their straight line segments to the sink. YEAST-FF (Farthest First) the farthest coordinators from the sink are the first to create their straight line segments to the sink. YEAST-BC (Best Combination) this variation of the YEAST algorithm checks all possible combinations of straight lines and chooses the combination that provides the shortest Euclidean distance to create the routing tree. This approach is optimal, since it finds the routing tree of lowest cost.

In Energy conserving routing algorithm [7] for WSN consists of two phases routing setup and data transmission phase. In routing setup phase we divide the SNs into several scheduling sets and they work alternatively. So the sensors do not have to be active all the time thereby energy. In data transmission phase when choosing the next sensor to forward the data the distance from the BS to the sensor and its current energy level is taken into consideration to distribute the network power consumption among the sensors. When in network no more sensors have sufficient energy to run, it generates new scheduling sets automatically. Data collection contains the routing setup phase and the data transmission phase. During this phase, the scheduling sets check the energy levels of their member nodes when they finish working for a period. If there is any node whose next hop nodes are all with energy levels below a threshold value, the corresponding scheduling set is eliminated from the list. Periodically, the network constructs new scheduling sets with all the available nodes unless it cannot make even one scheduling set with all the available nodes.

Energy Efficient Load Balancing Clustering Algorithm (EELBC) [8] is minimum heap based clustering algorithm. In this algorithm restricted nodes (node able to communicate with only one gateway) is assigned to their corresponding gateway. Then the min heap is build based on the no. of SNs in the gateway. The gateway having min nodes is the root of min heap. Now the nearest node is assigned to gateway and heap is rearranging again .Then the nodes from open nodes (able to communicate with more than one gateway) is assigned to gateway and within its communication range. The procedure is repeated until all the SNs are allotted to their correct gateway. This algorithm satisfy load balancing property by assigning SN to that CH having minimum no. of SN and conserve energy by assigning node to nearest CH.

5. Conclusion

WSNs are used in lots of applications in which sensed information must be gathered from remote location. In this paper we have discussed some of the challenges, energy efficient computation protocols and have given the comparison of these protocols as shown in Table 1. As the requirement of energy efficiency for each WSN differ based on application, it is infeasible to design a protocol which satisfies the requirements in all scenarios of WSN. Although these energy conserving techniques look promising, there are still many challenges that need to be solved in WSN. Therefore, further research is necessary for handling these kinds of situations

Table 1: Energy Efficient Techniques in WSN: Analysis

Protocol	***Category***	***Features***	***Advantages***	***Limitation***
DECSA[3]	Clustering Aggregation	Multihop communication	Avoid direct communication . Good for unbalanced Energy consumption	Extra overhead
EBMA[4]	Clustering Duty cycling Aggregation	Multihop communication	No overhead for CH selection Time slots reduce probability of collisions. Balance energy consumption.	Limited mobile agents
EEBCDA[5]	Clustering Aggregation	Unequal cluster size	Balance energy dissipation based on CH rotation.	Extra overhead
Energy Conserving Routing Algorithm for WSN[6]	Routing Duty cycling	Multiple sink reduces transmission latency	Distribute power consumption by constructing scheduling sets. While selecting next node energy as well as distance from BS also taken into consideration	Nodes closer to sink carry more traffic
YEAST[7]	Clustering Aggregation	Hybrid approach	Adaptive Reduces redundancy by exploiting spatial correlation Adjust size of correlated region according to application requirements	Requires position information
EELBC[8]	Clustering	Less complexity	Load balancing	Extra overhead in rearranging heap

References

[1] Jamal N. Al-Karaki, Ahmed E. Kamal, "Routing Techniques in Wireless Sensor Networks: A Survey", in proceeding of IEEE Wireless Communications, 2004 Vol.11, pp.26-28.

[2] David Culler, Deborah Estrin and Mani Srivastava ," Overview of Sensor Networks", in proceedings of IEEE Vol. 37, No. 8, 2004, pp. 41-49.

[3] Zhu Yong, Qing Pei," A Energy-Efficient Clustering Routing Algorithm Based on Distance and Residual Energy for Wireless Sensor Networks", in proceedings of International workshop on information and electronics engineering ,Vol. 29, 2012, pp. 1882-1888.

[4] Kai Lin, Min Chen, Sherali Zeadally, Joel J.P.C. Rodrigues," Balancing energy consumption with mobile agents in wireless sensor networks",in proceedings of Future Generation Computer Systems, 2012, pp. 446-456.

[5] Jun Yuea, Weiming Zhang, Weidong Xiao, Daquan Tang, Jiuyang Tang," Energy Efficient and Balanced Cluster-Based Data Aggregation Algorithm for Wireless Sensor Networks", in proceedings of International Workshop on Information and Electronics Engineering,Vol.11,2012,pp.2009-2015.

[6] Leandro A. Villas, Azzedine Boukerche, Horacio A.B.F. de Oliveira, Regina B. de Araujo, Antonio A.F. Loureiro," A spatial correlation aware algorithm to perform efficient data collection in wireless sensor networks",in proceedings of Elsevier, 2012.

[7] Yuping Dong, Hwa Chang, Zhongjian Zou and Sai Tang ," An Energy conserving Routing Algorithm for Wireless Sensor Networks", in proceedings of International Journal of Future Generation Communication and Networking,Vol.4, No. 1,2011,pp.39-54.

[8] Pratyay Kuila, Prasanta K. Jana," Energy Efficient Load-Balanced Clustering Algorithm for Wireless Sensor Networks", in proceedings of 2nd International Conference on Communication, Computing & Security,Vol.6, No. 1,2012,pp.771-777.

Dr. Kavita taneja received her Ph.D (Computer Science) in 2011 from Kurukshetra University, M.Phil (Computer Science) in 2007 from Alagappa University and Master of Computer Applications in 2002 from Kurukshetra University. She is working as an Associate Professor in M.C.A at M.M.I.C.T & B.M; MM University Since Feb 2003.She has published 20 papers in National/International journals/conferences. She has reviewed manuscript at international level. Her teaching and interests include Simulation & Modeling, Mobile Adhoc Networks and Wireless Sensor Networks. She has bagged Best Paper Award, 2007 at international conference for the paper entitled "Dynamic Traffic Conscious Routing for MANET" at DIT Dehradun.

Amandeep kaur received her M.Tech (Computer Science) in 2011 from Maharishi Markandeshwar University and B.Tech in 2005 from Kurukshetra University. She is pursuing Ph.D (Computer Science) from Maharishi Markandeshwar University. She is working as Assistant Professor in CSE Department at MM University since August 2005.Her teaching and intersts includes Design and Analysis of Algorithms, Operating System and Wireless Sensor Networks.

An Improved Particle Swarm Optimization Algorithm based on Membrane Structure

Yanhua Zhong[1] , Shuzhi Nie[2]

**[1]Depart. Of Electronics and Information Technology,Jiangmen Polytechnic
Jiangmen, China**

**[2]Depart. Of Electronics and Information Technology,Jiangmen Polytechnic
Jiangmen, China**

Abstract

Presented a new hybrid particle swarm algorithm based on P systems, through analyzing the working principle and improved strategy of the elementary particle swarm algorithm. Used the particles algorithm combined with the membrane to form a community, particles use wheel-type structure to communicate the current best particle within the community. The best particles, as Representative, compete for the optimal particle of the higher level. Utilized the Objective Functions to test the designed algorithm performance, compared with other particle swarm optimization algorithms, the experiment results shown that the designed algorithm has better performance in seeking Optimization solution quality, robustness and convergence speed.

Keywords: *P system; Particle swarm; Hybrid intelligent algorithm*

1.Introduction

Particle swarm optimization (PSO) is a class of a stochastic optimization algorithm based on swarm intelligence. In PSO algorithm, each individual look as a particle in a d-dimensional search space, fly at a certain speed in the search space, dynamically adjust the speed according to themselves and companion flying experience. Each of the particles has an objective function to determine the adaptation value, particles search the optimal solution followed the current optimal particle in the solution spaces, find the optimal solution by iteration. In each iteration, particle by tracking individual extreme values and global extreme values to update myself, in the process of looking for the two extremes, the particle updates own speed and location according to the corresponding location [1].

PSO has easy to understand and implement, the global search ability and other characteristics, much broader field of science and engineering concern, has become the fastest growing intelligent optimization algorithms. As PSO's performance depends on the algorithm parameters, to overcome these shortcomings, national researchers have proposed various measures for improvement. These improvements aim is to improve the particle diversity, and enhance the global exploration ability of particle, or to improve local development capacity, and enhance the convergence speed and accuracy [2-4].

There are usually two kind of combination way: First, used other adaptive optimization shrinkage factor, inertia weight and acceleration constants; second, combined with PSO and other evolutionary algorithm operators or other technology. Juang C F. combined the ant algorithm and the PSO for solving discrete optimization problems; Robinson et al [4] and Juang the combination of GA and PSO are used to optimize the antenna design and recurrent neural network design [4-5]; Documents [6] divided the population into more sub-population, and then the different sub-populations using PSO or GA or independent evolution of hill-climbing; Naka et al [7] have studied that the selection operation of GAs introduced into the PSO according to select a certain rate of reproduction to copy better individuals; Angeline [8] introduced tournament selected into the PSO algorithm, according to the current location of individual fitness, to each individual and compared to several other individuals, and then compare the results to the whole group based on sorting, the best half of the particle in the current group replace the worst half of the position and speed of the position and speed, while preserving the memory of individuals of each individual the best position; El-Dib et al [9] proposed crossover operation on the particle position and speed; Higashi et al [10] introduced Gaussian mutation to PSO; Miranda [11] used the variation, selection and breeding a variety of operating at the same time update the formula in the neighborhood of the best locations, and inertia weight and acceleration constants; Zhang et al [12-14] using differential evolution operator to select the speed of update formula in the best position of particles; And Kannan et al [15] using differential evolution to optimize the PSO inertia weight and acceleration constants. Document [16] proposed a discrete quantum individual PSO; Document [17] proposed the quantum PSO that update particle position based on behavior of quantum.

In this paper, combine membrane computing ideas with standard PSO, form a membrane community solution space, the particle in the community will be quick convergence solution space effects. The best particles as Representative compete the optimal particle of the higher level. This can effectively prevent the fall into local minimum, premature convergence or stagnation.

2.Hybrid Algorithm Design

2. 1P System

Studied DNA computing for many years, inspired by biological cells, Gheorghe Păun proposed the membrane computing in 2000 [3], through dealing with compounds from layered structure of living cells to abstract the computing model. The model is called membrane systems or P system, A P system of n dimensional can be expressed as the following formula:

$$\Pi = (V, T, C, \mu, w_1, ..., w_m, (R_1, \rho_1), ..., (R_m, \rho_m)) \quad (1)$$

In formulary (1), V is alphabet, its elements are called objects; $T \subseteq V$ is output alphabet; $C \subseteq V - T$ is catalyst, its elements don't change in the evolutionary process, don't create new characters too, however, perform some evolution rules required its; μ is a membrane structure that contain m membranes, each membrane and its enclosed area show with a label set of H, $H = \{1, 2, \cdots, m\}$, then m is Π's dimensional; $w_i \in V^*$ ($1 \le i \le m$) show multiple sets that contain objects in the region i of membrane structure m, V^* is the collection of arbitrary character string composed of V's character; Evolution rules is a binary set of (u, v), Usually written as $u \to v$, u is the string of V^*, $v = v'$ or $v = v'\delta$, v' is the character string of collection $\{a_{here}, a_{out}, a_{in_j} | a \in V, 1 \le j \le m\}$, δ is the special characters don't belong to V, When a rule contains δ, membrane is dissolved while the implementation of the rules, regard the length of u as the rules of $u \to v$ 'radius, R_i ($1 \le i \le m$) is the finite set of evolution rules, each R_i is associated with the region of i in the membrane structure of μ, ρ_i is the second order in R_i, which is the preferential relations, Which is the preferential relation that implementation the rules of n.

In short, P system consists of three parts: hierarchical structure of membrane; the multiple set of objects; evolution rules.

2.2Particle Swarm Optimization

Particle swarm optimization (PSO) is initialized to a group of random particles (random solutions), and then iterate to find the optimal solution. In each iteration, the particles by tracking the two extreme to update their own, the first one is the particle itself to find the optimal solution, this solution is called individual extreme; the other one is the whole population to find the optimal current solution, this extreme is a global extreme. In addition you can also use part of the particles rather than the entire population as a neighbor; the extreme in all neighbors is the local minimum.

Suppose a D-dimensional target in the search space, there are N particles to form a community in which the i-th particle is represented as a D-dimensional vector

$$X_i = (x_{i1}, x_{i2}, \cdots, x_{iD}), \quad i = 1,2,\cdots,N \quad (2)$$

The first i particles "flying" velocity is also a D-dimensional vector, expressed as

$$V_i = (v_{i1}, v_{i2}, \cdots, v_{iD}), \quad i = 1,2,\cdots 3 \quad (3)$$

Each particle also maintains a memory of its previous best position, represented as

$$p_{best} = (p_{i1}, p_{i2}, \cdots, p_{iD}), \quad i = 1,2,\cdots,N \quad (4)$$

The optimal position can be search in whole group which can be represented as

$$g_{best} = (p_{g1}, p_{g2}, \cdots, p_{gD}) \quad (5)$$

When found these two optimal values, the particles update their speed and position according to the following formula (6) and (7).

$$v_{id} = w * v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \quad (6)$$

$$x_{id} = x_{id} + v_{id} \quad (7)$$

Where, c_1 , c_2 are learning factor, also known as the acceleration constant, r_1 , r_2 are random numbers in Range [0,1].

Equation (6) on the right consists of three parts, the first part of the "inertia (inertia)" or "momentum (momentum)" part, reflects the movement of particles "habit (habit)", the particle has the speed to maintain their previous trend; the second part of "cognitive (cognition)" part, reflects the historical experience of the particles on their own memory (memory) or memory (remembrance), the particle has the best location close to their historical trends; third part of "social (social) "part, reflects the particle collaboration and knowledge sharing between groups of historical experience, the particle with the group or neighborhood close to the historical best location trend.

According to experience, Usually take $c_1 = c_2 = 2$, $v_{id} \in [-v_{\max}, v_{\max}]$ is velocity of particle, r_1 , r_2

is random numbers in Range [0,1], $v_{\max}$ is a constant, set by the user to limit the speed of the particle.

2. 3Hybrid Algorithm

Particle swarm optimization based on Membrane computing, is assigned to different groups of particles in the membrane, a membrane is equivalent to the community, the membrane particles are selected the best of its specific particles and use the difference between him and the best neighbor vector instead of the standard PSO velocity update formula "part of individual consciousness." Topology of populations of membrane between each particle using wheel-type structure, can further improve the performance of the algorithm to avoid premature convergence phenomenon. In this structure, each particle with the best center particle exchange information in the membrane, the best particle in the community as representatives of the best particle in the higher level competition, the topology shown in Figure 1. With a knockout way to find this optimal solution to improve the speed of dissemination of information to avoid loss of information to better improve the efficiency of the algorithm.

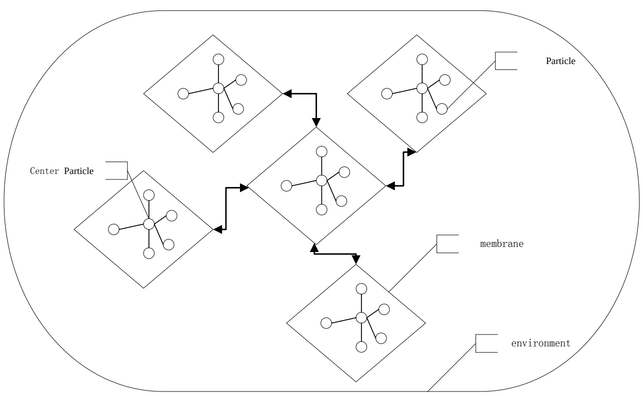


Fig. 1 Particle swarm based on P systems

Particle swarm optimization based on P system process is as follows:

- Step 1: the establishment of the structure contains m membranes, each membrane and the surrounding areas use label set $H=\{1,2,\cdots j\cdots,m\}$.
- Step 2: The particles were randomly assigned to the m -membranes.
- Step 3: initialize the particle swarm, including population size N^j, the number of j representatives of membranes, each particle position x^j_i and velocity V^j_i.
- Step 4: to assess the fitness value $F^j_{it}[i]$ of each particle;
- Step 5: For each particle, compare its fitness value $F^j_{it}[i]$ and individual extremum p^j_{best}（i）, if $F^j_{it}[i] > p^j_{best}(i)$, Then replace p^j_{best}（i） with $F^j_{it}[i]$;
- Step 6:the best particle p^j_{best} in the membranes j , compare its fitness value $F^j_{it}[i]$ and Global extremum g_{best} , if $F^j_{it}[i] > p^j_{best}$, Then re place g_{best} with $F^j_{it}[i]$;;
- Step 7: According to formula (6), (7), update the particle velocity v^j_i and position x^j_i.

Step 8: If the end conditions are met to elect the best particle, the algorithm ends, otherwise return to Step 4.

3.Algorithm test and analysis

To verify the performance of the algorithm, we proposed in this paper based on particle swarm optimization membrane structure (mPSO) with standard PSO algorithm (sPSO) [1], differential evolution particle swarm optimization [12,13], GA particle swarm optimization [6] for comparison. In the experiment, we selected five standard optimization functions to test these algorithms, each function uses 10-dimensional search space.

3. 1Test functions

1) F1 De Jone's function

The simplest test function is De Jong's function 1. It is also known as sphere model. It is continuos, convex and unimdal. Function definition:

$$f_1(\bar{x}) = \sum_{i=1}^{D} x_i^{\ 2}; x \in [-5.12, 5.12] \tag{8}$$

Global minimum: $f(\bar{0}) = 0;$

2) F2 Axis parallel hyper-ellipsoid

The axis parallel hyper-ellipsoid is similar to De Jong's function 1. It is also known as the weighted sphere model. Again, it is continuos, convex and unimodal. Function definition:

$$f_2(x) = \sum_{i=1}^{D} i x_i^{\ 2}; x \in [-5.12, 5.12] \tag{9}$$

Global minimum: $f(0) = 0;$

3) F3 Sum of different Powers

The sum of different powers is a commonly used unimodal test function. Function definition:

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

$$f_3(\bar{x}) = \sum_{i=1}^{D} |x_i|^{(i+1)} \quad ; x_i \in [-1,1] \qquad (10)$$

Global minimum: $f(0) = 0;$

4) F4 Rastrigin's function

Rastrigin's function is based on function 1 with the addition of cosine modulation to produce many local min-ima. Thus, the test function is highly multimodal. However, the location of the minima is regularly distributed. Function definition:

$$f_4(\bar{x}) = \sum_{i=1}^{D} 10 + x_i^{\ 2} - 10 \times \cos(2\pi x_i); x \in [-5.12, 5.12] \qquad (11)$$

Global minimum: $f(0) = 0;$

5) F5 Schwefel's function

Schwefel's function is deceptive in that the global minimum is geometrically distant, over the parameter space, from the next best local minima. Therefore, the search algorithms are potentially prone to convergence in the wrong direction. Function definition:

$$f_5(\bar{x}) = \sum_{i=`}^{D} -x_i \times \sin\sqrt{|x_i|}; x \in [-500, 500] \qquad (12)$$

Global minimum: $f(420.9687) = 418.9829;$

We use the standard initialization range, but the particular function to facilitate the display of its performance, its initialization to narrow the scope of [-1, 1]. For all of the function we use the 10-dimensional search space, Table I shows the five test function the initialization and the search space.

Table 1: Function initialization settings

test function	**Search domain**	**Initialization range**
F1 De Jone's function 1	[-5.12 , 5.12]	[-1 , 1]
F2 Sum of different Powers	[-1 , 1]	[-1 , 1]
F3 Rastrigin's function 6	[-5.12 , 5.12]	[-5.12,5.12]
F4Axis parallel yper-ellipsoid	[-5.12 , 5.12]	[-5.12,5.12]
F5 Schwefel's function 7	[-500,500]	[-500 ,500]

3.2Performance analysis

As the optimization result is random, for scientific comparison algorithms, we take the average of the results which independently test each test function 50 times. For each run, the end result of 1000 iterations to compare the accuracy of algorithm convergence. Can be found by observing, MPSO has better optimization ability for various functions, especially for multimodal function of search capability is much higher than other comparison algorithms. Table 4.2 shows Statistics for average optimal results of the algorithm which the five Objective functions are run 50 times independently.

Table 2: Convergence value of average

F	**Di**	**N**	**Gma**	**Average**			
	m		**x**	**S PSO**	**DEPSO**	**GAPSO**	**MPSO**
F1	10	50	1000	0.91232	9.3759e-023	0.0028096	6.8731e-005
F2	10	50	1000	1.0425e-007	7.1588e-022	1.0247	2.7028e-005
F3	10	50	1000	0.17698	2.7959e-036	3.3251e-007	9.0747e-008
F4	10	50	1000	90.653	43.751	14.467	0.23902
F5	10	50	1500	-3777.1	-2883.7	-2793.5	-4189.8

In the table, F is testing function ,N is the search space, N is running times, Gmax iterative times is 1000.

3.3Algorithm parameters analysis

Some of the parameter settings for MPSO, such as population size, may have an important effect on the performance of the algorithm, so experiment to verify the relationship between algorithm parameters and performance. The function F5 was be selected as the evaluation function to test the algorithm convergence precision under different parameter settings. Assuming population 30, particle number 30 in one membrane, Exchange particle 20 at a time, switching time 40 generations, the following is assuming other conditions remain unchanged following the case of change in one parameter experimental results are as follows:

- Number of particles within populations, population sizes was taken 30, 40, and 50.
- Population size (number of membranes), were taken twenty, thirty, forty populations.
- Exchange particle time, exchange time were taken 30 generations, 40generations, 50 generations.
- The number of particles each exchange, the number of exchange particles were taken 15, 20, 25.

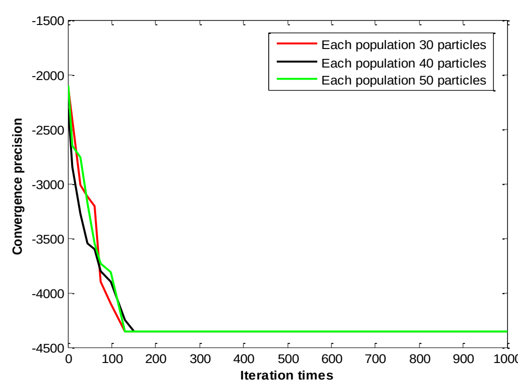


Fig. 2 Number of particles within populations

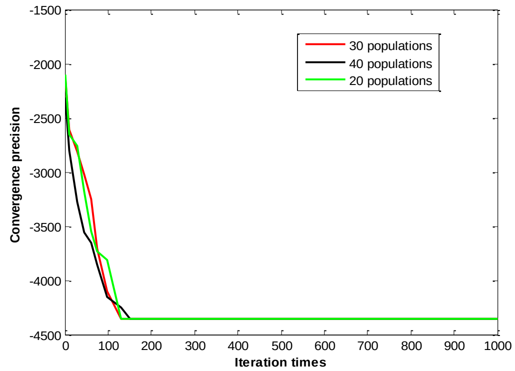


Fig. 3 Population size

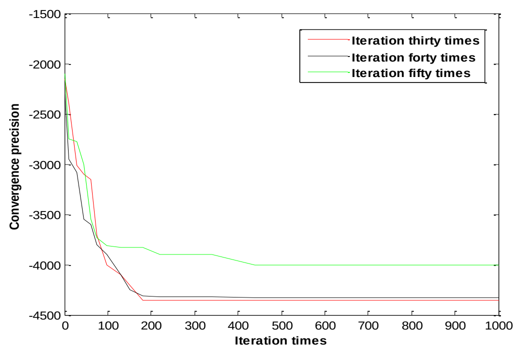


Fig. 4 Exchange particle time

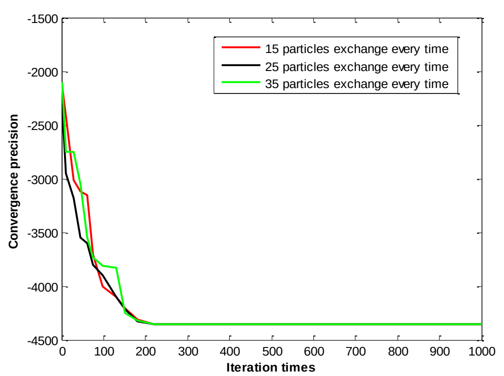


Fig. 5 The number of particles each exchange

By several key parameters for mPSO to be set different values, experimental results as follows: As show in the Figure 2 that the number of particles within populations has little effect on convergence accuracy. In each set 40 particles within a population have the best convergence results. Figure 3 shows the number of membrane was set in the experiment, the number of membrane were set to (20, 30, 40), the populations that was set 40 has obvious advantages.

As seen in Figure 4, the results of different exchange timing which were to take timing (30, 40, 50) generation. When iteration take to the 30 that the results is best, However, The algorithm cannot converge to the global minimum when the exchange timing was set 50. The number of particles each exchange are shown in Figure 5, When the particles were set in (15, 20, 25), the optimization effect is had little fluctuation, but each time exchange 15 particles is better.

4.Conclusion

In this paper, proposed a new hybrid particle swarm algorithm based on P system, which is through the analysis of elementary particle swarm algorithm the working principle and improvement a strategy. And the algorithm is excellent in finding extreme for multimodal function. But there is still much scope for improvement in this algorithm, population exchange between the particles are too simple, stupid, and tend to a single. Future improvements can be used with the guidance of the dynamic evolution of membranes exchange, according to the algorithm to optimize conditions for specific particle exchange operation. Hybrid evolutionary algorithm is one trend in the field of evolutionary algorithms, so that the advantages of different algorithms can be integrated and "better, faster, cheaper" to get the solution of the problem, it is a valuable research direction.

References

[1] Kennedy J, Eberhart R. Particle swarm optimization. in: Proceedings of the 4th IEEE International Conference on Neural Networks, Piscataway: IEEE Service Center, 1995, pp.1942 -1948.
[2] Zhang G. X,Liu C. X,Gheorghe M.,Pǎté F. Solving satisfiability problems with membrane algorithms. Proceedings of the Fourth International Conference on Bio-Inspired Computing: Theories and Applieations.2009:29-36.
[3] Fen Zhou, et al, .A Particle Swarm Optimization Based on P systems. 2010 Sixth International Conference on Natural Computation (ICNC 2010) .pp.3003-3007.
[4] Gh. Pˇ aun. Computing with membranes. Journal of Computer and System Sciences.2000, 1(61):108-143.
[5] Juang C F. A Hybrid of Genetic Algorithm and Particle Swarm Optimization for Recurrent Network Design. IEEE Transactions on Systems, Man and Cybernetics (Part B-Cybernetics), 2004, 34 (2): 997-1006.
[6] Krink T and Løvbjerg M, The Lifecycle Model: Combining Particle Swarm Optimization, Genetic Algorithms and Hill Climbers. In Proc. of PPSN 2002, pp. 621-630.
[7] S. Naka, T. Genji, T.Yura, and Y. Fukuyama. A hybrid particle swarm optimization for distribution state estimation. IEEE Trans. Power Syst, pp. 60-68, Feb. 2003.
[8] Angeline P.J.Using Selection to Improve Particle Swarm Optimization. In Proc of the IEEE Congress on Evolutionary Computation. Anchorage, USA, 1998,84-89.
[9] El-Dib, H. Youssef, M. El-Metwally, and Z. Osman. Load flow solution using hybrid particle swarm optimization. In

Proc. Int. Conf.Elect., Electron. Comput. Eng., Sep. 2004, pp. 742-746.

[10] Higashi N, Iba H. Particle Swarm Optimization with Gaussian Mutation. in Proc. of the 2003 Congress on Evolutionary Computation. 2003. 72289.

[11] V. Miranda and N. Fonseca. EPSO: evolutionary particle swarm optimization, a new algorithm with applications in power systems. in IEEE/PES Transmission and Distribution Conf. Exhibition Asia Pacific, Oct. 2002, vol. 2, pp. 745-750.

[12] W. Zhang and X. Xie. DEPSO: Hybrid particle swarm with differential evolution operator. in Proc. IEEE Int. Conf. Syst., Man, Cybern., Oct. 2003, vol. 4, pp. 3816-3821.

[13] H. Talbi and M. Batouche, Hybrid particle swarm with differential evolution for multimodal image registration. in Proc. IEEE Int. Conf. Ind. Technol., Dec. 2004, vol. 3, pp. 1567-1572.

[14] P. Moore and G. Venayagamoorthy. Evolving combinational logic circuits using particle swarm, differential evolution and hybrid DEPSO. Int. J. Neural Syst., vol. 16, no. 2, pp. 163-177, 2006.

[15] S. Kannan, S. Slochanal, and N. Padhy. Application of particle swarm optimization technique and its variants to generation expansion problem. ELSERVIER Electric Power Syst. Res., vol. 70, no. 3, pp.203-210, Aug. 2004.

[16] [S. Yang, M. Wang, and L. Jiao. A quantum particle swarm optimization. in roc. IEEE Congr. Evol. Comput., Jun. 2004, vol. 1, pp.320-324.

[17] J. Sun, Particle Swarm Optimization with Particles Having Quantum Behavior. in Proc. 2004 Congress on Evolutionary Computation, pp. 325-331.

IBCAV: Intelligent Based Clustering Algorithm in VANET

Mehrnaz Mottahedi[1], Sam Jabbehdari[1] and Sepideh Adabi[1]

[1]Computer Engineering Department, North Tehran Branch, Islamic Azad University, Tehran, Iran

Abstract

Study of Vehicular Ad-hoc Networks (VANETs) is among challenges facing researchers today. Nodes in such networks have a relatively high speed, while there is a great shortage of the time needed for data transmission between them as a result of node speed difference as well as persistent changing of the network. Consequently, routing algorithms and modes of access to the conductor is considered as a serious challenge in such networks. In this context, researchers have introduced various protocols, including cluster-oriented methods to overcome such challenges. In cluster-oriented methods, it is of vital importance to select a sustainable cluster with due regards to the high speed and density of nodes. The proposed method is called Intelligent Based Clustering Algorithm in VANET (IBCAV), The present project seeks to improve routing algorithms in VANETs by employing inter-layered methods, awareness of the existing traffic flow as well as combination of various factors on the basis of a smart method based on artificial neural network. Here cluster size, speed and density of nodes are the factors which have taken into account. Finally, our simulated results show that the IBCAV outperforms better than AODV, DSR and epidemic routing in terms of packet delivery ratio, end-to-end delays and throughput.

Keywords: *VANET, Routing, Clustering, artificial neural network*

1. Introduction

VANET is a special form of MANET which is a vehicle to vehicle and vehicle to roadside wireless communication network. It is an autonomous and self-organizing wireless communication network. The purpose of VANET is transferring data with high speed and minimizing the delay in a communication range. Due to the nature of mobile nodes in the VANET, the location is constantly changing. Given the distinct features of VANET, different problems are tackled by the researchers. The main part of research works has focused on routing algorithms. Since the wireless nodes in VANET, i.e., vehicles are faster than nodes in a usual MANET and the mobility patterns of vehicles are confined to road maps and thus more predictable, selecting a suitable routing algorithm is significant. Two main types of common routing protocols in VANET are Proactive and Reactive.

Proactive algorithms rely on the tables containing nodes data. Since vehicles are in a constant change of their positions accompanied by rapid change of network topology, therefore the routing data in those tables rapidly lose their validity. To keep validity of the huge amount of data should be communicated between the nodes, which cause such algorithms to use a high bandwidth. Proactive algorithms are also known as table-based algorithms. DSDV, OLSR and STAR are referred as the algorithms used for routing purposes, which as expected, use a high bandwidth [1,2]. Reactive algorithms, on the other hand, operate to detect a route only when there is a need to a route. AODV, DSR and TORA are algorithms that use this method for routing.

Disadvantage of the proactive method relates to its poor scaling system, while the problem with reactive method is that before transmission of the first packet it needs to detect the route, thus leading to longer transmission time and when the route detected from its original point of departure to its destination may disappears even before transmission of the first packet due to the high movement of the nodes. Meanwhile, active duration of a route declines with the number of hops. A route may fade away at the same speed it appeared [3,4].

The moving nature of nodes in VANETs and the fact that they are constantly changing positions, has made researchers use the location of the nodes in order to be able to communicate with the specific node for transmission of messages, an initiative which resulted in routing based on position. In routing based on node position, a device, for example, Global Positioning System can help pinpoint the position of every node as well as that of the destination node. GSR, GPSR are algorithms that use this method for routing [5,6].

Also due to the high mobility of nodes in VANET the topology is changing constantly, so in this situation to keep the communication between nodes all nodes must constantly sending packages to update their tables that have led to widespread storms in the network. The research on the issue of scalability in VANET where there are a high number of nodes, it has been shown that the clustering approach is an efficient solution to the scalability issue.

In clustering methods, the moving nodes are categorized into different groups and stand together in one single cluster. Nodes in each cluster can have different roles, including cluster head, member node, or gateway node. Cluster head acts as a coordinating agent of that cluster and is responsible for transmission of data and making arrangement for such data transmission between clusters. Gateway nodes establish communication between two clusters. The advantages of clustering method that encourage using such method: have less overhead as compared with non-clustering methods, making it easier to redeploy resources. Therefore, this method increases the system capacity. Since it does not need the data of all nodes in the network, this network is smaller than any other moving terminal. Also, clustering methods can be suitable for creating scale acceptability in the network. COIN, CBLR and CBR are algorithms that use this method for routing [7,8].

The rest of the paper is organized as follows: Section 2 introduces IBCAV. Section 3 relates the header selection and training an Artificial Neural Network with Genetic Algorithm. Section 4 presents simulation and experimental results. Finally, section 5 concludes the paper.

2. IBCAV

In the IBCAV some important factors are considered such as cluster size, velocity and density. In clustering methods, moving nodes are divided into different groups and stand together in one single cluster based on certain rules. The size of each cluster represents the number of vehicles in a street or highway within a communication range and form a cluster. Within each cluster a vehicle should be designated as cluster head. What makes a vehicle qualified to become a cluster head depends on a series of computations, time and consumption of network resources, that is why the size of cluster is of prime importance. In the case of a small cluster, the vehicle chosen as cluster head can rapidly leave the confines of the cluster, causing a recurrent algorithm of choosing new cluster head. In the case of a large cluster the vehicle chosen as cluster head stay there longer even though there might be another vehicle which could be more efficient as a cluster head. Therefore, it is of great importance that the cluster size should neither be too large nor too small.

Clustering techniques are usually employed in VANETs to reduce resources consumption and improve the efficiency of the network. Vehicles have specific role in clustering and only a few of them chosen as cluster heads carry or transmit packets of data. Thus, once vehicles are placed inside a cluster it is very important to select a cluster head. In the IBCAV, RSU is chosen as a cluster head if it is within the confines of the cluster. That is because RSUs have stronger processing capabilities than other nodes and since they are motionless, there is no need to change the cluster head. In general, they could provide better services as compared to other nodes. However, in the event RSU is not present within the confines of the cluster, a vehicle with slower speed is designated as a cluster head. It is assumed that this vehicle would stay longer within the confines of the cluster as compared with a vehicle with higher speed.

In addition to the vehicle with lower speed, other factors are also considered in the IBCAV, selecting a vehicle with lower speed that stands in an appropriate position as the cluster head. Appropriate position here means: density of vehicles around it is greater and to be in a spot that with due consideration of the moving direction of the cluster stay longer within the confines.

Following header selection, a Store-Carry-Forward concept is used to send a message. In fact, if cluster heads fail to stand within the communication range of one another, they store the message inside their buffer and as soon as a cluster head is in range, they transmit a Hello message to them. [9,10,11].

3. Header Selection

So far, key factors of clustering and header selection introduced and explained how each one helps to improve clustering process. Now in order to place all factors next to each other, need a smart function or algorithm to combine them. In the IBCAV, take up an artificial neural network for header selection [12]. By training it through an algorithm, can combine the effective factors in such a way that the network turns into an optimized one with due regards to the defined optimization criteria. The advantage of this method is that such optimization criteria allow VANETs to be specifically defined based on their physical environment. In addition, it would permit to train the artificial neural network to meet optimization criteria as closely as possible.

3.1. Training an Artificial Neural Network with Genetic Algorithm for choosing Header

For the purpose of training such a network by using genetic algorithm, can employ an artificial neural network with five inlets and one outlet channels. Such a network conforms to the figure.1.

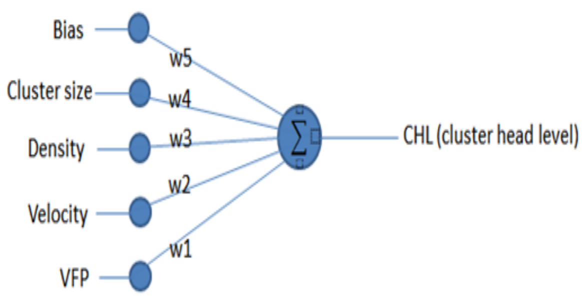


Figure.1: Scheme of the proposed artificial neural network

In the relation.1, (CHL) represents cluster head level. It determines the score that each vehicle inside the cluster has in order to become a cluster head. A vehicle with higher score could be selected as cluster head.

CHL= w5* Bias + w4*Cluster size + w3* Density + w2* Velocity +w1*VFP (1)

where VFP (Vehicle in cluster Flow Position) indicates the place of vehicle in relation to the cluster movement, velocity stands for the node speed and density stands for the related nodes connected in every direction. Bias represents the threshold for different assessments. Here bias is attached to fixed amount of 1. Therefore w5 weight determines the amount of bias. Cluster size stands for the size of cluster and the inlet connected to the neuron with fixed amount of 1.Therefore w4 weight determines the appropriate cluster size.
The issue is to find w1 to w5 in a manner that the network has the best output. But it is enough to determine the weights w1 to w4 for the network. It is necessary to note that w5 is considered as a bias network, is used as the threshold for evaluation and its value does not depend on the output of optimality criteria.

4. Simulation and Experimental Results

4.1 Simulation setting

After training the artificial neural network and employing it for the purpose of simulation, here present the result of our findings together with preliminary conditions and the parameters used.
The simulation was carried out using NS3 software [13]. Common and general parameters, default settings for all stimulations and the final results can be found in Table.1.

Table.1: Simulation Parameters

Simulation Parameters	Value
Number of simulation run time	1100
Highway Length	1040 m
Velocity	Varied between 0 to 29 m/s
Power Transmission	41.5 W
Number of nodes	Varied between 90 to 100

The working method is that the algorithm starts to work with a casual population composed of a population between 20 to 100. Then simulating units carry out simulation process and the clustering units choose suitable cluster head based on the weights specified in each repetition of the algorithm. Once simulation process ends, the algorithm calls the cross over or mutation operations and error function to change the population and direct searching to the optimal outlet. To adjust weights by using algorithm a genetic composition as explained before, a working current similar to figure.2 has been considered. In the process of adjustment of artificial neural network weights, the number of repetition for the used genetic algorithm, which is very important, can be obtained experimentally.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

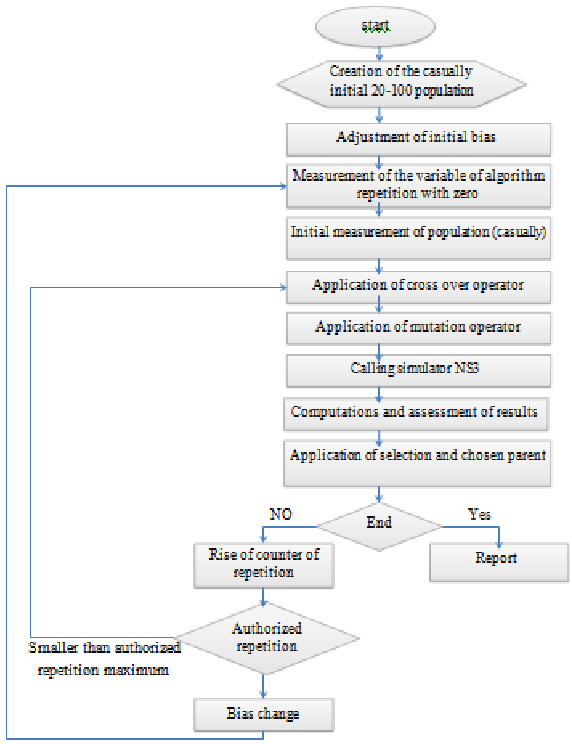


Figure.2: Training of artificial neural network based on the IBCAV

4.2. Simulation Results and Discussions

Cluster head election affects control packets and usage of networks resources. The better operation of the clustering method and cluster head election, the lesser will be the number of operations for cluster head election or change of cluster head. Therefore, overhead and used resources of the network will decline. It is shown in the figures.3-4 that the IBCAV has effectively decreased the number of this operation taking into account these factors.

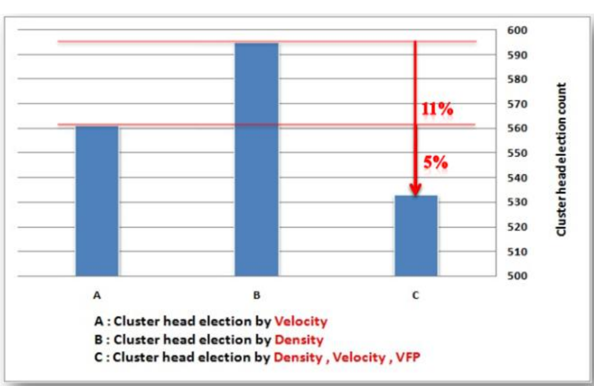


Figure.3: The number of cluster head election in the IBCAV

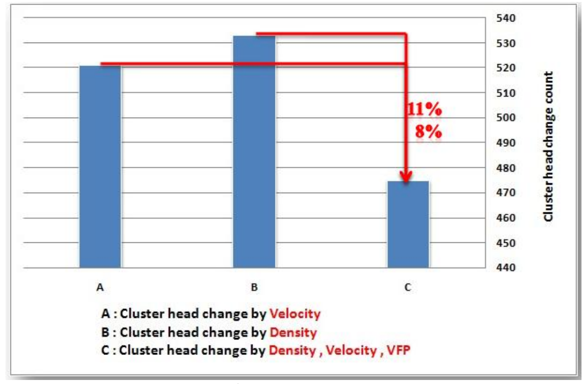


Figure.4 : The number of cluster head switching in the IBCAV

In the IBCAV, after the cluster head is chosen, a Store-Carry-Forward concept is used to send a message. In fact, when cluster heads are outside the communication range of the cluster head that intends to send message, it stores the message inside its buffer and as soon as a cluster head enters into its communication range it transmits a Hello message for that cluster head.

By taking this rule into account, based on figure.5 packet delivery ratio stand in a range of 82.66 % while, failure to use this rule, the ratio stands in a range of 22.66%.

The packet delivery ratio is defined as the percentage of packets that successfully reach the receiver nodes.

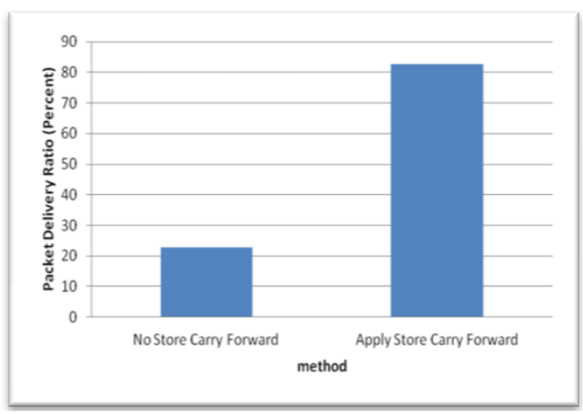


Figure.5: Packet Delivery Ratio in the IBCAV with presence and absence of Store-Carry-Forward

In the simulation process mentioned so far, vehicles speed varies between 0 to 29 m/s. Figure.6 shows the effects of higher speeds.

As it can be seen, the higher the speed of vehicles the lower becomes the packet delivery ratio. The decline in the simulation carried out is estimated at 50%. Figure.6 implies the appropriate speed of vehicles to be something between 0 to 29 meter per second. Because, according to the figure, with changes in speed between 0 to 29, packet delivery ratio will slightly changes and is nearly fixed and equal to 90%.

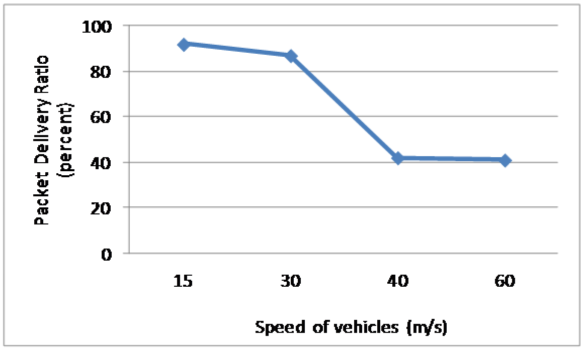


Figure.6: Packet Delivery Ratio in the IBCAV with different speeds

Figure.7 refers to the impact of the number of vehicles on the packet delivery ratio. As it can be seen the higher number of vehicles causes packet delivery ratio to increase. Simulation puts such rise at about 17%.

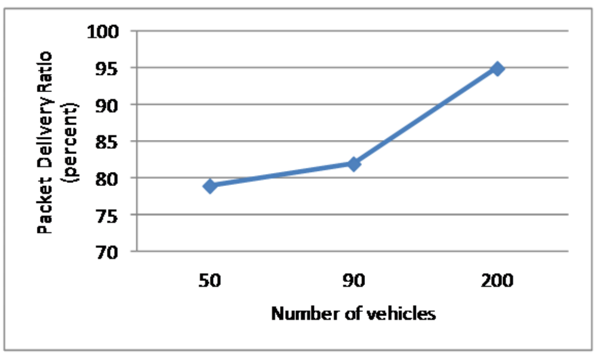


Figure.7: Packet Delivery Ratio in the IBCAV with considering the number of vehicles

This section presents some simulation results of the performance evaluation of the IBCAV in VANET and compares it with AODV, DSR and epidemic routing.
The performance of these algorithms is evaluated in terms of packet delivery ratio, end-to-end delay and throughput. The packet delivery ratio is defined as the percentage of packets that successfully reach the receiver nodes. The end-to-end delay is defined as the time between a packet being sent and being received. Throughput is the average rate of successful packet delivery over a communication channel.

Packet Delivery Ratio- Figure.8 illustrates the packet delivery ratio. IBCAV performs well. AODV performs worst due to its dropping data packets for which there is no valid route. In the IBCAV theses rates are about 18%, 17%, 2% higher than AODV, DSR and epidemic routing respectively.

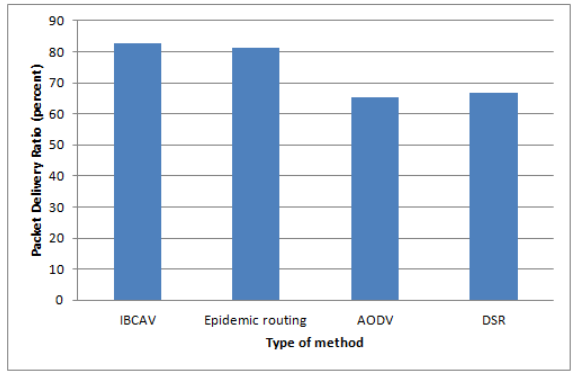


Figure.8: Packet Delivery Ratio in comparison with common algorithms

End-to-End Delay- The comparison of the end-to-end delay is shown in figure.9, DSR shows largest delay due to inefficient manner of handling route failure. On the other hand, IBCAV shows the lowest delay. The measurements in the IBCAV are about 17%, 24%, 11% lower than AODV, DSR and epidemic routing respectively.

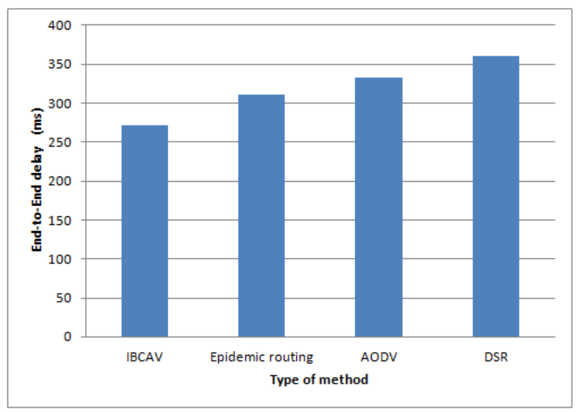


Figure.9: End-to-End Delay in comparison with common algorithms

Throughput- In figure.10 throughput in the IBCAV is compared with AODV, DSR and epidemic routing. Throughput in the IBCAV in comparison with other method is about 37%, 41% and 16% higher respectively.

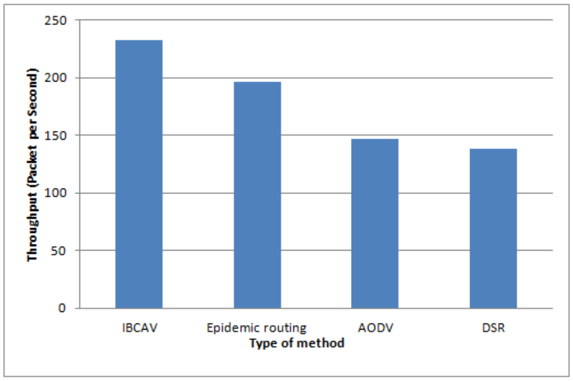


Figure.10: Throughput in comparison with common algorithms

5. Conclusions

In brief, simulation results indicate that in the IBCAV a combination of cluster size, velocity and density altogether, as compared to a manner in which each of the said elements is considered separately for header selection can, on one hand, reduce cluster head selection operation followed by a reduced usage of network resources and on the other hand with the decline of end-to-end delay, throughput is increased. Moreover such results reflect that the IBCAV outperforms in comparison to AODV, DSR and epidemic routing. In the IBCAV control packets of network decreases and packet delivery ratio increases.

References

[1] Sh. XU, S. Lee, "A Direct Routing Algorithm with Less Route Built Throughput", 2012 International Conference on Traffic and Transportation Engineering (ICICN 2012).
[2] U. Nagaraj, Dr. M. U. Kharat, P. Dhamal, "Study of Various Routing Protocols in VANET", International Journal of Computer Science & Technology, IJCST Vol. 2, Issue 4, 2011.
[3] K. Pandy, A. Swaroop, "A Comprehensive Performance Analysis of Proactive, Reactive and Hybrid MANETs Routing Protocols", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 3, 2011.
[4] R. Kumar, M. Dave, "A Comparative Study of Various Routing Protocols in VANET", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, 2011.
[5] C. Tee, A. C. R. Lee, "Survey of position based routing for Inter Vehicle Communication system", in Distributed Framework and Applications, First International Conference, pp.174-182, 2009.
[6] Ch. Durga, A. Anitha, Dr. K. Krishna, "An Efficient Protocol for Intelligent Transportation in Vehicular Adhoc Networks", 2011, http://www.ijaest.iserp.org. All rights Reserved.
[7] B. Ramakrishnan, "Performance Analysis of AODV Routing Protocol In Vehicular Ad-hoc Network Service Discovery Architecture", RPN Journal of Systems and Software, VOL. 2, NO. 2, 2012.
[8] M. Rezaee, M. Yaghmaee, "Cluster based Routing Protocol for Mobile Ad Hoc Networks", 2009.
[9] B. Ramakrishnan, Dr. R. S. Rajesh, R. S. Shaji, "CBVANET: A Cluster Based Vehicular Adhoc Network Model for Simple Highway Communication", Int. J. Advanced Networking and Applications Volume: 02, Issue: 04, (2011).
[10] S. Teshima, T. Ohta, E. Kohno, Y. Kakuda, "A Data Transfer Scheme Using Autonomous Clustering in VANETs Environment", 2011 IEEE DOI 10.1109/ISADS.
[11] Y. OHTA, T. OHTA, E. Kohno, Y. KAKUDA, "A Store-Carry-Forward-based Data Transfer Scheme Using Positions and Moving Direction of Vehicles for VANETs", 2011 Tenth International Symposium on Autonomous Decentralized Systems.
[12] H. F. Leung, H. K. Lam, S. H. Ling, K. S. Tam, "Tuning of the Structure and Parameters of a Neural Network Using an Improved Genetic Algorithm", IEEE transactions on neural networks, VOL. 14, NO. 1, pp.79-89, 2003.
[13] M. H. Rehmani, "Personal Notes on the Network Simulator NS-3", 2010.

Mehrnaz Mottahedi received the M.Sc. degree in computer engineering from the department of computer engineering, North Tehran Branch, Islamic Azad University in 2012. Her research interests include Information Security, MANETs, VANETs and Wireless Sensor Networks.

Sam Jabbehdari currently working as an assistant professor at the department of Computer Engineering in IAU (Islamic Azad University), North Tehran Branch, in Tehran, since 1993. He received his both B.Sc. and M.S. degrees in Electrical Engineering Telecommunication from K.N.T (Khajeh Nasir Toosi) University of Technology, and IAU, South Tehran branch in Tehran, Iran, in 1988, through 1991 respectively. He was honored Ph.D. degree in Computer Engineering from IAU, Science and Research Branch, Tehran, Iran in 2005. He was Head of Postgraduate Computer Engineering Department IAU-North Tehran Branch during 2008-2010. He also has written "Advanced Topics in Computer Networks" book in Persian Language (Tehran, Classic, 2009). His current research interests are Scheduling, QoS, MANETs, Wireless Sensor Networks and Grid Computing Systems.

Sepideh Adabi received the B.Sc., M.Sc. and Ph.D. degrees in computer and software engineering from the Islamic Azad University, Tehran, in 2006, 2008 and 2012, respectively. She is currently an assistant professor in the Department of Computer Engineering at Islamic Azad University-North Tehran Branch in Tehran, Iran. Her research interests include distributed systems, grid and cloud computing, and resource management. She is a member of the IEEE.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Design of Secondary Bootloader for Embedded System based on DSP

Jie Yan[1,2], Xiaosu Xu[1,2], Lihui Wang[1,2], Yiting Liu[1,2]

**[1] School of Instrument Science and Engineering, Southeast University
Nanjing, 210096, China**

**[2] Key Laboratory of Micro Inertial Instrument and Advanced Navigation Technology, Ministry of Education
Nanjing, 210096, China**

Abstract

The paper desirables the secondary bootloader based on the TMS320C6713B which connected with external 16 bit width FLASH memory. The secondary bootloader design is one of the difficulties in the project development of the embedded system based on DSP. The paper briefly introduces the EMIF and FLASH chip features. Then the hardware interface design is given. The linker command options are used to create boot copy table and the secondary bootloader for large-scale program code is designed. The last, the paper introduces the process of all the design steps. The result shows that the paper offers the method of the secondary bootloader is convenient and useful. This method not only can be used in the TMS320C6713B but also can be used in the other TMS320C6000s.

Keywords: *Digital Signal Processor(DSP), secondary Bootloader, FLASH memory.*

1. Introduction

With the development of electronic technology, the digital signal processor (DSP) has gained rapid development. TI has created TMS320C6000 series digital signal process with advanced very long instruction word (VLIW) DSP core that makes the chip with superior performance. TMS320C6000s are widely used in radar, audio data processing, images data processing and other demanding real-time data processing field. In the DSP system development, complex signal processing is necessary, so that need to use large amounts of data and algorithm program. However, the ROM spaces in the DSP chip so little that it cannot store large program and data. In this case, program code and data need to be stored in the external ROM. When the DSP reset, program code and data which stored in the external ROM need to be moved from external ROM to the DSP chip internal.

The first task to consider is how to copy the program code and data which are stored in the external ROM to DSP chip internal RAM. The different chips have the different ways to boot. Take TMS320C6000s for example, they offer three types of boot configuration.1) No boot process. 2) ROM boot process. 3) Host boot process. The most commonly selected boot configuration is the ROM boot process.

Take C671x series DSP chip for example, when ROM boot is selected as the boot configuration, 1k byte of code and data will automatically be copied from CE1 to address 0 by the Enhanced Direct Memory Access (EDMA) following the release of RESET while the CPU is stalled. When the program code exceeds 1k byte, a special code which called the secondary bootloader should be developed to copy the additional section of code not copied by the ROM boot.

In practice, the program code often exceeds 1k byte, so we should develop the secondary bootloader code. When reset the system, 1k byte code automatically copied by the EDMA from CE1 space in the address 0x90000000 to DSP internal RAM in the address 0x0000000. When this 1k byte code executed, the CPU will copy the remaining code from the CE1 space in the address 0x90000400 to DSP internal RAM in the address 0x00000400, and then begins to run the main function.

2. Hardware Design

The secondary bootloader for the hardware has contact with two parts of the system. The first part is DSP chip, which used in the system is TMS320C6713B. The second part is FLASH chip. In the system, the FLASH chip model is SST39VF800A produced by SST.

2.1 TMS320C6713B and SST39VF800A

The TMS320C6713B device is a floating-point DSP, which based on the high-performance. Operating at 225MHz, the C6713B delivers up to 1350 million float-point operation per second (MFLOPS), 1800 million instructions per second (MIPS) [1].

The C6713B integrated 32-bit external memory interface (EMIF), and the external space capacity of 64MB. The external space divided into 4 independent spaces such as (CE0~CE3) which can be different control. The data bus width is 32 bit, and also support 8 / 16bit registers are read / write. The EMIF support a glueless interface to a variety of external devices, including SDRAM, SBSRAM, ROM, FIFOs [2].

The SST39LF800A is produced by SST. The chip writes (Program or Erase) with single 3.3V power supply. When the chip is in the active module, the active current is typical 20mA. When the chip is in the standby mold, the standby current is typical 3μA. The address bus width is 19 bits, and the data bus is 16 bits [3].

2.2 System hardware interface design

The DSP bootloader mainly associates with DSP EMIF module and FLASH (SST39VF800A). The data bus (D0~D15) in the FLASH correspondingly connect with the data bus (ED0~ED15) in the C6713B. The address bus (A0~A18) in the FLASH correspondingly connect with the address bus (EA2~EA20) in the C6713B. $\overline{\mathrm{CE1}}$ pin in the C6713B connects with the corresponding $\overline{\mathrm{CE}}$ in the FLASH. $\overline{\mathrm{CE}}$ pin is used for device selection. When $\overline{\mathrm{CE}}$ is high, the chip is deselected, otherwise the chip is selected. $\overline{\mathrm{AOE}}$ pin in the C6713B connects with $\overline{\mathrm{OE}}$ pin in the FLASH. $\overline{\mathrm{OE}}$ is the output control and is used to gate data from the output pins. The data bus is in high impedance state when either $\overline{\mathrm{CE}}$ or $\overline{\mathrm{OE}}$ is high. $\overline{\mathrm{AWE}}$ pin in the C6713B connects with $\overline{\mathrm{WE}}$ pin in the FLASH. $\overline{\mathrm{WE}}$ is input control and is used to latch data from the input pins. A command is written by asserting $\overline{\mathrm{WE}}$ low while keeping $\overline{\mathrm{CE}}$ low. Figure1 shows the EMIF to SST39LF800A interface.

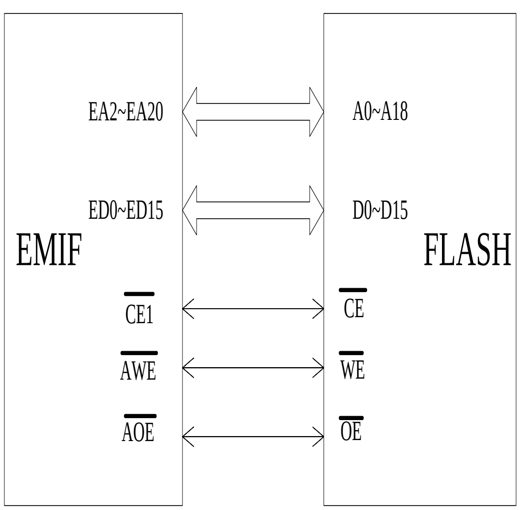


Figure1. EMIF To SST39LF800A Interface

In this system, the FLASH chip data bus width is 16 bit, so we should set the appropriate EMIF registers to 16 bit width. However, the EMIF default bus width is 32 bit. When EMIF connect with 16bit FLASH, EMIF automatically changes the double 16 bit data read from FLASH into a 32 bit data. Each read operation of EMIF is always carries in 32 bit data. The output of the address will automatically shift in order to ensure that the memory read operation in the 16 bit width. Therefore, when EMIF connects with 16 bit width data bus, the EMIF address automatically left one, and the high bit will be automatically removed from discarded [4].

3. Secondary Boot Program Theory

When system resets, the EDMA will automatically copy 1k byte program code from external FLASH. In this moment, the CPU has not yet started work. When EDMA has been copied all 1k byte code from CE1 to address0, the CPU begins to work. The code copied by the EDMA is called secondary bootloader code [5]. Once the project has been built and linked, the secondary bootloader code should be written. The secondary bootloader is typically written in assembly language because the C environment is not initialized at boot-time. The secondary bootloader always achieve following several features:

The first point
Configure the PLL: This step is recommended for C6713B that has software programmable PLL in order to configure the registers in PLL. After configure PLL, the CPU begins to work, and executes the remaining code.

The second point
Configure the EMIF: In order to achieve the glueless connect with external SDRAM, FLASH, the EMIF registers such as EMIF Global Control Register (GBLCTL), EMIF CE Space Control Register (CECTL) and so on.

The third point
Copy the initialized section code from the FLASH: This step achieve the function that copies the code from FLASH which the address space occupy the CE1 in the address 0x90000400 to DSP internal run address 0x00000400.

The fourth point
Call_c_int00: Secondary bootloader copies the complete application to its runtime memory. At completion, the secondary bootloader code branches to _c_int00

The last
Go to Main: After Call_c_int00, the system can support the C environment, the C program can be copied to DSP Main Program and run.

Figure2 shows start-up sequence of application which uses secondary bootloader.

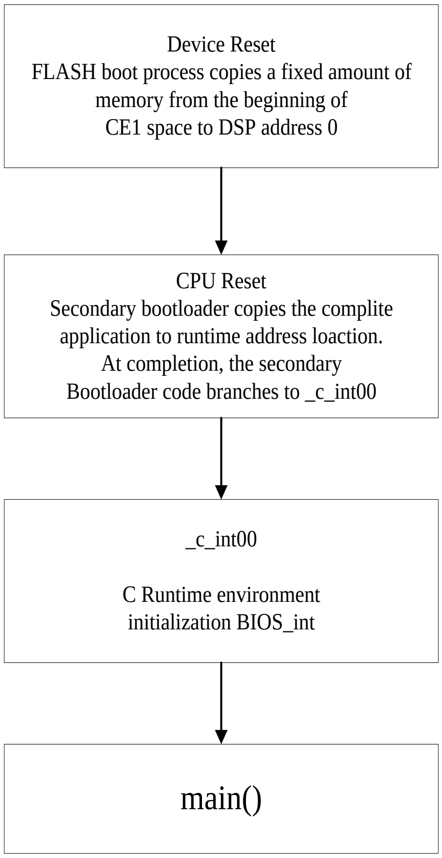


Figure2. Start-up Sequence of Application

4. Secondary Bootloder Copy Table

The secondary bootloader copies the contents of memory sections from its load address to its run address using a section copy table. The copy table contains entries for all the memory sections that need to be copied from their load address to their run address. Each table entry contains information describing the size of the section of memory, the destination address or address from where the section will execute, and the source address or the address where the section was loaded.

There are a number ways to create the section copy table, 1) Inspecting the map file. If this method is chosen to create the copy table, then the appropriate value of the size of the section, the destination address of the section, and the source address of the section must be manually filled in for that entry. 2) Using the –boot option in the hex conversion utility. The hex conversion utility provided with Code Composer Studio provides a more convenient method for creating the section copy table by automatically building the copy table when the appropriate options are specified in the hex conversion utility command file. 3) Using linker options.

This embedded system based C6000 DSP uses the linker options to create the section copy table. This method is simple to use, so selects this way to create copy table. The linker options include three parts such as LOAD_START, RUN_START, and SIZE. The user gives the definition of all parts. In the following code is the definition in the link2.cmd file [6].

```
ROMS
{
  FLASH:org= 0x9000000 ,
  Len = 0x40000,
  Rowwidth = 8,
}

.SECTIONS
{
  .boot_load : load = FLASH_BOOT,
               run =BOOTRAM

  .text : load = FLASH_REST,
          run = IRAM

   LOAD_START(_text_ld_start),

   RUN_START(_text_rn_start),

   SIZE(_text_size)

  .cinit      >       FLASH_REST
  .const      >       IRAM
  .stack      >       IRAM
  .bss        >       IRAM
  .data       >       IRAM
  .far        >       IRAM
  .switch     >       IRAM
  .sysmem     >       IRAM
  .cio        >       IRAM

  }
```

In the following code is the definition about the copy table in the boot_c671x.s62 file.

```
copyTable:

  ; count
  ; flash start (load) address
  ; ram start (run) address

      ;; .text
        .word _text_size
        .word _text_ld_start
        .word _text_rn_start

  ;; end of table
        .word 0
```

```
.word 0
.word 0
```

When function project has built successfully in the CCS, boot_c671x.s62 file and link2.cmd file will be load to the project in order to generate COFF format file.

5. Programming Flash

Code Composer Studio (CCS) comes with several utilities that help with flashing applications into ROM. For C6000s a Flash programmer tool is offered in the CCS. A hex conversion utility is also provided. Because the Flash programmer works only with the hex format, the COFF format file obtained from CCS must be converted to .hex file through the hex conversion utility [7] [8].

The following procedure describes how to create an application to program into flash. First of all, build the target project in order to generate the .out file. Secondly, use the hex conversion utility to create a .hex file from the .out file [9]. The last, DSP program the flash with the .hex file. Figure 3 shows the flash programming sequence.

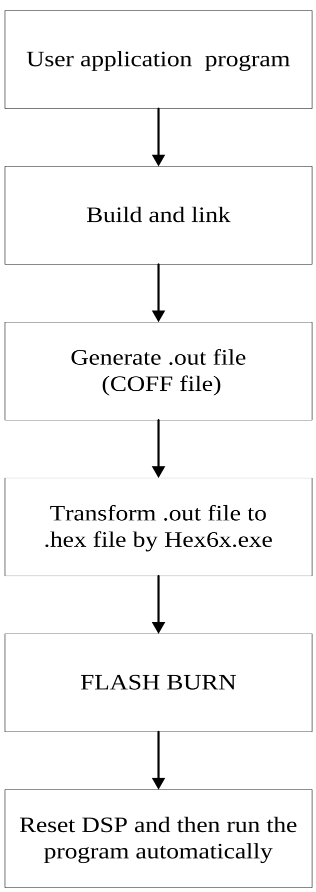


Figure 3. Flash Programming Sequence

The boot.cmd code lists in the following which can convert the .out file to .hex file.

```
BOOT.out
-map BOOT.map
-a
-image
-zero
-memwidth 8

ROMS
{
FLASH: org = 0x90000000,
len = 0x40000,
romwidth = 8,
files = {BOOT.hex}
}

SECTIONS
{

.boot_load
.text
.cinit
.pinit
.const
.switch

}
```

The SST39LF800A FLASH supports sector and block erase operation [10]. The sector size is 4k byte, and the block size is 64k byte. Flash erase operation is on the basis of sector and block. Once FLASH programming, user should erase the chip, and then program the code into the Flash. The Sector Erase operation is initiated by executing a six-byte command sequence with Sector-Erase command (30H) and sector address (SA) in the last cycle. The Block-Erase operation is initiated by executing a six-byte command sequence with Block-Erase command (50H) and block address (BA) in the last bus cycle. Any commands issued during the Sector or Block-Erase operation are ignored.

6. Conclusions

As is known to all, in the development of the embedded system which is based on DSP, the internal ROM is so little to storage too much user code. Hence, a second-level bootloader should be created by user, which is a difficult task in the DSP development. The paper which is based on TMS320C6713B describes an easy and convenient way to create a secondary bootloader. This method uses ROM boot process and uses linker options to create the section copy table. The experiment results show that the method offered by this paper is effective and convenient. The method is not only could be used in C6713B device also can be used in other C6000 devices.

7. Acknowledgment

The project was supported by following funds: National Natural Science Foundation of China (Nos. 51175082, 61203192, 61273056), Natural Science Foundation of Jiangsu Province (BK2012326), Open Research Fund of State Key Laboratory of Transient Optics and Photonics, Chinese Academy of Sciences (SKLST201108).

References

[1] Texas Instruments Incorporated, “TMS320CB713B FLOATING-POINT DIGITAL SIGNAL PROECSSOR”, June, 2006.

[2] Texas Instruments Incorporated, “TMS320C6000 DSP External Memory Interface (EMIF) Reference Guide”, April, 2008.

[3] Silicon Storgae Technology, Inc, “2 Mbit/ 4Mbit / 8Mbit (x16) Multi-Purpose Flash”, 2002.

[4] SanHengXing Technology, TMS320C6713 DSP Theory and Application, BeiJing Electronic Industry Press，2009.

[5] Texas Instruments Incorporated, “Creating a Second-Level Bootloader for FLASH Bootloading on TMS320C6000 Platform with Code Composer Studio”, May, 2006.

[6] Zhang guolong, Xu xiaosu, “Implemention of Secondary Bootloader for Large-scale Embedded System”, Computer Engineering, vol. 36, No.13, July, 2010, pp 219-221.

[7] Vijayarajeswaran, R., “DSP Implementation of a Power Factor Correction Strategy for BLDC Motor Drive” , International Journal of Computer Science Issues, v 9, n 1 1-3, January 2012, pp 215-220.

[8] Radhakrishnan N.and Ramaswamy M. “DSP implementation of Fuzzy based Power Quality enhancement strategy for IPOS converter fed drive”, International Journal of Computer Science Issues, v 9, n 2 2-2, 2012, pp 301-305.

[9] Suthar, A.C., Vayada, Mohammed and Patel, C.B. et al, “Hardware Software co-simulation for Image Processing Applications”, International Journal of Computer Science Issues, v 9, n 22-2, January 2012, pp 560-562

[10] Chen daiyuan, “External FLASH Memory’s Bootloader System for C6000,” Telecommunication Engineering, vol. 49, No5, May, 2009, pp. 86-88.

Jie Yan, male, is now a Ph.D. graduate in the School of Instrument Science and Engineering, Southeast University, Nanjing, China. His main research interest is Embedded Systems, SINS/GNSS integrated navigation.

Xiaosu Xu, male, received his Ph.D. degree in the School of Automation, Southeast University, Nanjing, China, in 1991. He is a professor in the School of Instrument Science and Engineering at Southeast University, also is the director of Key Laboratory of Micro Inertial Instrument and Advanced Navigation Technology, Ministry of Education, China. His current research interests include inertial navigation, integrated navigation.

Lihui Wang, male, received BS, MS, and Ph.D. degrees, all from Harbin engineering university, China, in 2002, 2005, and 2009, respectively. From 2009, he joined post-doctoral research center in Southeast University, China, and now he is an associate professor in southeast university. He has been engaged in developing fiber optic gyroscopes (FOG) and fiber optic current sensor (FOCS). His major research interests are error analysis of fiber optic sensors, application of FOG in strapdown inertial navigation system and FOCS in electric power system.

Yiting Liu, male, received his master degree in detection technology and automatic equipment form Wuhan instititude of technology, Hubei,China. And now, he is a Ph.D. graduate in the School of Instrument Science and Engineering, Southeast University, Nanjing, China. His main research field is embedded system and inertial navigation.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

The Affect of Multi-source Time Delay to the Human Balance Based on Robust Control Model

Hongrui Wang [1], Kun Liu[2] and Jinzhaung Xiao[3]

[1] **Institute of Electrical Engineering, Yanshan University**
Qinhuangdao, Hebei 066004, PR China

[2] **Institute of Electrical Engineering, Yanshan University**
Qinhuangdao, Hebei 066004, PR China

[3] **College of Electronic and Information Engeering, Hebei University**
Baoding, Hebei 071000, PR China

Abstract

Time-delay is a prevalent phenomenon in human body during standing and locomotion, and it is one of the main factors to affect human balance ability. The purpose of this study is to design a human balance control model, and to analyze the human possible posture control mechanism with multi-source delay considered. Based on an inverted pendulum with the ankle joint model, we derived and proved that the human body existed multi-source time delay, and then designed a robust state feedback controller with an observer for the generalized system, which is with delay state and delay control inputs in continuous time. This controller can be considered as the central nervous system (CNS) for the human balance controller mathematical model. Finally, applying simulation software, we obtained the balance regulation kinematics responses to anterior and posterior (A/P) surface translations, and discussed the effect on the dynamic response of the human body to different kinds of the time-delay. The robust controller can stable the human system within larger interference and larger time-delay. This paper provides a useful method for analyzing the impact of the multi-source time delay on the human balance system.

Keywords: *human balance, multi-source time delay, robust controller, surface translation.*

1. Introduction

The phenomenon of time-delay happens prevalently in human body during standing and locomotion, and it is one of the main factors to affect human balance ability [1]. In reality, accident or aging may result in the increase of delay time, and the specific performances include slow motion, reaction time extension or sensory organ insensitivity [2, 15]. Since the human body is not statically stable, maintaining upright posture requires continuous action of tonal adjustments in the antigravity muscles [15]. The time-delay which comes from different organs of the body, has adverse effects to balance ability of the human body [4, 9]. So it is necessary to analyze various delays impacted on body balance capacity objectively and efficiently.

Engineering models have been developed that effectively describe aspects of human balance [3,4]. Some scholars [5] model the human body as a single segment, single joint inverted pendulum that rotates about the ankle joint; Jiang and Kimura [6] model it as a more complex dynamic model including five joints; Gawthrop, Loram, and Lakie [14] model it as an inverted pendulum that is balanced by an active muscle working in series with a tendon; Robert J. and Schilling, Senior purposed a phase-locked loop model of the response of the postural control system to periodic platform motion[13].

Standing posture control strategies now are commonly believed to be a fundamental motor skill learned by the central nervous system (CNS) [9]. The behavior of the human postural control system was approximated by various systems such as feed forward/feedback[14, 11], linear (P/PD/ PID) control[2] [5] [6], optimal control [8, 15], and predictive control [10]. In the previous research of human postural control system, many researchers have considered the impact of delay to balance ability [5, 8]. Qu Xingda established a human body mathematical model with the approximation time delay in the method of Taylor series expansion [8], as the time delay constant from body sense organs. In analyzing human delay model, John Milton, Juan Luis Cabrera, and Toru Ohira et al attempted to set the controller delay in paper [11]. However, their human body model was oversimplified about the important delays. And they neither analyzed the potential larger delay from patients, nor considered the effect delay from multiple organs human on human balance ability. Even many studies ignored the effects of time delays in biomechanical applications of the inverted pendulum to human balance control.

Based on the inverted pendulum with ankle joint model, we derivate the kinetic equation of human musculoskeletal model, and proposed a robust controller as the central nervous system (CNS) for the human balance controller mathematical model. Then the controller parameters K were calculated by using Matlab LMI ToolBox. Finally, applying simulation software, we obtained the balance regulation kinematics responses to anterior and posterior (A/P) surface translations, and discussed the effect on the dynamic response of the human body to different kinds of the time-delay.

2. Model formulations

The balance maintaining control system can be assumed as a continuous time feedback control system [5]. The neural controller with an upright reference position can sense the error when body sway from it, and then sends commands to various muscles to keep the body upright [7].

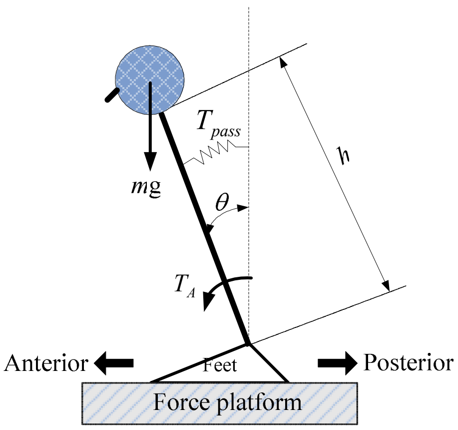


Fig. 1. A inverted pendulum model. The model consists of links with ankle joint. *m* represents the total mass of the inverted pendulum, *h* the distance from ankle joint to the total center of mass (CoM) of the pendulum. T_{pass} is the viscous damping coefficient.

The dynamical equation for the inverted pendulum model of the body musculoskeletal was given by:

$$J\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = mgh\sin(\theta) + T_{\mathrm{A}} + T_{pass} + T_d$$
$$T_{pass} = k_e\theta + k_v\dot{\theta} \tag{1}$$

where, θ the was the sway angle; the input T_A was the total ankle torque, which was produced by CNS; m was mass of the musculoskeletal segment; g was the acceleration of gravity; h is the distance from the ankle joint to COM; J was the moment of inertia of the musculoskeletal segment; T_{pass} was the passive torque from joint viscous and elastic; viscous and elastic coefficient were k_v and k_e respectively; and T_d was the disturbance torque. Without loss of generality, we set $\sin\theta \approx \theta$ to simplify the human body system, for the angular scope is less than $\pm 5^\circ$ during stable posture regulation [6].

The dynamical equation (1) did not consider any delay. Actually, the time delay is one of the most significant factors which would affect the standing stability of the human body [8,9,11]. The main delays include: the sensory delay, the controller delay and the musculoskeletal movement delay [8, 9]. In order to study how the effect of the various delays affect the human balance ability, we represented the musculoskeletal movement delay as τ_1, and controller delay as τ_2. Especially, the sensory delay can be regarded as a part of controller delay here, but we did not ignore it.

Considering the effect of multi-source delay, the human body model dynamical equation should be expressed as:

$$J\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = mgh\theta(t-\tau_1) + u(t-\tau_2) + T_{pass} + T_d$$
$$T_{pass} = k_e\theta + k_v\dot{\theta}$$
$$0 \le \tau_i \le d_0, i = 1,2 \tag{2}$$

In view of the dynamical function (2) and the delay conditions derive from the above, we considered the human body as a contnuous time linear system with time delays:

$$\dot{\mathrm{x}}(t) = \mathbf{A}\mathrm{x}(t) + \mathbf{A}_{\mathrm{d}}\mathrm{x}(t-d_0) + \mathbf{B}u(t-d_0))$$
$$z(t) = \mathbf{C}\mathrm{x}(t)$$
$$\mathrm{x}(0) = 0, t \in [-d_0, 0] \tag{3}$$

where, $\theta = x_1, \dot{\theta} = x_2$, $\mathbf{x}(t) = [x_1(t) \quad x_2(t)]$ was the state vector; T_d was a non-determined partial, and it could be omitted in order to facilitate the analysis; $z(t)$ is the system output, which was given by test data; and the matrix C was obtained by regulation;

$$\mathbf{A}=\begin{bmatrix}0 & 1\\ \frac{k_e}{J} & \frac{k_v}{J}\end{bmatrix}\ \mathbf{A_d}=\begin{bmatrix}0 & 0\\ \frac{mgh}{J} & 0\end{bmatrix},\ \mathbf{B}=\begin{bmatrix}0\\ \frac{1}{J}\end{bmatrix}$$

3. CNS Controller design

The sense organs of the human body can be considered a state observer, and it provided state information to CNS for motion control. We proposed a continuous time controller with an observer as human CNS:

$$\begin{aligned}\dot{\hat{\mathbf{x}}}(t) &= \mathbf{A}\hat{\mathrm{x}}(t)+\mathbf{A}_{\mathrm{d}}\hat{\mathrm{x}}(t-d_0)+\mathbf{B}\,\mathrm{u}(t-d_0))\\ &\quad +L(z(t)-\hat{z}(t))\\ \hat{z}(t) &= \mathbf{C}\hat{\mathrm{x}}(t)\\ \mathrm{u} &= \text{-}\mathrm{K}\hat{\mathrm{x}}(t)\end{aligned} \tag{2}$$

where, L was the observer gain, K is the constant gain matrix. and state error was defined by:

$$e(t)=x(t)-\hat{x}(t) \tag{3}$$

Substituting the observer system equation (2) to (3), we got:

$$e(t)=x(t)-\hat{x}(t)=(\mathbf{A}-LC)e(t)+\mathbf{A}_d E(t-d_0) \tag{4}$$

Substituting the control law $\mathrm{u}=\text{-}\mathrm{K}\hat{\mathbf{x}}(t)$ into the time-delay system state error (3) and (4) yields

$$\begin{aligned}\dot{\hat{\mathrm{x}}}(t) &= \mathbf{A}\hat{\mathrm{x}}(t)+\mathbf{A}_d\hat{\mathrm{x}}(t-d_0)+\mathbf{B}K\hat{\mathrm{x}}(t-d_0))\\ &\quad+\mathbf{LC}e(t)\\ \dot{e}(t) &= \dot{x}(t)-\dot{\hat{x}}(t)=\mathbf{A}\mathrm{x}(t)+\mathbf{A}_d\mathrm{x}(t-d_0)\\ &\quad-\mathbf{A}\hat{\mathrm{x}}(t)+\mathbf{A}_d\hat{\mathrm{x}}(t-d_0)+\mathbf{LC}(x(t)-\hat{x}(t))\end{aligned} \tag{5}$$

The human body system (1) can be asymptotically stable by the controller (4) with observer, if the $\mathbf{A_L}=\mathbf{A}-\mathbf{LC}$ is stable, and there exist positive-definite matrices P, Q such that

$$\begin{bmatrix}\mathbf{H_1} & \mathbf{LC} & \mathbf{A_dP^{-1}} & \mathbf{0}\\ \mathbf{(LC)^T} & \mathbf{H_2} & \mathbf{0} & \mathbf{QA_d}\\ \mathbf{P^{-1}A_d} & \mathbf{0} & -\mathbf{P^{-1}} & \mathbf{0}\\ \mathbf{0} & \mathbf{A_dQ} & \mathbf{0} & -\mathbf{Q}\end{bmatrix}<0 \tag{6}$$

where some terms are defined as follows:

$$\mathbf{H_1}=\mathbf{P^{-1}A^T}+\mathbf{AP^{-1}}+\mathbf{P^{-1}}$$

$$\mathbf{H_2}=\mathbf{(A-LC)^TQ}+\mathbf{Q(A-LC)}+\mathbf{Q},$$

the state feedback gain $\mathbf{K}=\mathbf{B^TP}$.

Proof. A candidate Lyapunov functional is defined as:

$$\begin{aligned}V(t)&=V_1(t)+V_2(t)+V_3(t),\\ V_1(t)&=\hat{\mathrm{x}}^T(t)\mathbf{P}\hat{\mathrm{x}}(t),\\ V_2(t)&=e^T(t)\mathbf{Q}e(t),\\ V_3(t)&=\int_{t-d_0}^{t}\hat{\mathrm{x}}^T(s)P\hat{\mathrm{x}}(s)+e^T(s)\mathbf{Q}e(s))\,\mathrm{d}\,s.\end{aligned} \tag{7}$$

Taking the derivative of the Lyapunov functional (7) along the solution of Eq. (5) yields

$$\begin{aligned}\dot{V}_1(t)&=\dot{\hat{x}}^{\mathrm{T}}(t)\mathbf{P}\hat{x}(t)+\hat{x}^{\mathrm{T}}(t)\mathbf{P}\dot{\hat{x}}(t)\\ &=\hat{x}^{\mathrm{T}}(t)(\mathbf{A^TP}+\mathbf{PA})\hat{x}(t)+\hat{x}^{\mathrm{T}}(t)(\mathbf{PA_d}-\mathbf{PBK})\\ &\quad\hat{x}(t-d_0)+\hat{x}^{\mathrm{T}}(t-d_0)(\mathbf{A_d}^{\mathrm{T}}\mathbf{P}-\mathbf{B}^T\mathbf{K}^T\mathbf{P})\dot{\hat{x}}(t)\\ &\quad+e^{\mathrm{T}}(t)(\mathbf{LC})^{\mathrm{T}}\mathbf{P}\hat{x}(t)+\hat{x}^{\mathrm{T}}(t)\mathbf{PLC}e(t)\\ \dot{V}_2(t)&=\dot{e}^T(t)\mathbf{Q}e(t)+e^{\mathrm{T}}(t)\mathbf{P}\dot{e}(t)\\ &=e^{\mathrm{T}}(t)(\mathbf{A}-\mathbf{LC})^T\mathbf{Q}+\mathbf{Q}(\mathbf{A}-\mathbf{LC})e(t)+\\ &\quad e^{\mathrm{T}}(t)\mathbf{QA_d}e(t-d_0)+e^{\mathrm{T}}(t-d_0)\mathbf{A_d^TQ}e(t)\\ \dot{V}_2(t)&=\hat{x}^T(t)\mathbf{P}\hat{x}(t)-\hat{x}^{\mathrm{T}}(t-d_0)\mathbf{P}\hat{x}(t-d_0)+\\ &\quad e^{\mathrm{T}}(t)\mathbf{Q}e(t)-e^{\mathrm{T}}(t-d_0)\mathbf{Q}e(t-d_0)\end{aligned} \tag{8}$$

$$\dot{V}(t)=\dot{V}_1(t)+\dot{V}_2(t)+\dot{V}_3(t) \tag{9}$$

Then, substituting the state feedback gain $K=\mathbf{B^TP}$ to equation (9),

$$\begin{aligned}\dot{V}(t)&\le\hat{x}^{\mathrm{T}}(t)(\mathbf{A^TP}+\mathbf{PA}+\mathbf{P})\hat{x}(t)+\\ &\quad\hat{x}^{\mathrm{T}}(t)\mathbf{PLC}e(t)+\hat{x}^{\mathrm{T}}(t)\mathbf{PA_d}\hat{x}^T(t-d_0)\\ &\quad+e^{\mathrm{T}}(t)(\mathbf{LC})^TP\hat{x}(t)+e^T(t)[(\mathbf{A}-\mathbf{LC})^T\mathbf{Q}\\ &\quad+\mathbf{Q}(\mathbf{A}-\mathbf{LC})+\mathbf{Q}]e(t)+e^{\mathrm{T}}(t)\mathbf{QA_d}e(t-d_0)\\ &\quad+\hat{x}^{\mathrm{T}}(t-d_0)\mathbf{A_d^TP}\hat{x}(t)+\hat{x}^{\mathrm{T}}(t-d_0)(-\mathbf{P})\\ &\quad\hat{x}(t-d_0)+e^{\mathrm{T}}(t-d_0)\mathbf{A_d^TQ}e(t)-\mathbf{Q}\end{aligned} \tag{10}$$

Let $\boldsymbol{\varphi}(t)=\begin{bmatrix}\mathbf{P}\hat{x}(t)\\ e^T(t)\\ \mathbf{P}\hat{x}^T(t-d_0)\\ e^T(t-d_0)\end{bmatrix}^{\mathrm{T}}$ we can obtained the function

$$\dot{V} \le \boldsymbol{\varphi}^T(t)\begin{bmatrix} \mathbf{H_1} & \mathbf{LC} & \mathbf{A_d P^{-1}} & \mathbf{0} \\ \mathbf{(LC)^T} & \mathbf{H_2} & \mathbf{0} & \mathbf{QA_d} \\ \mathbf{P^{-1}A_d} & \mathbf{0} & -\mathbf{P^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A_d Q} & \mathbf{0} & -\mathbf{Q} \end{bmatrix}\boldsymbol{\varphi}(t) \tag{11}$$

where, $\mathbf{H_1} = \mathbf{P^{-1}A^T} + \mathbf{AP^{-1}} + \mathbf{P^{-1}}$, $\mathbf{H_2} = \mathbf{(A - LC)^T Q} + \mathbf{Q(A - LC)} + \mathbf{Q}$. The matrix inequality (6) can make the $\dot{V} \le 0$, so the system could be stable by the controller, proof completed.

Anthropometric parameters were calculated according a student in Yanshan University ($m = 72\text{kg}$, $h = 0.9\text{m}$, $J = 77.8\text{kg}\cdot\text{m}^2$), $k_e = 0.8$, k_v=4.0[18], the state parameters are:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0.01 & 0.052 \end{bmatrix}$$

$$\mathbf{A}_d = \begin{bmatrix} 0 & 0 \\ 8.2 & 0 \end{bmatrix}$$

$$\mathbf{B}_1 = \mathbf{B}_2 = \begin{bmatrix} 0 \\ 0.013 \end{bmatrix}$$

Then, applying Matlab LMI Toolbox[19], the solution of Eq.(16) can be solved. We got the result of

$$K = \begin{bmatrix} -1372.5 \\ -319.1 \end{bmatrix}.$$

4. Model simulation and result analysis

The modeling and simulations were performed in matlab/simulink(the MathWork inc., Natick, MA, USA). Anthropometric parameters were calculated according to the same student in Yanshan University, whose parameters were used to design the controller gains *K*.

4.1 Kinematics responses to surface translations

A sigmoid signal was used to translate the platform 1cm in 100 ms at a peak velocity of 35 cm/s in posterior direction at the 10^{th} second after the simulation starting [20]. This stimulus was regarded as a pulse signal to the soles of the feet as fig.2.

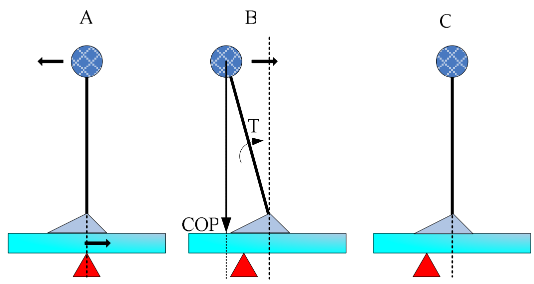


Fig.2. Kinematics response to posterior surface translation. The red triangle represents a ground reference. The blue rectangle represents the motion platform. Short black arrow indicates the direction of movement. A) The platform backward translations, so that the body center of gravity comes forward. B) Platform stopped movement in 100 ms, and the human body started to restore the balanced posture. C) The body returns to a balance posture.

More specifically, we added the interference value $\theta(10) = \delta \cdot s$ and $\dot{\theta}(10) = 0$ to current human musculoskeletal motion state $x(t)$ at the 10^{th} second. Here, δ was defined by $\delta \approx h^{-1}$ according to the individual body. The sudden motion of platform will destroy the equilibrium state of human body, so the robust controller will dynamically regulate musculoskeletal to keep upright standing posture. We obtained the balance adjustment process data, and solved COP with the formula $COP = h \cdot \sin(\theta)$. $w(t)$ was the Gaussian noise, with zero mean and unity variance. Meanwhile, we tested the student in reality, and recorded the COP data.

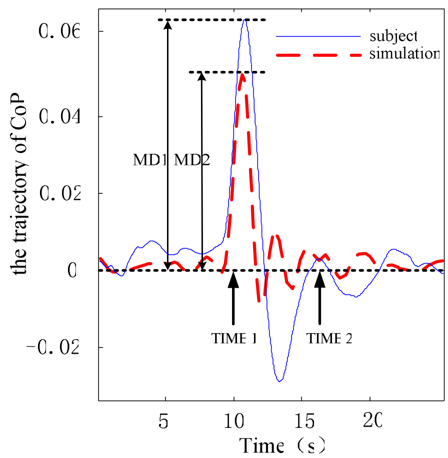


Fig.3 human balance recovery process. The horizontal axis indicates the simulation time; the vertical axis represents the trajectory of the COP in the anterior and posterior direction. Blue line shows the trajectory of the COP from the subject; the red dotted line is the trajectory of the COP from the human simulation model. At the moment of TIME1, platform started to move rearward. At the moment of TIME2, the body returned to balance state. The regulation time of the body is $\Delta\text{T} = \text{TIME2 - TIME1}$.

The simulation results and the experimental data were shown in Fig.3. It illustrates the model could be approximated to simulate the process of the transformation of the center of gravity of the human balance adjustment process, and there are some basic conclusions: (1) COP maximum distance MD2 is small, compared with the magnitude of the real human body data MD1. (2) The shock period of the results is larger, compared with the real data. (3) The damping of the oscillation process in the simulation is larger, and amplitude attenuation rapidly. (4) The time adjusted to the equilibrium state is consistent with real data.

4.2. Impacts of the different time-delay to the balance index

Furthermore, another task of the study was to analysis the affect of the different time delay to the human, and to investigate the relationship between the time delay and the balance ability performance. We simulated the system by varying the values of state delay or controller delay in the model, and let $w(t)=0$ to avoid the interference affect the results of the simulation.

Case 1

Simulation parameters: the level of motion platform backwards movement 10mm; controller delay τ_2 : 0.2s; the state delay changing the amount of 0.1-0.5s.

Case 2

Simulation parameters: the level of motion platform backwards movement: 10mm; state delay τ_1 : 0.2s; controller delay changing the amount of 0.1-0.5s.

The simulation results are list in Table 1 and Table 2. S-delay, C-delay, RT and MD mean state delay, controller delay, regulation time and maximum distance of COP respectively.

Table.1 Simulation Results about Case 1.

S-delay τ_1 (s)	RT (s)	MD (cm)
0.1	5.9	4.6
0.2	6.8	4.7
0.3	7.8	3.9
0.4	12.3	3.8
0.5	13.1	3.7
0.6	×	×

Table.2 Simulation Results about Case 2.

C-delay τ_2 (s)	RT(s)	MD(cm)
0.1	7.0	4.2
0.2	6.8	4.7
0.3	5.9	5.8
0.4	4.9	6.8
0.5	7.2	7.6
0.6	×	×

Nearly all the regulation time is within the ranges from 5.8s to 13.1s, and maximum distance of COP is within the ranges from 4cm to 7.6cm. In the simulation results, × means the system becomes unstable. In other words, the student may not withstand a delay more than 0.6s. The average changing rate of the simulation results obtained with the formula:

$$\varphi=\left(\sum_{2}^{N} a_i - a_{i-1}\right)\Big/(N-1) \tag{17}$$

where, φ average changing rate; a_i is the sample value, and N is the sample number. Fig. 4 shows the correlations between the simulation results and time-delay respectively.

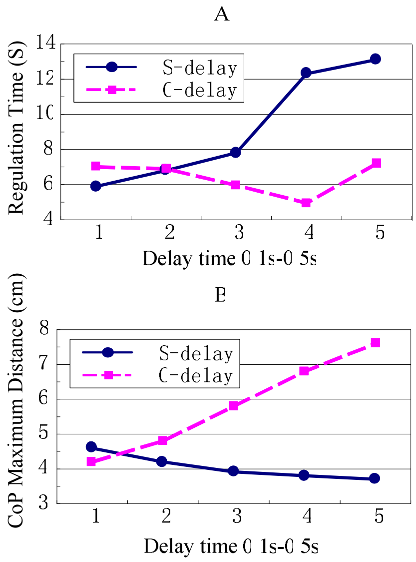


Fig.4 (A) the curve of regulation time about various delay. The horizontal axis indicates the S-delay and the C-delay, values were 0.1-0.5s; The vertical axis indicates the RT; Blue Line represents case1; red line represents case2. (B) The curve of the maximum distance COP about various delay. The horizontal axis indicates the S-delay and the C-delay, values were 0.1-0.5s; The vertical axis indicates the COP MD; Blue Line represents case1; red line represents case2.

From Fig. 4, we could conclude as following. The state delay τ_1 is proportional to regulation time, and the average change rate $\varphi=1.85$, however, controller delay τ_2 do not have the obvious effect to regulation time, whose average change rate $\varphi=0.1$. Fig.4.b shows the curves of the maximum COP distance affected by both different delays. Controller delay τ_2 has a significant positive relationship with COP maximum distance, and the average change rate is $\varphi = 0.85$. In addition, with the state delay increasing, the maximum distance COP is reduced, where the average changing rate is -0.225. The magnitude is small enough to be ignored.

5. Conclusions and Discussions

The purpose of the study was to develop a balance control model based on robust controller, which can reflect the human system with serious multi-source time delay in reality. The robust controller, which has a fixed gain vector K and a observer, can keep a system to stable with larger delay under strong interference. Of course, this is just a functional approximation, when the human body is assumed to be a black box. There are many deficiencies in the hypothesis model of human body, such as a larger gain K may result in ankle moment beyond the capability of the joints, but these deficiencies will be refined in our future studies. Anyway, the presented human balance control model indeed established a quantitative relationship between the changing trends of balance ability and multiple delays which exist in human body.

We designed the musculoskeletal model with muscle viscous force, which was related to the angular velocity of the ankle joint. The muscle viscous force is delay independent, but there are different time-varying delays within inertial force, gravity, and the controller inputs. By analyzing the above simulation results, we can visually see the controller time delay is proportional to maximum COP distance responding to the stimulation. According to the principle of human balance, The COP changing distance can not be out of the plantar stability region, otherwise the body will either fall down or step forward to maintain a balance posture. Therefore, we can conclude that: Human balance capacity is mainly affected by the impact of the amount of the controller delay. Larger controller delay can make the body lose standing balance, and the controller delay must be in a reasonable range. This conclusion also verified the conclusions in paper [7, 8].

Large state delay may be due to illness or disability, which is proportional to regulation time of human recovery stable posture. It means that human body need a large response time for the muscle activation. Some researchers [21, 22] used the hill model to study the relationship between bio-electrical signals and muscle forces, in which it assumed that the muscle activation characteristics satisfy an S-shaped curve. This derivative of the S function can be represented as the speed of human muscle activation. For athletes, the state delay determines the accuracy of posture in a short time. If one has a faster muscle activation speed, the state delay will decrease, which means he/she has a better performance in balance regulation. This performance could be improved through exercise.

References

[1] VDK Herman, J Ron, K Bart, G Henk, "A multisensory integration model of human stance control" Biological Cybernetics vol.80, No.5, 1999, pp.299-308.
[2] PA Fransson, M Magnusson, R Johansson, "Analysis of adaptation in anteroposterior dynamics of human postural control" Gait and Posture Vol. 7, No. 1, 1998, pp.64-74.
[3] Kuo, AD, "An optimal control model for analyzing human postural balance." IEEE Transactions on Biomedical Engineering, Vol. 42, No. 1, 1995, pp.87-101.
[4] Kuo, AD, "An optimal state estimation model of sensory integration in human postural balance" Journal of Neural Engineering, Vol. 2, No. 3, 2005, pp.S235-S249.
[5] K Masani, AH Vette, MR Popovic, "Controlling balance during quiet standing: Proportional and derivative controller generates preceding motor command to body sway position observed in experiments" Gait and Posture, Vol.23, 2004, pp.164-172.
[6] Y F Jiang and H Kimura, "A PID Model of Balance Keeping Control and Its Application to Stability Assessment in Proc. IROS", 2006, pp.5925-5930.
[7] Jo S, Massaquoi, SG, "A model of cerebellum stabilized and scheduled hybrid long-loop control of upright balance" Biological Cybernetics, Vol. 91, No.3, 2004, pp.188-202.
[8] Qu X, Nussbaum, MA, Madigan, ML, "A balance control model of quiet upright stance based on an optimal control strategy" Journal of Biomechanics, Vol.40, No.16, 2007, pp.3590-3597.
[9] Yoshiyuki A, Yuichi T, "Model of Postural Control in Quiet Standing: Robust Compensation of Delay-Induced Instability Using Intermittent Activation of Feedback Control" Plos One July 8, 2009.
[10] Peter G, Ian L, Martin L, "Predictive feedback in human simulated pendulum balancing" Biological Cybernetics, Vol.101, 2009, pp.131-146.
[11] John M, Juan LC, et al, "The time-delayed inverted pendulum: Implications for human balance control" CHAOS Vol.19, No.026110, 2009, pp.1-12.
[12] P. Gatev, S. Thomas, T. Kepple, M. Halett, "Feedforward ankle strategy of balance during quiet stance in adults" J. Physiol, Vol.514, No.3, 1999,pp.915–928.
[13] Robert J, Schilling S, Charles JR, "A Phase-Locked Loop Model of the Response of the Postural Control System to Periodic Platform Motion" IEEE Transactions on neural systems and rehabilitation engineering, Vol.18, 2010, pp.274-283.
[14] ID Loram, PJ Gawthrop, M Lakie, "The frequency of human, manual adjustments in balancing an inverted pendulum is constrained by intrinsic physiological factors " The Journal of Physiology, Vol.577,2006, pp.417-432.
[15] H C Diener, J Dichgans, et al, "On the role of vestibular, visual, and somatosensory information for dynamic postural

control in humans" Progress in Brain Res Vol.76, 1988, pp.253-262.
[16] Kreinder E, Jameson A "Conditions for nonnegativeness of partitioned matrices" IEEE ransactions on Automatic Control Vol.17, pp.147-148.
[17] Boyd SP, Ghaoui LE, et al, "Linear matrix inequalities in system and control theory" Philadelphia, PA:SIAM, 1994.
[18] Yasuyuki S, Taishin N, Maura C, Pietro M, "Intermittent control with ankle, hip, and mixed strategies during quiet standing: A theoretical proposal based on a double inverted pendulum model" Journal of Theoretical Biology, Vol.310,No.7, 2012, pp.55-79.
[19] Gahinet P, Nemirovski A, et al, "LMI control toolbox user's guide"The Mathworks, Inc, 1995, pp. 8.1-8.35.
[20] Sharon MH, Joyce F, Fay BH, "Control of Stance During Lateral and Anterior/Posterior Surface Translations" IEEE Transactions on rehabilitation engineering, Vol.6, No.1, 1998, pp.32-42.
[21] Rositsa TR, Boris IP, "Sensitivity of predicted muscle forces to parameters of the optimization-based human leg model revealed by analytical and numerical analyses" Journal of Biomechanics Vol.34, No.10, 2001, pp.1243–1255.
[22] Ahmet E, Scott M, Walter H, "Model-based estimation of muscle forces exerted during movements" Clinical Biomechanics Vol.22, 2007, pp.131-154.

Hongrui Wang received the M.S. degree from Department of Industry Automation, Northeast Heavy Machinery Institute, China, in 1981. He received the Ph.D. degree in 2002 from the College of Mechanical Engineering, Yanshan University(YSU), China. Since 1981, he has been with the Institute of Electrical Engineering of YSU, where he is currently a professor of YSU. His research interests include parallel mechanical design and control, rolling mill control, intelligent system design and analysis.

Kun Liu received the B.S. degree and the M.S. degree in Hebei University(HBU), China, in 2008 and 2011 respectively. And he is working for the Ph.D. degree in the Institute of Electrical Engineering of Yanshan University (YSU), China, from 2011 to today. Now his research interests include rehabilitation robot, human balance control strategy, human movement system modeling.

Jinzhaung Xiao received the B.S. degree in the Department of Automatic Control, Hebei University(HBU), China, in 2000, the M.S. degree and Ph.D. in the Institute of Electrical Engineering of Yanshan University(YSU), China, in 2003 and in 2009 respectively. Now, he is an associate professor in the Department of Automation, HBU, China, and Post-doctorial Fellow in Universiti Malaysia Pahang, Malaysia.His research interests include rehabilitation robot, parallel robotic control, fault diagnosis, fault tolerant control.

Mobile Application of Water Meter Recorder Based on Short Message Service Transmissions Using Windows Mobile Platform

Anom-Manuaba [1], Darma-Putra [2], Ika-Marini [3]

[1] **Department of Information Technology, Udayana University
Bali, 80361, Indonesia**

[2] **Department of Information Technology, Udayana University
Bali, 80361, Indonesia**

[3] **Department of Information Technology, Udayana University
Bali, 80361, Indonesia**

Abstract

The rapid development of technology nowadays has major impact to the development of cellular technology. This development led to a new wide range of smartphone. The growth of life nowadays requires people to work more quickly, so that they can use the time more effective and increase the performance. The process which done manually takes a lot more time than the process which done automatically, because the process which done manually have a higher risk of error than the process which done automatically. The process that are still done manually is recording the amount of customer water consumption in PDAM (Regional Water Company). This problem can be solved by creating mobile application that can record the water meter and then automatically send the data of the customer and the amount of water use directly to the computer server and calculated automatically. This application can solve the problem in recording the water meter.

Keywords: *Mobile Phone, Water Meter, SMS.*

1. Introduction

The rapid development of technology nowadays has a major impact in the development of mobile technology, one of them was the birth of mobile phone. Mobile phone is an electronic telecommunication device that has the same basic capabilities of conventional fixed-line telephone, but it can be carried anywhere (portable) and no need to be connected to the telephone network by wire (wireless). The development of mobile phones led to a wide range of smartphone. The smartphone is a mobile phone that has a high level of ability with computer-like functions. The rapid development of the smartphone gave birth to the new various of operating system.

Operating system (OS) is a collection of software that manages computer hardware resources and provides common services for computer programs. The operating system is a vital component of the system software in a computer system. Application programs require an operating system to function. [8] Mobile operating system, also referred to as mobile OS, is the operating system that operates a smartphone, tablet, PDA, or other digital mobile devices. Modern mobile operating systems combine the features of a personal computer operating system with touchscreen, cellular, Bluetooth, WiFi, GPS mobile navigation, camera, video camera, speech recognition, voice recorder, music player, Near field communication, personal digital assistant (PDA), and other features. [9]

New technology that comes along the development of mobile technology is an SMS Gateway. SMS gateway is a telecommunications network facility for sending or receiving Short Message Service (SMS) transmissions to or from a telecommunications network that supports SMS. Most messages are eventually routed into the mobile phone networks. [1]

The development of modern life requires everyone to work more quickly so the time can be used efficiently, so the performance should be improved. Enhancement performance should be accompanied by work faster and more accurately so that the work can be done with a single process. The process which done manually takes a lot more time than the process which done automatically, because the process which done manually have a higher risk of error than the process which done automatically.

One of the process that are still done manually is recording the amount of customer water consumption in PDAM (Regional Water Company). PDAM is one of regional business unit that enganged in the distribution of clean water to the public. PDAM available in each province and districts throughout Indonesia. The process of recording the water meter are still done manually by writing every water meter in each customer house and then recapitulated manually.

The problems that arise from the process of recording the water meter is the officer still recorded on paper and then manually recapitulated in the computer. This process takes a lot of time and have a high risk of errors that officer repeatedly checks to reduce the risk of error, this reduces the performance of the officer. This problem can be overcome by creating a mobile application that can record the water meter and then automatically sent via SMS messages containing customer data and water meter to the computer server to recapitulate and counting automatically, so that the officer only need to check and record water meter at customer house and then the next process will be done by the system automatically.

2. Short Message Service (SMS)

SMS is a communication service standardized in the GSM mobile communication systems, it can be sent and received simultaneously with GSM voice, text and image. This is possible because whereas voice, text and image take over a dedicated radio channel for the duration of the call, short messages travel over and above the radio channel using the signaling path. [2]

SMS contains some meta-data:
1. Information about the senders (service center number, sender number)
2. Protocol information (protocol identifier, data coding scheme)
3. Timestamp

SMS messages do not require the mobile phone to be active and within range, as they will be held for a number of days until the phone is active and within range. SMS transmitted within the same cell or to anyone with roaming capability. The SMS is a store and forward service, and is not sent directly but delivered via an SMS Center (SMSC). SMSC is a network element in the mobile telephone network, in which SMS is stored until the destination device becomes available. Each mobile telephone network that supports SMS has one or more messaging centers to handle and manage the short messages. [3]

3. Overview Of The System

Water meter record application embedded on a mobile phone that used by officers in the recording water cauge data. Fig. 1 describes the system overview of mobile application in recording water meter.

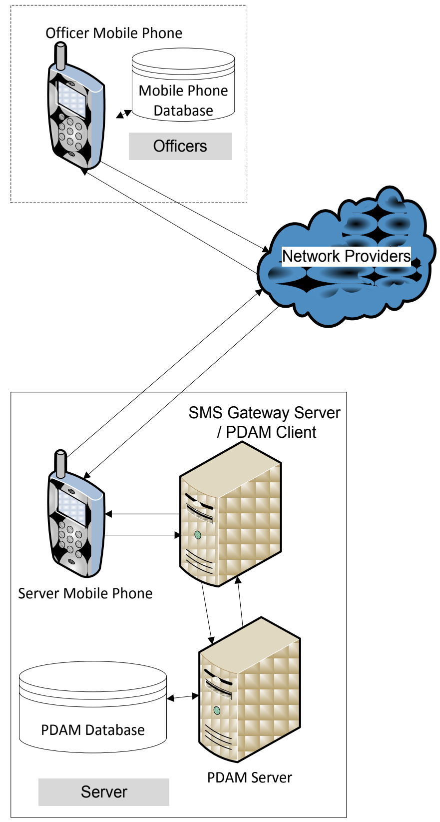


Fig 1. Overview of the system

Usability of the officer mobile phone that have been embedded by this water meter recording application are as follows:

a. Storing the water meter data that has been written by the officer to the mobile phone database.
b. Editing the water meter data that have previously entered to the mobile database.
c. Deleting the water meter data that have previously entered to the mobile database.
d. Capture images water meter.
e. Saving an image file on the phone memory.
f. Sending data to the mobile phone server in the form of SMS, excluding images. The image file is sent separately through mobile devices with direct connection to the computer.
g. Displays information and notifications.
h. Changing the settings that contain destination numbers to send SMS

Usability of server mobile phones and SMS Gateway server are as follows:

a. Receiving SMS from officer mobile phone
b. Checking the format of the incoming data.
c. Sending data to PDAM server.
d. Sending notifications message to officer mobile phone.

Usability of PDAM database and PDAM server are as follows:

a. Calculating the incoming data.
b. Executing queries.
c. Storing data after calculation.

The customer data delivered to the computer server periodically. Data is sent when all the customer data have been recorder and are entered the data delivery period. If passing the data delivery period and not all the customer data is recorded, the delivery will remain to be done and the customer data that not yet been recorded, will be sent later.

Fig. 2 describes the overview of the process of sending message from officer mobile phone to server mobile phone.

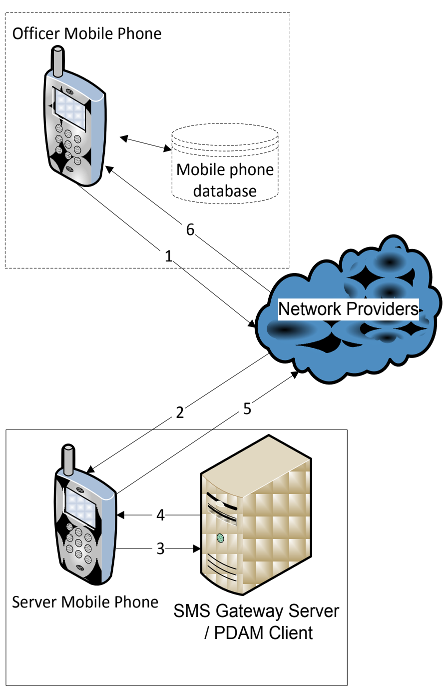


Fig 2. Overview of the process of sending the message from officer mobile phone to server mobile phone

Description of the Fig. 2 above is as follows:

1. Officer mobile phone sending message over the GSM / CDMA network provider.
2. GSM / CDMA mobile phones network providers continue to the server mobile phone.
3. Server mobile phone sending message to the server SMS Gateway computer which checks the received message then sent to PDAM server.
4. After the message has been received, the server SMS Gateway computer sending a notification.
5. Server mobile phone sending notification message over the GSM / CDMA network provider.

GSM / CDMA mobile phones network providers continue to the officer mobile phone.The delivery format adapted with the PDAM water meter, the data that sent is customer id, water meter condition and water meter value. From these data SMS format can be made like this.

```
customer_id
#water_meter_condition
#water_meter_value
```

There is kind of water meter conditon, such as 1 for a good condition, 2 for moderate condition, and 3 for the poor condition. For example:

```
1#1#101
```

From the example data above, there's only one customer data, but an SMS has 160 characters at maximum, so this application dynamically combine several customer data into one single SMS until reach 160 characters. If it reaches maximum 160 characters the rest of data will be sent afterward, depending on how much customer data to be sent. Example of several customer data that sent in one single SMS:

```
1#1#101*2#3#45*3#1#167*4#2#322
*5#1#1278*6#1#453*7#1#1537*8#1#1500
*9#1#178*10#2#145*11#1#111*12#3#145
*13#1#67*14#2#222*15#1#278*16#1#543
*17#1#237*18#1#2390
```

Messages will be sent after officers pressing the send button in the application. The message from officer mobile phone will be decrypted before storing into the database. After descrypted, the data will be like the early customer data format, like customer id, water meter condition and water meter value. The descrypted data will be stored temporarily in the PDAM database. The data from officer mobile phone after descryption will be look like the Table 1 below.

Table 1: Temporary table in the PDAM database

Customer id temp	Water meter condition temp	Water meter value temp	Officer id temp
1	1	101	pt3
2	3	45	pt3
3	1	167	pt3
4	2	322	pt3
5	1	1278	pt3
6	1	453	pt3
7	1	1537	pt3
8	1	1500	pt3
9	1	178	pt3
10	2	145	pt3
11	1	111	pt3
12	3	145	pt3
13	1	67	pt3
14	2	222	pt3
15	1	278	pt3
16	1	543	pt3
17	1	237	pt3
18	1	2390	pt3

Officer id obtained from checking the phone number of officer who listed in the database. Officer id which is the sender, recorded in the temporary table. The image file only used for checking and recording evidence. The image file is sent via officer mobile phone and directly connected to the computer PDAM. Fig. 3 describes the overview process of image file synchronization.

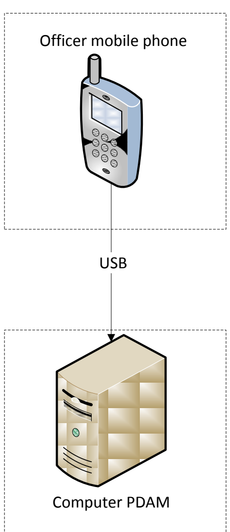


Fig 3. Overview process of image file synchronization

The image file is sent through a direct connection. The process will be done through synchronization of data and information on the officer mobile phone and the PDAM computer. The image file storage folder in the mobile phone sent to the PDAM computer and stored in the folder that has been provided as a collection of image files.

Context diagram in Fig. 4 below shows some of the entities involved in the water meter recording application.

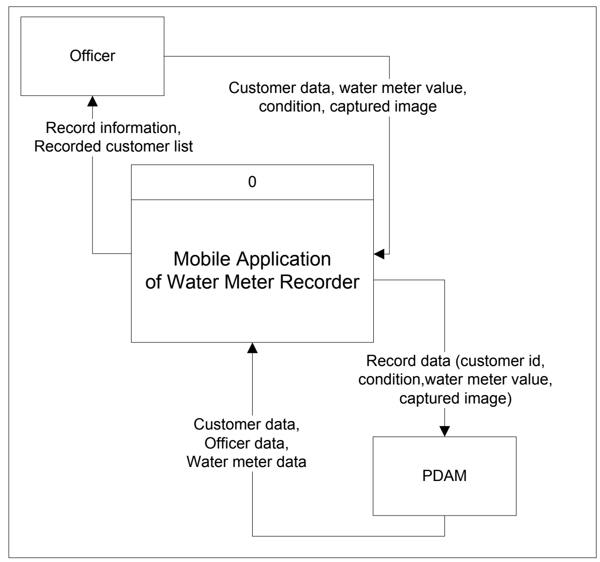


Fig 4. Context Diagram

Entities involved in context diagram in Fig. 4 above are as follows:

1. Officer
   Officers are system users who have full access to the system. Officers can insert data, editing data, deleting data and synchronize data.
2. PDAM
   PDAM is a company that has the authority to provide officer data, water meter data and customer data as the first recipient of the information from water meter record application before it is continued to the customer.

4. Experiments and Results

Trial is an important stage, because with the trial the programmer will know the error in the system. This PDAM water meter record application is made for the record officer in the process of recording the PDAM water meter every month to obtain the amount of customer water usage. Next will explain the trial of PDAM water meter record application.

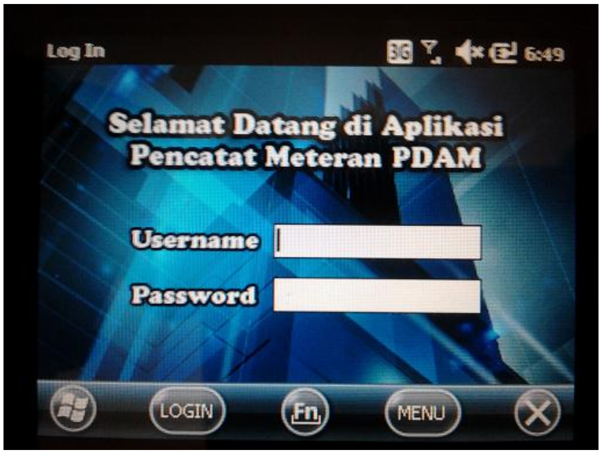


Fig 5. Login form

Fig. 5 above is the login used by officer who have permissions in the form of username and password to enter the application system. Username and password checked according username and password which listed on tb_officer, as in the Table 2 below.

Table 2: Officer table

Officer id	User name	Pass word	Phone number	Name
pt1	sby	sby	+6281917395177	Susilo
pt2	budi	budi	+6281805564983	Budiono
pt3	yusuf	yusuf	+6281999430972	Yusuf

After the login is successful, then the officer will enter the main menu. For example, officer who logged in are "Yusuf Kalla", it can be seen as the Fig. 6 below.

Fig 6. Main menu form

Fig. 6 above is the view of main menu in the PDAM water meter record application system. Can be seen "Yusuf Kalla" on the main menu form is a officer who logged in. There are seven menu, that is record, edit, view, send, list, setting and logout. Each menu have different function that will be explain as follow.

The first menu is record menu functioning as a place to input the customer water consumption data. For example, customer with a customer id "20510", which looks like Fig. 7 below.

Fig 7. Record form where customer id "20510"

Fig. 7 is a record page functioning as a place to input the customer water meter information. In this page occur an information such as customer id, customer name, customer address, class, area and the amount of water usage last month. The officer only input the amount of water usage that month, the water meter condition and capture the image of the water meter. Customer data as Fig. 7 above, taken from tb_customer as the Table 3 below.

Table 3: Customer table where customer id "20510"

Customer id	Name	Address	Class	Area
20510	KU Dusun Ketima	Dsn. Ketima	A	P022119

The information that appears on the record form where customer id "20510" can be seen as the Table 3 above. The process of capturing the water meter image can be seen as Fig. 8 below.

Fig 8. Capturing image of water meter

Fig. 8 show the process of capturing water meter image. Capturing water meter image provide evidence that the recording actually has been done to the customer. After capturing the image, image name will appear according to customer id, month and years of recording.

Fig 9. The record page that have been filled

Fig. 9 show that all of the record data where customer id “20510” were obtained and are ready to be stored. After the save button of the menu is pressed, data will be stored into the database automatically. If the data stored in the database is successful, it can be seen as the Table 4.

Table 4: Record table where customer id “20510”

Customer id	Date	Condition	Value old	Value new
20510	12/12/2012	1	0	1530

Editing can be done in editing page, which looks like Fig. 10 below.

Fig 10. Edit form where customer id “20510”

Fig. 10 is an edit page for editing the data where customer id “20510” that has been previously entered. Only editing the data that input within this month, such as the amount of water usage this month, water meter condition, and capturing image. After save button on the menu is pressed, the data will be saved into database. Deleting data can also be performed on this page by pressing the delete button. After the data in the database is updated successful, can be seen as the Table 5 below.

Table 5: Record table after updated where customer id “20510”

Customer id	Date	Condition	Value old	Value new
20510	12/12/2012	3	0	1531

Customer data where customer id "20510" changed after editing. As shown in the Table 5 above.

Fig 11. Tampilan data pelanggan dengan id “20510”

Fig. 11 is the view data page used to review the recording that have been done before. Search is performed based on customer id, month and record year. Information presented in the form in water meter condition, amount of water usage last month, this month and image file name.

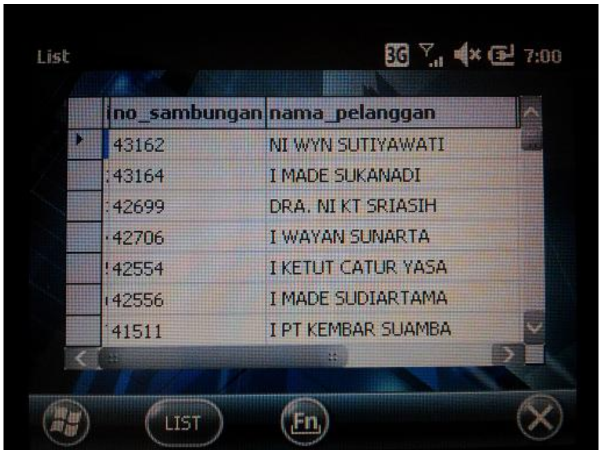


Fig 12. Sequence of recording list

Fig. 12 show list page that contains the sequence of the customer that should be record by the officer. It intend to help the officer so it's easier to choose which customer that should be noted first. Example of sequence data of recording list, can be seen as the Table 6.

Table 6: Sequence of recording list

Sequence	Customer id	Name
1	43162	Ni Wyn Sutiyawati
2	43164	I Made Sukanadi
3	42699	Dra. Ni Kt Sriasih
4	42706	I Wayan Sunarta
5	42554	I Ketut Catur Yasa
6	42556	I Made Sudiartama
7	41511	I Pt Kembar Suamba

Fig 13. Setting form

Fig. 13 show the page where the record date interval is set and setting the server destination number where the data should be sent by officer.

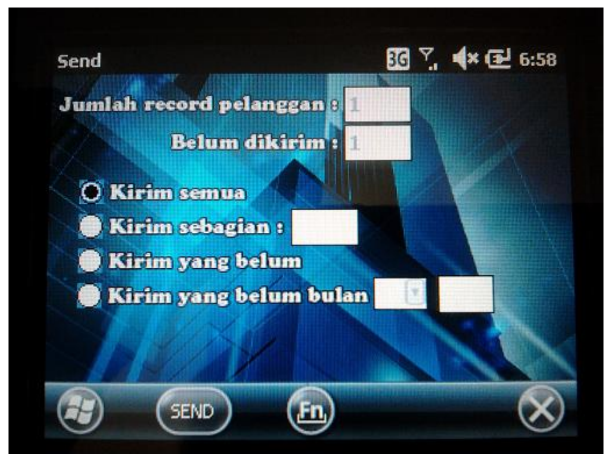


Fig 14. Menu send form

Fig. 14 is the image of send page that consist of four delivery menu, such as send all, send partially, send the delayed data and send the delayed data by selecting the month. Delivery is done by choose the delivery menu and push the send button. If the delivery is succed, a notification will appear as shown below.

Fig 15. Message sent successfully

Fig. 15 show the notification in the send page that message sent successfully. Data received on the server computer and directly decrypted then stored to the database, it can be seen as Fig. 16 below.

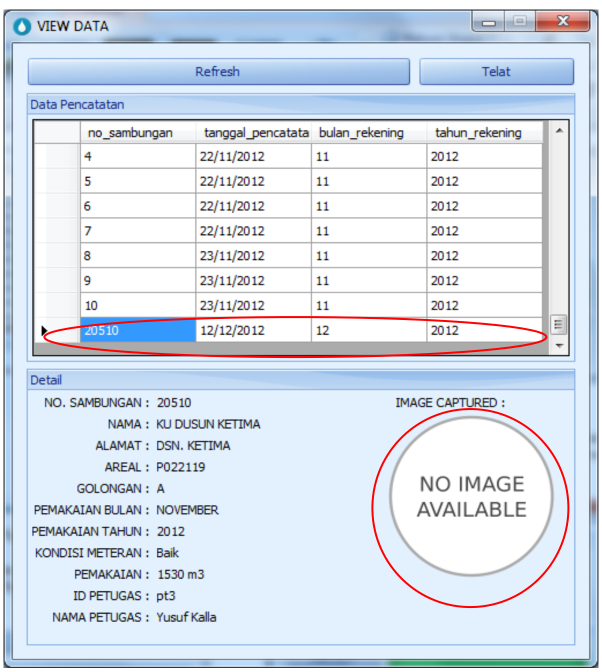


Fig 16. Record view in computer server

Fig. 16 shows that the data where customer id “20510” has been entered into a computer server database, but the picture has not yet available because the image file synchronization has not been done.

The next process is the synchronization that done on the server computer. Synchronization process can be seen as below.

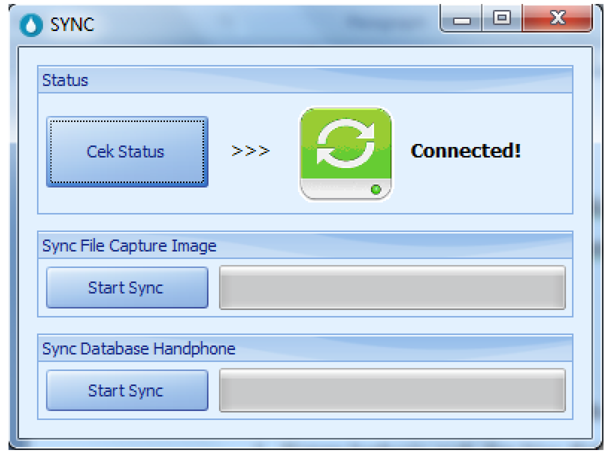


Fig 17. Sync form in computer server

Fig. 17 show the synchronization process that done on the server computer. There are two types of synchronization that can be done, the image file synchronization and database synchronization. Image file synchronization process can be seen as below.

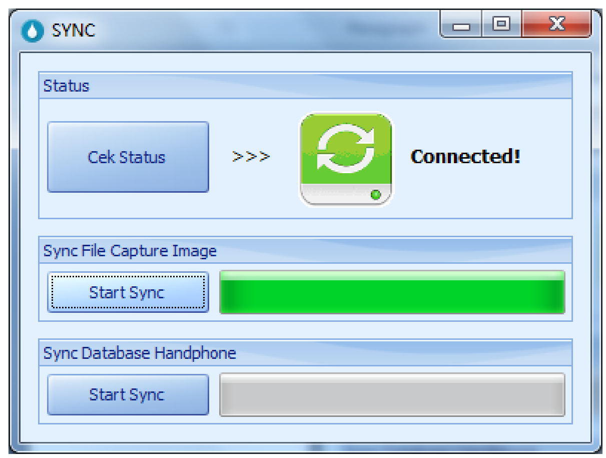


Fig 18. Image files synchronization process in computer server

Fig. 18 show the image file synchronization process. Images file in the officer mobile phone copied to server computer. After synchronization process, the image file has been copied to the server computer, it can be viewed as Fig 19 below.

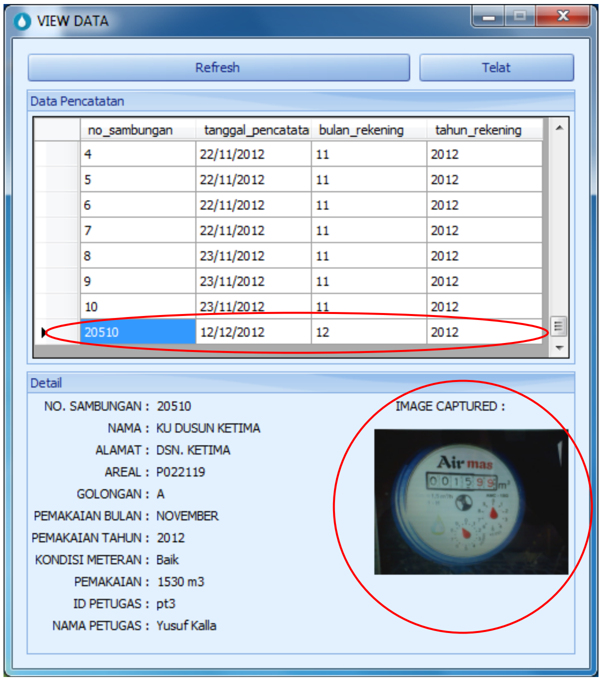


Fig 19. Image file has been copied to the computer server

Fig. 19 shows that the image file where customer id "20510" has been copied to the server computer. This can be seen from the picture above where the water meter image has been seen on the page record view.

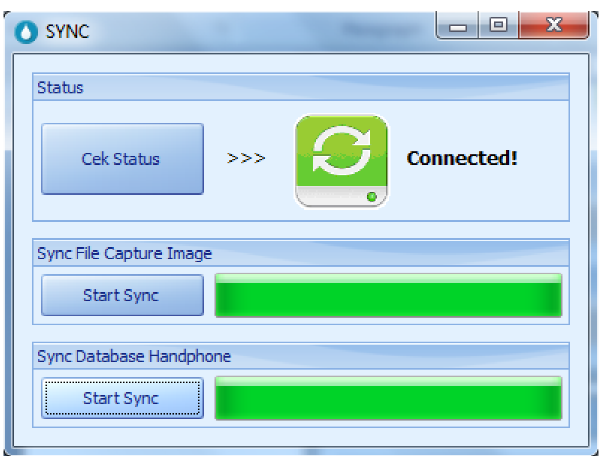


Fig 20. Database synchronization process

Fig. 20 show the database synchronization process. Database in the server computer copied to officer mobile phone.

5. Conclusions

Mobile application of water meter recorder are recorded all the information involved in the recording process, such as the officer, the water meter, the customer, the record date, and the name of the photo file when doing the documentation. Mobile application first stores all the information in the officer mobile phone database before it is sent via SMS transmission to the server.

Acknowledgments

Our thanks to Department of Information Technology Udayana University, Bali, who has helped organize this research.

References

[1] Veena K.Katankar et. al, Short Message Service using SMS Gateway, (IJCSE) International Journal on Computer Science and Engineering.Vol. 02, No. 04, 2010, 1487-1491

[2] M. M. Tarek et. al, SMS Spam Filtering Technique Based on Artificial Immune System, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, 2012.

[3] G. Le Bodic, Mobile Messaging Technologies and Services SMS, EMS and MMS, Alcatel.France: John Wiley & Son Ltd, 2005.

[4] Li. Bo and Gyu Im. Eul, Smartphone, promising battlefield for hackers, Journal of Security Engineering. Vol. 8, No. 1, 2011

[5] Powell. Gavin, Beginning Database Design. Indianapolis: Wiley Publishing Inc, 2006.

[6] Teorey. Toby, Lightstone. Sam, and Nadeau. Tom, Database Modeling & Design: Logical Design, San Francisco: Morgan Kaufmann Publishers, 2006.

[7] Arregui. Francisco, Cabrera Jr. Enrique, and Cobacho. Ricardo, Integrated Water Meter Management. London: IWA Publishing, 2006

[8] Ranganathan. Prakash and Nygard. Kendall, Time Synchronization In Wireless Sensor Networks: A Survey, International Journal Of UbiComp (IJU), Vol. 1, No. 2, April 2010.

[9] Ningning Wu, Ming guang Wu, Siguo Chen, 2008, Real-time Monitoring and Filtering System for Mobile SMS, pp. 1319-1324.

[10] PDAM : http://www.pdamtabanan.com

[11] Operating system : http://en.wikipedia.org/wiki/Operating_system

[12] Mobile operating system : http://en.wikipedia.org/wiki/Mobile_operating_system

I Dewa Nyoman Anom Manuaba is a student at Department of Information Technology Udayana University, and now working her research for S.Ti. degree.

I Ketut Gede Darma Putra received his S.Kom. degree in Informatics Engineering from Sepuluh November Institute of Technology, his MT. degree in Electrical Engineering from Gajah Mada University and his Dr. degree in Electrical Engineering from Gajah Mada University. He is lecturer at Electrical Engineering Department (major in Computer System and Informatics) of Udayana University and lecturer at Information Technology of Udayana University.

Ni Made Ika Marini Mandenni received her S.T. degree in Informatics Engineering from Adhi Tama Institute of Technology and her M.Kom. degree in Informatics Engineering from Sepuluh November Institute of Technology. She is lecturer at Information Technology of Udayana University.

The Analysis of Comparison of Expert System of Diagnosing Dog Disease by Certainty Factor Method and Dempster-Shafer Method

Eka-Setyarini[1], Darma-Putra[2] and Adi-Purnawan[3]

**[1] Department of Information Technology, Udayana University
Bali, 80361, Indonesia**

**[2] Department of Information Technology, Udayana University
Bali, 80361, Indonesia**

**[3] Department of Information Technology, Udayana University
Bali, 80361, Indonesia**

Abstract

Expert system is one of branches of artificial intelligence that studies how to "adopt" an expert way of, inferring from a number of facts, and making decision. This paper presents a comparison between two methodologies, Certainty Factor Method and Dempster-Shafer Method to identify diseases of the dog. Providing proper health care can be done by knowing common dog diseases and being aware of appropriate prevention and treatment. In this paper used 74 physical symptoms of illness to find 17 types of common diseases of dogs. Five options are given to answer questions of calculations using each method: no, quite sure, pretty sure, sure, and definitely sure. The accuracy of the analysis of each method was tested by assessing the results of each analysis method based on the given user enter. The results of the analysis are correct when judged from the point of view of experts.

Keywords: *Diseases of Dog, Certainty Factor Method, Dempster-Shafer Method.*

1. Introduction

Dogs are animals which can adapt easily and can be a good friend to human, so that many of them are cared. High interest to get a dog cause many keepers need information on how to protect and care of their dog easily without going to a clinic or veterinarian. Some of the most common and serious dog diseases have been made less common through vaccines. However, these diseases threaten a dog that lacks proper immunization. [dog] In addition, there are many dog owners who do not pay attention to their pets health because much cost of bringing dog to the vet and the existence of veterinarian itself which is still lack. Carelessness in maintaining a dog can result in losses to the breeder and the surrounding communities. One of the examples is rabies disease which is transmitted very quickly and even can cause death to humans or animals. Some dog diseases can also be easily transmitted to humans. Veterinary experts, in this case, has the ability to analyze the symptoms of the disease, but to overcome any problems the dog caretaker constrained by time and the widespread of deployment to the animal.

Expert System is a reasoning system that performs comparable to or better than a human expert within a particular domain.[12] The basis of expert system is how to move the knowledge possessed by an expert into a computer and how to infer or make decisions based on that knowledge. Storing the knowledge into computer needs a database of knowledge (Knowledge Base) that is database modeling determined in advance. Expert systems are used by doctors to help with evidences that are hard to diagnose and to suggest preventive measures or measures self care where even human experts have difficulty [1].

There are several approaches that can be used in building an expert system. One approach that seems right for the case of diagnosis of disease is by using reasoning with uncertainty. This is because of many of the conditions or circumstances that are not entirely certain when diagnosing a disease. Expert systems created to help people in decision making that must be designed with the ability to cope with these uncertain domains [2]. A number of approaches can be used relating to this uncertainty, such as Certainty Factor (CF). In this method, in expressing the degree of confidence, a value called the certainty factor (CF) to assume a degree of belief in an expert to the data is used. Certainty Factor introduces the concept of belief and disbelief. For reason about degrees of certainty, CF is used. The second method will be used is Dempster-Shafer method. Dempster-Shafer method is based on two ideas, the first idea is to obtain a degree of belief for one question from subjective probabilities for a related question, and the second is the rule for combining belief which is based on a degree of evidence. There are three important functions in the Dempster-Shafer theory, namely: basic probability assignment function (BPA or m), Belief function (Bel), and the Plausibility function (Pl). Both of these approaches will be used, which is then compared which diagnosis result has the highest degree of accuracy close to the diagnosis of an expert.

The importance of information about the disease on dogs is based on the symptoms of the disease and ways to overcome, so this paper will discuss the application of expert system to diagnose the disease on dogs by Dempster-Shafer method and Certainty Factor that can later be used to reduce and minimize the risk of death. With the hope of the expert system can help as a veterinarian to identify diseases of the dog.

2. Methodology

The overview diagram of this research is shown in Fig. 1.

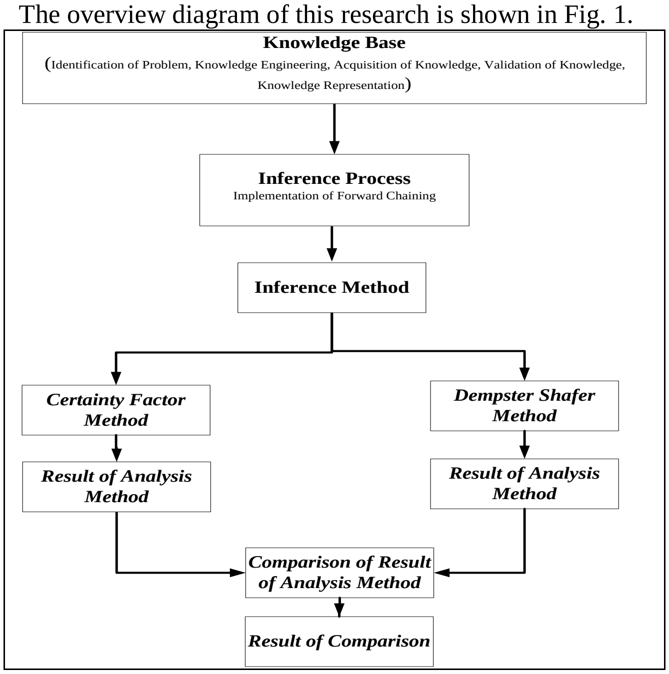


Fig. 1 General Overview System

2.1 Knowledge Base

In this paper, the problem identification process is carried out which will be made the expert system of it. In this case problem tried to explore is the disease in dogs. After a literature study of this problem is taken, the problems seen in this disease of dogs is that it can no longer be underestimated, because there are diseases that can cause death in dogs and even death in humans such as rabies which has been widespread.

Knowledge Engineering is a knowledge base development activity that will be integrated into the expert system. This knowledge base will be used as a basis of knowledge of the system in doing the reasoning. The stages performed in knowledge engineering in an expert system to diagnose the disease in dogs use Dempster-Shafer method and the Certainty Factor method; it consists of knowledge acquisition process, knowledge validation process, knowledge representation process, and conclusions or inferences.

In the process of knowledge acquisition, a knowledge engineer will attempt to gather relevant knowledge by expert systems to be built. Collection of knowledge is done by studying books and other references relating to the problems of disease in dogs, but it is also carried an interview with the vet to get an explanation and more knowledge about the disease. The knowledge gained is then transformed by the knowledge engineer into the knowledge base. Basically the process of acquiring this knowledge is the bridge between the expert and the knowledge base.

This process is the stage where the knowledge base that has been built on the knowledge acquisition phase will be validated by experts whether the knowledge that has existed was appropriate and proper or not. If it turns out that knowledge is not appropriate yet, then re-acquisition of knowledge will be done by the addition or reduction in the knowledge base to a knowledge base that is valid and in accordance with the approval of experts. If the expert has given its approval that the existing knowledge base is the right one then the process continues to the stage of knowledge representation.

Having knowledge successfully acquired and validated by experts, knowledge which has acquired needs to be organized and arranged in a configuration with a particular format or representation. The goal is to make the system easier to access the existing knowledge, so it can be used as a basis for finding a solution. A popular knowledge representation method is production rules and frames. In the expert system to diagnose the disease in dogs is used a production rule from the tree model as a knowledge representation method.

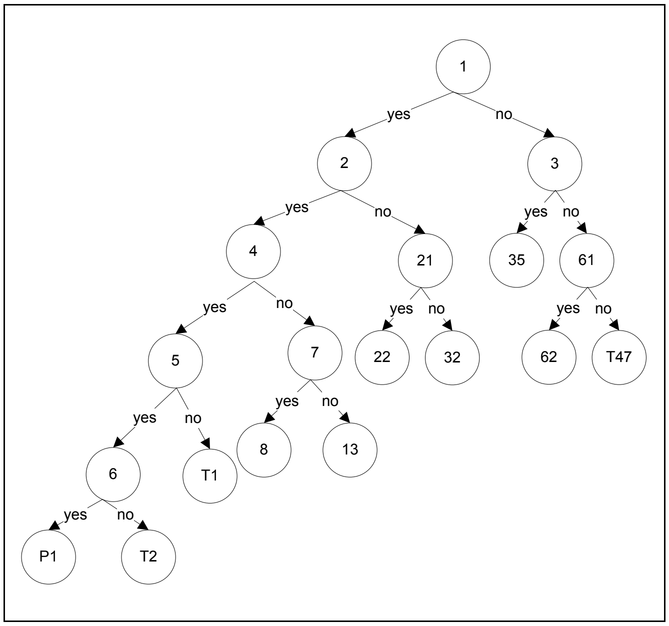


Fig. 2 Modeling knowledge base of diseases of dog with tree

Description of the tree above:
1 : Does your dog look sluggish or feeble?
2 : Is your dog's appetite decreasing?
4 : Is your dog having yellowish (mucous membrane jaundice)?
5 : Is your dog experiencing dehydration (increasing thirst)
6 : Is your dog issued peeing blood?
P1 : Your dog is diagnosed having leptospirosis disease.
21 : Is your dog's attitude and behavior changing?
22 : Is your dog's tail folded in?
32 : Is your dog experiencing diarrhea with yellow to green?
7 : Is there a liquid or viscous rheum in the nose (flu) on your dog?
8 : Is there any rheum on your dog's eyes?
13 : Is there any thickening on your dog's skin?
T1 : Your dog is may be diagnosed having leptospirosis disease.
T2 : Your dog is may be diagnosed having leptospirosis disease.
3 : Is your dog having a fever (body temperature is above 37 ° C)
35 : Is your dog experiencing bloody diarrhea?
61 : Is your dog experiencing to be itchy?
62 : Is your dog's skin experiencing redness?
T47 : No disease found.

It can be seen on modeling base knowledge of dog disease with tree diagram like the in Fig. 2, notation "yes" means node (evidence) fulfill the node (evidence) above it; "no" means the node (evidence) doesn't fulfill the previous node (evidence) above it. Hypotheses P1 will be achieved if there is evidence 1,2,4,5 and 6. Each node represents certain evidence for each condition "yes" and "no", so that there will be no node leading to the same evidence. In diagnose session, this means each answer is different from the user; therefore, it leads to different question.

2.2 Certainty Factor Method

Certainty Factor is introduced by Shortliffe Buchanan in making MYCIN (Kusumadewi, 2003). Certainty Factor (CF) is a clinic parameter value which is given by MYCIN to show the measurement of belief. Certainty Factor (CF) shows the measurement of certainty to a fact or rule.
The using of certainty factor is done for: [2]

1. Determining the measurement of belief to the early fact which will be given by every user,
2. Determining the measurement of belief to conclusion or decision which is obtained from the rule; experts determine this value to the rule,
3. Determining the measurement of belief to facts and result which is obtained along the process of reasoning from the result of the rule execution,
4. Adjusting the measurement of belief to fact or result which is obtained from the different rule but producing the same conclusion.

Certainty factor is defined as below:

CF [h,e]=MB[h,e]-MD[h,e] ....(1)

With:
CF [h,e] = Certainty Factor
MB [h,e] = Measurement of Belief to hypothesis *h*, if given evidence *e* (between 0 and 1)
MD [h,e] = Measurement of Disbelief to hypothesis *h*, if given evidence *e* (between 0 and 1)

The following is the description of some combinations of Certainty Factor to any conditions:

1. Certainty Factor for single premise rules:
   CF (H, E) = CF (E)*CF (rule)
   = CF(user)*CF(expert)...................(2)
2. Certainty Factor for multiple premise rules:
   CF (A AND B) = Minimum (CF (a), CF (b)) * CF (rule)
   CF (A OR B) = Maximum (CF (a), CF (b)) * CF (rule)
3. Certainty Factor for similar concluded rules:
   CF COMBINE(CF1,CF2)= CF1 + CF2*(1-CF1)........(3)

Calculating process of belief percentage is done by the following steps:

1. Determining rule based on basic knowledge
2. Solving rule which has multiple premise rule to be single premise rule

3. Calculating value of CF expert with the value of mb and md in every symptoms using equation (1)
4. Calculating value of CF in every rule using equation (2) to get the value of CF in every rule
5. Then the value of CF is combined to equation (3)

Example for the experiment using Certainty Factor Method:
First experiment is performed using Certainty Factor Method. User will be given five choices of answers to answer every question such as the following table.

Table 1: Certainty Value for User

No	0
Quite Sure	0,4
Pretty Sure	0,6
Sure	0,8
Definitely Sure	1

Value 0 shows that the patient does not experience symptoms of the disease which is asked by the system. The more the patient sure that he is indeed experiencing the symptoms, so that the larger the percentage of total belief that the results will be obtained.

Rule 1:
IF Look lethargic / weak
AND decreased appetite
AND Mucous membranes jaundice (yellowish)
AND Dehydration (increased thirst)
AND Bloody urine
THEN Leptospirosis

First step, the expert determines the CF value of every evidences as follows:
CFexpert(look lethargic / weak)
= mb – md
= 0,7 – 0,1 = 0,6
CFexpert(decreased appetite)
= mb – md
= 0,6 – 0,1 = 0,5
CFexpert(Mucous membranes jaundice (yellowish)
= mb – md
= 0,7 – 0,1 = 0,6
CFexpert(dehydration (increased thirst))
= mb – md
= 0,6 – 0,1 = 0,5
CFexpert(bloody urine)
= mb – md
= 0,8 – 0,1 = 0,7

For example the user choosing answers as follows:
Look lethargic / weak: Maybe = 0,4
Decreased appetite: Probably= 0,6
Mucous membranes jaundice (yellowish): Maybe = 0,4
Dehydration (increased thirst): Probably = 0,6
Bloody urine: Certainly = 0,8
Early rule which has 5 premises is changed into rule which has single premise to be:
Rule 1.1
IF Look lethargic / weak
THEN Leptospirosis
Rule 1.2
IF Decreased appetite
THEN Leptospirosis
Rule 1.3
IF Mucous membranes jaundice (yellowish)
THEN Leptospirosis
Rule 1.4
IF Dehidration (increased thirst)
THEN Leptospirosis
Rule 1.5
IF Bloody urine
THEN Leptospirosis

Then the CF value of the rules is calculated using the equation below:
CF (H, E) = CF (E)*CF (rule)
= CF (user)*CF (expert)

CF 1.1 = 0,4 * 0,6 = 0,24
CF 1.2 = 0,6 * 0,5 = 0,30
CF 1.3 = 0,4 * 0,6 = 0,24
CF 1.4 = 0,6 * 0,5 = 0,30
CF 1.5 = 0,8 * 0,7 = 0,56

Combine the value CF 1.1 to value CF 1.2 using the formula below:
CF COMBINE (CF1,CF2)= CF1 + CF2*(1-CF1), so it becomes
CF COMBINE (CF 1.1, CF 1,2) = 0,24 + 0,30*(1-0,24)
= 0,46 = CFold
Combine CFold to CF 1.3 as follows:
CF COMBINE (CFold,CF 1.3) = 0,46 + 0,24*(1-0,46)
= 0,58 = CFold

Combine CFold to CF 1.4 as follows:
CF COMBINE (CFold,CF 1.4) = 0,58 + 0,30*(1-0,58)
= 0,70 = CFold
Combine CFold to CF 1.5 as follows:
CF COMBINE (CFold,CF 1.5) = 0,70 + 0,56*(1-0,70)
= 0,86

Belief percentage = CF COMBINE * 100%
= 0,86 * 100%
= 86%
So belief system percentage against leptospirosis disease is 86% according to the answer given by the user.

2.3 Dempster-Shafer Method

Dempster-shafer method is first introduced by Dempster, who did an experiment to an uncertainty model with probability range as a single probability. Then in 1976, Shafer published the Dempster theory in a book entitled Mathematical Theory of Evident. A way to reason about degrees of belief is provided by Dempster-Shafer method. [1]
Generally, Dempster-Shafer is written in an interval: [3]

[Belief, Plausibility]......................(4)

Belief (Bel) is measurement of evidence power in supporting a proposition assemblage. If it is worth 0 (zero), it indicates that there is no evidence; if it is worth 1, it shows that there is certainty. According to Giarratano and Riley, the function of belief can be formulated as:

$$Bel(X) = \sum_{Y \subseteq X} m(Y) \qquad \text{.....................(5)}$$

While *Plausibility* (Pls) is denoted as:

$$Pls(X) = 1 - Bel(X') = 1 - \sum_{Y \subseteq X'} m(X') \qquad \text{.............(6)}$$

Where:
Bel(X) = Belief (X)
Pls(X) = Plausibility (X)
M(X) = mass function of (X)
m(Y) = mass function of (Y)

In Dempster-Shafer theory, the set of the universe of discourse of a set of hypotheses given the notation θ , where it is assumed that the hypotheses used is grouped to an individual environment. To show how big the belief of evidence to a certain hypotheses is called probability density function which is given notation (m).

In the application of expert system in a disease, there is some evidence which will be used to uncertainty factor in taking decision for diagnosis of a disease. To solve that some evidence in Dempster-Shafer theory, it is used a rule known as Dempster's Rule of Combination.

$$m1 \oplus m2(Z) = \frac{\sum_{X \cap Y = Z} m1(X)m2(Y)}{1 - \sum_{X \cap Y = \theta} m1(X)m2(Y)}. \qquad \text{......................(7)}$$

Where:

$m1 \oplus m2(Z)$ = mass function of evidence (Z)
$m1(X)$ = mass function of evidence (X)
$m2(Y)$ = mass function of evidence (Y)
K = the amount of evidential conflict

Calculating process of belief percentage is done by the following steps:
1. Determining rule based on basic knowledge
2. Determining value of density (m) and m(θ) = 1- m1{x} using equation (9)
3. Calculations using Dempster's Rule of Combination using equation (11)
4. Determining the greatest density with Max { m {}}

Example for the experiment using Dempster-Shafer Method :

In the case of diagnostics, user answered five evidence, include look lethargic or weak, decreased appetite, mucous membranes jaundice (yellowish), dehydration (increased thirst), and bloody urine.
Rule:
IF Look lethargic / weak
AND Decreased appetite
AND Mucous membranes jaundice (yellowish)
AND Dehydration (increased thirst)
AND Bloody urine
THEN Leptospirosis

Known:
θ = {P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, P16, P17}

There is evidence el which support the hypotheses P1,P2,P3,P4,P5,P6,P7,P8,P10 with m = 0,60, so it can be written as follows:
m_1 { P1,P2,P3,P4,P5,P6,P7,P8,P10 } = 0,60
m_1 {θ} = 1-0,60 = 0,40

Then, there is evidence e2 which supports P1,P2,P3,P4,P5 with m = 0,60, so it can be written as follows:
m_2 {P1, P2, P3, P4, P5 } = 0,60
m_2 {θ}= 1-0,60 = 0,40

First, perform calculation to the first evidence (el) and the second evidence (e2) using equation *Dempster's Rule of Combination* to overcome a number of evidence showed up and get a new m value, as follows:

Table 2: Combination e1 and e2

	{ P1,P2,P3, P4,P5 } 0,60	θ 0,40
{ P1,P2,P3,P4, P5,P6,P7, P8,P10} 0,60	{ P1,P2,P3, P4,P5 } 0,36	{ P1,P2,P3,P4, P5,P6,P7, P8,P10 } 0,24
Θ	{ P1,P2,P3,P4,P5	θ

	}	
0,40	0,24	0,16

By using the *Dempster's Rule of Combination* equation, found value of m_3 as follows:
m_3 {P1, P2, P3, P4, P5} = (0,36+0,24)/(1-0) = 0,60
m_3 {P1,P2,P3,P4,P5,P6,P7,P8,P10} = (0,24)/(1-0) = 0,24
m_3 {θ} = (0,16)/(1-0) = 0,16

Then, emerge the third evidence (e4) which supports P1 with m = 0,70, so it can be written as follows:
m_4 {P1} = 0,70
m_4 {θ} = 1-0,70 = 0,30

Conducted calculation using *Dempster's Rule of Combination* equation to get m_5 as follows:

Table 3: Combination with e4

	0,70 { P1 }	0,30 θ
{ P1,P2,P3, P4,P5 } 0,60	{ P1 } 0,42	{ P1,P2,P3, P4,P5 } 0,18
{ P1,P2,P3,P4, P5,P6,P7, P8,P10 } 0,24	{ P1 } 0,17	{ P1,P2,P3,P4, P5,P6,P7, P8,P10 } 0,072
Θ 0,16	{ P1 } 0,112	θ 0,05

So m_5 can be as followed:
m_5 {P1} = (0,42+0,17+0,112)/(1-0) = 0,702
m_5 {P1, P2, P3, P4, P5} = (0,18)/(1-0) = 0,18
m_5{P1,P2,P3,P4,P5,P6,P7,P8,P10}=(0,072)/(1-0)= 0,072
m_5 {θ} = (0,05)/(1-0)= 0,05

Then conducted the fourth evidence (e5) which supports P1 with m = 0,60, so it can be written as follows:
m_6 {P1} = 0,60
m_6 {θ} = 1-0,60 = 0,40

Conducted the calculation using *Dempster's Rule of Combination* equation to m_7 as follows:

Table 4: Combination with e5

	0,60 { P1 }	0,40 θ
{ P1 } 0,702	{ P1 } 0,42	{ P1 } 0,28
{ P1,P2,P3, P4,P5 } 0,18	{ P1 } 0,108	{ P1,P2,P3, P4,P5 } 0,072
{ P1,P2,P3,P4, P5,P6,P7,P8,P10} 0,072	{ P1 } 0,043	{ P1,P2,P3,P4, P5,P6,P7,P8,P10} 0,029
θ 0,05	{ P1 } 0,024	θ 0,016

So m_7 can be as followed:
m_7 {P1} =(0,42+0,108+0,043+0,024+0,28)/(1-0) = 0,875
m_7 {P1, P2, P3, P4, P5} = (0,072)/(1-0)= 0,072
m_7{P1,P2,P3,P4,P5,P6,P7,P8,P10}=(0,029)/(1-0)= 0,029
m_7 {θ} = (0,016)/(1-0) = 0,016

After that, the next evidence showed up is e6 which supports P1 with m= 0,70, so it can be written as follows:
m_8 {P1} = 0,70
m_8 {θ} = 1-0,70 = 0,30

Conducted calculation using *Dempster's Rule of Combination* equation to get m_9 as follows:

Table 5: Combination with e6

	0,70 { P1 }	0,30 θ
{ P1 } 0,875	{ P1 } 0,62	{ P1 } 0,26
{ P1,P2,P3,P4,P5 } 0,072	{ P1 } 0,0504	{ P1,P2,P3,P4,P5 } 0,0216
{ P1,P2,P3,P4, P5,P6,P7,P8,P10 } 0,029	{ P1 } 0,0203	{ P1,P2,P3,P4, P5,P6,P7,P8,P10 } 0,0087
Θ 0,016	{ P1 } 0,0112	θ 0,0048

So m_9 can be as followed:
m_9 {P1} = (0,53+0,0504+0,0203+0,0112)= 0.962
m_9 {P1, P2, P3, P4, P5} = (0,0216)/(1-0)=0,0216
m_9{P1,P2,P3,P4,P5,P6,P7,P8,P10}= (0,0087)/(1-0)= 0,0087
m_9 {θ} = (0,0048)/(1-0)= 0,0048

From the result of the calculation of probability value above, obtained the biggest density is m_9 { P1 } that is 0,962. Thus, its can be concluded that *user* most likely suffered P1 that is Leptospirosis with belief percentage is 0,962 * 100 % = 96,2 % according to answer which is given by *user*.

3. Experiments and Results

In the experiment, Fig. 3 shows early view of disease diagnose that will be done. System will show the first question which is always being the root of the tree of decision. Five choices of answer will be provided to answer each question; no, quite sure, pretty sure, sure, and definitely sure. Each choice of answer has quality that will be used later in the process of calculation of each method. Answer "no" means user's dog do not experience symptom which was asked by the system. The higher is rate of certainty of the user toward symptom that is experienced, so that the higher is the percentage of certainty which will be produced.

Fig. 3 System shows the first question

After the system shows the first question such as in Fig. 3, so that the user has to answer one of five choices of answer that is given by the system, it can be seen in Fig. 4

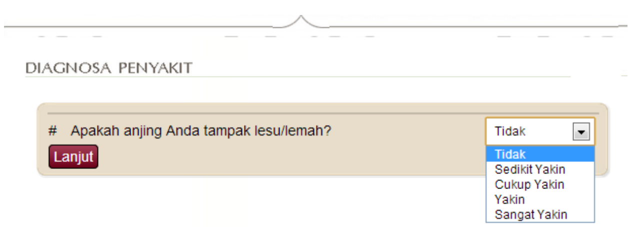


Fig. 4 Five possible answers

As the first example, if the user chooses "quite sure" on the first question, so that there will be the next question as showed in Fig. 5. The next question that will be shown is based on base knowledge modeling which has been made in a tree model. If on the first question the user answered "no", the next question which will be shown will not be the same as the question when the user answers "quite sure". In fig. 5, it is shown the next question when user answers "quite sure" and in fig. 6 will show the next question when user answers "no".

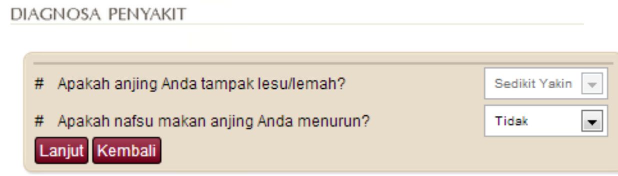


Fig. 5 The user chooses "quite sure" on the first question

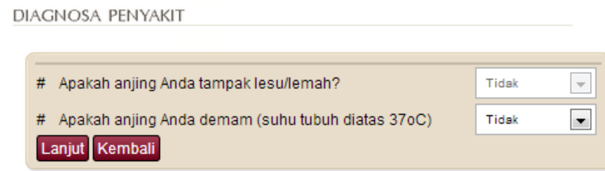


Fig. 6 The user chooses "no" on the first question

The following questions showing up depend on each answer from the user. For each different answer, there will also be different that will be shown. Each question refers to a conclusion that is based on the rules that has been determined.

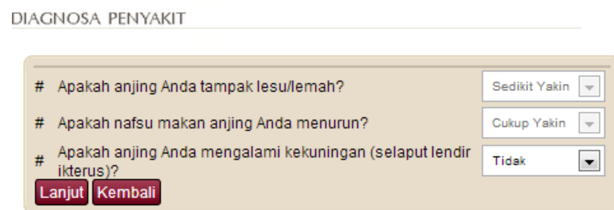


Fig. 7 The following question

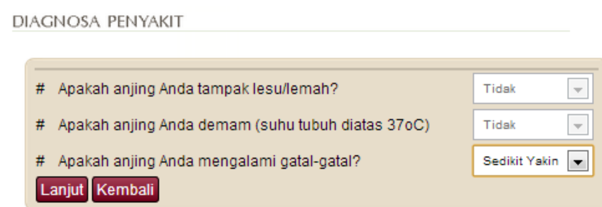


Fig. 8 The following question

System will stop showing questions if the system has found a conclusion. The conclusion is in a form of diagnose result given by the system from each user's answer. The conclusion shown is kind of disease that suffered by the user's dog and certainty value produced from Certainty Factor and Dempster-Shafer method calculation. It can be seen in Fig. 9. The system has been success in giving the diagnose result (conclusion).

Fig. 9 The diagnose result

Fig. 10 The diagnose result

Based on the Fig. 9 and Fig.10 above, there was difference result showed by using Certainty Factor and Dempster-Shafer Method. differences seen by the result of belief presentage which obtained by the calculation of each method.

Certainty factor method has simpler calculation than dempster-shafer method because the calculation in certainty factor method just involving values from mb and mb from each question. Whereas, on dempster shafer method, the calculation is done by considering value from each disease possibilities.

Dempster-shafer method, in determining percentage result of belief value, pays attention on value of whole variables used which are combined based on equation of dempster's rule of combination, so that there will be produced values from the calculation which is more varied and more accurate. The writer concludes that dempster-shafer is one of good uncertainty value completion ways in determining percentage of belief value.

System will not find the kind of disease suffered by the user's dog if the user answers "no" in each question. It can be seen in Fig. 11.

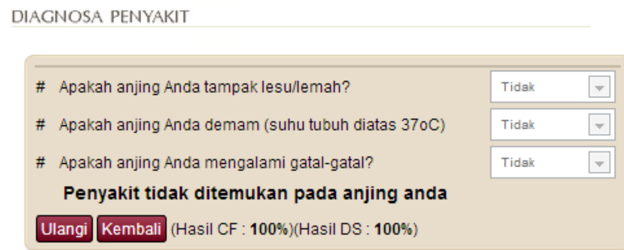


Fig. 11 No disease found

4. Conclusions

There is comparative relevant in Certainty Factor and Dempster-Shafer method in this paper. Certainty factor method has simpler calculation than dempster-shafer method. Dempster-Shafer method is better than Certainty Factor because in determining the results of the belief percentage consider the value of all variables used in combination by equation Dempster's Rule of Combination resulting calculation values are more varied and more accurate. Based on this data, the writer concludes that dempster-shafer is one of good uncertainty value completion ways in determining percentage of belief value.

The accuracy of the analysis of each method was tested by assessing the results of each analysis method based on the given user enter. The results of the analysis are correct when judged from the point of view of experts. Used 10 cases will be assessed and tested to the experts, namely Drh. Ary as a veterinarian.

Table 6: Accuracy of Result

Case	Percentage of Accuracy Using Certainty Factor (%)	Percentage of Accuracy Using Dempster-Shafer(%)
1	80	85
2	65	70
3	70	75
4	70	70
5	90	90
6	75	80
7	70	65
8	70	75
9	65	75
10	80	85
avg	73	76

Based on the value of the accuracy of the above could be said that the analysis method is good enough.

Acknowledgments

Our thank goes to Department of Information Technology Udayana University, Bali, who has helped organize this research's Indonesia.

References

[1] Coppin Ben. 2004. Artificial Intelligence Illiminated. Canada : Jones and Bartlett Publishers, Inc.
[2] Smith Barbara S.., 1990, uncertainty Reasoning and Representation: A Comparison of Several Alternative Approaches, Master Thesis, Graduate Program, Rochester Institute of Technology Department of Computer Science.
[3] Ali, Tazid, Dutta Palash, 2012, Methods to Obtain Basic Probability Assignment in Evidence Theory, International Journal Of Computer Applications, Vol. 38, No. 4
[4] Sadiq Rehan, Kleiner Yehuda, Rajani Balvant, 2006, Estimating Risk of Contaminant Intrusion in Water Distribution Networks using Dempster-Shafer Theory of Evidence, National Research Council Canada, Vol. 23, No. 3
[5] Prihatini Manik, Darma Putra. 2012. Fuzzy Knowledge-based System with Uncertainty for Tropical Infectious Disease Diagnosis. IJCSI International Journal of Computer Issues, Vol.9, Issue 4, No 3, July 2012
[6] Mahoney J. Jeffrey, Mooney Raymond J., 1994, Comparing Methods for Refining Certainty-Factor Rule-Bases, Appears in Proceedings of the Eleventh International Workshop on Machine Learning, pp. 173-180Fakhry. Hussein, "A Fuzzy Logic Based Decision Support System for Business Situation Assessment and e-Business Models Selection", Journal of Communications of the IIMA, Vol. 10, Issue 4, 2010.
[7] Sasikumar M., Ramani S., Raman S Muthu, Anjaneyulu KSR., Chandrasekar R., 2007, A Practical Introduction to Rule Based Expert Systems, New Delhi : Narosa Publishing House
[8] Wu Huadong, Siegel Mel, Stiefelhagen Rainer, JieYang. Sensor Fusion Using Dempster-Shafer Theory. IEEE Instrumentation and Measurement Technology Conference Anchorage, AK, USA, 21-23 May 2002
[9] Chen, C.H., Khoo, L.P., Yan, W., 2002, A strategy for acquiring customer requirement patterns using laddering technique and ART2 neural network, Advanced Engineering Informatics, 16, 229–240
[10] Hughes, A.M., 1994, Strategic database marketing, Chicago: Probus Publishing Company.
[11] Carpenter, G., Grossberg, S., 1987, ART2: stable self-organization of pattern recognition codes for analog input patterns. Appl Optics 26: 4919–30.
[12] Fausset, L., 1993. Fundamental of Neural Network : Architecture, Algorithms, and Application, Precentice Hall
[13] Carpenter, G., Grossberg, S., The art of adaptive recognition by a self-organising neural network, Computer 1988, 21, 77-88
[14] Rousseeuw, P.J., "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis", J. Comp App. Math, Vol. 20, 1987, pp. 53-65

Ni Putu Eka Setyarini studied Information Technology in Departement of Information Technology Udayana University since August 2008, and now woking her research for S.Ti. degree in Information Technology.

Dr. I Ketut Gede Darma Putra, S.Kom. MT received his S.Kom degree in Informatics Engineering from Institut Teknologi Sepuluh Nopember University, his MT. degree in Electrical Engineering from Gajah Mada University and his Dr. degree in Electrical Engineering from Gajah Mada University. He is lecturer at Electrical Engineering Department (major in Computer System and Informatics) of Udayana University, lecturer at Information Technology Department of Udayana University.

I Ketut Adi Purnawan, ST., M.Eng received his ST. degree in National Development University "Veteran" Yogyakarta, his M.Eng. degree in Informatics Engineering from Gajah Mada University and his Dr. degree in Electrical Engineering from Gajah Mada University. He is lecturer at Electrical Engineering Department (major in Computer System and Informatics) of Udayana University, lecturer at Information Technology Department of Udayana University.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Analysis of Static and Dynamic Behavior of T-shape Beam Reinforced by External Prestressing Tendon

Dinghai Li[1,2], Xiong Xiong[1] , Keyue Zhang [1]and Wei Tang[1]

[1] Southwest Jiaotong University-Emei
Sichuan,Emei-614202, China

[2] Sichuan Provincial Key Lab of Procsee Equipment and control -Zigong

Abstract

External prestressing has become a primary method for strengthening existing concrete beam and has been increasingly used in the construction of newly erected ones, particularly railroad bridges in recent years. In order to evaluate the effect of this method, the static and dynamic behavior of a T-frame beam reinforced by external prestressed strengthened concrete beam was analyzed by 3D finite element method, and the field test study was also made. The study was carried out to further investigate the simply supported reinforced prestressed concrete beam strengthened by external prestressing through theory analysis and experiment.

Keywords: *External Prestressing, Reinforcement, Numerical Simulation, Displacement, Concrete Compressive Stress.*

1. Introduction

With the development of transportation facilities, in order to adapt the need of market economic development, the railway excavates existing line potential and enhances transport capacity through operating heavy haul trains and improving train speed while accelerating to build new lines at the same time. Especially after the implementation of tow transport and container traffic, the safety capacity of many old bridges are insufficient because the bridges on existing lines are designed and built according to the old design specification. There are many problems universally existing such as the bridge of displacement caused by the vibration of vertical and transverse are more larger, crack resistance and dynamic coefficient of some key parts do not meet the operation security and so on; and these problems are relevant to too many times of revising specification[1]. Railway bridge design specification has been carried on many times of revision combined with the actual problem, which causes technical standards used for design of concrete beam on existing lines to have certain differences. Therefore, although design loads of most bridges adopt China railway standard -22 level and part of them are China railway standard -26 level, part of them also adopt downgrading and load standard is China railway standard -21 level, and even lower of China railway standard -20 level. Especially early specifications have inadequate understanding for prestressed long-term loss of prestressed concrete beam (shrinkage and creep of concrete and relaxation of prestressed steel), thus vertical displacement, vertical frequency and crack resistance of existing beam can't meet the requirements according to current standard to test[2-4]. Therefore, reinforcing the old bridges to satisfy current requirement of 200 km/h is one of the main tasks of bridge construction.

The national railway goes through large-scale acceleration for six times. For the early design of prestressed concrete simply supported T beam, the problems of vertical impact load significantly increasing, obvious beam vertical vibration, deficient concrete crack resistance, Train ride comfort greatly reducing will come out when the speed of trains is 200km/h and that of freight trains is 80km/h. Vertical impact load of the trains increases, and especially the influence to locomotive with prefix Z is most significant. Because high-speed locomotives of new type like DF11G are hanging two-shipper, axle load is 230kN, unsprung mass is bigger, and impact load after acceleration has also increased. When the Z-prefix speed-increasing train passes the bridge, vertical displacement of the bridge is significantly increased. At the same time, significant vertical vibration occurs on the bridge [4][5].

2. vehicle-bridge dynamical analysis

The constant force moves to the right with a constant velocity on the simple supported beam as shown in Figure1[6-9], the mobile vehicle inertial characteristic was ignored and mobile vehicle was regarded as a constant force. The variable coefficient vibration equation was avoided by this way. From zero time to time t, fore moves from left end supporting to right at υ(see Fig.1). From vibration analysis the vibration differential equation of the beam was represented by the formula (1):

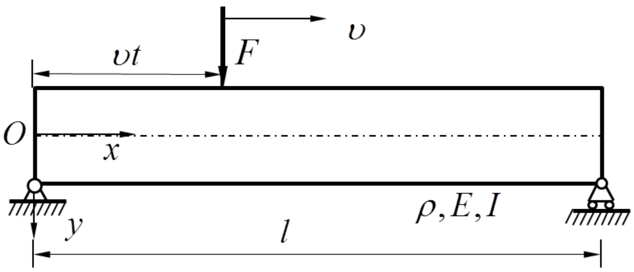


Fig.1. The constant force through the simple supported bridge with constant speed

$$EI\frac{\partial^4 y}{\partial x^4}+\rho A\frac{\partial^2 y}{\partial t^2}=p(x,t) \quad (1)$$

in which E=modulus of elasticity of beam; I= moment of inertia for transformed cross section taken about the centroid; A= transformed area of beam cross section and ρ= material density. Assume that the dynamic displacement of the forced vibrations $y(x,t)$ can be expressed as series form of vibration mode:

$$y(x,t)=\sum \varphi_n(x)q_n(x) \quad (2)$$

in which $\varphi_n(x)$ = simply supported beam displacement = $\sin(\frac{n\pi x}{l})$. The forced vibration equation can be gotten by putting the formula (2) into the formula (1) and the orthogonality of vibration mode.

$$\begin{aligned}&EI\sum q_n(t)\frac{d^4\varphi_n(x)}{dx^4}\\&+\rho A\sum \ddot{q}_n(t)\frac{d^2 q_n(x)}{dt^2}=p(x,t)\end{aligned} \quad (3)$$

Using the standardization of vibration mode, the formula (3) times $\varphi_k(x)$, then integral in the range 0 to l, the left of item exist only when $k = n$, others was zero. Leading to form:

$$\begin{aligned}&EIq_n(t)\int_0^l \varphi_n(x)\frac{d^4\varphi_n(x)}{dx^4}dx\\&+\rho A\frac{d^2 q_n(t)}{dt^2}\int_0^l \varphi_n^2(x)dx\\&=\int_0^l \varphi_n(x)p(x,t)dx\end{aligned} \quad (4)$$

Simplified formula (4), leading to forced vibration differential equation:

$$\ddot{q}_n(t)+\omega_n^2 q_n(t)=Q_n(t) \quad (n=1,2\cdots N) \quad (5)$$

In which

$$\omega_n^2=\frac{EI\int_0^l\left[\frac{d^2\varphi_n(x)}{dx^2}\right]^2 dx}{\rho A\int_0^l \varphi_n^2(x)dx},$$

$$Q_n(t)=\frac{\int_0^l p(x,t)\varphi_n(x)dx}{\rho A\int_0^l \varphi_n^2(x)dx}$$

So constant moving single constant force, generalized excitation force

$$\begin{aligned}Q_n(t)&=\frac{\int_0^l F\delta(x-\upsilon t)\varphi_n(x)dx}{m\int_0^l \varphi_n^2(x)dx}\\&=\frac{2F}{\rho Al}\sin(\Delta_n t) \quad (n=1,2\cdots N)\end{aligned} \quad (6)$$

in which $\delta(\xi)$ Draco function, it is defined as:

$$\delta(\xi)=\begin{cases}1 & \xi=0\\ 0 & \xi\neq 0\end{cases} \quad (7)$$

The forced vibration equation can be gotten by putting the formula (6) into the formula (5).

$$\begin{aligned}&\ddot{q}_n(t)+\omega_n^2 q_n(t)\\&=\frac{2F}{\rho Al}\sin(\Delta_n t) \quad (n=1,2\cdots N)\end{aligned} \quad (8)$$

When the time is 0, the system is in a static state, the formula (8) can be obtained.

$$q_n(t)=\frac{2F}{\rho Al}\frac{1}{\omega_n^2-\Delta_n^2}(\sin\Delta_n t-\frac{\Delta_n}{\omega_n}\sin\omega_n t) \quad (9)$$

In which $\omega_n^2=\left(\frac{n\pi}{l}\right)^2\sqrt{\frac{EI}{\rho A}}$ = the natural frequency of simply supported beam and $\Delta_n=\frac{n\pi\upsilon}{l}$ = the generalized excitation frequency of moving load.

Therefore, the dynamic stress of the beam is expressed as:

$$\begin{aligned}&y(x,t)=\frac{2F}{\rho Al}\sin(\frac{n\pi x}{l})\\&\sum\frac{1}{\omega_n^2-\Delta_n^2}(\sin\Delta_n t-\frac{\Delta_n}{\omega_n}\sin\omega_n t)\end{aligned} \quad (10)$$

Where formula (10) the first item in the brackets represent a forced vibration the second one expressed the free vibration.

3. Establishment of finite element model

This paper adopts large and universal finite element software Ansys to conduct analog computation[10][11], solid 45 element to model for T beam and constraint processing form which is basically in agreement with the reality, that is, the panel point of horseshoe part ux = uy = uz =0, the other part of horseshoe uy = uz =0; the prestressed bars uses link8 to imitate, the impacts of other stirrup on beam shall take the way of improving concrete elastic modulus (E) to achieve. The reinforcement materials mainly used are as follows: adhesive plate, external cable (including calculating steering position of external cable and the position of anchoring in the upper). Calculate compressive stress and vertical displacement of structural concrete of T beam before and after reinforcing above materials. Finite element model include that bridge and gravel and rails like shown in Fig.2. Fig.4 and Fig.5 show the typical constitutive law of materials for steel reinforcement and prestressing steel, concrete[12-15].

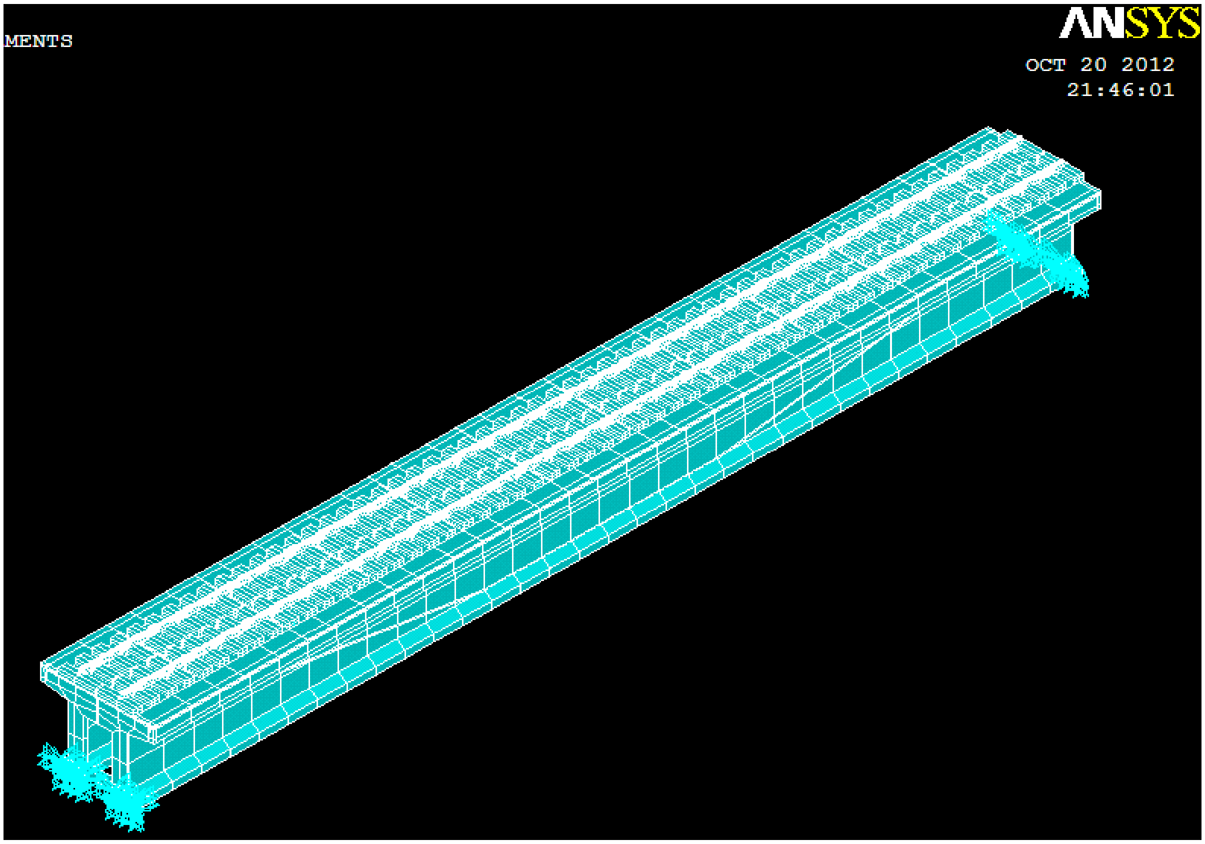


Fig.2. finite element model

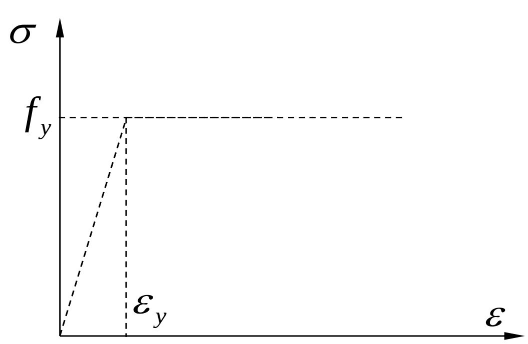


Fig.3. Constitutive law for steel reinforcement

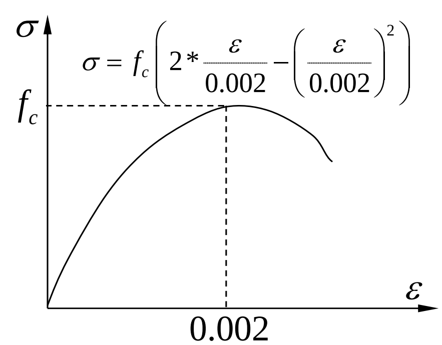


Fig.4. Constitutive law for concrete

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

4. External prestressed strengthened concrete beam calculation

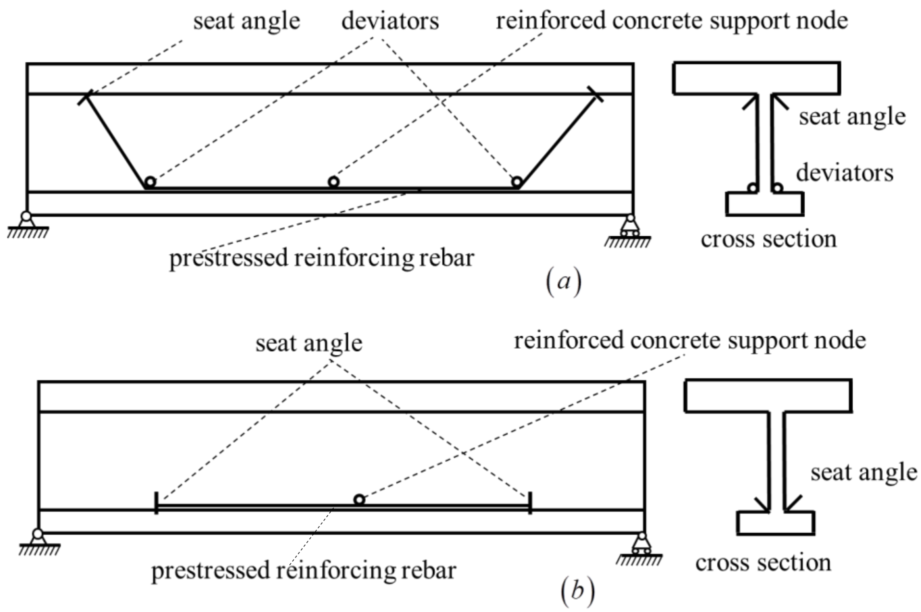
4.1 Advantages and disadvantages of common reinforcement methods

Common reinforcement methods in engineering: reinforcement of continuous beam, thicken and widen of upper and lower flange, adhesive plate of lower flange, adhesive fiber of lower flange, external prestressed steel wires. The two forms of continuous reinforcement and thicken and widen of upper and lower flange need to interrupt traffic. This goes against with maintaining the existing traffic order and shall not be used in practice. Adhesive fiber and plate reinforcement adept to the existing reinforced concrete bridge structure. Corrosion of reinforcement and other reasons cause the weakening of primary structural bearing capacity. They are the methods to recover primary structural design or increase the structural bearing capacity. This method has some defects, the improvement of primary structural bearing capacity is not obvious and it cannot play a part obviously in improving the concrete crack resistance. Thus, the two reinforcement methods are seldom adopted or only used in the reinforcement of reinforced concrete structure. The effect of external prestress wire on beam reinforcement is most obvious, its characteristics after reinforcement are:

(1) Structure weight remains unchanged and the bearing capacity can be increased dramatically;
(2) The prestressing steel reinforcement allows negative vertical displacement of the bridge structure and inverted arch to some extent occurs on the beam;
(3) The construction is convenient and quick and economic benefits are significant;
(4) The construction process has little impact on the traffic;
(5) The construction does not damage original structure and almost does not change the bridge substructure space;
(6) Prestress can be adjusted according to the need, and prestressed tendon changes.

4.2 External prestressed reinforcement beam calculation

In the external prestressed reinforcement of simply supported girder, the main four layout forms are shown in Fig.5.



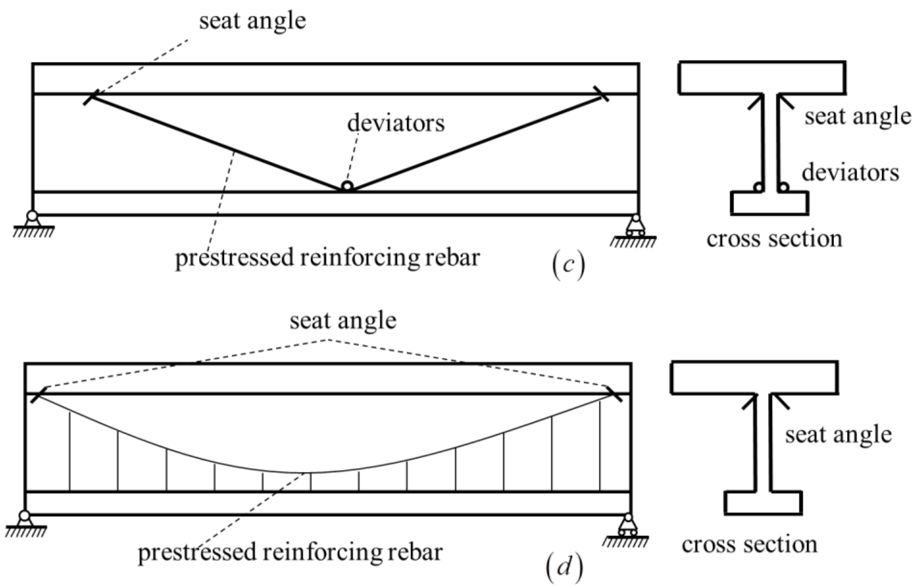


Fig.5. Arrangement of external reinforcement

If the four reinforcement layout forms shown in Fig.5 are adopted and prestressed tensile stress is taken as the strength limit of 0.6*f*, 0.65*f*, 0.7*f*, 0.75*f* and 0.8*f*, the calculation results of vertical displacement of the mid-span and the compression stress of the concrete on the lower side are shown in table 1and table 2.

Table 1. The mid-span vertical displacement of the four reinforcement layout forms

Ultimate multiple	0.6	0.65	0.7	0.75	0.8
Distance of 3.2m(in:mm)	22.03	21.15	20.12	19.94	19.65
Distance of 4.8m(in:mm)	22.63	22.01	21.23	20.32	19.98
Distance of 8.0m(in:mm)	23.33	22.63	21.93	21.08	20.11
Distance of 12m(in:mm)	22.24	21.90	21.05	20.02	20.06

Table 2. The lower side concrete's compression stress of the four reinforcement layout forms

Ultimate multiple	0.6	0.65	0.7	0.75	0.8
Distance of 3.2m(in:MPa)	5.353	5.558	5.761	5.965	6.17
Distance of4.8m(in:MPa)	5.380	5.587	5.793	6.000	6.206
Distance of 8.0m(in:MPa)	5.008	5.200	5.386	5.574	5.763
Distance of 12m(in:MPa)	5.138	5.338	5.537	5.737	5.936

For the reinforcement layout form shown in Fig.5 (a), if the distance between deviators is: 3.2 m, 4.8 m, 8 m and 12 m and prestressed force is 0.6*f*, 0.65*f*, 0.7*f*, 0.75*f* and 0.8*f*, the calculation results of the mid-span displacement and the compression stress of the concrete on the lower side are shown in table 3 and table 4.

Table3. The mid-span vertical displacement of the four reinforcement layout forms

Ultimate multiple(*f*)	0.6	0.65	0.7	0.75	0.8
Figure 2. (a) reinforcement mode(in:mm)	21.02	20.75	20.01	19.72	19.45
Figure 2. (b) reinforcement mode(in:mm)	20.87	20.68	19.93	19.69	19.43
Figure 2. (c) reinforcement mode(in:mm)	20.73	20.56	19.82	19.61	19.35
Figure 2. (d) reinforcement mode(in:mm)	21.95	21.23	20.22	19.90	19.67

Table 4. The lower side concrete's compression stress of the four reinforcement layout forms

Ultimate multiple(*f*)	0.6	0.65	0.7	0.75	0.8
Figure 2. (a) Reinforcement mode(in:MPa)	5.385	5.590	5.795	6.005	6.210
Figure 2. (b) Reinforcement mode(in:MPa)	5.425	5.635	5.844	6.054	6.264
Figure 2. (c) Reinforcement mode(in:MPa)	5.444	5.655	5.866	6.077	6.287
Figure 2. (d) Reinforcement mode(MPa)	5.208	5.405	5.686	5.904	6.105

The analysis of prestressed force value, deviator position and end anchorage position is provided as follows:
As shown in table 1 and table 2, the compressive stress of concrete in four kinds of external prestress arrangement modes in fig.5 are linearly increased; a and b increases more quickly, but b is anchored on the end of the beam; so it shall not be taken in to consideration as construction for strengthening will influence the traffic.

As shown in table 3 and table 4, the compressive stress of concrete in four kinds of external prestress arrangement modes in fig.5 are linearly increased; that is when not taking the steering angle into consideration, the influence of position of deviators is not big. But in order to prevent heavier stress concentration in beam body, the position of deviators shall be set on the diaphragm plate or on the diaphragm plate directly. In this way, stress concentration of the beam body can be reduced.

For the reinforcement layout form shown in Fig.5 (a), if the distance between deviators is 12 m and relative height of anchorage position ranges from 580 mm to 1,780 mm, the calculation results of the compression stress of the concrete on the lower side are shown in Fig.6.

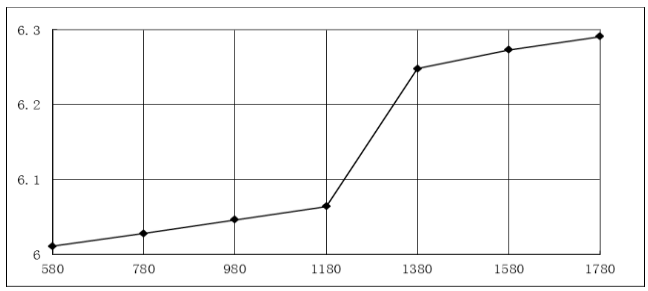


Fig.6 . Horizontal section bar length to 12 m and in different height anchorage reinforcement mode side concrete compressive stress

Fig.6 shows that when the ends are anchored at different relative height, the compressive stress of concrete on the downside of beam body is 6.011MPa at least and 6.291MPa at most with 4.6 % increased which means the position of according point is selectable in actual strengthening according to the convenience of construction.

4.2 Calculation results and experimental analysis comparison

The mode of fig.5(a) is selected and used as actual reinforcing mode, relevant parameters are that intermediate length is 12m and prestress is 1445MPa.The anchor point position of reinforcement shall be determined according to actual position and meet the construction requirements [16][17]. The field experimentation conditions are vibrational state under the function of operating loads after bridge beam is reinforced on JiaoJi line. The vertical maximum amplitude of actual measured beam is 1.01mm and 0.96mm respectively when the speed of the truck is 77 km/h and 74. 9 km/h. They are 1.9mm and 2.1mm which is far less than the requirements of 26.4mm in “Temporary Provisions of Mixed Passenger and Freight Railway of New 200 Kilometers per Hour” after considering dynamic coefficient of live load. It is consistent with the calculation of this paper basically.

5. Conclusions

The research of longitudinal reinforcement of simply supported girder is to ensure that vehicle passes through the bridge safely and that passengers are comfortable. The calculated and experimental results show that external prestressed reinforcement applied on the beam allows fully utilized concrete compression stress and qualified indicators of the bridge.

6. Acknowledgments

The study described in this paper was supported by the open foundation of Sichuan Provincial Key Lab of Process Equipment and control (project number: GK201111) and the Central University Fundamental Research Funds for young teachers in the Hundred Talents Program (project number: SWJTU2011BR048EM).

References

[1] Hongxia Duan, Shouju Li, Yingxi Liu, "A Nonlinear Two-stage Damage Identification Method for Steel Frames", IJACT: International Journal of Advancements in Computing Technology, Vol. 3, NO. 4, 2011, pp. 109-120.

[2] Nie Jianguo, Wang Yuhang, "Experimental research on fatigue behavior of RC beams strengthened by steel plate-concrete composite technique”, Journal of building structures, Vol32 NO.2, 2012, pp.1-9.

[3] Wang Xiaorong, "Reinforcement calculation deduction and enlightenment of railway concrete beams carbon fiber cloth", Vol37 NO.20, 2011, pp.1-9.

[4] Xu Xuedong,“ Both passenger and cargo collinear railway speed of 200 km/h bridge related technical countermeasures”, Railway engineering, NO.6, 2006, pp.1-3.

[5] Niu Bin, Li Xinmin, Ma Lin,“ Speed increase reinforcement design for the prestressed concrete beam of Shahe extra-long bridge on Beijing-Shanhaiguan railway Line”, china railway science,Vol26 NO.6, 2004, pp.65-70.

[6] Jing Ji, Wenfu Zhang etc., "Analysis and Comparison on Dynamic Characteristics of the Bridge Subjected to Moving Load Based on ANSYS", Journal of Convergence Information Technology (JCIT), Vol7, NO.8, 2012, P159-168.

[7] T. Kocaturk, M. Simsek, “Dynamic analysis of eccentrically prestressed viscoelastic Timoshenko beams under a moving harmonic load”, Computers and Structures, 84 (2006) P.2113–2127.

[8] M. Simsek, T. Kocaturk, “Free and forced vibration of a functionally graded beam subjected to a concentrated moving harmonic load”, 90 (2009), P.465–473.

[9] Mandhapati P. Raju ,“Parallel Computation of Finite Element Navier-Stokes codes using MUMPS Solver”, IJCSI International Journal of Computer Science Issues, Vol. 4, No. 2, 2009, P.20-24.

[10] Wenddabo Olivier Sawadogo, Noureddine Alaa and Blaise Somé, “Numerical simulation of groundwater level in a fractured porous medium and sensitivity analysis of the hydrodynamic parameters using grid computing: application of the plain of Gondo (Burkina Faso)”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, 2012, P.227-235.

[11] Mohammad(Behdad) Jamshidi, Qaem Jahangiri etc, “Identify and simulation a furnace of steam boiler based on a new fuzzy modeling approach.pdf”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012, P452-458.

[12] Chee Khoon Ng, Kiang Hwee Tan, "Flexural behaviour of externally prestressed beams. Part I: Analytical model", Engineering Structures,28,2006, pp.609-621.

[13] Chen Ke, Song Jianyongetc.,"3D Finite Element Simulation and Mechanical Behavior Analysis for External Prestressed Bridges", Journal of Highway and Transportation Research and Development,Vol29,NO.2,2012,pp.77-81.

[14] Ahmed Ghallab, "Calculating ultimate tendon stress in externally prestressed continuous concrete beams using simplifed formulas", Engineering Structures,46(2013),pp.417-430.

[15] Chee Khoon Nga, Kiang Hwee Tan,"Flexural behaviour of externally prestressed beams. Part I: Analytical model",Engineering Structures,28(2006),pp.609-621.

[16] Yin Zihong, Li yuanfu, “Structural Deformation Forecasting Based On Support Vector Regression Trained By Particle Swarm Optimization Algorithm”, AISS: Advances in Information Sciences and Service Sciences, Vol.4, No.5, 2012 , pp.130-136.

[17] Yin Zihong, Li yuanfu, “Structural Deformation Forecasting Based On Support Vector Regression Trained By Particle Swarm Optimization Algorithm”, AISS: Advances in Information Sciences and Service Sciences, Vol.4, No.5, 2012, pp.130-136.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Comparisons and Countermeasures of Physical Status among 40-49 Different Groups in Weifang

Xiaoli Chen

School of Sports, Weifang University
Weifang, 261061, China

Abstract

In order to properly evaluate physical health status of adults in Weifang city, shape, physique, physical function and physical quality indicators of adults and the differences between urban people and rural people were probed by means of random sampling, measurement and comparative analysis. The results show that: urban non-manual adults were higher and heavier than urban manual ones and rural ones, while urban manual adults were higher and heavier than rural ones; urban manual adults had higher lung capacity and rest pulse rate than urban non-manual ones and rural ones; urban manual male adults had greater grip strength than urban non-manual ones and rural ones, while rural female adults had greater grip strength than urban manual and non-manual female ones; seated forward bends of urban non-manual adults were better than urban manual ones and rural ones. In general, there are not significant differences between physical quality of rural adults and urban ones.

Keywords: *adult, physique, physical function, physical quality.*

1. Introduction

In order to enrich and improve Chinese physique monitoring system and database, to learn the status and change of Chinese physique, to provide scientific basis for the national fitness program, according to "Sports Law of People's Republic of China", "National Program of Action for Fitness in China" and the "National Physical Fitness Monitoring Provisions", the 2nd national physical fitness monitoring was decided to carry out in 2005. Fitness was the basis of the overall quality of workers. The physical quality of the people was related to the socialist modernization, so studying the physical condition of the people of Weifang was important practical significant for national program of action for fitness.

To learn about the characteristics and conditions of adults' physical quality in Weifang city, in 2005, the city sports bureau organized and launched a large-scale monitoring of the adult physical quality. Reliable data were obtained; based on those data, the results were analyzed and compared; recommendations were given for the existing problems, what provided scientific basis for adults' fitness program, and serviced economic construction and social development.

2. Research Object and Methods

2.1 Research Object

Adults aged 40-49 in Weifang city were randomly selected for the study, every 5 years span was divided into an age group. 1,230 adults were measured, while the effective samples were 1200, which included 600 male and 600 female. For convenience, we assumed that the symbol A denoted urban non-manual adults, the symbol B denoted urban manual adults and the symbol C denoted rural adults.

2.2 Research Methods

Based on "national physical quality monitoring data entry card in 2005", the adults' physique, physical function and physical quality were monitored. We have used two physique indexes height (cm) and weight (kg), two physical function indexes lung capacity (ml) and rest pulse rate (times), two physical quality indexes grip strength (kg) and sit and reach (cm) to analyze and to compare the different of each groups' physical quality.

Literature data, questionnaire, the actual measurement and other methods were used. SPSS11.0 for Windows statistical package was used to process data.

3. Results

Test results of adults aged 40-49 in different groups and different stages (Table 1-4)

Table 1 The comparison of male adults aged 40-44 in different groups of Weifang city

Index	A		B		C	
	X	SD	X	SD	X	SD
Height(cm)	171.	6.07	169.	5.94	168.	5.48

	6		6		9	
Weight(kg)	74.5	11.31	75.1	10.34	70.7	11.61
Lung Capacity(ml)	4045	824	3735	703	3689	651
Rest Pulse Rate (times)	77.9	9.35	75.2	9.61	74.8	8.13
Grip strength (kg)	51.3	6.57	50.9	7.60	49.9	6.79
Seated forward bends (cm)	8.5	8.36	9.9	8.51	9.8	7.59

Index	t_{AB}	t_{AC}	t_{BC}
Height(cm)	0.24	0.13	0.25
Weight(kg)	1.12	2.15*	2.13*
Lung Capacity(ml)	0.26	2.66**	2.20*
Rest Pulse Rate (times)	0.23	0.24	0.33
Grip strength (kg)	3.10*	0.26	0.27
Seated forward bends (cm)	3.09**	3.04**	0.24

*p<0.05, ** p<0.01

Table 2 The comparison of male adults aged 45-49 in different groups of Weifang city

Index	A		B		C	
	X	SD	X	SD	X	SD
Height(cm)	170.8	5.91	170.1	5.42	167.7	4.74
Weight(kg)	74.9	10.97	71.1	10.34	69.1	11.11
Lung Capacity(ml)	3700	668	3568	605.7	3545	630
Rest Pulse Rate (times)	78.7	8.76	75.4	7.86	76.5	8.72
Grip strength (kg)	49.6	7.25	48.0	7.02	47.5	6.97
Seated forward bends (cm)	6.7	8.91	8.1	8.44	8.4	8.02

Index	t_{AB}	t_{AC}	t_{BC}
Height(cm)	0.13	2.11*	2.02*
Weight(kg)	0.45	1.23*	2.12*
Lung Capacity(ml)	0.13	0.16	0.12
Rest Pulse Rate (times)	0.09	2.15*	2.13*
Grip strength (kg)	3.12*	3.43**	0.12
Seated forward bends (cm)	3.01*	3.11*	0.22

*p<0.05, ** p<0.0

Table 3 The comparison of female adults aged 40-44 in different groups of Weifang city

Index	A		B		C	
	X	SD	X	SD	X	SD
Height(cm)	159.3	4.68	159.5	5.67	158.5	5.20
Weight(kg)	62.6	8.60	62.5	8.72	64.5	9.66
Lung Capacity(ml)	2895	402.1	2867	470.1	2694	441.2
Rest Pulse Rate (times)	75.7	7.33	76.1	8.59	74.9	7.54
Grip strength (kg)	31.2	4.59	31.2	5.12	31.8	6.33
Seated forward bends (cm)	11.9	5.81	10.6	7.76	12.6	6.90

Index	t_{AB}	t_{AC}	t_{BC}
Height(cm)	1.21	1.26	2.23*
Weight(kg)	0.12	0.25	0.35
Lung Capacity(ml)	0.12	2.23*	0.39
Rest Pulse Rate (times)	0.23	2.45*	0.13
Grip strength (kg)	2.26*	0.34	0.11
Seated forward bends (cm)	0.67	0.27	2.34*

*p<0.05, ** p<0.01

Table 4 The comparison of female adults aged 45-49 in different groups of Weifang city

Index	A		B		C	
	X	SD	X	SD	X	SD
Height(cm)	159.9	4.68	158.8	4.74	158.3	4.52
Weight(kg)	64.8	8.05	65.7	8.73	64.2	10.80
Lung Capacity(ml)	2806	533.1	2675	451	2589	453

Rest Pulse Rate (times)	75.7	7.85	74.8	9.06	74.7	9.08
Grip strength (kg)	30.6	4.44	29.8	4.69	30.4	6.01
Seated forward bends (cm)	10.9	7.36	11.9	7.67	10.7	6.15

Index	t_{AB}	t_{AC}	t_{BC}
Height(cm)	0.23	0.33	0.25
Weight(kg)	2.13*	2.11*	3.23**
Lung Capacity(ml)	2.11	2.18*	0.34
Rest Pulse Rate (times)	0.23	2.26*	0.12
Grip strength (kg)	2.12*	2.16*	0.23
Seated forward bends (cm)	0.12	0.34	0.55

$*p<0.05$, $** p<0.01$

4. Results and Discussion

4.1 Analysis and comparisons of adults' morphology in different groups in Weifang City

Height levels reflect the longitudinal growth of human bones which are influenced by age, environment, living standards, physical exercise and other factors.

It can be seen from Table 1 and 2 that different groups of men aged from 40 to 44 do not have significant differences. However, height levels of men in other age groups do have significant differences except the urban men and urban non-manual laborers. Tables 3 and 4 show that height levels of women aged 45-49 do have significant differences except height levels of the urban female laborers and farmers aged 40-44. This demonstrates that the nutritional status is good in the city. However, because urban non-manual labors aged from 45 to 49 suffer relatively great pressure and their exercise time is relatively small. Height levels between them and farmers do not have significant difference.

The living conditions of urban manual workers are good, their pressure is small, and so they have higher heights when compared with farmers. Weight is an important indicator of changes which reflects the human skeleton, muscle, fat, the weight gain that muscle, in general cases the growth of weight demonstrates muscle growth and improved nutritional status. It can be seen from Table 1 and 2 that in the two stages male adults in urban and urban non-manual laborers do not have significant differences, while each of them do have significant differences with farmers respectively. Table 3 and 4 show different groups of adult women aged 40-44 do not have significant differences. In the 45-49 periods significant differences can be found and urban laborers are heavier than both farmers and urban non-manual which indicates that women in the urban non-manual workers pay more attention to exercise, but the urban manual workers pay no attention. According to some materials, body weight of adults increases with the increases of fat weight components of muscle in a certain age, but up to a certain age, fat body weight is relatively stable or even decrease, and these increases of body weight are mainly the increases in fat.

The increase of body weight of adult female aged after 40 should be considered as the growth of fat. This situation may increase the risk of diabetes, heart disease and stroke and other diseases, so due attention should be paid to the phenomenon which the weight of adult females increases with increasing age.

4.2 Analysis of functional indexes between different groups adults in Weifang City

Vital capacity can reflect the size of the chest, lung elasticity and respiratory muscle strength and physical activity level. During the decline of the body's natural process of breathing both the respiratory function declines and the increases of diseases can be reflected by changes in vital capacity.

According to data analysis of Table 1 and 2 adult males in the 40-44 years old and urban non-manual laborers do not have significant differences, urban laborers and farmers have a significant difference and the urban non-manual workers and farmers have a very significant difference. All urban non-manual workers, urban laborers and farmers aged from 45-49 do not have a statistically significant difference. From the view of mean statics constitutions of both urban non-manual men and women are better than manual workers. These shows that manual labor is not a substitute for physical activity, manual labor can only exercise a part of the function which can not improve the full body, so continuous and comprehensive exercises should be adopted to improve physical condition.

The purpose of measuring the pulse is to understand the quiet resting heart rate which is one of the important physiological indicators used for prevention and treatment of cardiovascular disease. Heart rate reflects the person's cardiovascular reserve and is also an important early performance in the common cardiovascular disease. From Table 1 to 4 it can be seen that 40 to 44 year-old males of different groups do not have significant difference; while in the 45-49 years old urban non-manual workers and urban laborers and farmers respectively have a significant difference.

In the two phases in this urban non-manual women and female farmers have significant differences and others have non-significant difference which demonstrates manual work influence health. From the view of mean values point urban workers' mean values were higher than those of farmers, but with the increase of age, the mean pulse of adult men and women decline due to the development of the cardiovascular system tend to stability and maturation. Along with the gradually reduce in the level of metabolism, basal metabolic rate begins to decline, the body's physiological function starts to decline from the peak, which results in the decreased resting pulse.

4.3 Analysis of adults' physical indicators in different groups in Weifang City

Grip strength reflects the strength of forearm and hand muscles, data from Table 1 to 4 shows that both men and women of urban and urban non-manual laborers aged 40-44 years old have significant differences, the others have no significant difference. In the 45-49 year-old stage urban non-manual laborers have significant differences with urban laborers and a very significant difference with farmers, yet there are no significant differences between manual workers.

From the view of mean value, followed from high in the end are urban manual workers, farmers and urban non-manual workers which overall shows manual labor is conducive to the growth of upper body strength, but the upper body strength of farmers is weak due to the poor living conditions in rural areas. Sit and Reach reflect the torso, waist, hip and other parts of the joints, muscles, ligaments, flexibility and stretching targets, which have a close relationship people's daily life and movement.

Table 1-4 show that in the stage of male adults aged 40-44 urban non-manual laborers and farmers have very significant differences respectively, and urban non-manual workers have significant difference with farmers and urban laborers. Female adults in stage of 45-49 do not have significant difference, but in the phase of 40-44 urban laborers and farmers have a significant difference, which indicates that urban women pay more attention to physical exercise and nutritional health than men.

5. Conclusions and Suggestions

5.1 Conclusions

(1) Targets based on the analysis of morphological, functional indicators and physical indicators it shows that physical conditions of urban people is better than farmers due to the differences of the social environment, living, productive labor, nutrition and physical activity between urban and rural areas.
(2) Both urban and rural areas of the body shape, physical function and physical aspects, there is a lack of physical exercise is not scientific and reasonable. Both the urban and rural people have defects in the body shape, physical function and physical aspects, which can be contributed to unscientific and unreasonable exercise.

5.2 Suggestions

Sub-health has caused close attention as a hot issue by the whole society now. The results above objectively reflect that middle-aged people among 40-49 years old should focus on disease prevention and fitness and positively control sub-health. According to the above results, recommendations for the existing differences between different groups in Weifang City are put forward:

(1) Physical exercise should be determined by the results of physical and then scientific understand of your physical condition and carry out physical exercise. According to the relevant data scientific and reasonable activities can adjust the sub-health status, restoring health, cheer spirit. Therefore sub-health people need exercise prescription.

(2) In the development of social economy and improving people's living standards publicity fitness guide should also be strengthened, the government should be organized and planned to carry out the National Fitness, and effectively improve the national physique.

(3) The government should strengthen the construction of venues, provide protection for people's exercise, take great advantage of existing venues, and organize more high level competition to enrich the cultural life, put enhancing people's health first, and focus on social benefits.

References

[1] The State Sports General Administration Division, the National Monitoring Centre for National Physical Fitness. National Physical Fitness Monitoring in 2005, 2005.
[2] Priya R, Chitra M T and Elizabeth Sherly, "Categorization Of Social Networks Based On Multiplicity Constraints", International Journal of Computer Science Issues, Vol. 9, Issue 2, No 2, 2012,pp.533-537.
[3] W.H.Xing, et al. Physical Measurement and Evaluation. Beijing Sport University Press, 1986.
[4] Y.Chen, et al.Comparison of adult physical constitution in Zhenjiang, Jiangsu Province. Shanxi Normal University, 2006.
[5] J.J.Shi.Analysis of physical condition and influencing factors of adults in Zhejiang province. Zhejiang Sports Sciences, 2001.
[6] Ashraf Abbas M. Al-Modwahi, Behrang Parhizkar, Arash Habibi Lashkari,"Web-based AR Advertising & Branding for Proton Company", International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, 2012,pp.149-158.
[7] J.J.Chen. Survey on Adult Constitution in Hangzhou city. Beijing Sport University, 2005.
[8] W.D. Lu.SPSS for Windows statistical analysis. Electronic Industry Press, 2002.

On Postponed Production Optimization for Two-stage Supply Chain in Mass Customization

Guangxing Wei[1], Danping Xiong[2]

[1]**School of Management, Chongqing Jiaotong University**
Chongqing, 400074, China

[2] **School of Management, Chongqing Jiaotong University**
Chongqing, 400074, China

Abstract

Cost models were built for both traditional supply chain environment and mass customization environment. Then, the queuing theory of extending M/M/1 model is used to optimize the total cost. Through simulations, the lowest cost for both situations was explored, and the results show that retailer's ordering strategy as well as the location of the CODP has significant influence on the total cost. It can provide theoretical support for manufacturers to locate the CODP point correctly, and also provide a basis for making a decision in implementing postponed strategy for manufacturers.

Keywords: *Postponed production, Supply chain, Cost Optimization, Queuing Theory, Customer order decoupling point*

1. Introduction

In the current market environment, consumer demand is becoming personalized and diversified increasingly with large fluctuations. In the traditional mode of production, supply chain node enterprises plan and arrange production according to market demand forecast and order information, so there are a lot of uncertainty factors. Postponement production is an effective strategy to achieve mass customization, which can delay the final customization activity until the certain order is received, and by doing this, the postponement strategy can provide the differentiated, low-cost, high-quality final products for final market, and improve the economy profit for each supply chain node enterprise (Su, 2005). As a well-known fashion brand, Levi is facing the rapid change of market demand and the shortening life cycle, it moves the CODP (Customer Order Decoupling Point, CODP) to the end of production process, that is to say, Levi produce the standard product in the first place, and then finish the customization activity according to the demand for color and size of pants, which can reduce the inventory level and the uncertain risk.

There have been many researches focusing on the postponement strategy. The postponed production is applied by retailer to increase the profit, but the manufacture is ignored (Yang et al., 2009). The inventory managing strategy model is set in supply chain applying postponed strategy, but the total production cost was not considered (Shao, 2004 and Ji, 2009). Jin et al. (2011) proposed the method of “Kanban” to achieve customization, but they didn’t analyze whether the Kanban system can reduce the cost of manufacture in the quantitative way. Lee and Tang (1999) implemented the postponed production by restructure the product portfolio and set an simple model by combing the profit, cost and product strategy, and the model was an general postponed production model to achieve customization, but in their model, they thought the manufacture need different processes to obtain the customization after the CODP (Customer Order Decoupling Point). Hua (2007) optimized the flexible production system by queue theory. Van (1998), Krajewski (2005) and Rao (2007) research the postponed production in terms of simple customization and mass customization, but the key to their problem is whether the manufacture can reduce cost by customized postponement. Huang et al. (2008) compared the cost change before and after the CODP, but they ignored the payment of customer waiting cost to keep the customer purchasing the product. Dan (2009) set the cost optimization model for the two-stage supply chain made up of retailer and manufacture applying postponed production, and Li (2010) set the similar model and CODP orientation model, but both of them didn’t consider the cost change before and after the CODP only for the manufacture, i.e. they didn’t compare the cost change in different condition in applying postponed production or not applying the postponed strategy. Due to postponed production will lead to additional packaging, storage, loading and unloading, as well as the need to increase the storage space to store custom parts and general intermediate products, and may need to introduce flexible equipment to complete the customized manufacturing at the location close to the customer, which makes postponed production cost may be higher than non-postponed production, so it is necessary to analyze and compare the cost under non-postponed production and postponed production in the supply chain made up of the manufacturer and retailer.

This paper will focus the two-stage supply chain made up of a manufacturer and a retailer based on the research of Dan (2009) and Li (2010), and set the cost model when the manufacture in postponed production and without the postponed production (non-postponed production) respectively in the first place. And then we will adopt the extended M/M/1mode in queue theory to simulate the total postponed production cost in supply chain and obtain the result that the order strategy of retailer and the optimal location of CODP will take important influence on the total cost of supply chain. Finally, the sensitive analysis of order strategy of retailer (safe inventory quantity and order quantity) and the optimal location of CODP on the total cost is provided.

2. The Basic Mode of Supply Chain Cost

2.1 Notation

(1) The notation relative to manufacture

λ_1 , $m_1(r)$, $v_1(r)$, $h_{1,M}(r)$ and $E[I_{1,M}](r)$ is the unit average reaching ratio, unit average manufacture cost, average holding cost of processing product, average inventory cost and expected inventory level respectively. μ_1 is the expected demand of standard product. $\lambda_{2,k}$, $m_{2,k}(r)$, $v_{2,k}(r)$, $E[T_{2,k}](r)$ and $w_{2,k}$ is the average reaching ratio, unit average manufacture cost, average holding cost of processing product, expected production time and unit waiting cost of the K th final product , and $\mu_{2,k}$ is the expected demand of the K th final product. $F(r)$ denote the investment cost of implementing postponed production.

(2) The notation relative to retailer

b_k , $d_k(r,s,Q)$, h_{R_k} , $I_{R_k}(r,s,Q)$, a_k and $f_k(r,s,Q)$ denote the unit order cost, average order frequency, unit inventory cost, average inventory level, unit shortage cost and average shortage ratio of the retailer respectively.

2.2 The model under the non-postponed production mode

Assumption

(1) The manufacture will face the uncertain market demand and produce N kinds of customized products belonged to the same product family while keeping some safe inventory. Each kind of product need time T to be finished. The customer demand of the K th customized product is a random variable whose average value is $\lambda_{2,k}$, and the variant value is $\sigma^2_{2,k}$. The average demand amount of total customized products is $D=\sum_{k=1}^{N}\lambda_{2,k}$, where $\lambda_{2,k}$ is the unit demand reaching ratio of the K th product.

(2) The retailer face the final market directly, and the order strategy (s,Q) will be adopted, and P_i is the probability when the inventory is i ($i=0,1\ldots\ldots s+Q$).

(3) Only the form postponement is considered, but not the time and place postponement, so the transportation time between two enterprises can be ignored. The location of CODP can be denoted as the ratio of production time of standard product to the total production time, i.e. $r(0\le r\le 1)$, and $r=0$ denote the non-postponement strategy.

Model

For quantitative research of cost optimization of production cost of supply chain, we consider the total cost in an order cycle including manufacture cost $C_{M_1}(r,s,Q)$ and retailer cost $C_{R_1}(r,s,Q)$. Under the non-postponement strategy, the cost model is as

$$C_{T_1}(r,s,Q)=C_{M_1}(r,s,Q)+C_{R_1}(r,s,Q) \tag{1}$$

The total cost of manufacture in a production cycle should include manufacture cost, holding cost of processing product and inventory cost. The cost of retailer should include order cost, inventory cost and shortage cost. So $C_{M_1}(r,s,Q)$ and $C_{R_1}(r,s,Q)$ can be denoted as

$$C_{M_1}(r,s,Q)=\sum_{k=1}^{N}\lambda_{2,,k}(r)m_{2,k}(r)+\sum_{k=1}^{N}\frac{\lambda_{2,k}(r)}{\mu_{2,k}(r)}v_{2,k}(r)+\sum_{k=1}^{N}h_kQ_k \tag{2}$$

$$C_{R_1}(r,s,Q)=\sum_{k=1}^{N}b_kd_k(r,s,Q)+\sum_{k=1}^{N}h_{R_k}I_{R_k}(r,s,Q)+\sum_{k=1}^{N}a_kf_k(r,s,Q) \tag{3}$$

As a result, the formula (1) can be denoted as:

$$C_{T_1}(r,s,Q)=\sum_{k=1}^{N}\lambda_{2,,k}(r)m_{2,k}(r)+\sum_{k=1}^{N}\frac{\lambda_{2,k}(r)}{\mu_{2,k}(r)}v_{2,k}(r)+$$

$$\sum_{k=1}^{N} h_k Q_k + \sum_{k=1}^{N} b_k d_k(r,s,Q) + \sum_{k=1}^{N} h_{R_k} I_{R_k}(r,s,Q) + \sum_{k=1}^{N} a_k f_k(r,s,Q) \tag{4}$$

$$\text{s.t.} \quad r = 0 \tag{5}$$

$$0 \le s \le Q \tag{6}$$

$$0 < \frac{\lambda_{2,k}(r)}{\mu_{2,k}(r)} < 1 \tag{7}$$

In the formula (4)： $\lambda_{2,,k}(r)m_{2,k}(r)$, $\frac{\lambda_{2,k}(r)}{\mu_{2,k}(r)}v_{2,k}(r)$ and $h_k Q_k$ is corresponding to manufacture cost of the K th customized product, holding cost of processing product and inventory cost. $b_k d_k(r,s,Q)$, $h_k I_{R_k}(r,s,Q)$ and $a_k f_k(r,s,Q)$ denote the order cost, inventory cost and shortage cost of the K th customized product. Constraint (5) denotes the initial stage of CODP in non-postponement, i.e. $r = 0$, and Constraint (6) denote the order quantity should be bigger than safety inventory under order strategy (s,Q) . Constraint (7) denotes the production rate is bigger than order reaching rate, so there is no backlog order.

2.3 The model under the postponed production mode

Assumption

(1) The manufacture will face the uncertain market demand and produce N kinds of customized products belonged to the same product family by postponed production. The investment cost that the manufacture invest in the equipment so as to implement postponed production is $F(r)$. Each kind of product need time T to be finished. The production time can be divided into two parts: before CODP, the pushed production process needs time T_1 , and after CODP, the pulled production process needs time T_2 . The customer demand of the K th customized product is a random variable whose average value is $\lambda_{2,k}$, and the variant value is $\sigma_{2,k}^2$. The average demand amount of total customized products is $D = \sum_{k=1}^{N} \lambda_{2,k}$.

(2) Each kind of product need time T to be finished. The production time can be divided into two parts: before CODP, the pushed production process needs time T_1 , and after CODP, the pulled production process needs time T_2 . The location of CODP can be denoted as the ratio of production time of standard product to the total production time, i.e. $r(0 \le r \le 1)$.

(3) The retailer implement zero inventory strategy, and he order product from manufacture according to customized customer order, and all the products produced by manufacture will sold by retailer.

Model

The cost model under postponement strategy is as

$$C_{T_2}(r,s,Q) = C_{M_2}(r,s,Q) + C_{R_2}(r,s,Q) \tag{8}$$

The manufacture should consider the total cost in a production cycle including investment cost, manufacture cost before and after CODP, holding cost of processing product and inventory cost. Retailer should consider the order cost and customer waiting cost, so $C_{M_2}(r,s,Q)$ and $C_{R_2}(r,s,Q)$ can be denoted as

$$C_{M_2}(r) = F(r) + \sum_{k=1}^{N} \lambda_{2,k}(r)m_1(r) + \sum_{k=1}^{N} \lambda_{2,k}(r)m_{2,k}(r) + \frac{\lambda_1(r)}{\mu_1(r)}v_1(r) + \sum_{k=1}^{N} \frac{\lambda_{2,k}(r)}{\mu_{2,k}(r)}v_{2,k}(r) + h_{1,M}(r)E[I_{1,M}](r) \tag{9}$$

$$C_{R_2}(r) = \sum_{k=1}^{N} b_k \lambda_{2,k}(r) + \sum_{k=1}^{N} \lambda_{2,k}(r) w_{2,k}(r) E[T_{2,k}](r) \tag{10}$$

The formula (8) can be optimized and denoted as

$$C_{T_2}(r) = F(r) + \sum_{k=1}^{N} \lambda_{2,k}(r)m_1(r) + \sum_{k=1}^{N} \lambda_{2,k}(r)m_{2,k}(r) + \frac{\lambda_1(r)}{\mu_1(r)}v_1(r) + \sum_{k=1}^{N} \frac{\lambda_{2,k}(r)}{\mu_{2,k}(r)}v_{2,k}(r) + h_{1,M}(r)E[I_{1,M}](r) + \sum_{k=1}^{N} b_k \lambda_{2,k}(r) + \sum_{k=1}^{N} \lambda_{2,k}(r) w_{2,k}(r) E[T_{2,k}](r) \tag{11}$$

$$\text{s.t.} \quad 0 \le r \le 1, \tag{12}$$

$$0 < \frac{\lambda_1(r)}{\mu_1(r)}, \frac{\lambda_{2,k}(r)}{\mu_{2,k}(r)} < 1 \tag{13}$$

$$s = 0, \quad Q = 1 \tag{14}$$

In formula (11): $F(r)$ is the investment cost, $\lambda_{2,k}(r)m_1(r)$ and $\lambda_{2k}(r)m_{2,k}(r)$ is the manufacturing cost before and after CODP, $\frac{\lambda_1(r)}{\mu_1(r)}v_1(r)$ and $\frac{\lambda_{2,k}(r)}{\mu_{2,k}(r)}v_{2,k}(r)$ is the holding cost of processing product before and after CODP. $h_{1,M}(r)E[I_{1,M}](r)$ is the expected inventory cost of standard product, $b_k \lambda_{2,k}(r)$ is the order cost of the K th product, and $\lambda_{2,k}(r)w_{2,k}(r)E[T_{2,k}](r)$ is the customer waiting cost provided by retailer to keep customer.

Constraint (12) denotes the location of CODP can be at random stage in the whole production process. Constraint (13) denotes the production intension of manufacture is bigger than the reaching ratio of customer demand, which indicate that the overstock condition will not be appeared. Constraint (14) denotes the zero inventory strategy of retailer, safety inventory is zero and only one product is ordered at one time.

3. M/M/1 model

3.1 Assumption

Demand for the product can be regarded as basic to meet independence, common characteristics and obey Possion distribution. The production rule adopts the first come first serve, and the second stage of manufacture can be seen as a M/M/1 queuing system. For simple analysis, we will add some assumptions:

(1) All the customized products are belonged to the same product family K, and thus each customized products are similar and have the same production time, so we can assume that $\lambda_{2,k}$, $\mu_{2,k}$, $m_{2,k}$, $v_{2,k}(r)$, $E[T_{2,k}](r)$ can be equal λ_2, μ_2, $m_2(r)$, $v_2(r)$ and $E[T_2](r)$ respectively.

(2) With the location of CODP moving to the end of production process, more process will be standard and modular, and the manufacture center finishing more standard semi-finished products need be more flexible, leading the increasing investment cost. We can assume $F(r)$ is the simple increasing function of r. The customized products of product family are not the high additional-value product, and the incremental process is a continuous and even process of production time, i.e. $m_1(r)$, $h_{1,M}(r)$ are the simple increasing functions of r, and $m_2(r)$ is the simple decreasing functions of r.

(3) The unit inventory cost of product in retailer and manufacture is same, so for random product, b_k, h_k, a_k, d_k, I_{R_k} and f_k is equal as b, h, a, d, I_R and f respectively.

3.2 Model

We can obtain the efficiency index before and after CODP according to the research of Buzacott and Shanthikumar (1993), such as order frequency, average inventory, shortage rate, expected inventory of standard semi-finished product and expected production time in customized production process as following:

$$\frac{\lambda_2(r)}{\mu_2(r)} = \frac{(1\text{-}r)\mathrm{DT}}{\mathrm{N}} \tag{15}$$

$$d(r,s,Q) = \lambda_2 P_{s+1} = \frac{(1+\frac{1}{(1\text{-}r)\mathrm{DT}})^s}{\mathrm{T}(1\text{-}r)[1+\frac{Q}{(1\text{-}r)\mathrm{DT}}(1+\frac{1}{(1\text{-}r)\mathrm{DT}})^s]} \tag{16}$$

$$I_R(r,s,Q) = \sum_{i=0}^{s+Q} iP_i = \frac{\{(1+\frac{1}{(1\text{-}r)\mathrm{DT}})^s[\frac{Q^2+2sQ+Q}{2(1\text{-}r)\mathrm{DT}}\text{-}Q]+Q\}}{1+\frac{Q}{(1\text{-}r)\mathrm{DT}}(1+\frac{1}{(1\text{-}r)\mathrm{DT}})^s} \tag{17}$$

$$f(r,s,Q) = \lambda_2 P_0 = \frac{D}{\mathrm{N}[1+\frac{Q}{(1\text{-}r)\mathrm{DT}}(1+\frac{1}{(1\text{-}r)\mathrm{DT}})^s]} \tag{18}$$

$$E[I_{1,M}](r) = s_M - \frac{\rho_1(1-\rho_1^{s_M})}{1-\rho_1} = s_M - \frac{rDT-(rDT)^{s_M+1}}{1-rDT} \tag{19}$$

$$E_{T_2}(r) = \frac{1}{\mu_2-\lambda_2} = \frac{T(1-r)}{1-DT(1-r)} \tag{20}$$

s_M is the safety factor relative to inventory before CODP.

The queuing model under non-postponed production

The queuing model under non-postponed production can be extended as following:

$$C_{T_1}(r,s,Q) = Dm_2(r) + Dv_2(r) + NQh + Nbd(r,s,Q) + NhI_R(r,s,Q) + Naf(r,s,Q) \tag{21}$$

$$\text{s.t.} \quad r = 0 \tag{22}$$

$$0 \le s \le Q\,, \quad s \in Z^+ \tag{23}$$

$$0 < DT < 1 \tag{24}$$

In the formula (21): the first three items denote the manufacture cost, holding cost of processing product and inventory respectively. The forth item denote the order cost, fifth item denote the holding cost and the sixth item is the shortage cost. Constraint (22) denotes the location of CODP can be at random stage in the whole production process. Constraint (23) denotes the zero inventory strategy of retailer, safety inventory is a positive integer.

Constraint (24) denotes the production intension of manufacture is bigger than the reaching ratio of customer demand, which indicate that the overstock condition will not be appeared.

The queuing model under postponed production

The queuing model under postponed production can be extended as following:

$$C_{T_2}(r) = F(r) + Dm_1(r) + Dm_2(r) + rDTv_1(r) + (1\text{-}r)DTv_2(r) + h_{m_1}(r)E[I_{1,M}](r) + Db_k + DwE_{T_2}(r,s,Q) \quad (25)$$

$$\text{s.t.}\quad 0 \le r \le 1 \quad (26)$$

$$s = 0\,,\quad Q = 1 \quad (27)$$

$$0 < DT < 1 \quad (28)$$

$$v_1(r) < h_1(r) < v_2(r) \quad (29)$$

$$\frac{\partial F(r)}{\partial r} > 0\,;\quad \frac{\partial m_1(r)}{\partial r} > 0\,;\quad \frac{\partial h_1(r)}{\partial r} > 0\,;\quad \frac{\partial m_2(r)}{\partial r} < 0 \quad (30)$$

In formula (25), $F(r)$ is the investment cost caused by customization, the second and third item is the manufacturing cost before and after CODP, the fourth and fifth item is the holding cost of semi-finished product before and after CODP, the sixth item is the inventory cost before CODP, seventh item is the order cost of retailer, and the eighth item is the customer waiting cost provided by retailer to keep customer.

4. Simulation

For further research, we will apply MATLAB software to simulate the two models under postponed production mode and non-postponed production, so as to compare the cost in the two different production modes. By referring to the data in research of Rietze (2006), the parameter value of production cost is as following: $F(r) = 12 + 0.5r$, $m_1(r) = r$, $m_2(r) = 5 - r$, $D = 1000$, $T = 0.0008$, $v_1(r) = 0.4r$, $v_2(r) = 2\text{-}0.4r$, $h_M(r) = 0.5r$, $s_M = 2$, $N = 8$, $w = 0.5$, $h_R = 0.5$, $b = 0.2$, $a = 0.6$.

4.1 The simulation expression under non-postponed production

The cost simulation is shown in figure 1, and the axis X denotes the safety inventory, Y denotes the order quantity of retailer, and Z denote the total cost of supply chain. The cost in non-postponed production is relative to the inventory strategy of retailer, and when the safety inventory and order quantity change, the total cost of supply chain will change, and it will increase as they increase. The minimum cost is7085.9, the safety inventory is 1, and the order quantity is 10.

4.2 The simulation expression under postponed production

The cost simulation is shown in figure 2, X denote the CODP location and Y denote the total cost of supply chain. The total cost of supply chain is a decreasing function of r .

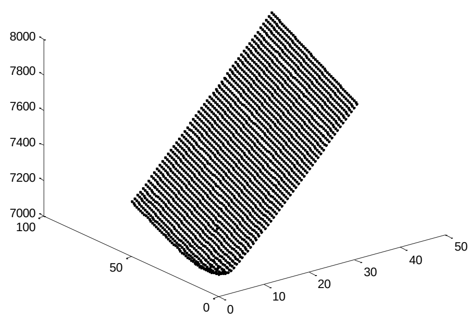


Fig.1 Cost change of supply chain in non-postponed production

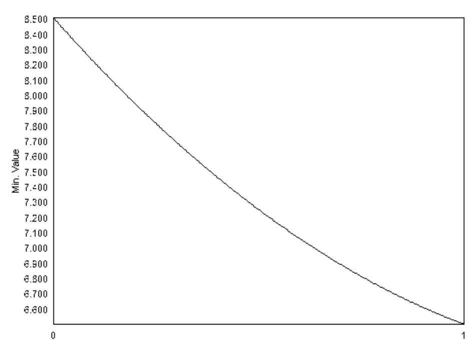


Fig.2 Cost change of supply chain in postponed production

4.3 Cost analysis in two production modes

The influence of order strategy of retailer on the total cost

From the simulation result, the optimal inventory strategy of retailer is $s = 1$ and $Q = 10$, so we can let $s = 1,2\ldots\ldots6$, and analyze the total cost of supply chain. The obvious factor influencing the total cost of supply chain is the order strategy of retailer, and the total cost change with the order strategy, i.e. safety inventory and order quantity. When $S = 1$ and $Q = 10$, the minimum cost of supply chain is7085.9. Given the same order quantity, if the safety inventory level is higher, then the total cost of supply chain will be higher, and the inventory cost of retailer will increase. When the safety inventory level is lower, the total cost of supply chain will decrease with order quantity in advance and then increase with order quantity increasing, because when the order quantity is small, the order frequency of retailer will increase, which will lead to the increment of order cost exceeding the saving of decreased inventory, so after the optimal point with minimum cost, as the order frequency deceases in a order cycle, the saving of order cost is smaller than the increment of inventory cost, so the total cost will increase as order quantity increase. When the safety inventory is high, under the given inventory level, the total cost of supply chain will increase as order quantity increase.

The influence of CODP on the total cost

From the simulation, the optimal location of CODP is at the end of production process, so we can let $r = 0, 0.2, 0.4, 0.6, 0.8, 1$ to analyze the total cost of supply chain.

In table 2, under postponed production, the key factor influencing the cost of supply chain is CODP location, and when the CODP location moves to the end of production process, the customer waiting cost will decrease, and the increment of other cost of supply chain is smaller than the saving caused by customer waiting cost, so the total cost of postponed production system will decrease and postponed production can reduce the cost of supply chain. By comparing the table 1 and table 2, the minimum cost and maximal cost under non postponed production mode is 7085.9 ($s=1$, $Q=10$) and 7313.5($s=1$, $Q=1$) respectively, and the corresponding data under postponed production is 6506.3 ($r=1$) and 8513.8 ($r=0$). The cost under postponed production is not always smaller than the cost under non postponed production, for example, when $r \le 0.4$, the cost under postponed production will be bigger than 7313.5, which is bigger than 7313.5($s=1$, $Q=1$) under non postponed production, so the manufacture should choose the right CODP so as to reduce the cost of supply chain and improve the efficiency of production.

Table 1: The influence of order strategy of retailer on the total cost of supply chain under non-postponed production

C_{T_1} / s / Q	1	2	3	4	5	6
1	7313.5	-	-	-	-	-
2	7190.5	7155.2	-	-	-	-
3	7144.3	7118.1	7107.6	-	-	-
4	7122.1	7101.7	7094.2	7093	-	-
5	7110.5	7094.1	7088.5	7088.2	7090.3	-
6	7104.3	7090.8	7086.7	7087.1	7089.4	7092.7
7	7101.5	7090.1	7087.1	7087.9	7090.5	7093.9
8	7100.8	7091.1	7088.8	7090	7092.8	7096.2
9	7086.2	7093.2	7091.6	7093	7095.9	7099.4
10	7085.9	7096	7094.9	7096.6	7099.6	7103.2
11	7086.4	7099.4	7098.7	7100.6	7103.7	7107.3
12	7087.5	7103.3	7102.9	7105	7108.1	7111.8

Table 2 : The influence of CODP on the total cost of supply chain under postponed production

r	0	0.2	0.4	0.6	0.8	1
C_{T_2}	8513.8	7947.8	7466.1	7065.8	6745.9	6506.3

5. Conclusion

This paper set the basic model and extended M/M/1 queuing model under non-postponed production and postponed production mode respectively. By solving the model and compare the total production cost, we found the order strategy of retailer is the key factor influencing the total cost of supply chain under non postponed production, so we can implement effective order strategy, such as make sure the right safety inventory level and order quantity to control the cost within a lower level. Under the postponed production, CODP location will bring the obvious influence on the production cost of manufacturer, and if the manufacturer can't choose the right CODP location in postponed production mode, then the production cost will exceed the cost of non-postponed production mode. So the manufacturer should make right decision on the postponed production and choose the right CODP location according to the characteristic of product and market demand. Besides, we only consider the form postponement, the combination of form postponement and time postponement should be considered into the model in the future research.

References

[1] J. C. P. Su, Y. L. Chang, and Ferguson Mark, "Evaluation of postponement structure to accommodate mass customization" [J], Journal of Operations Management, Vol. 23, No. 3/4, 2005, pp. 305-318.

[2] Z. L. Yang, and Y. M. Tian, "A study of partial backorder policy under random demand "[J], Journal of Systems & Management, Vol. 18, No. 6, 2009, pp. 686-691.

[3] X. F. Shao, and J. H. Ji, "Research on impact of postponement strategies on inventory cost in mass customization" [J], System Engineering Theory Methodology Application, Vol. 13, No. 5, 2004, pp. 385-389.

[4] X. J. Xu, and W. L. Zhou, "Simulation Study on Virtual Inventory Decision Making" [J], Industrial Engineering Journal, Vol. 12, No. 5, 2009, pp. 76-81.

[5] Q. Jin, and Z. Zhang, "A visualized e-kanban system for small batch JIT production" [J], Industrial Engineering Journal, Vol. 14, No. 3, 2011, pp. 137-145.

[6] H. L. Lee, and C. Tang, "Modeling the costs and benefits of delayed product differentiation" [J], Management Science, Vol. 43, No. 1, 1997, pp. 40-53.

[7] Z. S. Hua, Flexible manufacturing systems and flexible supply chain—Modeling optimization [M], Beijing: Science Press, 2007.

[8] H. R. I. Van, "Logistics and virtual integration: postponement, outsourcing and the f1ow of information"[J], International Journa1 of Physical Distribution & Logistics Management, Vol. 28, No. 7, 1998, pp. 508-523.

[9] L. Krajewski, J. C. Wei, and L. L.Tang, "Responding to schedule changes in build-to-order supply chains"[J], Journal of Operations Management, Vol. 23, No. 5, 2005, pp. 452-469.

[10] K. Rao, B. Dan, and Y. Liu, "Research on cost optimization model of manufacturer implementation postponement strategy in mass customization"[J], Computer Integrated Manufacturing Systems, Vol. 13, No. 9, 2007, pp. 1665-1671.

[11] Y. Y. Huang, and S. J. Li, "Suitable application situations of different postponements approaches: Standardization vs. Modularization"[J], Journal of Manufacturing Systems, Vol. 27, No. 2, 2008, pp. 111-122.

[12] K. Rao, B. Dan, and Y. Liu, "Cost optimization model of supply chain implementation postponement strategy in mass customization"[J], Computer Integrated Manufacturing Systems, Vol. 15, No. 2, 2009, pp. 287-291.

[13] D. D. Li, and J. X. Wang, "Customer order decoupling point position for mass customization"[J], Journal of Qingdao University (Natural Science Edition), Vol. 23, No. 2, 2010, pp. 89-92.

[14] S. M. Rietze, "Case Studies of Postponement in the Supply Chain"[D], Cambridge: Massachusetts Institute of Technology. 2006.

[15] Noureddine Labiad, Zitouni Beidouri and Otmane Bouksour, "Evaluation of coordination contracts for a two stage Supply Chain under price dependent demand", International Journal of Computer Science Issues, Vol. 9, No 1, 2012, pp. 151-159.

[16] Benaissa Ezzeddine, Benabdelhafid Abdellatif and Benaissa Mounir, "An Agent-based framework for cooperation in Supply Chain", International Journal of Computer Science Issues, Vol. 9, No 3, 2012, pp. 77-84.

GuangXing Wei Chongqing Province, China. Birthdate: Aug, 1977. Work as professor in Management School of Chongqing Jiaotong University, and research interests on mass customization and supply chain management.

Danping Xiong Jiangxi Province, China. Graduate student in Management School of Chongqing Jiaotong University, and research interests on mass customization and postponement strategy.

Production Cost Optimization Model Based on CODP in Mass Customization

Yanhong Qin[1], Yuanfang Geng[2]

**[1]School of Management, Chongqing Jiaotong University
Chongqing, 400074, China**

**[2] School of Management, Chongqing Jiaotong University
Chongqing, 400074, China**

Abstract

The key for enterprises to implement the postponement strategy is the right decision on the location of Customer Order Decoupling Point (CODP) so as to achieve the scope economics of mass customization and scale economics of mass production fully. To deal with production cost optimization problem of postponement system based on various situation of CODP, a basic model of production cost and its M/M/1 extended model are proposed and compared so as to optimize the overall production cost of the postponement system. The production modes can be classified as MTS (make to stock), ATO (assemble to order), MTO (make to order) and ETO (engineering to order) according to the inventory location, and the postponed production system considered here includes manufacturing cost, semi-finished inventory cost and customer waiting cost caused by delaying delivery. By Matlab simulation, we can compute the optimal location of CODP in each production mode, which can provide some management insight for the manufacturer to decide the right production mode and utilize the resources efficiently.

Keywords: Mass customization, Postponed production, Production cost, Customer order decoupling point

1. Introduction

With the market changing frequently, more and more companies are starting to adopt mass customization (MC) to provide a larger degree of product customization to fulfill the various demands of increasingly differentiated market segments. The advanced information technique can also provide more companies with more possibility to interact with customers and obtain individual customer requirements effectively (Huang, 2008). Postponement strategies allow a company to be flexible in developing different versions of the product as needed, to meet changing customer needs, and to differentiate a product or to modify a demand function. As many researchers pointed out, the customer order decoupling point (CODP) technology as a technique of postponement is an effective way to achieve MC, and many firms made out the different location of CODP to revise their supply chain so as to meet various requirements of different customized extents, by doing this it can meet exact requirement and lessen the often-painful effects of upstream order magnification, i.e. "bullwhip effect" (Alessandro, 2009). CODP means customer order decoupling point, that is to say, the place in the supply chain where the customer requirements are permitted to penetrate up the value chain of supply before differentiation of the product is started (Ricardo, 2001). The manufacturing strategy planned based on the forecast started to customized production driven by customer order at CODP, i.e. CODP is related to manufacturing strategy with a distinction is necessary between pre-CODP and post-CODP operations, since these have fundamentally different characteristics. The post CODP means the product is linked to a specific customer order. Before CODP, the mass production can be implemented to achieve the high production efficiency and scale economy, and post CODP, the flexible manufacture process should be adopted to route greater volumes of compatible products through a fixed asset (Lee, 1996 and Lee, 1997). So in postponed production condition, there are two production stages, i.e. mass production before CODP and customized production after CODP, and the location of CODP is very important for the enterprise in terms of production cost.

Italian clothing company Benetton adopted the postponement strategy to delay the CODP in the production process until to the end production process, and thus to reduce the risk of uncertainty demand. There have been many theories and literatures on postponement strategy. The postponed production is applied by retailer to increase the profit, but the manufacture is ignored (Yang et al., 2009). The inventory managing strategy model under the centralized and decentralized decision mode is set in supply chain applying postponed strategy, but the total production cost was not considered (Shao and Ji, 2004). Jin et al. (2011) proposed the method of Kanban to achieve customization, but they didn't analyze whether the Kanban system can reduce the cost of manufacture in the quantitative way. Sun (2010) implemented the postponed production by restructure the product portfolio and set an simple model by combing the profit, cost and product strategy, and the model was an general postponed production model to achieve customization, but in their model, they thought the manufacture need different processes to obtain the customization after the CODP (Customer Order Decoupling Point). Hua (2007) optimized the flexible production system by queue theory. Van (1998), Krajewski (2005) and Rao (2007) research the postponed production in terms of simple customization and mass customization, but the key to their problem is whether the manufacture can reduce cost by customized postponement. Huang et al. (2008) compared the cost change before and after the CODP, but they ignored the payment of customer waiting cost to keep the customer purchasing the product. Dan (2009) set the cost optimization model for the two-stage supply chain made up of retailer and manufacture applying postponed production, and Li (2010) set the similar model and CODP orientation model, but both of them didn't consider the cost change before and after the CODP only for the manufacture, i.e. they didn't compare the cost change in different condition in applying postponed production or not applying the postponed strategy.

Garg (1997) studied the postponed production model with many potential CODPs. Swaminathan (1998) found out that the general components and modules can reduce the inventory cost of the postponed production system, but when the cost of design and manufacture is too high, it is uneconomic to implement the postponed production. Ma (2002) studied the multi-assembly postponed production problem under assemble to order mode, and he analyzed the relation between the general components /modules and production cost. Tibben and Bassok (2005) found out the postponed production can reduce the total production cost based on the inventory model when the postponed production was applied. Su (2005) set the model of time postponement and form postponement, and analyzed the implementation condition of these two postponement strategies. Dan (2009) studied the cost optimization model under mass customization for manufacture and supplier, and analyzed the influence of CODP on the systematic production cost and the key factors on the CODP location. Yang (2010) studied the cost evaluating model of mass customization based on CODP locating, and set the cost model reflecting the influence of multi-CODP on the assemble products.

Most of these researches focused on the postponed production system with fixing CODP location and without considering the production leading time or fixing the leading time, few referred to the multi-CODP location problem. Based on these researches, this paper will consider the manufacturing cost, semi-finished inventory cost and customer waiting cost caused by delaying delivery to set the cost optimization model of basic model its M/M/1 extended model so as to optimize the overall production cost of the postponement system. The production modes can be classified as MTS (make to stock), ATO (assemble to order), MTO (make to order) and ETO (engineering to order) according to the inventory location, and the postponed production system considered here includes manufacturing cost, semi-finished inventory cost and customer waiting cost caused by delaying delivery. By Matlab simulation, we can compute the optimal location of CODP in each production mode, which can provide some management insight for the manufacturer to decide the right production mode and utilize the resources efficiently.

2. The production modes under mass customization

The key to carry out postponement strategy in mass customization is CODP. Before CODP, enterprise mainly adopt push supply chain which mainly forecast market demand and implement large-scale production, and the goal is to improve supply chain efficiency and reduce the cost of the supply chain. After CODP, enterprise can mainly adopt pull supply chain which depend on customer orders and implement small-scale processing, and the goal is to improve the response speed of the supply chain as well as the ability to provide customers with customized products and service.

In general, supply chain activities include supply, design, manufacture, assembly and retail. The position of CODP is changeable. The more upstream of CODP in the supply chain, the more obvious pull supply chain, and the higher the degree of customer participation. The more downstream of CODP in the supply chain, the more obvious push supply chain, the lower the degree of customer participation. The different position of CODP lead to different modes of production, this is a new understanding of these four production modes based on cost and customization. As shown in Figure 1, the vertical axis represents the cost, and the horizontal axis represents customization level. When other conditions are same, scale advantage plays a very important role in reducing the production cost, but it also reduces customization level. At this point, the cost and customization level are in a low level, enterprises will adjust the position of CODP according to product characteristic and market demand characteristic, every points will result in corresponding changes about the costs and customization level. As shown in Figure 1, from MTS (Make to stock) to ETO (Engineer to order), the costs and customization level increase or improve gradually.

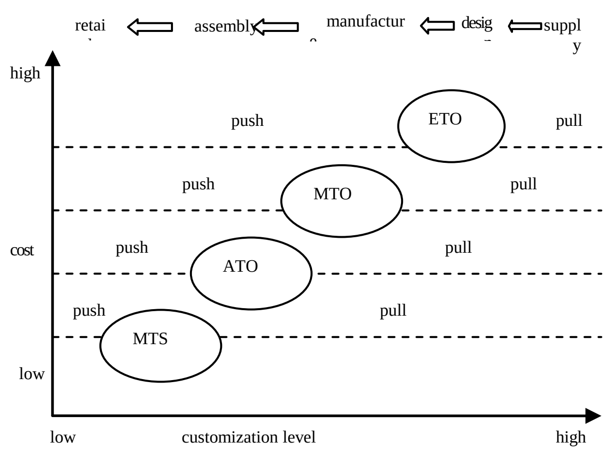


Fig.1 Four production mode based on different positioning of CODP

Under MTS, CODP is positioned in retail activity. Enterprises forecast products fully depend on market demand and arrange production according to their own inventory. In the supply chain, there is not the participation of customers from the supply to the retail, customers select their favorites in the final product, the entire supply chain can quickly and timely response to customer demands, and the costs is low, but the available choice of customers is very limited. Under this production mode, customization level is very low, and most of the products are popular consumable. Under ATO, CODP is positioned in assembly, enterprises produce large-scale standardized components and assemble existing standardized components and modules to meet customer order requirements after receiving customer orders, customers participate in assembly aspects. Under this production mode, production cycle is reduced and the response time to customers is prolonged, costs and customization level are both improved relatively. Enterprises can avoid overstock or out of stock, and they can also meet customers demand quickly. Most products are about electronic equipment, automobiles, etc. Under MTO, CODP is positioned in manufacture activity, enterprises arrange production plans with purchased components and modules to meet the needs of customers, customers participate in manufacture aspects. Under this production mode, advantages of scale economy is gradually weaken, the response time to customers are longer, cost and customization level are relatively higher. Enterprises will arrange production only receiving customer orders, not to make any inventory. Most products are about aircraft, ships, etc. Under ETO, CODP is positioned in design activity, enterprise entirely design product according to customer order, and then proceed to the procurement, production and other activities, and customers participate in design aspects from the beginning, enterprises meet the demand of customer service at the greatest degree. Under this production mode, production costs and customization level are highest. Such enterprises generally have a high degree of product design management capabilities, and most products are about complex structure, such as special test equipment, generator sets, etc (Qin and Geng, 2012).

3. The basic model under postponed production mode

3.1 Problem description

The manufacture will face the uncertain market demand and produce N kinds of customized products belonged to the same product family by postponed production with the same total production time S. Each final product will go through two production stage, i.e. mass production and customized production, and the first stage is the modular semi-finished production pushed by forecast information, and the second stage is the customized production pulled by order. r denote the customer order penetration point, i.e. CODP, which is key to implement postponed production and it will link the two production stages. r $(0 \le r \le 1)$ can stand for the ratio of production time in stage 1 to total production time, and $r=0$ can indicate the manufacture to choose the ETO production mode, and all the production activities are driven by orders, but $r=1$ indicate the MTS should be applied, and all the products are produced according to the forecasting information and stored as inventory. The bigger value of r denotes the longer mass production time and shorter customized production time.

3.2 Assumption

(1) Whether to implement the postponed production will not influence the transportation time, so we will not consider the transportation cost. The location of CODP can be denoted as the ratio of production time of standard product to the total production time, i.e. $r(0 \le r \le 1)$.

(2) The customer demand of the K th customized product is a random variable observing Poisson distribution whose average value is λ_k , and the variant value is σ_k^2 . λ_k means the average demand quantity of customized product K in unit time.

(3) The manufacture will produce some product and keep the inventory according to the forecast information of the market demand, and the produce intension will be bigger than the reaching ratio of customer demand, so there will be no shortage cost.

(4) For the CODP divide the production into two stages and in the second stage, the final product will delivered to customer directly once it is finished driven by order, so the there is no inventory in the second stage, and we only consider the buffer inventory at CODP.

3.3 Denotation and model

To better study the influence of CODP location on the production cost quantitatively, the production cost includes only the manufacture cost $M(r)$, inventory cost $H(r)$ and customer waiting cost $W(r)$. For the order delivery needs some lead time, and the longer the lead time is, the longer the waiting time is, leading to the more waiting cost, so time is the measure standard of the waiting cost. For the production cost can be divided as two parts: the manufacturing cost $M_1(r)$ in the stage 1 of mass production and the manufacturing cost $M_2(r)$ in stage 2 of customized production. So we can denote the production cost in postponed production as:

$$Z(r) = M_1(r) + M_2(r) + H(r) + W(r) \tag{1}$$

In the formula, the relative denotation in $M(r)$ is as following: $\lambda_{2,k}$ is the unit demand reaching ratio of the K th customized product, $c_1(r)$ is the unit average production cost in stage 1 when the CODP location is at r and $c_{2,k}(r)$ is the unit average production cost of the K th product in stage 2 when the CODP location is at r . The relative denotation in $H(r)$ is: $h_1(r)$ average inventory cost of semi-finished inventory in stage 1, $E[I_1](r)$ is the expected inventory amount of semi-finished inventory. So the cost optimization model can be transformed as:

$$\min Z(r) = \sum_{k=1}^{N} \lambda_{2,k} c_1(r) + \sum_{k=1}^{N} \lambda_{2,k} c_{2,k}(r) \\ + h_1(r) E[I_1](r) + \sum_{k=1}^{N} w_{2,k} \lambda_{2,k} ET_{2,k}(r) \tag{2}$$

$$s.t. \quad 0 \le r \le 1 \tag{3}$$

Each term in formula (2) is corresponding to the cost item in formula (1), and it means the total production cost in postponed production system. The first item in (2) is the manufacture cost in stage 1, the second item is the customization cost in stage 2, the third term is inventory cost of semi-finished product at CODP, and the forth item is the customer waiting cost in stage 2. Constraint (3) denotes the location of CODP can be at random stage in the whole production process.

4. The M/M/1extended model of postponed production

4.1 Assumption

(1) The production process is a value increasing process, so $c_1(r)$ is simple increasing function of r , and it will increase with delaying CODP to the downstream of production process, i.e. $\Delta M_1(r) = M_1(r+1) - M_1(r) \ge 0$. Besides, the inventory cost is increasing in the production process, so $h_1(r)$ is all simple increasing function of r , $\Delta H(r) = H(r+1) - H(r) \ge 0$.

(2) The customized products of product family are not the high additional-value product, and the incremental process is a continuous and even process of production time and $c_{2,k}(r)$ is the simple decreasing functions of r , i.e. $\Delta M_2(r) = M_2(r+1) - M_2(r) \le 0$.

(3) Not all the customer would like to wait some time for the customized product, and the longer the waiting time, the bigger the lost order, so $w_{2,k}$ is decreasing functions of r , i.e. $\Delta W(r) = W(r+1) - W(r) \le 0$.

(4) All the customized products are belonged to the same product family, and thus each customized products are similar and have the same production time, so we can assume that k , $\lambda_{2,k}$, $c_{2,k}(r)$, $w_{2,k}$, $ET_{2,k}(r)$ can be equal to λ_2 , $c_2(r)$, w and $ET_2(r)$ respectively.

(5) In the random time zone, the probability of customer demand for the certain customized product is dependent on the length of time zone, but it independent of the terminal of time zone, and if the time zone is small enough, the probability that the number of customer demand for the certain customized product is more than twice can be ignored. Besides, in each independent time zone, the demand of product is independent.

(6) The production process in stage 1 and stage 2 is smoothly, and the production time and customer order is independent.

(7) The manufacture will adopt the first come first service rule to finish the order, so the production system can be seen as an M/M/1 queuing system.

The better your paper looks, the better the Journal looks. Thanks for your cooperation and contribution.

4.2 Denotation and model

h is the basic inventory in stage 1, D is the average total demand and $D=\sum_{k=1}^{N}\lambda_{2,k}$, $\rho_1(r)$ is the amount of processing product, and $\rho_{2,k}(r)$ is the amount of processing product of the k th final product.

We can obtain the efficiency index before and after CODP according to the research of Buzacott and Shanthikumar (1993). According to the assumption of extended model, the manufacture intension of standard semi-finished product is $\rho_1(r)$ before CODP, and the production intension of customized product is $\rho_{2,k}(r)$. So we can get the expected inventory in stage 1 and expected production time in stage 2 as following:

$$\rho_1=\lambda_1\mu_1^{-1}=rDS;$$

$$\rho_2=\lambda_2\mu_2^{-1}=(1-r)DS;$$

$$E[I_1]=h-\rho_1(1-\rho_1{}^h)(1-\rho_1{}^{-1})$$

$$=h-[rDS-(rDS)^{h+1}](1-rDS)^{-1}$$

$$ET_2=(\mu_2-\lambda_2)^{-1}=N(1-r)S[1-D(1-r)S]^{-1}.$$

So the production cost model can be transformed as

$$\min Z(r)=Dc_1(r)+Dc_2(r)+h_1(r)[h-\frac{rDS-(rDS)^{h+1}}{1-rDS}]+\frac{NwD(1-r)S}{1-D(1-r)S} \tag{4}$$

$$s.t.\quad 0\le r\le 1 \tag{5}$$

$$0\le DS\le 1 \tag{6}$$

$$\frac{\partial c_1(r)}{\partial r}>0;\frac{\partial h_1(r)}{\partial r}>0 \tag{7}$$

$$\frac{\partial c_2(r)}{\partial r}<0 \tag{8}$$

$$h\in Z^+ \tag{9}$$

The formula (4) denotes the total production cost of postponed production in mass customization. In formula (4), the first item denotes the manufacturing cost in mass production of stage 1, the second item denotes the manufacturing cost in customized production of stage 2, the third item denotes the inventory cost in stage 1, and the forth item denotes the customer waiting cost. Constraint (5) denotes the location of CODP can be at random stage in the whole production process. Constraint (6) means that the overstock condition will not be appeared, because the production intension of manufacture is bigger than the reaching ratio of customer demand. In constraint (7) can be obtained easily by assumption (1), i.e. $c_1(r)$ is simple increasing function of r and $h_1(r)$ is all simple increasing function of r. Constraint (7) means that the product value is increasing with production time in stage 1, and the inventory cost of standard semi-finished product is increasing with its value. Similarly, constraint (8) means the product value in stage 2 is increasing with production time, when the production time in stage 2 is shortening, the product value increment will decrease. Finally, constraint (9) indicates the safety stock is an integer.

5. Simulation and computation

For further research, we will apply MATLAB software to simulate the influence of four production modes on the production cost and the optimal CODP location in each production mode. According to the case data in research of Rietze (2006), the value of each parameter is as following: $h=3$, $h_1(r)=0.6r$, $D=1000$, $S=0.0009$ and $N=10$. For the data in this research, the unit inventory cost and customer waiting cost has little influence on the total production cost, so we only consider the effect of unit manufacturing cost on the optimal CODP location. In the following simulation figures, the horizontal axis denote the CODP location, and with the CODP moving to the end of production process, the location on the axis is inclined to moving toward the right side. The vertical axis denotes the manufacturing cost.

5.1 The optimal CODP location in MTS

The products under MTS are about a large class of low complex standardized products, costs and delivery time are key factors in delivery process, which requires enterprises to produce and store large amounts of standardized products, such as the largest food and beverage manufacturer Wahaha, the species has been involved in the drinking water, carbonated drinks, tea drinks, milk drinks and other beverages industries to meet the diverse needs of customer. In the MTS, $c_1(r)=2r^2$, $c_2(r)=3-4r$.

$$Z(r)=1000(2r^2)+1000(3-4r)+1.8r-\frac{0.6r[0.9r-(0.9r)^4]}{1-0.9r}+\frac{4.5(1-r)}{1-0.9(1-r)}$$

The simulation result is shown in fig 2.

From fig 2, the manufacturing cost is decreasing with CODP moving to the end, and the optimal CODP location with the minimum cost is r=1 , i.e. all the production process is mass production with scale economics. For easy decision on the CODP location, the sensitive analysis of $c_1(r)$ and $c_2(r)$ is shown in table 1.

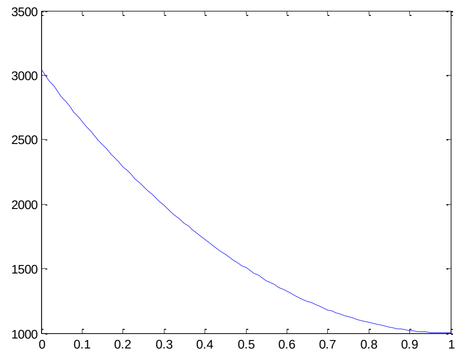


Fig.2 The trend of production cost changing under MTS

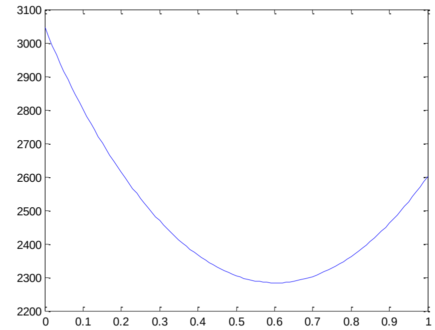


Fig. 3 The trend of production cost changing under ATO

From table 1, the production process in stage 1 is an value increasing process, i.e. $c_1(r)$ is an increasing function, its value will increase with CODP moving to the end process, and the manufacturing cost $M_1(r)$ will increase too. $c_2(r)$ is a decreasing function, its value will decrease with CODP moving to the end process, and the manufacturing cost $M_1(r)$ will decrease too. When $r \le 1$, c_1 will increase with CODP moving to the end process, but c_2 will decrease. For all the production process is mass production, there exits obvious scale economics, so $M(r)$ is decreasing, and $Z(r)$ is decreasing.

5.2 The optimal CODP location in ATO

When ATO is adopted, customization level is relatively low and the cost is relatively high. Under ATO, enterprises combine different standardized modules into a limited variety of products to meet customer demand, although choice of customers is limited, but it allows enterprises to produce mass standard parts and components to get the advantage of economies of scale, and it also allows enterprises to get components by outsourcing to provide a reasonable price to the customer, such as computer industry, enterprises use the advantages of manufacturing resources around the world to produce parts and to assemble in places close to the sale of land, For example, Dell put their main focus on co-operation with suppliers in order to share information with suppliers, purchase and organize production after receiving customer orders. Under ATO, $c_1(r) = 2r^2$ and $c_2(r) = 3 - 2.4r$.

$$Z(r) = 1000(2r^2) + 1000(3 - 2.4r^2)$$
$$+1.8r - \frac{0.6r[0.9r - (0.9r)^4]}{1-0.9r} + \frac{4.5(1-r)}{1-0.9(1-r)}$$

The simulation result is shown in fig 3.

In figure 3, the cost curve is concave in CODP location, and the optimal CODP location with minimum cost is r=0.6 . For the CODP is at the downstream of production process, so the mass production will take more ratio before CODP, and the sensitive analysis of $c_1(r)$ and $c_2(r)$ is shown in table 2.

From table 2, when $r \le 0.6$, c_1 will increase with CODP moving to the end process, but c_2 will decrease with CODP moving to the end process. For the mass production process is longer than the customization process, there exits some scale economics, so $M(r)$ is decreasing, and $Z(r)$ is decreasing. So for the electronic products, the optimal CODP location is r=0.6 .

5.3 The optimal CODP location in MTO

When MTO is adopted, customization level is relatively high, the cost is relatively low. Enterprises store a lot of raw materials and parts in order to obtain scale advantages, and the rest of the activities will be delayed until the customer orders. Amaretto is the cream liqueur produced in Italy, it can take different customer strategies according to different market demand. In the face of international market, the enterprise delay packaging and labeling, but in the local market, timely delivery does not allow the enterprise postpone any link. Procter & Gamble has adopted "Network order management" online system to complete the order request of the end retail link, whenever and wherever retail customers can be directly connected to the P & G. Under ATO, $c_1(r) = 2r^2$ and $c_2(r) = 3 - 1.2r$.

$$Z(r) = 1000(2r^2) + 1000(3 - 1.2r) +$$
$$1.8r - \frac{0.6r[0.9r - (0.9r)^4]}{1-0.9r} + \frac{4.5(1-r)}{1-0.9(1-r)}$$

The simulation result is shown in fig 4.

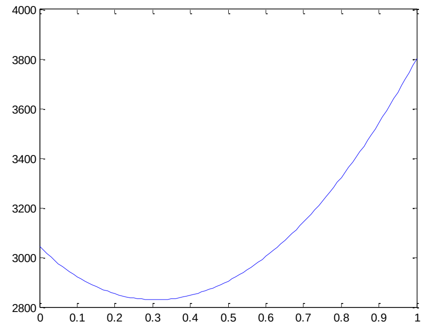


Fig. 4 The trend of production cost changing under MTO

In figure 4, the cost curve is concave in CODP location, and the optimal CODP location with minimum cost is 0.3. For the CODP is at the upstream of production process, so the mass production will take less ratio of the total production process, and the sensitive analysis of $c_1(r)$ and $c_2(r)$ is shown in table 3.

From table 3, when $r \le 0.3$, c_1 will increase with CODP moving to the end process, but c_2 will decrease. For the mass production process is shorter than the customization process, there exits few scale economics, so $M(r)$ is decreasing, and $Z(r)$ is decreasing. So for Furniture products, the optimal CODP location is r=0.3 .

5.4 The optimal CODP location in ETO

When ETO is adopted, customization level and costs are very high. The components have standard interfaces, so enterprise would assemble timely according to customer needs, and they have their own suppliers and distributors, and take flexible supply chain structures based on the expectations of the customer strategy. Toyota's production accounts for a substantial advantage in the automotive market driven by customer orders, and Trainer aircraft manufacturer Macchi aircraft can produce a variety of aircraft for special training according to the special needs. Under ETO, $c_1(r) = 2r^2$ and $c_2(r) = 3 + 0.5r^2$.

$$Z(r) = 1000(2r^2) + 1000(3 + 0.5r)$$

$$+1.8r - \frac{0.6r[0.9r - (0.9r)^4]}{1 - 0.9r} + \frac{4.5(1-r)}{1 - 0.9(1-r)}$$

The simulation result is shown in fig 5.

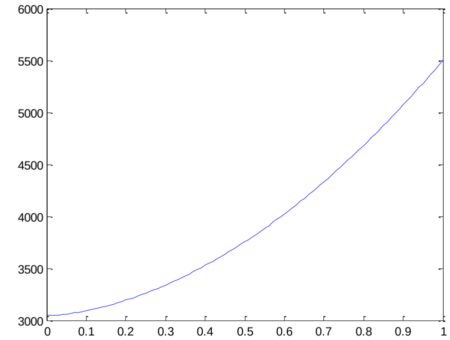


Fig. 5 The trend of production cost changing under ETO

In figure 5, the cost curve is increasing function of CODP location, and the optimal CODP location with minimum cost is 0. For the CODP is at the threshold of production process, so the whole production process is driven by customer order, and the sensitive analysis of $c_1(r)$ and $c_2(r)$ is shown in table 4.

For the mass production process is shorter than the customization process, there exits few scale economics, so $M(r)$ is decreasing, and $Z(r)$ is decreasing. For more technical, more complex structure of the product, the optimal CODP location is r=0 .

From table 3, c_2 will decrease with CODP moving to the end process. For the whole production process is, there is no scale economy, so $M(r)$ is increasing, and $Z(r)$ is increasing in the whole production process.

6. Conclusion

This paper studied the cost optimization problem under considering manufacturing cost, inventory cost of semi-finished product and customer waiting cost, and analyzed the optimal CODP under four production modes, i.e. MTS, ATO, MTO and ETO. The Matlab was applied to analyze the cost optimization under each production mode. When the products are a large class of low complex standardized products, such as Swiss coffee and the largest food and beverage Wahaha, the MTS will be chosen and the optimal CODP is r=1 . When the products are electronic products of high customization level and low customization cost, e.g. Dell and GOME, the ATO will be chosen and the optimal CODP location is r=0.6 .When the products are Furniture products, the optimal CODP location is r=0.3 and the MTO will be applied. When the products are more technical, more complex structure of the product, e.g. aircraft and machine, ETO will be applied and the optimal CODP location is r=0 .So manufacture can choose the right production mode and make right decision on the optimal CODP according to its own characteristics and market demand characteristics. Besides, the shortness in this paper is that the basic cost model and its M/M/1extended model did not consider the other cost factors, such as holding cost of processing product, investment cost caused by redesign for the postponed production, and so on, which can be future research direction.

Table 1: The influence of unit manufacturing cost on the optimal CODP location under MTS

r	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
c_1	$0.4r^2$	$0.6r^2$	$0.8r^2$	r^2	$1.2r^2$	$1.4r^2$	$1.6r^2$	$1.7r^2$	$1.8r^2$	$1.9r^2$	$2r^2$
c_2	$3+0.3r$	3	$3-0.3r$	$3-0.6r$	$3-r$	$3-1.4r$	$3-2r$	$3-2.5r$	$3-3r$	$3-3.5r$	$3-4r$
M	3000.0	3006.0	2972.0	2910.0	2792.0	2650.0	2376.0	2083.0	1752.0	1389.0	1000.0
Z	3045.0	3027.5	2985.2	2919.0	2798.5	2654.8	2379.5	2085.6	1753.8	1390.0	1000.3

Table 2 The influence of unit manufacturing cost on the optimal CODP location under ATO

r	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
c_1	$0.4r^2$	$0.6r^2$	$0.8r^2$	$1.1\,r^2$	$1.4r^2$	$1.7r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$
c_2	$3+0.3r$	3	$3-0.3r$	$3-0.6r$	$3-1.1r$	$3-1.7r$	$3-2.4r$	$3-2.8r$	$3-3.2r$	$3-3.6r$	$3-4r$
M	3000.0	3006.0	2972.0	2919.0	2784.0	2575.0	2280.0	2020.0	1720.0	1380.0	1000.0
Z	3045.0	3027.5	2985.2	2928.0	2790.5	2579.8	2283.5	2022.6	1721.8	1381.0	1000.3

Table 3 The influence of unit manufacturing cost on the optimal CODP location under MTO

r	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
c_1	r^2	$1.5r^2$	$1.8r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$
c_2	$3+0.3r$	$3-0.2r$	$3-0.7r$	$3-1.2r$	$3-1.6r$	$3-2r$	$3-2.4r$	$3-2.8r$	$3-3.2r$	$3-3.6r$	$3-4r$
M	3000.0	2995.0	2932.0	2820.0	2680.0	2500.0	2280.0	2020.0	1720.0	1380.0	1000.0
Z	3045.0	3016.5	2945.2	2829.0	2686.5	2504.8	2283.5	2022.6	1721.8	1381.0	1000.3

Table 4 The influence of unit manufacturing cost on the optimal CODP location under ETO

r	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
c_1	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$	$2r^2$
c_2	$3+0.5r$	$3-0.3r$	$3-0.7r$	$3-1.1r$	$3-1.6r$	$3-2r$	$3-2.4r$	$3-2.8r$	$3-3.2r$	$3-3.6r$	$3-4r$
M	3000.0	2995.0	2932.0	2820.0	2680.0	2500.0	2280.0	2020.0	1720.0	1380.0	1000.0
Z	3045.0	3016.5	2945.2	2829.0	2686.5	2504.8	2283.5	2022.6	1721.8	1381.0	1000.3

References

[1] Y. Y. Huang, S. L. Li, "Suitable application situations of different postponements approaches: Standardization vs. modularization", Journal of Manufacturing Systems, Vol.27, No.2, 2008, pp.111-122.

[2] B. Alessandro, Marta Z, "Evaluation of product customization strategies through modularization and postponement", International Journal of Production Economics, Vol. 120, No. 3, 2009, pp. 205-220.

[3] E. Ricardo, K. Bardia, "Evaluation of Supply Chain Structures through Modularization and Postponement", European Journal of Operational Research, Vol.124, No. 3, 2001, pp. 495-510.

[4] H. Lee, "Effective inventory and service management through product and process redesign", Operations Research, Vol.44, No.1, 1996, pp. 151-158.

[5] H. Lee, C. Tang, "Modeling the costs and benefits of delayed product differentiation", Management Science, Vol. 43, No.1, 1997, pp. 40-53.

[6] Z. L. Yang, Y. M. Tian, "A study of partial backorder policy under random demand", Journal of Systems & Management, Vol. 18, No.6, 2009, pp. 686-691.

[7] X. F. Shao, J. H. Ji, "Research on impact of postponement strategies on inventory cost in mass customization", System Engineering Theory Methodology & Application, Vol. 13, No. 5, 2010, pp. 385-389.

[8] J. Qing, Z. Zhong. "A visualized e-kanban system for small batch JIT production", Industrial Engineering Journal, vol. 14, No. 3, 2011, pp. 137-145.

[9] S. W. Sun, "The discussion about push-pull supply chain based on delay tactics", Storage Transportation & Preservation of Commodities, vol. 32, no. 9, 2010, pp.103-105.

[10] Z. S. Hua. "Flexible manufacturing systems and flexible supply chain: Modeling optimization". Beijing, Science Press, 2007.

[11] Van Hoek R I. "Logistics and virtual integration: postponement, outsourcing and the f1ow of information", International Journa1 of Physical Distribution & Logistics Management, vol. 28, no. 7, l998, pp.508-523.

[12] Krajewski L, Wei J C, Tang L L. "Responding to schedule changes in build-to-order supply chains", Journal of Operations Management, vol. 23, no.5, 2005, pp.452-469.

[13] R. Kai, D. Bin, L. Yu. "Research on cost optimization model of manufacturer implementation postponement strategy in mass customization", Computer Integrated Manufacturing Systems, vol. 13, no. 9, 2007, pp. 1665-1671.

[14] R. Kai, D. Bin, L. Yu. "Research on cost optimization model of manufacturer implementation postponement strategy in mass customization", Computer Integrated Manufacturing Systems, vol. 13, no. 9, 2007, pp. 1665-1671.

[15] D. Bin, R. Kai, Li H. "Cost optimization model of supply chain implementation postponement strategy in mass customization", Computer Integrated Manufacturing Systems, vol. 15, no.2, 2009, pp. 287-291.

[16] D Li., J. X. Wang. "Customer order decoupling point position for mass customization", Journal of Qingdao University (Natural Science Edition), vol. 23, no.2, 2010, pp. 89-92.

[17] Garg A, Tang C. "On postponement strategies for product families with multiple points of differentiation", IIE Transactions, vol. 29, no.8, 1997, pp. 641-650

[18] S. Ma, W. Wang, L. Liu "Commonality and postponement in multistage assembly systems", European Journal of Operational Research, vol.142, no.3, 2002, pp. 523-528.

[19] Tibben-Lembke R S, Bassok Y. "An inventory model for delayed customization: a hybrid approach", European Journal of Operational Research, vol. 165, no.3, 2005, pp. 748-764.

[20] C.P. Su, Y. Chang, M. Ferguson. "Evaluation of postponement structure to accommodate mass customization", Journal of Operations Management, vol. 23, no.3, 2005, pp. 305-318.

[21] S. L. Yang, S. F. Zhou T. F., Li. "Research on the Estimation of Mass Customization Production Cost Based on CODP", Forecasting, vol.13, no.6, 2010, pp.62-66.

[22] S. L. Yang, S. F. Zhou T. F., Li. "Research on the Effects of Multi-CODP on Production Cost of Processing and Assembling Products-Perspective of Assemble to Order", Operations Research and Management Science,vol.18, no.3, 2010, pp. 144-150.

[23] Y. H.Qin, Y. F. Geng. "Matching Model Analysis of Supply Chain Structures and Production Mode Under Mass Customization", International Journal of Electronics Communication and Computer Engineering, vol. 3, no.3, 2012, pp. 455-459.

[24] Mohamed Ramzi Bouzid, Naoufel Kraiem, "An Empirical Evaluation of a Reuse Based Approach for ERP Customization", International Journal of Computer Science Issues, Vol. 9, No. 3, 2012, pp. 17-26.

[25] Kompan Chomsamutr, Somkiat Jongprasithporn, "Optimization Parameters of tool life Model Using the Taguchi Approach and Response Surface Methodology", International Journal of Computer Science Issues, Vol. 9, No. 1, 2012, pp. 120-125.

Yanhong Qin Chongqing Province, China. Work as professor in Management School of Chongqing Jiaotong University, and research interests on mass customization and supply chain management.

Yuanfang Geng Jiangxi Province, China. Graduate student in Management School of Chongqing Jiaotong University, and research interests on mass customization and postponement strategy.

The Design of Statistic System in the Country Enterprise based on SAS and GIS

Li Xiaoling[1,2], Yuan Jimin[3,4]

[1] **School of Computer Science and Technology, Chengdu University**
Chengdu, 610106, China

[2] **University Key Laboratory of Pattern Recognition and Intelligent information Processing**
Sichuan Province, 610106,China

[3] **School of Computer Science and Technology, Panzhihua University, ,**
Panzhihua, 617000, China

[4] **University Key Laboratory of Electrical and electronic technology**
Sichuan Province,610106, China

Abstract

Statistics plays an important role in social and economic development. Traditional data collection is too laborious, time-consuming and ineffective. Its statistic is not correct and can't share the data. Aiming at the requirement for regional economy statistics of Pan Zhihua, this paper designed an analytical system based on regional economy statistic and achieved the analytical function of regional economic statistics. Spatial analysis method was used to integrate the special visual effect of GIS and geographic analysis function into the general statistical database, and the distributed GIS was utilized to release the data, provide browsing, searching and analytical function of the space data for the users. This analytical system accomplishes data share through internet and has the advantages of intuition, convenience and lower management cost. Most of all, it reduces the loss of the tax source after putting into operation and supplies the government's decision with the reasonable evidence.

Keywords: SAS, GIS, Statistics, Enterprise,Database.

1. Introduction

This document is set in 10-point Times New Roman. If absolutely necessary, we suggest the use of condensed line spacing rather than smaller point sizes. Some technical formatting software print mathematical formulas in italic type, with subscripts and superscripts in a slightly smaller font size. This is acceptable. At the moment, governments at all levels strengthen the statistic of regional economy data. For example, the statistic of regional enterprises and their production can analyze the energy consumption and unit energy of the enterprises. The GDP statistic of the enterprises is the proof of paying taxes. In the face of various statistic data, data collection has a tremendous workload, a bad function for the common use, the lower utilization ratio and inconspicuous decision of the government management. Traditional statistic method hardly gets timely, accurate and reliable data.

The SAS System provides a powerful programming language and components called procedures that allow us to perform many different kinds of analysis and data management functions, as well as produce many different types of text-based and graphical presentation output. Combined with other features of the SAS System, the SAS language and its procedures make possible an unlimited variety of applications from general-purpose data processing to highly specialized analyses in diverse application areas[1,2]。SAS/GIS software provides an interactive geographic information system within the SAS System. A geographic information system (GIS) is a tool for organizing and analyzing data that can be referenced spatially; that is, it is data that can be tied to physical locations. Many types of data have a spatial aspect, including demographics, marketing surveys, customer addresses, and epidemiological studies[3-5]. A GIS helps to analyze the data in the context of their location. For example, if we need to evaluate population data for census tracts, we could view the information in tabular format. However, consider how much easier and more effective it would be to view the demographic information in the context of the geography of the tracts. When viewing information that has a spatial component, we may find it easier to recognize relationships and trends in these data if we view the information in a spatial context. As the further application of GIS and the fast development of object-oriented programming and the component

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

technology[4-6], the synthesis method of spatial analysis in many economic activities has no difficulty to realize the technology. The spatial management function easily achieves the analysis and application of the regional economy. Therefore, based on the background of the whole country's enhancing the statistic, the paper uses secondary development module and visual programming language to develop the analytical system of regional economy, realize common data management, operation, graphic processing functions etc., strengthen statistic analysis and optimize the functions of spatial and location analysis.

2. System Developing Platform and Structure

According to the thought of system expansibility and the need of regional economy SAS system, this system uses the systematical structure of C/S. Database uses relation database, SQLServer2000 and stores spatial data, business data, metadata etc. Logic application layer consists of bottom data layer, middle Public application components and upper application components. Bottom database engine of spatial data (Supermap SDX) and data source management database engine components (ADO.NET) compose the bottom data layer. The upper application components include thematic application components and GIS application components (Supermap components). Components accomplish the mutual use through COM interface. Implementation level achieves the user's dynamic interactions and various functions of the system through user interface（GUI）. Detailed structure is illustrated in Chart1.

GIS exploring platform selects component object platform, (Component GIS) SuperMap Objects. SuperMap Objects is the geographic information system combing GIS with component technology. It provides GIS functions from symbol customization to visual map, from map collection to building spatial data and from two to third dimension so as to fully use the reusability of the component to raise the efficiency and the quality[7-9].

3. Basic functions of system

Regional economy SAS system mainly uses GIS to statistically analyze regional economy data and its result is visualized on the chart. It is convenient for users to explore the spatial connection of regional economy development effectively. Meanwhile, it presents different types of statistical charts. The system is able to supply the functions of attribute data entry and editing, management of database, spatial statistic analysis, spatial query analysis, information report and cartography export. The functions of the system are illustrated in Fig.2.

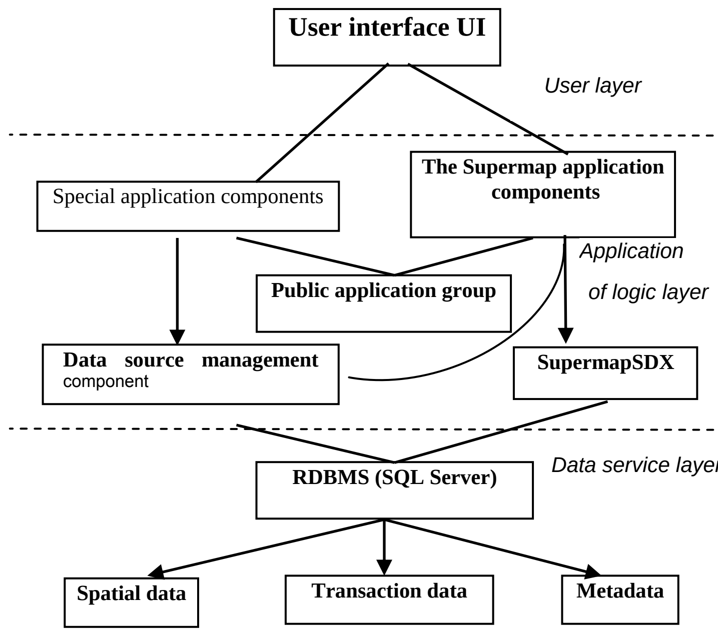


Fig.1 The system architecture based on GIS

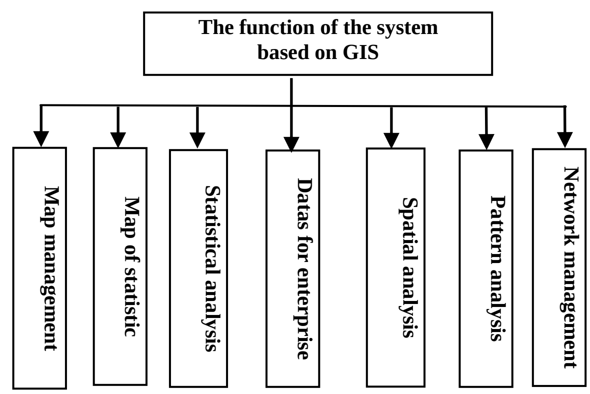


Fig. 2 The function of SAS based on GIS.

4. Database design

They SAS/GIS software uses two basic types of data which are spatial data and attribute data. The spatial data describes the location, shape, and interrelationships of map features. The attribute data provide information that relates to the map features. The spatial data represent point features, line features and an area features of map. To represent point, line, and area features in the map, SAS/GIS software defines the following topological features in the spatial data. The spatial data coordinate space can be represented in any numeric units even those that include arbitrary values. Coordinates that are stored as longitude and latitude values have a maximum usable precision of about one centimeter. Representations of map features are implemented with one or more chains, attribute data are all other data that are related to map features in some way, including the data to analyze in the context of the map. Attribute data can be stored in the spatial database.

System data includes various statistical data tables, such as administrative region table, industrial and mining enterprises table, investment table, industry information table, tax table etc.

4.1 Administrative Region Table

The administrative region table mainly introduces its geographic location, general survey of the cities and so on.

4.2 Industry Information Table

Industry information table is made up of industry gross product, scale structure, main economic benefit indices. Main economic benefit indices enable to define the fields of regional codes, industry value-added rate, total asset contribution rate, asset-liability rate, the turnover rate of circulating assets, industry cost profit rate, product sale ratio and so on.

4.3Fixed Asset Investment Information Table

Fixed Asset Investment Information Table consists of the following tables: regional codes, fixed asset investment of the whole country, basic construction investment etc. Fixed asset investment of the whole country defines the fields of national economy, collective economy, individual economy, jointly operated economy etc.

5. Main Function Module Design of System

The system provides the modules of management of regional economy data, spatial query, calculation of statistic analysis, graphic analysis chart, space etc.

5.1 Data management of regional enterprise economy statistic

Regional enterprise economy statistic data includes categories of enterprises, products, energy consuming, sales etc. Because of a huge mass of statistic data, regional enterprise economy statistic data and spatial data are respectively input and stored. At first, input economy statistic data to a sequential document, edit and store in the relevant file of database.

5.2 Modules of Spatial query and statistic analysis

Spatial query includes buffer query, crossing query, the interaction query of chart data and attribute data etc. The system supports the statistical function and finishes the statistical function of regional economy data. For instance, those statistic information, such as total number, mean, standard deviationetc. At the same time, the system carries out the functions of spatial statistic analysis and reflects the spatial connection of economic development. These statistical data showing on the geographic base helps to make a correct decision about the development tendency of the whole region.

5.3 Geographic Analysis Module

According to different requirements, diverse charts express economy data of every administrative region. It usually adopts pie chart, point density diagram, histogram, line chart etc. Drawing the graphs of the same region at different periods visually reveals the law of regional economy development. Parts of codes are as following:

```
public SLThematicMap(Panel chartContainer,
ObservableCollection<ObservableCollection<object>>
table)
    { this.chartContainer = chartContainer;
      this.table = table;
    }  #endregion
    #region
    Dictionary<string, DataSeries> dicDataSeries = new
Dictionary<string, DataSeries>();
    DataSeries GetDataSeries(string sTitle)
    {   if (!dicDataSeries.ContainsKey(sTitle))
      {
        RenderAs renderAsEnum; // out parameters
        dicDataSeries.Add(sTitle, new DataSeries()
        {RenderAs =
Enum.TryParse<RenderAs>(ChartRenderAs, true, out
renderAsEnum) ? renderAsEnum : RenderAs.Column,
          XValueType = ChartValueTypes.Auto,
          Name = sTitle,
          LegendText = sTitle,
          //YValueFormatString = "#",
          ShowInLegend = ChartShowInLegend,
        });
        if (dicDataSeries[sTitle].RenderAs ==
RenderAs.Pie) // (pie chart tip
        { if (chartAxisYTitle.Contains("quantity "))
            dicDataSeries[sTitle].ToolTipText =
"#AxisXLabel -- #YValue 个(占#Percentage%)";

          else if (chartAxisYTitle.Contains("amount"))
```

```
            dicDataSeries[sTitle].ToolTipText =
"#AxisXLabel -- #YValue(#Percentage%)"; }
            else { }
        }
        return dicDataSeries[sTitle];
    }
```

Parts of query service codes are as following:

```
 void multipleQuery_Click(object sender,
RoutedEventArgs e)
    {
        MLKQueryRoot mqr = new MLKQueryRoot();
        mqr.multiQueryAction = true;
        mqr.AdvancedQuery();
    }
public SpatialResult(Dictionary<string, string>
pDictionary, double x, double y, double bufferRadius,
Graphic queryGraphic)
        : this()
    {
        _QueryGraphic = queryGraphic;
        loadData(pDictionary, x, y, bufferRadius);
    }

    /// <param name="pDictionary"
    public SpatialResult(Dictionary<string, string>
pDictionary)
        : this()
    {
        _pDictionary = pDictionary;
        loadData(_pDictionary);
    }
public SpatialResult(Dictionary<string, string>
pDictionary, Envelope rectangle, Graphic queryGraphic)
        : this()
    {
        _QueryGraphic = queryGraphic;
        loadData(pDictionary, rectangle);
    }
```

6. Conclusions

The operation result of the statistic is illustrated inFig.3 and Fig.4. The operation result of management information in the basic units is shown in Fig.5. At present, this system has Applied Statistics for Management and got a good result.

Fig.3 the statistics result

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Fig.4 the statistics result

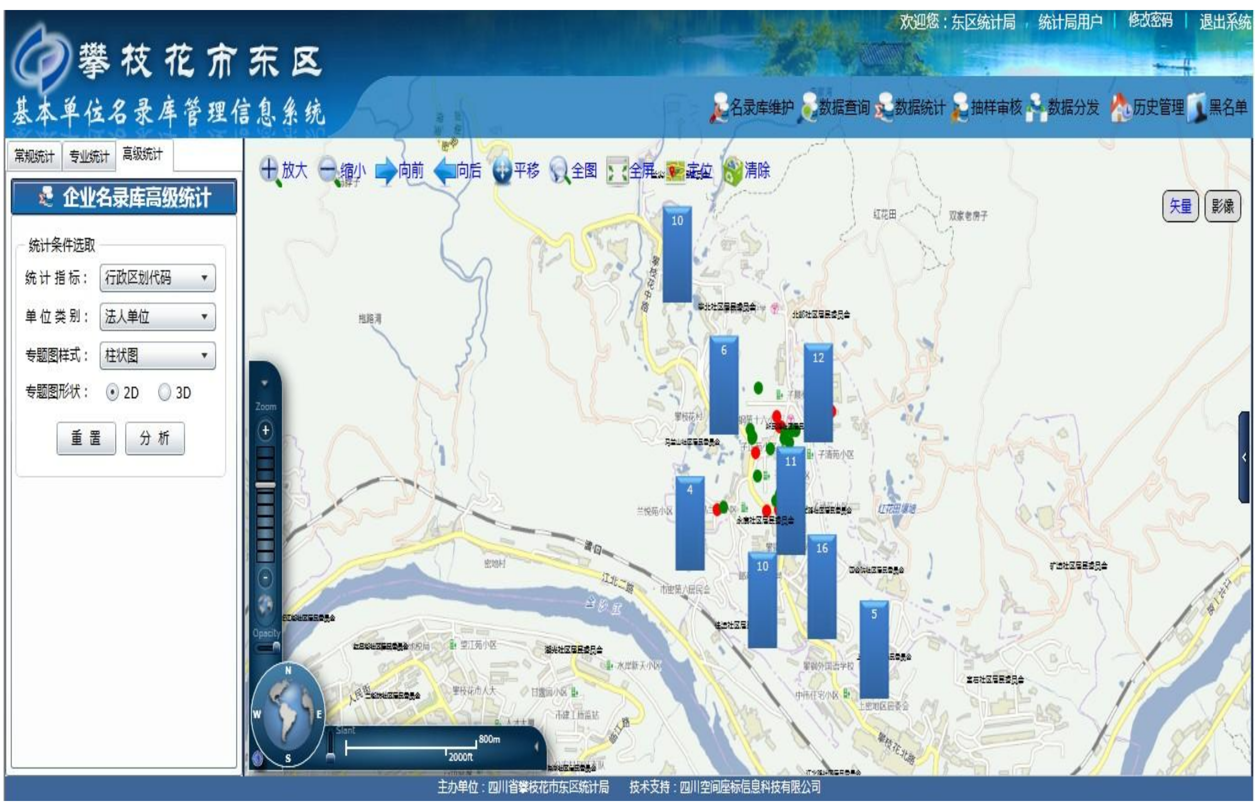


Fig.5 the statistics result based on GIS

References

[1] WANG Jia-yao, MIAO Guo-giang, and CHENG Yi, "The Data acquisition of Spatial Information System", HYDROGRAPHIC SURVEYING AND CHARTING, Vol.24, No.2, 2004, pp.1-4.

[2] M. F. Goodchild. "Geographic Information Systems and Disaggregate Transportation Modeling", Geographical Systems, Vol.(5),No. 1, 1998, pp. 9–17.

[3] LAN heng xing,WU Faquan, and ZHOU Chenghu, etc, "Spatial hazard analysis and prediction on rainfall-induced landslide using GIS", Chinese science bulletin, Vol.48,No.7,2003,pp.703-708.

[4] Jones, C.B. A. I. Abdelmoty and G. Fu, "Maintaining onto logies for geographical information retrieval on the web On The Move to Meaningful Internet Systems", Lecture Notes in Computer Science, Vol. 288,No. 8, 2003,pp.934-951.

[5] B. Zheng, W. C. Lee, and D. L. Lee, "Search K Nearest Neighbors on Air", InProc. MDM, 2003, pp. 181 –195.

[6] M. Essid, O. Boucelma, F-M. Colonna and Y. Lassoued, "Query processing in a Geographic 3 Mediation System", In Proceedings of the 12th ACM International Workshop on Geographic Information System, ACM-GIS, Washington, 2004.

[7] Jianbo Sui, Xiao Wang, and Xiuwan Chen, "Design and Application of the Flood Loss Evaluation System Based on GIS and RS". 2005 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2005), vol.2,2005.

[8]XIAO Gen-ru, CHENG Peng-gen, and PAN Hai-yan, "The design and implementation of regional economic statistics analysis system based on GIS", Science of surveying and mapping,Vol.33, No.1,2008, pp.182-184.

[9] Saira Saleem Pathan, Omer Rashid, Ayoub Al-Hamadi and Bernd Michaelis, "Multi-Object Tracking in Dynamic Scenes By Integrating Statistical and Cognitive Approaches", International Journal of Computer Science Issues, Vol. 9, No. 4, 2012, pp. 180-189.

Li Xiaoling Received M.S. degree from University electronic science and technology for circuit and system in 2002, and became a professor in electronic technology and computer science in 2009.She is teaching in Chengdu University and University Key Laboratory of Pattern Recognition and Intelligent information Processing, Sichuan Province. Her main research direction is image & video feature extraction, signal detection and processing.

Yuan Jimin Received PhD from University electronic science and technology for Measuring and Testing Technologies and Instruments in 2009, and became a professor in electronic technology and computer science in 2008. He is teaching in Pan Zhihua University. His main research direction is the computer application and neural network.

Embedded Wireless Fingerprint Exam Affair Management

Yuan Yuan[1], Peng Min[2]

**[1] Computer and Engineering Department, Sichuan College of Architectural Technology
Deyang,618000,China**

**[2] Sichuan College of Architectural Technology
Deyang, 618000,China**

Abstract

This thesis studies the design method of fingerprint identification module based on embedded system, which can realize the design of wireless fingerprint examination management system.This method reduces the hardware cost by building hardware environment whose platform is ARM9 microprocessor and MBF200 fingerprint sensor-integrated module. And, under the enviroment of Linux, adopting framework for embedded fingerprint identification system and introducing relevant technology about fingerprint identification, the design of distributed fingerprint identification system could be realized, which is low-cost and reliable. Meanwhile, by combining wireless network communication with Ethernet communication, it can maintain real time data exchanege with central data bank. This system adopts SQL Server database to manage usres' baisc information and information data of their fingerprints' characteristics and runs reliably. Comprehensive tests and practical application proved that fingerprint images collected by this system are more clear and less fuzzy. After being processed, the outlines of images are clear and have outstanding features, so that it is easy to classify and identify fingerprints. This system could be applied in the management of large-scale examinations and remote network examination. It also improves traditional checking way which identifies person by his holdings, and is the future direction of examination management.

Keywords: *exam affair management; fingerprint identification system; embedded system; wireless communication; Hardware Design.*

1. Forward

With the high development of computer and Internet, there is a higher demand of the veracity, safety and practicability of the human identification. Intelligence authentication technology based on biometrics. In many biometric identification technology, fingerprint is one of the earliest and most widely used[1]. As fingerprint has the characteristics of universality, uniqueness, and fixity, fingerprint identification technology has gradually replaced traditional identification methods based on marks and numbers, and now is widely applied in many business, such as Internet, banks, finance, medical treatment and security check. So we are considering to apply fingerprints identification into current exam system.Currently, in examinations we primarily use admission ticket, student's identity card or ID card to inspect examinees' identity. However, misconducts, such as forging admission tickets or ID cards to hire someone to take exams for her or him, happen all the time. In order to eliminate such cheating activities, we need to approve the real identity of examinees, which is an urgent problem that needs to be solved in exam management[2]. Because human bodies cannot be copied, people resort to biological identification technology, and hope we could use this technology to deal with challenges current exam system are facing. And fingerprints are the most obvious external characteristic and they are universal, unique and easy to get.

2. Principle of Fingerprint Identification System

Major headings are to be column centered in a bold font without underline. They need be numbered. "2. Headings and Footnotes" at the top of this paragraph is a major heading. The center of this system is fingerprint identification system. Fingerprint identification technology is a kind of biometrics that reads fingerprint images from image acquisition equipment, extracts feature data from fingerprint images obtained by identification software, and then identifies all people's identity according to the results got from matching algorithms[3]. Fingerprint identification system mostly concerns three steps: preprocess, minutiae extraction, and feature match.

The principle of fingerprint identification system is as Chart 1:

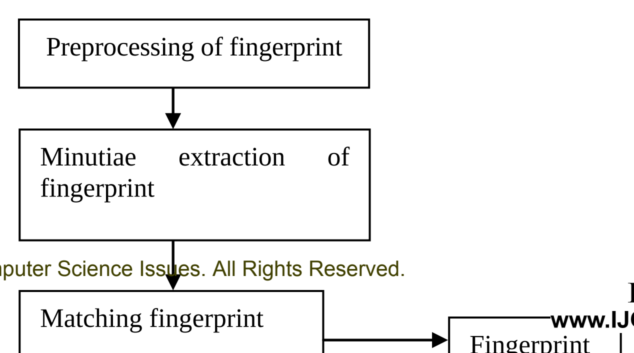


Fig. 1 Fingerprint identification system

Actually the process of fingerprint identification is rather complicated. The whole process includes acquisition of fingerprint image, preprocessing of fingerprint image, calculating fingerprint direction and filtering, binary image, thinning fingerprint threads, picking-up characteristic, and fingerprint match[4]. Considering requirements in practical application, this process would also include processing and transporting data and controlling external equipments so that a series of jobs can be done, such as human-computer dialogue, reading data in fingerprint database, and outputting matching results.

2.1 Preprocessing of fingerprint image

We need to enhance the images of fingerprint that we read from fingerprint read equipment. The key problem we need to deal with is preprocessing fingerprint image whose purpose is to gain images that fit characteristic from inputted gray image. This is the first step of fingerprint identification processing. Generally, those images gained from sensor would contain noises and distortions in fingerprint images because of various reasons[5]. So it is an important content in fingerprint identification that we should eliminate those noises and distortions by preprocessing and turn them into some standard forms, which would make the characteristic acquisition and identification more easier.

The main steps of preprocessing images including: gray-value normalization, image segmentation, filter, image enhancement, binaryzation, thinning etc. The purpose of preprocessing is to improve the quality of input fingerprint image so as to increase accuracy of minutiae extraction[6]. Original fingerprint images usually would have problems, such as noise pollution, disconnection or ambiguity of ridge line, so we need to enhance images (by using filter technology) to improve quality. Because the characteristics of fingerprint only consists in the form of ridge line, and ridge lines have different darkness and width, we could turn them into thin ridge lines that have the same gray-value and are one pixel width by using normalization, binaryzation, and thinning, thus we can perform minutiae extraction.

2.2 Minutiae extraction of fingerprint image

The most frequently used way of minutiae extraction is feature extraction, because it is the most widely applied matching method which is based on position, type and direction of feature characteristics. Minutiae extraction which we use in reality is also based on feature characteristics. Fingerprint minutiae include endpoint, bifurcation, center-point, triangulation point etc. The endpoint and bifurcation are the most common features in fingerprint and they are relatively stable, so in this essay we also adopt the two features[7]. Fingerprint is composed by the crisscross of ridges and valleys. After fingerprint images' binaryzation, we can see this structure clearly. After analysis, we can find out that endpoints of fingerprint ridges are corresponding with same bifurcations of fingerprint valleys and bifurcations of fingerprint ridges are corresponding with same endpoints of fingerprint valleys[8]. Thus we only need to extract endpoints of two thinning images to extract all endpoints and bifurcations of fingerprint.

2.3 Minutiae extraction of fingerprint image

Feature matching of fingerprint is the last step of fingerprint identification system, and is also the most important foundation to evaluate the function of the whole system. Compare newly input minutiae eigenvalue with minutiae eigenvalue in fingerprint database, find the most similar fingerprint as the result and output it[9]. This is the process of fingerprint validating identification and it is also the ultimate propose of fingerprint identification system. Because of the influence of various factors, it is very likely that minutiae module might be different when the same fingerprint are put in. So, as long as the minutiae module of input fingerprint is similar with the one in database, we think they are a match[10]. Then comes to the problem of measures. Usually, we describe the matching result as match measure. When match measure is greater than a certain threshold, we think the two fingerprints match; otherwise, when match measure is lesser than this threshold, we think they don't match. The number of threshold is usually set based on some factors, such as experience, by man.

3. Hardware Design Of Exam Affair Mangement

This system could read human fingerprints' image from fingerprint reader equipment, initially process original images to make them clearer, and build fingerprints' characteristic data by fingerprint identification software. Then the software could find data point that we called nodes from fingerprints. These data are usually called templates[11]. With computers' vague comparison, we could compare the templates of two fingerprints, calculate their similarities, and finally get a match result of two fingerprints.

The design of hardware is centered with microprocessor UP-NETARM 2410-S that includes ARM core. The server hardware of fingerprint identification are made up with Ex-patulous memory Flash, SDRAM, and external expansion fingerprint sensor MBF200. And we could use wireless Internet technique and Ethernet to build system hardware.
The design of system hardware is as follows in Chart 2.

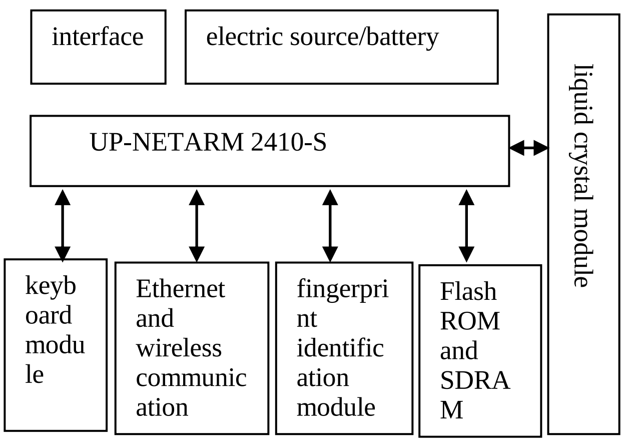


Fig. 2 System Hardware Structure

3.1 Fingerprint Identification Module

Ordinary fingerprint identification module is usually composed of fingerprint acquisition module, system center module, data storage module and output display module. The hardware circuit of this module is centered with ARM9 microprocessor, while peripheral circuit primarily includes fingerprint identification module, data storage module Flash, hardware calendar& clock equipment DS1320, power supply/circuit, reset circuit and clock circuit.
The structure of the system hardware is as Chart 3:

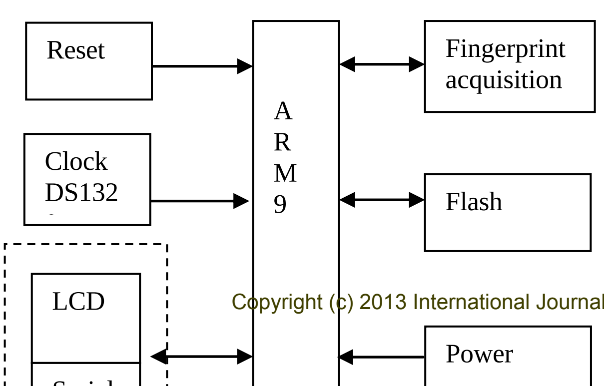


Fig. 3 Hardware Design Of Fingerprint Identification Module

ARM9E is a kind of integrated processor including micro-controller, DSP, Java function. It intensifies the ability of digital processing processing, and applies to conditions that should combine DSP and micro-controller[12]. Meanwhile, it extends Thumb signal and DSP into ARM instruction, and possesses EmbededICE-RT logic which fits the need of time system in a better way. ARM9 is a high-performance and low-power consumption microprocessor that is widely used in embedded system.
We take MBF200 fingerprint sensor chip, produced by Fujitsu Ltd, as the acquisition module to perform acquisition of fingerprint images. This fingerprint sensor adopts standard CMOS technology, contains eight A/D converter, works with wide range voltage input between 3.3V to 5V, can automatically check whether fingerprint has reached sensor, and realizes online acquisition.
Data storage module includes a SRAM of 512KB and a FLASH of 4MB. SRAM is used to store acquired data of fingerprint images and temporary data when program is running; the FLASH of 4MB is used to store system applications and minutiae database. The identification results would be output through LCD.

3.2 Analysis on Other Related Technology

3.2.1 Analysis on Ethernet and Wireless Communication

There is a key USB at the port of ARM development board, which could insert WLAN cards that follow 802.1l WLAN protocol. When relevant equipment Drives are loaded in Linux, the data transportation of wireless fidelity could be completed. After the WLAN card that connects to ARM development board registered to AP access point, the WLAN cards between AP and ARM port could form wireless fidelity. In order to obtain more resources, we can get access from wireless to wire LAN and Internet.

3.2.2 LCD Display Module

The LCD display screen in UP-NETARM 2410-S development platform could be compatible with many kinds of LCD[13]. We could use a five-inch 256 color screen or an eight-inch 16bit true color screen, and at the same time provide a 24bit port.

3.2.3 Keyboard Module

The keypad in UP-NETARM 2410-S development platform uses ATMEGA8 single chip to control two PS2 ports and keypad on board, and the two PS2 could connect to PC keyboard and mouse.

3.2.4 FLASH ROM and SDRAM

FLASH ROM ： the chip of FLASH ROM is CMOS FLASH of 16Mb, which is used to store applications, such as the kernel of Linux operating system, fingerprint acquisition & identification, and communication, and file system that is used to support all kinds of service[14].

SDRAM ： the existence of embedded operating system needs some dynamic RAM. Operating system, after being decompressed, would bootstrap from FLASH ROM to SDRAM[15], resident memory; simulate memory to hard disk space, so that we could save files on it like the way we did on hard disk; meanwhile, users' stack and service data are also put in SDROM.

4. Design of System Software

Normally, when the system is electrified, we should perform an initialization operation to the whole acquisition system, including the initialization operation of microprocessor and the peripheral on target board, so that we could perform configuration parameter[16]. After initialization operation, we could start to test if there is any finger on sensor. If there is, we begin to acquire fingerprint images. When build fingerprint database, acquired data would be stored in image SDRAM. System pulls out data from SDRAM. If the quality of acquired images is bad, new fingerprint data would be acquired. Because processing fingerprint data concerns much algorism and high computation, ARM9E system adds some enhanced processors to perform attached instructions of typical DSP algorism conducting ability. After a fingerprint image is required, inform DSP to perform data processing[17]. In this phase, data in image RAM would first be partitioned, and imaged would go through some operations, such as preprocessing, and minutiae extraction. Fingerprint data information which has been processed would be stored in fingerprint database constructed by FLASH team, and finally results would be output through USB[18]. When comparison is needed, fingerprint could be taken out from database, and the software can compare the extracted result and acquired fingerprint data to determine whether they are match and draw a conclusion.

Software of this system are consisted of two parts mainly: PC server software and clients software. The flow chart is under below:

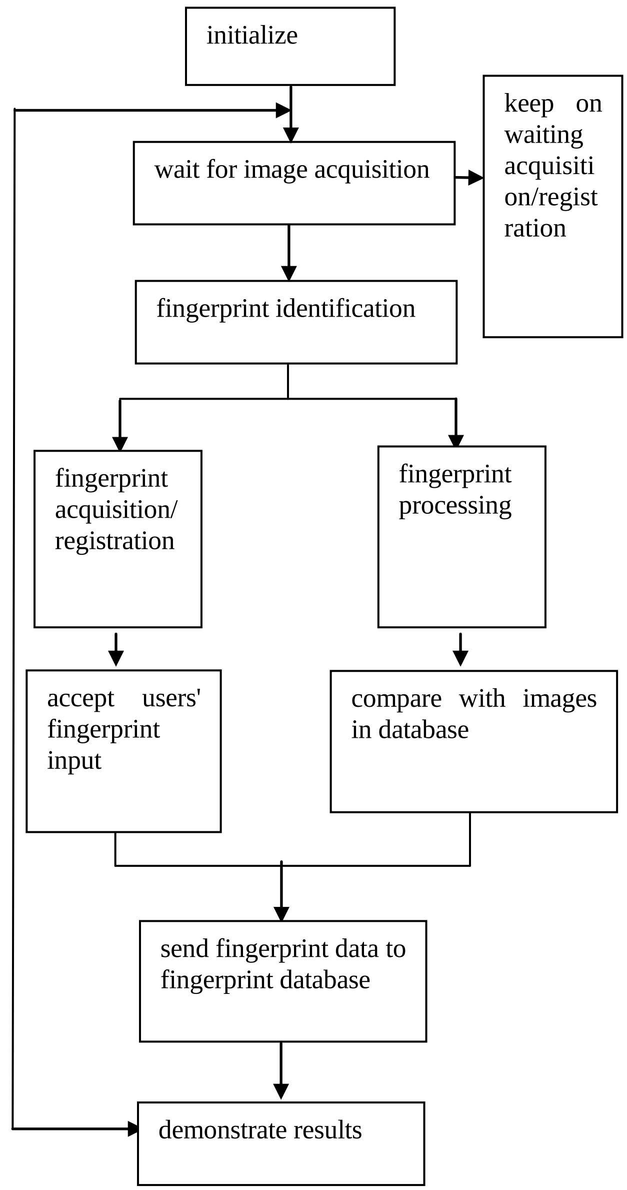


Fig. 4 Flow Chart of Software Module

Server runs on PC host, chooses Windows XP operating system, and adopts ACCESS database at server. When client sends fingerprint data, asking to register, fingerprint data from client would be saved in database. When client asks to identify fingerprint, compare fingerprints that need to be compared with fingerprints in database, and then send the result back to client.

Client runs on development board, adopts Linux operating system which has been clipped. And it keeps running since the system starts. First, client would connect to PC host through wireless or wire network. After this connection, wait for users to choose operation[19]. If users want register fingerprint, then ask them to input fingerprint. If users want to identify fingerprint, also ask them to input fingerprint; after input, data, if legitimate, would be transported to server to deal with through wireless or wire network. Client would only need to wait. Finally, this program would show the received results on the liquid crystal display of the development board.

5 Test and Running

We use two classrooms, one as exam-room and another one as control center. Including: fingerprint acquisition module, UP-NETARM2410-S development board, PC host (with Windows XP system and Access database), and WLAN card.

5.1 Testing Data and Analysis

Having finished the circuit of control board hardware ARM9, we embed Linux operating system, and then compile applications under operating system[20]. The main functions include: connect Internet, transport fingerprint information to Internet to realize network function, initialize and reset the equipment, register fingerprint information, manage fingerprint information database, match fingerprint information and other managing functions.

Start PC server and then client to see if two ports can be connected; operate program, transport data, and observe whether or not server and client can correctly send and receive data.

Follow testing procedure and data to run the test. The results are in Table 1:

Table 1:Testing Data and Analysis

testing fingerprint	*expected registration result*	*expected reaction time*	*expected checking time*	*expected reaction time*
	succeed	<1S	succeed	<1S
	succeed	<1S	succeed	<1S
	succeed	<1S	fail	<1S
	fail	<1S		

Testing results show that our equipment and program have basically achieved predesigned standards.

When acquired fingerprints are a bit obscure, normal registration could be performed. But these fingerprints could fail in checking because of deviation. If acquired fingerprints are really obscure and the program cannot even identify, registration would fain because system would deny registration. But these incidents rarely happen, only when fingers are covered with mud while fingerprints are being acquired.

6 Conclusions

To adapt the need that fingerprint acquisition sensor and fingerprint identification system are developing towards being small and embedded, this essay put forward a design plan about embedded fingerprint identification system based on ARM9 processor, and also introduces basic methods of acquisition, processing, extraction, and matching of fingerprint. At the same time, this essay illustrates the hardware structure and software flow of this fingerprint identification system[21]. Judging from final fingerprint identification results, we can tell this set of design plan is efficient and has reached our expectations.

Embedded wireless fingerprint exam affair management system is suitable for current exam information management system. When it's being used, users can independently complete the comparison of fingerprint and information inquiry without connecting to other equipments and computers[22]. Examinees only need to sweep fingerprint acquisition equipment with his or her fingers, then identities could be confirmed. They do not need to show a lot of credentials which is convenient for both invigilators and examinees. Real running results demonstrate this system of great practical value.

References

[1] Li Xueqiao, Design and Realization on One-finger Exam Affair Management Based on ARM; Microcomputer Applications Vol. 22, No. 9, 2006, pp.32-34

[2] Roli Bansal, Priti Sehgal and Punam Bedi, “Minutiae Extraction from Fingerprint Images - a Review”, International Journal of Computer Science Issues, Vol. 8, No. 5, 2012, pp. 74-85.

[3] Cuipeng,“The ARM9-based Design of the Fingerprint Identification System”, Electronics& Packaging,V01.10,No.7,2010,pp.28-30.

[4] Sun Tianze, Yuan Wenju, Zhang Haifeng, Embedd Design and Development Guide on Linux Drive_Based on ARM9 Processor, Beijing: Publishing House of electronics industry,2005.

[5] Wu kuanming,Bus Principle and application system design,Beijing: Beihang University press,2005

[6] MENDOZA P,VIA,J,”Developing CAN based network on RT-Linux”,IEEE, Vol. 2, No. 8, 2001, pp.161-167.

[7] Wang Limin, Chen Shuangjiao, Yan Xiaolin, etc, Development and Practice of ARM9 Embedded System, Beihang University Press, 2008.

[8] Ham L．DiotM multi,”signature with distinguished signing authoriti”,Electronics Lett, Vol. 35, No. 4, 1999, pp. 294-295.

[9] Zeng Jianyi, “Fingerprint Identificalion”, Techniques of automation and Applications, Vol. 26, No. 8, 2007, pp. 74-76.

[9] ke Yuanbin, The Study of Fingerprint Acquisition System Based on DSP,Wuhan University of Technology, 2006.

[10] Wang Tianmiao, Design and Example Development of Embedded System, Beijing: Tsinghua University press,2004.

[11] Xu Guohui, "Development and Progress of Fingerprint Verification Technology and Product”,Communications Technology, Vol. 42, No. 8, 2009, pp. 88-90.

[12] Liu Zhengrong, Zhang Zhichao and Xu Zhengshan,Explanation on Embedded Linux Application Development, Beijing:Machinery Industry Press,2004.

[13]Wang Kejun, “Research and Design of Automatic Fingerprint Identification System”, M.S.thesis, Harbin Engineering University, Harbin ,zhongguo,2003.

[14] Ren Zhiguo , “Design of fingerprint acquisition system based on S3C2410 and MBF200”, Electronic Design Engineering,Vol. 17, No. 10, 2009 , pp.36-38.

[15] Zhang Yujin, Image Engineering, beijing:Tsinghua University Press, 2000.

[16] Zeng Tong, “Design and Realization of Fingerprint Identification System Based on Embedded Processor”, Electronic Design Engineering,Vol. 18, No. 4, 2010 , pp.119-121.

[17] Ma zhongmei,Embedded processor architecture and Application, Beijing: Beihang University press,2002

[18] Li Xiaoling, Design and Realization on the Wireless Fingerprint Examination Management System Based on ARM; Journal of Sichuan Normal University(Natural Science), V01.32,No.4, 2009, pp. 541-544

[19] Zhou Ligong, Basci Course on ARM Embedded System, Beijing, Beihang University Press, 2005.

[20] Xu Jia, Zhang Minrong and Ma Yuebiao,” D esign and Research of the Fingerprint Identification System Based on ARM on ARM WinCE Platform”, Computer Applications and Software,Vol. 27, No. 6, 2010, pp. 251-253.

[21] You Linru,and Ding Fei,”Improved fingerprint feature extraction algorithm based on ARM”,Computer Applications,Vol. 27,No. 1,2007, pp.172-173.

[22] Josphineleela Ramakrishnan, Ramakrishnan malaisamy, “Performance Measurement and Method Analysis (PMMA) for Fingerprint Reconstruction”, International Journal of Computer Science Issues, Vol. 9, No. 3, pp. 59-62.

Yuanyuan: female, born in Chengdu, Sichuan province, lecturer, master, research direction is application of embedded technology.
Teaching reform of microcomputer theory and application, Northwestern Polytechnical University , Vol.32,No.9,2008,
Design and Realization of Fingerprint Identification System Based on Embedded Processor Icci,2012,
The Design and Realization of Dynamic Vehicle Scheduling Management System,Value engineering, Vol.12,No.9,2012,
The educational reform and research of the course of computer organization’s principle, Vol. 9, 2009,
Analysis on Overall Design of Campus Network。Journal of Nanjing Normal University,Vol.30,2007

Peng Min：
Teaching reform of given and drains off water of project，Shanxi architecture,2009.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Designing Ontology Schema and Data Instance for Nutrition Domain

Meena Unni[1] and Dr. K. Baskaran[2]

[1]**Research Scholar, Department of Computer Science, Karpagam University**
Coimbatore *641 021,* India

[2]**Associate Professor, Department of Computer Science, Government College of Technology**
Coimbatore *641 013,* India

Abstract

The web has become a main research area. There is a large amount of information available on the web that is to be discovered, shared and utilized. Most of the information available on the web is unstructured and hence cannot be understood by machines. Semantic web is a step forward in this direction where data can be understood by both humans and machines. By encouraging the inclusion of semantic content in web pages, the Semantic Web aims at converting the current web dominated by unstructured and semi-structured documents into a "web of data". Ontology plays a critical role for the Semantic Web, and it is necessary to understand ontology in order to fully appreciate the idea of the Semantic Web. Ontology is defined as a common set of terms that are used to describe and represent a domain. It defines the terms used to describe and represent an area of knowledge. This paper explains in detail the development of ontology in the nutrition domain using Noy and McGuiness method.

Keywords: *Ontology, Semantic Web, Taxonomy, OWL, Class Hierarchy*

1. Introduction

Ontology is a formal explicit specification of a shared conceptualization [1]. Formal implies the fact that the ontology should be machine understandable. Explicit implies that the concepts and their constraints are unambiguously defined in the ontology. Shared implies that ontology should capture agreed knowledge. Conceptualization implies creating abstract model in the domain by precisely identifying relevant concepts

Ontology is domain specific, and it is used to describe and represent an area of knowledge. A domain is simply a specific subject area or area of knowledge, such as the area of photography, medicine, real estate, and education. Ontology contains terms and the relationships among these terms. Terms are often called classes, or concepts, and these words are interchangeable. The relationships between these classes can be expressed by using a hierarchical structure: super classes represent higher level concepts and sub-classes represent finer concepts. The finer concepts have all the attributes and features that the higher concepts have. Ontology also represents properties which [2] describe various features and attributes of the concepts, and they can also be used to associate different classes together. Therefore, the relationships among classes are not only super class or sub-class relationships, but also relationships expressed in terms of properties.
Ontology comprises of ontology schema and instance. Ontology schema defines the terms and their relationships in the ontology whereas data instance focuses on augmenting the ontology with instances of concepts and properties. This part of the ontology can be kept in the knowledge base.

2. Designing Ontology

The methodology that is adopted in building the ontology is a modified version of the Agile Software Development proposed methodology called Methontology [3] which is a method to develop ontology from scratch.

Once all the information regarding vitamin A is collected, concepts of the ontology and their relationships were defined based on the collected information. Next step is to formalize the ontology by expressing concepts in a class subclass hierarchy. Then relationships between those classes and restrictions were added wherever applicable.

According to Noy and McGuinness [4], the following phases are to be followed to develop ontology.

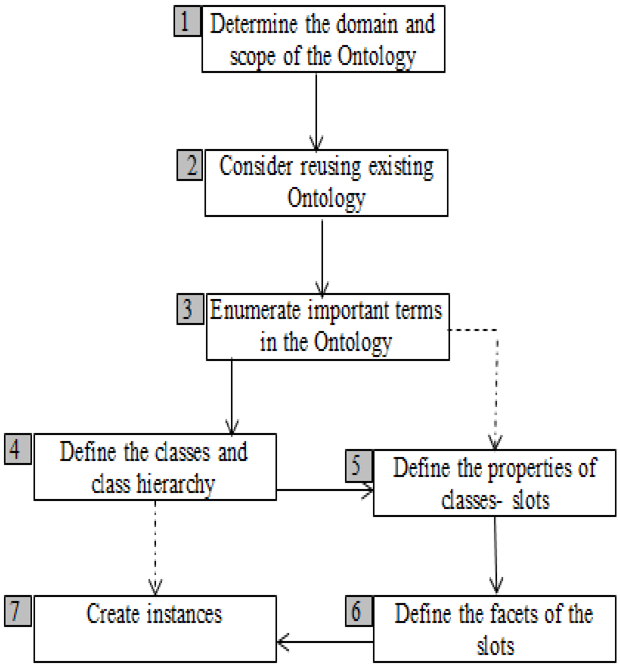


Fig. 1 Ontology development phases

2.1 Determine the domain and scope

Development of ontology starts with defining its scope and domain. There is need to study reasons for using the ontology. Considering the kinds of questions the ontology is supposed to answer is an important criteria while developing the ontology.

The ontology that is being developed for this work is that of Vitamin A. Representation of vitamin A for different dietary sources and its effect on human beings is the domain of the ontology. This ontology can be used to suggest diet with varying levels of vitamin A content. The concepts describing different types of diet both from plants and animals, effects of deficiency or excess intake of vitamin A on humans will be main focus of the ontology. Vitamin A requirement for different categories of humans, other diseases which may hinder the vitamin a effects too will figure into the ontology.

One of the ways to determine the scope of the ontology is to sketch a list of questions that a knowledge base based on the ontology should be able to answer, competency questions [5].

While the development of vitamin A ontology, the following competency questions have been considered.

What are the effects of deficiency of Vitamin A?

What is the recommended dietary allowance of vitamin A for different categories of people?

Which foods are rich in Vitamin A?

What are the different forms of Vitamin A?

How much vitamin A is there in each dietary source?

Does presence of other disease affect Vitamin A in the body?

Judging from this list of competency questions, the ontology will include information on various dietary sources, its effect on human, requirement amount of people, affect of diseases.

2.2 Consider reuse

There are ontologies available from third party that provides a starting point. If suitable ontology already exists, then it can be extended for the required domain. Research was carried out to find any existing ontologies on vitamin A, but could not locate any. Hence this ontology was developed from scratch.

2.3 Enumerate Key terms

It is useful to list all the terms that is likely to appear in the ontology. Classes, relationship between them and instances can be listed. For the chosen domain, important terms include Forms of Vitamin A, Effects on Human, Deficiency, Prevents, animal food, plant food, deficiency, overdose, person, Male, Female, Child, Interaction, Tolerable Upper limit, Causes, Enhances, May Result, Suffers from etc. Fundamentally, it is useful to obtain a comprehensive list of terms without worrying the overlap between the term and concept of the property, or concept includes slot or class.

Developing the class hierarchy and defining properties of concepts (slots) are closely interdependent. These two steps are also the most important steps in the ontology design process.

2.4 Define taxonomy

Once the key terms are identified, they have to be organised into a superclass-subclass hierarchy, which is also known as taxonomy. Subclasses specialise their super classes. There are several approaches in the development of class hierarchies [6].

Class hierarchy can be developed using top-down, bottom-up or combinational approach.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

- Top-down development starts with defining the most general concepts in the domain and then deriving more specific concepts under it. In order to create ontology on Vitamin A, general concepts like diet, person, and effects can be identified. Then subclasses can be created for these concepts. For example *DietarySource* can be further categorised as *AnimalSource* and *PlantSource*. This work uses top down developmental approach.

- Bottom-up development starts with defining the specific classes, the leaves of the hierarchy and then all these classes can be grouped to form more general concept. For example classes *Fruit, Vegetable, Lentil* can be created. Then a common superclass *PlantSource* for these two classes is created, which in turn is a subclass of *DietarySource*.

- Combination development combines both topdown and bottom-up approach. Some classes are created using top down approach and others are created using bottom up approach.

Top	**Level 1**	**Level 2**	**Level 3**	**Level 4**
Thing	DietarySource	AnimalSource	Dairy	Butter Cheese Curd Milk
			Egg Fish Meat	
		PlantSource	Fruit Lentil Vegetable NutAndSeed WholeGrain	
	EfectOnHuman	Deficiency	EyeDisease SkinDisease OtherDisease	
		ExcessIntake WithinLimit		
	Person	Infant1-3		
		Child	Between4-8 Between9-13	
		Female	Lactating Pregnant Women	
		Male		
	Interaction	PositiveInteraction NegativeInteraction		
	FormsOf VitaminA	PreformedVitaminA ProvitaminACaretenoid		
	Supplement			

Table. 1. The different levels of the Vitamin A taxonomy represented in a tabular form

Table 1shows top level, level1, level2 level 3 and level of the proposed ontology in the nutrition domain

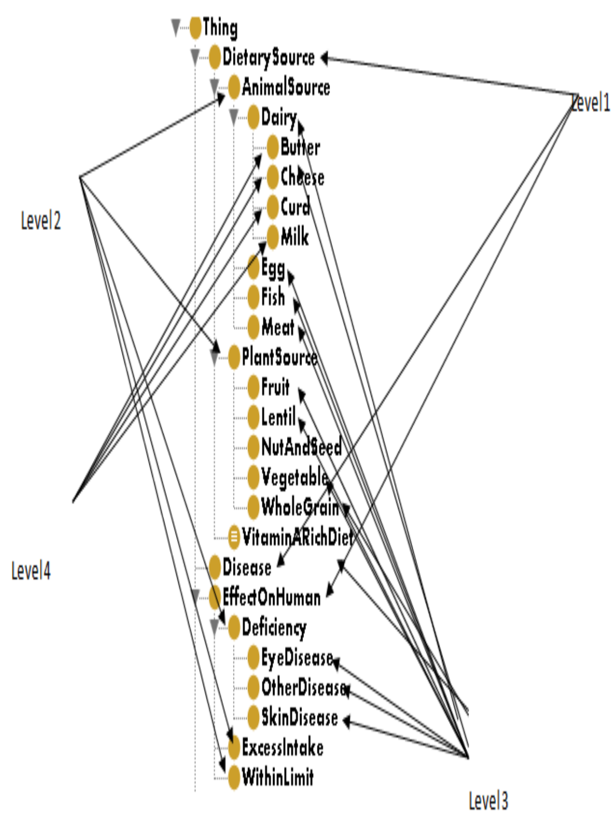


Fig. 2 Different levels of taxonomy represented using ontology editor

Fig 2 represents different levels of taxonomy using ontology editor. Ontology editor is a useful tool to represent information in the class subclass hierarchical structure.

Vitamin A ontology is designed which is used by every aspect of the system, in the information extraction, inference and retrieval phases. An iterative development process in the ontology engineering phase is followed till a stable ontology containing nearly classes and appropriate properties are obtained.

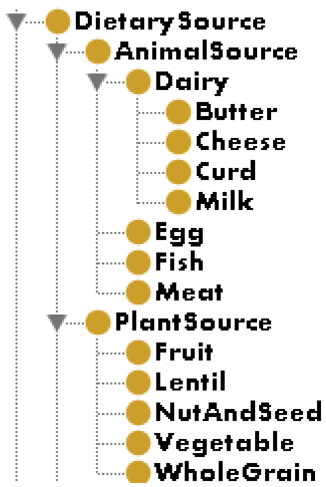


Fig. 3 Figure showing class and subclass under DietarySource

Figure shows class subclass hierarchy for *DietarySource*. This view shows two types of class, primitive and defined class. *VitaminARichDiet* is a defined class; all other classes are primitive classes. A class that only has necessary conditions is known as a Primitive Class. A class that has at least one set of necessary and sufficient conditions is known as a Defined Class.

2.5 Define the properties of the classes

A property defines relationship between object. If the relationship is between two individuals, it is an object property. Data property defines relationship between an individual and data values. *Enhances*, *IsEnhancedby*, *Causes* are examples of object properties. *Causes* defines relationship between *Dietarysource* and *ExcessIntake*. Whereas *HasQuantity*, *HasVitaminAinIU* are examples of data property. *HasVitaminAinIU* gives the quantity of vitamin A in an instance say plant food carrot.

For Example

```
<owl:ObjectPropertyrdf:about="#Causes">
<rdfs:domainrdf:resource="#DietarySource"/>
<rdfs:rangerdf:resource="#ExcessIntake"/>
<owl:inverseOfrdf:resource="#IsCausedBy"/>
</owl:ObjectProperty>
```

The above code defines an object property *Causes* has *DietarySource* has its domain, and has *ExcessIntake* as range. Inverse property of *Causes* is *IsCausedBy*. If one attribute has many domains then its domain will be intersection of its entire domain [7].

The classes, to which a property is attached, are called the domain of the slot. Allowed classes for properties of type Instance are often called a range of a property.

```
<owl:ObjectProperty rdf:about="#SuffersFrom">
      <rdf:type
rdf:resource="&owl;FunctionalProperty"/>
      <rdfs:comment rdf:datatype="&xsd;string"
         >This slot indicates the disease suffered
 by a particular person</rdfs:comment>
      <rdfs:range rdf:resource="#Disease"/>
      <rdfs:domain rdf:resource="#Person"/>
   </owl:ObjectProperty>
```

The above code defines an object property *SuffersFrom* which has *Person* has its domain, and *Disease* as range. Object property *SuffersFrom* defines relationship between the class *Person* and *Disease*.

The OWL code given below explains how an instance *John* who is a member of the class *Male* is related to *CysticFibrosis* which is a member of the class *Disease*.

```
<owl:Thingrdf:about="#John">
<rdf:typerdf:resource="#Male"/>
<SuffersFromrdf:resource="&www;CysticFibrosis"/>
</owl:Thing>
```

For each property, there is need to determine which class it describes. These properties become slots attached to classes. Thus, some of the slots of *vegetable* class are *hasVegType*, *HasLycopene*, *HasName*, *HasQuantity* etc. All subclasses of a class inherit the slot of that class. All the slots of the class *person* will be inherited by all its subclasses like *Women*, *Male*, *Child*.

2.6 Define the facets of a class

A facet of a property explains the value type, allowed values, the cardinality and other features of the values the property can take (Noy and McGuinness, 2000).

For example consider the property *HasVitaminAinIU*. The value type of this is 'Float' and cardinality is one. That is *HasVitaminAinIU* is a slot with value type Float. A slot can have multiple values and the values are instances of the class.

Slot cardinality: Slot cardinality defines how many values a slot can have. Single cardinality allows only one value for the slot whereas multiple cardinality allows more than one value for the slot. *HasVitaminAinIU* is a single cardinality slot. Slot value type describes what types of values can fill in the slot. There are number of value types like Boolean, integer, float, string, XML, Literal, URI to name a few. *HasVitaminAinIU* has value type as Float where as *HasName* has value type as string.

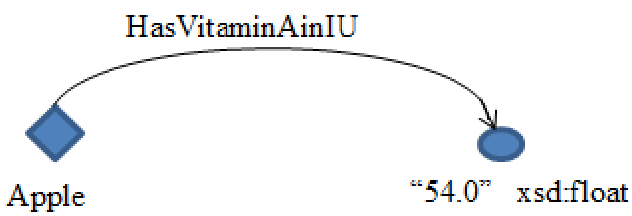


Fig. 1 Datatype property linking the individual Apple to the data literal '54.0.', which is of type float

The same information is represented in OWL code as shown below.

```
<Fruit rdf:about="#Apple">
<HasVitaminAinIU
rdf:datatype="&xsd;float">54.0</HasVitaminAinIU>
</Fruit>
```

2.7 Define instances

This is the last step in creating ontology. Instances can be manually typed or retrieved from another source like database or web page. Defining an individual instance of a class requires first choosing a class, secondly creating an individual instance of that class, and finally filling in the slot values.

To create an individual instance of *Cabbage*, the class under which it has to be created should be determined. Later slot values can be filled. Since *Cabbage* is a type of *Vegetable* which in turn is a *PlantSource* which is under *DietarySource*. Different slot values are as follows.

HasAlphaCarotene	33.0
HasName	Cabbage
HasQuantity	100
HasLuteinZeaxanthin	30.0
HasUnitOfMeasurement	gram
HaLycopene	0.0
HasVitaminAinIU	98.0
HasPlantType	Annual
HasBetaCarotene	42.0
HasVegType	Leaf
HasRetinolActivityEquivalent	5.0
HasBetaCryptoxanthin	0.0
HasColor	Green
HasMainConstituent	Cabbage

The OWL code for representing the attributes of cabbage is given below.

```
<owl:Thing rdf:about="#Cabbage">
      <rdf:type rdf:resource="#Vegetable"/>
      <HasLycopene
rdf:datatype="&xsd;float">0.0</HasLycopene>
      <HasBetaCryptoxanthin
rdf:datatype="&xsd;float">0.0</HasBetaCryptoxanthin>
      <HasQuantity
rdf:datatype="&xsd;int">100</HasQuantity>
      <HasLuteinZeaxanthin
rdf:datatype="&xsd;float">30.0</HasLuteinZeaxanthin>
      <HasAlphaCarotene
rdf:datatype="&xsd;float">33.0</HasAlphaCarotene>
      <HasBetaCarotene
rdf:datatype="&xsd;float">42.0</HasBetaCarotene>
      <HasRetinolActivityEquivalent
rdf:datatype="&xsd;float">5.0</HasRetinolActivityEqui
valent>
      <HasVitaminAinIU
rdf:datatype="&xsd;float">98.0</HasVitaminAinIU>
      <HasPlantType
rdf:datatype="&xsd;string">Annual</HasPlantType>
      <HasMainConstituent
rdf:datatype="&xsd;string">Cabbage</HasMainConstitu
ent>
      <HasName
rdf:datatype="&xsd;string">Cabbage</HasName>
      <HasColor
rdf:datatype="&xsd;string">Green</HasColor>
      <HasVegType
rdf:datatype="&xsd;string">Leaf</HasVegType>
```

<HasUnitOfMeasurement
rdf:datatype="&xsd;string">gram</HasUnitOfMeasurement>
</owl:Thing>

3. Conclusion

This paper describes designing ontology using Noy and McGuiness steps. Ontology on Vitamin A is created and it is instantiated with data values. Determining the domain and scope of the ontology is the foremost step while developing ontology. Defining taxonomy by identifying the key terms is the next important step. Representing the properties and facets of the class culminates the creation of ontology schema. Once the schema is created, instances are added to the ontology. Queries can be performed over the ontology to retrieve instances satisfying the given condition. This also results in validating the rules and logics of the ontology.

References

[1] Dieter Fensel · HolgerLausen · Axel Polleres Jos de Bruijn · Michael Stollberg · Dumitru Roman John Domingue, Enabling Semantic Web Services, springer, 2007

[2] Liyang Yu, 2011, A developers guide to the semantic web, Springer-Verlag Berlin Heidelberg,

[3]M Fernandez-Lopez, A Gomez-Perez, N Juristo.1997.Methontology: from ontological art towardsontological engineering, Spring Symposium on Ontological Engineering of AAAI. Stanford University, California. 1997.

[4] Noy, N., McGuinness, D.L, Ontology Development 101: A guide to creating your first ontology, Stanford Medical Informatics Technical Report No SMI-2001-0880, 2000

[5] Gruninger, M. and Fox, M.S. (1995). Methodology for the Design and Evaluation of Ontologies. In: Proceedings of the Workshop on Basic Ontological Issues in KnowledgeSharing, IJCAI-95, Montreal

[6] "DL Query Tab", p. 1, http://protegewiki.stanford.edu/ index.php/ DLQueryTab . Accessed on May 20th 2009.

[7] Heming Gong, Jianyi Guo ―Research and Building and Reason-ing of Travel Ontology‖ IEEE-2008.

A New Buy-back Contract Coordinating Dual-channel Supply Chain under Stochastic Demand

Guangxing Wei, Qiang Lin and Yanhong Qin

School of Management, Chongqing Jiaotong University, Chongqing, China

Abstract
The conflict between the manufacturer and the retailer except the double marginalization is an important issue in order to coordinate the dual-channel supply chain. In the general case of the non-linear stochastic demand which also is affected by the sales effort of the retailer, this paper designs a new buy-back contract to coordinate the dual-channel supply chain. On the base of developing the Stackelberg game model between the manufacturer and the retailer, the value of every parameter for the new buy-back contract, which can coordinate the dual-channel supply chain, is achieved respectively. The numerical experiment shows that the effort of the retailer can promote the sales amount of the retailing channel and the direct selling, and increase the profits of the manufacturer and the retailer simultaneously, while the overmuch effort of the retailer isn't good anymore.
Keywords: *Supply Chain Coordination, Dual-Channel, New Buy-Back Contract, Stochastic Demand, Stackelberg Game.*

1. Introduction

As the development of electronic pay technology, the manufacturers had the opportunities and abilities to sell to customer directly except the common retailing channel. More and more manufacturing firms attached importance to build their own online direct channel. Many famous brand manufactures, such as Hewlett-packard, Nike, Dell, IBM, Apple, have added their own online direct channel (Tsay and Agrawal, 2004). In China, the electric appliance manufacturers like Haier, Lenovo, also have their own direct marketing channel. In this context, the conventional single retail channel supply chain model changes to dual-channel supply chain model. Manufacturers might get benefit from their own direct channel. On the one hand, they can be closer to customers and understand their real demand information better. On the other hand, they can reduce the cost, expand the market shares, enhance his competitive power, increase the revenues, and avoid be dominated by the retailers. However, this will inevitably causes marketing competition between suppliers and retailers, and hence sharpen their conflict and contradiction. How to coordinate the dual-channel supply chain effectively in order to improve supply chain performance is an important theoretical and practical problem.

As a matter of fact, supply chain contract coordination has been research emphases. Cachon (2003) has reviewed and expanded a series of supply chain contracts, including wholesale price contract, buy-back contract, revenue sharing contract, quantity flexibility contract and quantity discount contract. The supply chain contract research progress also has been reviewed by Yang et al. (2006). There are many literatures on supply chain contract coordination with market demand influenced by the retailer's sale effort, which includes improving the shelf space layout (Wang, 2001), increasing the advertising investment (Taylor, 2002). For example, Cachon (2003) showed that the quantity discount contract can coordinate such supply chain model, while the constrained buy-back contract by constraining the quantity of buying back products can also coordinate the supply chain (Xu et al. , 2008). Suo et al. (2005) evaluated that the effect of the retailer's loss aversion on supply chain coordination. He et al. (2009) examined how the dependence level of demand to price influences the coordination of such supply chain. However, they mainly focus on the contract coordination of the traditional single retail channel supply chain.

There are many literatures in the dual-channel supply chain conflict and coordination also. Tsay and Agrawal (2004) indicated that although there are conflicts, the direct channel may also increase the retailer's profit. Dumrongsiri et al. (2008) analyzed the equilibrium conditions of the direct to increase the retailer's profit. Geng and Mallik (2007) discussed the equilibrium inventory strategies for increase the manufacturer and the retailer's profit simultaneously under multi-channel inventory competition environment. Takahashi et al. (2011) developed a new control policy for the two-echelon dual-channel supply chain with setup costs of production and delivery. Huang et al. (2012) examined how to adjust the price and the production plan so that the potential maximal profit is obtained under a disruptive demand. They found that the optimal production quantity had some robustness under a demand disruption. Li et al. (2011) analyzed the effect of supply chain members' risk preference on choosing of dual-channel supply chain operation mode. Chen et al. (2012) examined the coordination schemes for a dual-channel supply chain, found that a two-part tariff or a profit-sharing agreement, can coordinate the dual-channel supply chain and enable both the manufacturer and the retailer to realize win–win. Xu et al. (2010) designed a coordinating revenue sharing contract under linear demand, and Qu et al. (2010) designed a changed coordinating revenue sharing contract when the marker demand effected by the sales effort of the retailer. Yu and Liu (2012) found that the buy-back contract can improve the manufacturer and the retailer's profit but can not coordinate the dual-channel supply chain with the linear demand and joint promotion. Chen et al. (2011) pointed out that the system of the innovation compensation also could realize the development of supply chain members' profit, but cannot coordinate the dual-channel supply chain from the view of traditional distribution and online selling price competition. Wang et al. (2011) showed in the context of the switching fraction of non-linear demand that a revenue sharing contract of direct channel enable to decrease channel conflict, improve the channel system profit and coordinate the supply chain.

It is clear that the existing literatures about dual-channel supply chain coordination with linear demand and stochastic demand are focus on the study of revenue sharing contract, without considering the effect of retailer's effort on demand, and they fail to design a buy-back contract to coordinate the dual-channel supply chain.
In this paper, we consider the general case of the stochastic demand, which is affected by the sales effort of the retailer. On the base of developing the Stackelberg game model between the manufacturer and the retailer, we design a new buy-back contract coordinating the dual-channel supply chain. The value of every parameter for the new buy-back contract, which can coordinate the dual-channel supply chain, is achieved respectively. And by the numerical example, we analyze the effect of retailer's effort level on the sale quantity of retail channel and direct channel, and on the manufacturer and the retailer's profit.

2. Hypothesis and denotation

2.1 Hypothesis

As the literature practices (Cachon, 2003; Yang et al., 2006; Taylor, 2002; Qu et al., 2010), we makes the following hypothesis. The dual-channel supply chain includes only a manufacturer and a retailer, which both are the risk neutral. The manufacturer has the retailer and the direct channel. The retailer faces a newsvendor problem and only has one stochastic sales season. Before the sales season, the retailer has only one chance to order from the manufacturer. The market demand is non-linear stochastic, and influenced by the level of retailer's effort. The retailer's effort not only can increase the sales quantity of retail channel, but also can improve the direct channel's sales quantity. However, the manufacturer selling products through direct channel will reduce the retailer's sales quantity. So, the competition and conflict between dual-channel is inevitable. This is a special problem that makes the dual-channel supply chain difficult to coordinate beside the double marginalization. In addition, we assume that the retailer's penalty cost, the manufacturer's penalty cost, the direct channel's penalty cost and the product's net salvage value all are zero.

As two mutually independent decision makers, it is a typical Stackelberg game relationship between the manufacturer and the retailer. The manufacturer is the dominant one in the supply chain. The manufacturer first decides the direct price, wholesale price and buy-back price. Then, the retailer decides his effort level, his retail price. The common goal is to pursue own maximum expected profits.

2.2 Denotation

e: the retailer's effort level, represents all the retailer's effort activities;
$G(e)$: the retailer's effort cost of exerting effort e,where $G(0)=0$, $G'(e)>0$, $G''(e)>0$;
x: the total market demand, which is non-linear stochastic. According to Taylor (2002), let $x=\psi(e)\cdot\tau$, where $\psi(e)$ is non-negative, differentiable, $\psi'(e)>0$, $\psi''(e)\le 0$. τ is a random variable which independent to e,and whose distribution function is $F(\tau)$,density function is $f(\tau)$;
$F_X(x|e)$: the distribution function of total market demand, and it is differentiable, strictly increasing, denote $\bar{F}_X(x|e)=1-F_X(x|e)$;
$f_X(x|e)$: the density function of total market demand;
y : the demand of traditional retail channel, $y=\theta x$,where $\theta(0<\theta<1)$ presentation of the market share of retailer, θ is the function about e with $\theta'(e)>0$, $\theta''(e)<0$;
z: the demand of direct channel, $z=(1-\theta)x$;
$F_T(y|e)$:the demand distribution of retail channel;
$f_T(y|e)$:the demand density of retail channel;
$F_E(z|e)$:the demand distribution of direct channel;
$f_E(z|e)$:the demand density of direct channel;

w_m :the wholesale price;

b_m :the buy-back price, namely the price that the manufacturer pays the retailer for the remaining product, $b_m < w_m$;

p_T :the retail price;

p_E :the direct price;

c_m : the supplier's production cost per unit;

c_r :the retailer's marginal cost per unit;

c_E : the direct channel's marginal cost per unit;

q_T :the retailer's quantity, q_T is increasing in e;

q_E : the direct channel's quantity, q_E is decreasing in e.

3. The Basic Dual-channel Supply Chain Model

The manufacturer's profit includes the direct channel's profit under the dual-channel supply chain. The direct channel leads to competition and conflict between the supply chain members, the retailer's effort not only affect the retail channel, but also the direct channel. According to the above hypothesis, we can obtain that the distribution function of total market demand is $F_X(x|e) = F(\frac{x}{\psi(e)})$, the density function is $f_X(x|e) = \frac{1}{\psi(e)} f(\frac{x}{\psi(e)})$; The demand distribution function of retail channel is $F_T(y|e) = F(\frac{x}{\theta(e)\cdot\psi(e)})$, the density function is $f_E(z|e) = \frac{1}{(1-\theta(e))\cdot\psi(e)} f(\frac{x}{(1-\theta(e))\cdot\psi(e)})$.

Let $S_T(q_T, e)$ be the expected sales of retailer, and $S_E(q_E, e)$ be the expected sales of direct channel, $Z(q_T)$ be the transfer payment function. Under the above conditions, through simple calculation, we can obtain:

The retailer's expected sales is

$$S_T(q_T, e) = \int_0^{q_T} y f_T(y|e) d_y + \int_{q_T}^{\infty} q_T f_T(y|e) d_y$$

$$= q_T - \int_0^{q_T} F_T(y|e) d_y = \frac{3}{2} q_T - \frac{1}{2\theta(e)\cdot\psi(e)} q_T{}^2,$$

the expected sales of direct channel is

$$S_E(q_E, e) = \int_0^{q_E} z f_E(z|e) d_z + \int_{q_E}^{\infty} q_E f_E(z|e) d_z$$

$$= q_E - \int_0^{q_E} F_E(z|e) d_z = \frac{3}{2} q_E - \frac{1}{2(1-\theta(e))\cdot\psi(e)} q_E{}^2.$$

Therefore, the retailer's expected profit function under dual-channel is

$$\prod\nolimits_r(q_T, e) = p_T S_T(q_T, e) - c_r q_T - G(e) - Z(q_T)$$

$$= (\frac{3}{2} p_T - c_r) q_T - \frac{p_T}{2\theta(e)\cdot\psi(e)} q_T{}^2 - G(e) - Z(q_T) \tag{1}$$

the manufacturer's expected profit function is

$$\prod\nolimits_m(q_T, q_E, e) = Z(q_T) - c_m q_T + p_E S_E(q_E, e) - c_E q_E$$

$$= Z(q_T) - c_m q_T + (\frac{3}{2} p_E - c_E) q_E - \frac{p_E}{2(1-\theta(e))\cdot\psi(e)} q_E{}^2 \tag{2}$$

and the dual-channel supply chain's profit function is

$$\prod\nolimits_{sc}(q_T, q_E, e) = p_T S_T(q_T, e) - (c_r + c_m) q_T + p_E S_E(q_E, e) - c_E q_E - G(e)$$

$$= (\frac{3}{2} p_T - c_r - c_m) q_T - \frac{p_T}{2\theta(e)\cdot\psi(e)} q_T{}^2 + (\frac{3}{2} p_E - c_E) q_E - \frac{p_E}{2(1-\theta(e))\cdot\psi(e)} q_E{}^2 - G(e) \tag{3}$$

Consider the problem of the centralized decision making. By Eq. (3), for a given effort level, it can be obtained that there is $\partial^2 \prod_{sc}(q_T, q_E, e)/\partial q_T{}^2 = -p_T/\theta(e)\cdot\psi(e) < 0$ and $\partial^2 \prod_{sc}(q_T, q_E, e)/\partial q_E{}^2 = -p_E/(1-\theta(e))\cdot\psi(e) < 0$, this means that the dual-channel supply chain system exists only one optimal solution, i.e., the retailer's optimal order quantity q_T^* satisfies $q_T^* = \arg \frac{\partial \prod_{sc}(q_T, q_E, e)}{\partial q_T} = 0$.

According to the first-order condition of Eq. (3), the specific retailer's optimal order quantity is

$$q_T^* = \frac{\frac{3}{2} p_T - c_r - c_m}{p_T} \cdot \theta(e)\cdot\psi(e) \tag{4}$$

and the direct channel's optimal order quantity q_E^* satisfies $q_E^* = \arg \frac{\partial \prod_{sc}(q_T, q_E, e)}{\partial q_E} = 0$,i.e.,

$$q_E^* = \frac{\frac{3}{2} p_E - c_E}{p_E} \cdot (1-\theta(e))\cdot\psi(e) \tag{5}$$

In a similar way, the retailer's optimal effort level for a given q_T and q_E satisfies $e^* = \arg \frac{\partial \prod_{sc}(q_T, q_E, e)}{\partial e} = 0$.

According to the first-order condition of Eq. (3), the specific retailer's optimal effort level is

$$e^* = \arg[p_T \frac{\partial S_T(q_T, e^*)}{\partial e} + p_E \frac{\partial S_E(q_E, e^*)}{\partial e} - G'(e^*)] = 0 \tag{6}$$

4. The New Buy-back Contract

With a buy-back contract the supplier charges the retailer w_m ($w_m < p_T$) per unit purchased, but pays the retailer b_m ($b_m < w_m$) per unit remaining at the end of the selling season. The transfer payment between the manufacturer and the retailer is

$$\begin{aligned} Z(q_T, w_m, b_m) &= w_m q_T - b_m(q_T - S(q_T, e)) \\ &= b_m S(q_T, e) + (w_m - b_m) q_T \end{aligned} \tag{7}$$

However, according to Cachon (2003), the buy-back contract cannot coordinate the supply chain with effort-dependent demand. In the way of the general processing method, we introduce a contract parameter $\phi (0 < \phi < 1)$ (Qu et al., 2010), which presents the fraction of effort cost the retailer bears. So, $(1-\phi)$ is the fraction of the manufacturer. According to Eq. (1) and Eq. (7), the retailer's expected profit function is

$$\begin{aligned} \prod\nolimits_r (q_T, e, w_m, b_m) &= (p_T - b_m) S_T(q_T, e) \\ &\quad -(c_r + w_m - b_m) q_T - \phi G(e) \\ &= (\frac{3}{2} p_T - \frac{1}{2} b_m - c_r - w_m) q_T \\ &\quad -(\frac{p_T - b_m}{2\theta(e)\cdot\psi(e)}) q_T{}^2 - \phi G(e) \end{aligned} \tag{8}$$

By Eq. (2) and Eq. (7), the manufacturer's expected profit is

$$\begin{aligned} \prod\nolimits_m (q_T, q_E, e, w_m, b_m) &= b_m S_T(q_T, e) - (c_m - w_m + b_m) q_T \\ &\quad + p_E S_E(q_E, e) - c_E q_E - (1-\phi) G(e) \\ &= (\frac{1}{2} b_m - c_m + w_m) q_T - \frac{b_m}{2\theta(e)\cdot\psi(e)} q_T{}^2 \\ &\quad +(\frac{3}{2} p_E - c_E) q_E - \frac{p_E}{2(1-\theta(e))\cdot\psi(e)} q_E{}^2 - (1-\phi) G(e) \end{aligned} \tag{9}$$

By Eq. (3) and Eq. (7), the dual-channel supply chain's expected profit function is

$$\begin{aligned} \prod\nolimits_{sc} (q_T, q_E, e, w_m, b_m) &= p_T S_T(q_T, e) - (c_r + c_m) q_T \\ &\quad + p_E S_E(q_E, e) - c_E q_E - G(e) \\ &= (\frac{3}{2} p_T - c_r - c_m) q_T - \frac{p_T}{2\theta(e)\cdot\psi(e)} q_T{}^2 + (\frac{3}{2} p_E - c_E) q_E \\ &\quad - \frac{p_E}{2(1-\theta(e))\cdot\psi(e)} q_E{}^2 - G(e) \end{aligned} \tag{10}$$

On the other hand, consider the problem of the decentralized decision making. Because it is clear that $\partial^2 \prod_r (q_T, q_E, e, w_m, b_m) / \partial^2 q_T{}^2 = -(p_T - b_m) / \theta(e)\cdot\psi(e) < 0$, the retailer has an unique optimal order quantity q_T^{**}. Then, by Eq. (8), the retailer's optimal order quantity for a given effort level satisfies

$$q_E^{**} = \arg \frac{\partial \prod_m (q_T, q_E, e, w_m, b_m)}{\partial q_E} = 0,$$

i.e.,

$$q_T^{**} = \frac{\frac{3}{2} p_T - \frac{1}{2} b_m - c_r - w_m}{p_T - b_m} \theta(e)\cdot\psi(e) \tag{11}$$

And the retailer's optimal effort level for a given q_T satisfies $e^{**} = \arg \frac{\partial \prod_r (q_T, e, w_m, b_m)}{\partial e} = 0$, i.e.,

$$e^{**} = \arg(p_T - b_m) \frac{\partial S_T(q_T, e^{**})}{\partial e} - \phi G'(e^{**}) = 0 \tag{12}$$

By $\partial^2 \prod_m (q_T, q_E, e, w_m, b_m) / \partial q_E{}^2 = -p_E / (1-\theta(e))\cdot\psi(e) < 0$, the direct channel has unique optimal order quantity q_E^{**}. By Eq. (9), the direct channel's optimal order quantity for a given effort level satisfies

$$q_E^{**} = \arg \frac{\partial \prod_m (q_T, q_E, e, w_m, b_m)}{\partial q_E} = 0$$

i.e.,

$$q_E^{**} = \frac{\frac{3}{2} p_E - c_E}{p_E} (1-\theta(e))\cdot\psi(e) \tag{13}$$

5. The Dual-channel Supply Chain Coordination with New Buy-back Contract

The competition and conflict between the supply chain members will occur when the manufacturer adds a direct channel. To make the dual-channel supply chain achieves coordination, the coordination of the retailer's order quantity, effort level and the direct channel's order quantity must be realized at the same time. We can prove that as long as the value of every parameter for the new buy-back contract are appropriate, the dual-channel supply chain can achieve coordination. We can obtain the following Propositions respectively.

Proposition 1: The retailer's order quantity q_T can be coordinated, as long as the contract parameter w_m and b_m satisfies

$$w_m = b_m + c_m - \frac{c_r + c_m}{p_T} b_m \tag{14}$$

Proof: If the retailer's order quantity achieved coordination, $q_T^* = q_T^{**}$. By Eq. (4) and Eq. (11), Eq. (14) can be obtained through calculation. **Q.E.D.**

Proposition 2: The retailer's effort level e can be coordinated, as long as the contract parameter b_m and ϕ satisfies

$$1-\phi=\frac{b_m\frac{\partial S_T(q_T,e)}{\partial e}+p_E\frac{\partial S_E(q_E,e)}{\partial e}}{G^{'}(e)} \qquad (15)$$

Proof: If the retailer's effort level achieved coordination, $e^*=e^{**}$. By Eq. (6) and Eq. (12), Eq. (15) can be obtained through simple calculation. **Q.E.D.**

The Eq. (5) and Eq. (13) show that the direct channel optimal order quantity is relative to itself price, marginal cost and the retailer's effort level. When the retailer's effort level realizes coordination, the coordination of the direct channel order quantity will be achieved. So, we can receive the following conclusion:

Conclusion: The new buy-back contract can coordinate the dual-channel supply chain as long as the value of every contract parameter satisfies the certain condition.

6. Numerical experiment

In order to discuss the model and illustrate the conclusion more clearly, this section through numerical example analysis the optimal decision results of dual-channel supply chain members before and after channel coordination with the above new buy-back contract. Suppose a certain kind of product with the market characteristics: $\psi(e)=1000(1-e^{-1})$, $\theta(e)=\sqrt{e}/(\sqrt{e}+1)$, $x=1000(1-e^{-1})\cdot\tau$, where the random variable τ comply with uniform distribution at [0.5,1.5], $G(e)=50e^2$, $p_T=120$, $c_r=10$, $c_m=30$, $p_E=80$, $c_E=8$, $w_m=55$, $b_m=25$, $\phi=0.7$.

We put these parameters into the above model. And through Matlab software, we can obtain the optimal decision results of decision-makers under centralized decision and decentralized decision, and the value of every contract parameter when the supply chain achieve coordination, which is shown in the Table 1.

The results of Table 1 shows that the dual-channel supply chain system profit under centralized decision higher than buy-back contract model under decentralized decision before coordination. With the appropriate contract parameters value, the new buy-back contract make the system profit under decentralized decision equals to the system profit under centralized decision, which realize the dual-channel supply chain coordination. After the coordination, the retailer's optimal order quantity and effort level are increased, but the direct channel's optimal order quantity is reduced. According to Eq. (8), Eq. (9), Eq. (11) and Eq. (13), we can obtain the corresponding the retailer's order quantity, the direct channel's order quantity, the retailer's profit and the manufacturer's profit for a given effort level respectively. Thus, we can examine the effect of the retailer's effort level on the decisions and profits of the supply chain members by the obtained data, which is shown in the following Figure 1 and Figure 2 respectively.

Table 1: The optimal decision results of dual-channel supply chain

Dual-channel Supply chain	The new buy-back contract model under decentralized decision		The basic model
System index	Before coordination	After coordination	Centralized decision
e	10.93	11.39	11.39
q_T	753	821	821
q_E	295	292	292
w_m	55	(38,80)	/
b_m	25	(12,75)	/
ϕ	0.7	(0.49,1)	/
π_r	34389	(18372,45235)	/
π_m	32826	(22094,48957)	/
π_{sc}	67215	67329	67329

The results of Fig. 1 shows that with the increasing of the retailer's effort level, the retailer's order quantity increase gradually and tend to be stable. From the general view, the reason is that the greater effort level, the higher sales and order quantity. However, the total amount of the market is limited, and the retailer's order quantity can not continue to increase with the effort level. So, the order quantity maintaining in a stable state, otherwise, the retailer will face the risk of big increase in the number of inventory and unmarketable. The order quantity of direct channel appears decreasing gradually. But, the reduced scope becomes more and more small. The reason is that the retailer's also will increase its order quantity when improve effort level, but the retailer could not increase order quantity infinitely. So, the direct channel's order quantity tends to be stable. In this way, the entire supply chain's order quantity increases gradually and tends to be stable. The retailer can increase sales quickly in a short period by improving his own effort level, which makes the gap between its order quantity and the direct channel's order quantity maintain a bigger level, and be able to maintain a competition advantage. Thus, the change trend of order quantity of the whole supply chain is consistent with that of the retailer

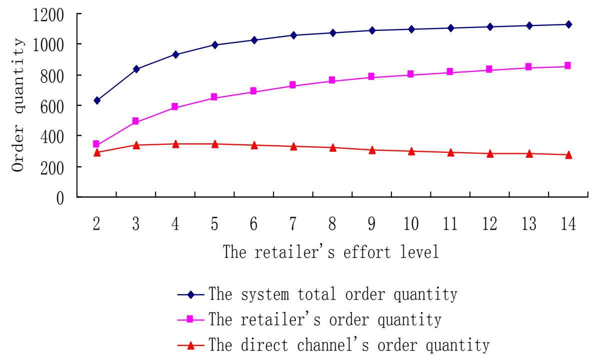


Fig. 1 The effect of the retailer's effort level on the order decisions.

Fig.2 shows that with the increasing of the retailer's effort level, the change trend of supply chain members' profit is consistent with the trend of each order quantity: the retailer's profit appears increasing slowly and tends to be stable, while the manufacturer's profit has a small increase and then drop slowly. As a result, the entire supply chain's profit appears decreasing obviously but moderately. It suggests that the bigger retailer's effort level does not mean the better. Once beyond the best effort level critical value, it not only cannot continue to increase his own profit, but also lessen that of the supplier and supply chain. Obviously, the manufacturer shall not accept this outcome. This also explains why the manufacturer is not willing to stimulate the retailer adds more effort through sharing more effort cost in reality.

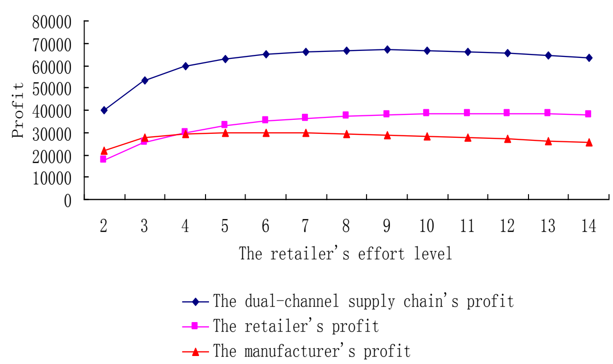


Fig. 2 The effect of the retailer's effort level on the profits.

7. Conclusions

In this paper, we develop Stackelberg game model of a simple two-echelon dual-channel supply chain with one supplier and one retailer, and design a new buy-back contract coordinating the dual-channel supply chain under a non-linear stochastic demand influenced by the retailer's sales effort. Because the traditional buy-back contract can not coordinate the supply chain with the effort-dependent demand, we introduce the effort cost sharing parameter into the buy-back contract, i.e., the supplier shares the fraction of the retailer's effort cost. Our results reveal that the new buy-back contract can coordinate the dual-channel supply chain as long as the contract parameters satisfy certain conditions. In the numerical analysis, we verify the theoretical analysis by calculating the specific optimal decision results of supply chain members under centralized decision and decentralized decision, and further examine the effect of retailer's effort level on the sale quantity of retail channel, direct channel, and on the manufacturer and the retailer's profit. The results of numerical analysis show that the appropriate increase of the retailer's effort level can promote the increase of the sale quantity of retail channel and direct channel simultaneously, and also the profit of the manufacturer and the retailer. However, as the continuous increasing of the retailer's effort level, the direct channel's order quantity and the manufacturer's profit decrease slowly down, and the increasing of the retailer's profit is also very limited. The research conclusions have certain realistic significance to supply chain members, while still have limitations. For example, it doesn't consider the effort level of the manufacturer while the manufacturer may also increase the direct channel's sales quantity through improve effort level, which is the next research direction.

References

[1] A. A. Tsay, and N. Agrawal, "Channel Conflict and Coordination in the E-commerce Age", Production and Operations Management, Vol. 13, No. 1, 2004, pp. 93-110.

[2] G. P. Cachon, Supply chain coordination with contracts, In: De Kok, A.G., Graves, S. (Eds.), Handbooks in Operations Research and Management Science, Amsterdam: Elsevier, 2003.

[3] D. L. Yang, Q. Guo, Y. He, and J. Y. Xu, "Review of Supply Chain Contracts", Chinese Journal of Management, Vol. 3, No. 1, 2006, pp. 117-123.

[4] Y. Wang, and Y. Gerchak, "Supply Chain Coordination When Demand is Shelf-space Dependent", Manufacturing and Service Operations Management, Vol. 3, No. 1, 2001, pp. 82-87.

[5] T. A. Taylor, "Supply Chain Coordination under Channel Rebates with Sales Effort Effects", Management Science, Vol. 48, No. 8, 2002, pp. 992-1007.

[6] Z. Xu, D. L. Zhu, and W. G. Zhu, "Supply Chain Coordination under Buy-back Contract with Sales Effort Effects", Systems Engineering-Theory & Practice, Vol. 28, No. 4, 2008, pp. 1-11.

[7] Y. He, X. Zhao, L. D. Zhao, and J. He, "Coordination a Supply Chain with Effort and Price Dependent Stochastic Demand", Applied Mathematical Modeling, Vol. 33, No. 6, 2009, pp. 2777-2790.

[8] H. S. Suo, J. C. Wang, and Y. H. Jin, "Supply Chain

Coordination with Sales Effort Effects and Impact of Loss Aversion on Effort Decision", Tsinghua Science and Technology, Vol. 10, No. 1, 2005, pp. 102-107.

[9] A. Dumrongsiri, M. Fan, A. Jain, and K. Moinzadeh, "A Supply Chain Model with Direct and Retail Channels", European Journal of Operational Research, Vol. 187, No. 3, 2008, pp. 691-718.

[10] Q. Geng, and S. Mallik, "Inventory Competition and Allocation in a Multi-channel Distribution System", European Journal of Operational Research, Vol. 182, No. 2, 2007, pp. 704-729.

[11] K. Takahashi, T. Aoi, D. Hirotani, and K. Morikawa, "Inventory Control in a Two-echelon Dual-channel Supply Chain with Setup of Production and Delivery", International Journal of Production Economics, Vol. 133, No. 1, 2011, pp. 306-314.

[12] S. Huang, C. Yang, and X. Zhang, "Pricing and Production Decisions in Dual-channel Supply Chains with Demand Disruptions", Computers and Industrial Engineering, Vol. 62, No. 1, 2012, pp. 70-83.

[13] S. J. Li, Z. G. Zhang, and Y. Huang, "Effects of Risk Aversion on Operational Mode in Dual-channel Supply Chain", Industrial Engineering and Management, Vol. 16, No. 1, 2011, pp. 32-36.

[14] J. Chen, H. Zhang, and Y. Sun, "Implementing Coordination Contracts in a Manufacturer Stackelberg Dual-channel Supply Chain", Omega, Vol. 40, No. 5, 2012, pp. 571-583.

[15] G. Y. Xu, B. Dan, and J. Xiao, "Study on Dual-channel Supply Chain Coordination with New Revenue-sharing Contract", Chinese Journal of Management Science, Vol. 18, No. 6, 2010, pp. 59-64.

[16] D. G. Qu, and Y. J. Guo, "Coordination of Supply Chain with Hybrid Distribution Channels When Retailer's Demand Relies on Its Sales Effort", Chinese Journal of Management Science, Vol.16, No. 3, 2008, pp. 89-94.

[17] A. M. Yu, and L. W. Liu, "Competing and Coordination Strategies for the Dual-channel under Stochastic Demand and Cooperative Promotion", Journal of Industrial Engineering and Engineering Management, Vol. 26, No. 1, 2012, pp. 151-154.

[18] S. Z. Chen, Z. K. Xiong, G. D. Li, and H. H. Wen, "Coordination Mechanisms Based on Strategic Innovative Compensation in Dual-channel Supply Chain", Journal of Industrial Engineering and Engineering Management, Vol. 25, No. 2, 2011, pp. 45-52.

[19] H. Wang, J. Zhou, Y. L. Sun, "Study on Decisions and Coordination of Dual-channel Supply Chain in Competitive Environment", Operations Research and Management Science, Vol. 20, No. 1, 2011, pp. 35-40.

[20] Wang Xiu-hong, “Research of Knowledge Supply Chain Coordination Mechanism based on the View of Fuzzy Demand”, International Journal of Computer Science Issues, Vol. 9, No. 5, 2012, pp. 43-47.

[21] Noureddine Labiad, Zitouni Beidouri, Otmane Bouksour, “Evaluation of coordination contracts for a two stage Supply Chain under price dependent demand”, International Journal of Computer Science Issues, Vol. 9, No. 1, 2012, pp. 151-159.

GuangXing Wei Chongqing Province, China. Birthdate: Aug, 1977. Work as professor in Management School of Chongqing Jiaotong University, and research interests on mass customization and supply chain management.

Qiang Lin Henan Province, China. Graduate student in Management School of Chongqing Jiaotong University, and research interests on supply chain management.

Optimization of Postponed Production Decision in Mass Customization

Yanhong Qin, Danping Xiong

School of Management, Chongqing Jiaotong University
Chongqing, 400074, China

Abstract

Whether to implement postponed production with higher production cost is puzzled by more manufacturers, and to solve this problem, we set the total cost model considering inventory cost and customer waiting cost in non-postponed production and postponed production respectively. Then, we apply the M/M/1 model in queuing theory to optimize the total cost, and by simulation and sensitivity analysis, we find out some important factors influencing whether to apply postponed production such as customer waiting cost, the product categories, and so on. Besides, if the postponed production is adopted, where is the best CODP (Customer order decoupling point) in the postponed production process, and these research results will provide theoretical support for the manufacturer to locate the CODP point correctly, and help the manufacturer to make decision on implementing postponed strategy.

Keywords: *Mass customization, Postponed production, Production cost, CODP*

1. Introduction

In the current market environment, with the changing speed of product categories accelerating, the market demands appeared more various. Postponed production can complete the customized order according to customer demand and finish the final delivery by delaying the customization work until the certain customer order is received, and thus the postponed production can decrease the delivery lead time, reduce the risk of outdated inventory, and meet the various customer demands (Su, 2005). Italian clothing company Benetton adopted the postponement strategy to delay the CODP in the production process until to the end production process, and thus to reduce the risk of uncertainty demand. There have been many theories and literatures on postponement strategy. The postponed production is applied by retailer to increase the profit, but the manufacture is ignored (Yang et al., 2009). The inventory managing strategy model is set in supply chain applying postponed strategy, but the total production cost was not considered (Shao, 2004 and Ji, 2009). Jin et al. (2011) proposed the method of “Kanban” to achieve customization, but they didn’t analyze whether the Kanban system can reduce the cost of manufacture in the quantitative way. Lee and Tang (1999) implemented the postponed production by restructure the product portfolio and set an simple model by combing the profit, cost and product strategy, and the model was an general postponed production model to achieve customization, but in their model, they thought the manufacture need different processes to obtain the customization after the CODP (Customer Order Decoupling Point). Hua (2007) optimized the flexible production system by queue theory. Van (1998), Krajewski (2005) and Rao (2007) research the postponed production in terms of simple customization and mass customization, but the key to their problem is whether the manufacture can reduce cost by customized postponement. Huang et al. (2008) compared the cost change before and after the CODP, but they ignored the payment of customer waiting cost to keep the customer purchasing the product. Dan (2009) set the cost optimization model for the two-stage supply chain made up of retailer and manufacture applying postponed production, and Li (2010) set the similar model and CODP orientation model, but both of them didn’t consider the cost change before and after the CODP only for the manufacture, i.e. they didn’t compare the cost change in different condition in applying postponed production or not applying the postponed strategy.

To perfect the postponed strategy, based on the research of Dan (2009) and Li (2010), this paper will focus on the manufacture and set the cost model when the manufacture in postponed production and without the postponed production (non-postponed production) respectively in the first place. And then we will adopt the extended M/M/1mode in queue theory to simulate the postponed production cost of manufacture to find out the condition when the manufacture can apply postponed production mode and the optimal CODP location in postponed production. Finally, the sensitive analysis of factors influencing the production cost of manufacture will be analyzed.

2. The model under the non-postponed production mode

2.1 Denotation

The manufacture will face the uncertain market demand and produce N kinds of customized products belonged to the same product family while keeping some safe inventory. Each kind of product need time T to be finished. The customer demand of the K th customized product is a random variable whose average value is λ_k , and the variant value is σ_k^2 . The average demand amount of total customized products is $D=\sum_{k=1}^{N}\lambda_k$, where λ_k is the unit demand reaching ratio of the K th product, μ_k is the expected demand of the K th product, m_k is the unit average production cost, v_k is the unit holding cost of K th product inventory, h_k is the average inventory cost and $E[I_k]$ is the expected inventory amount of the K th product.

2.2 Assumption

① The manufacture will produce some product and keep the inventory according to the forecast information of the market demand, and the produce intension will be bigger than the reaching ratio of customer demand, so there will be no shortage cost and customer waiting cost.

② Whether to implement the postponed production will not influence the transportation time, so we will not consider the transportation cost.

2.3 Model

Under non-postponed production mode, the manufacture should consider the total cost in a production period including manufacture cost, holding cost of processing product, the inventory cost, etc.

$$Z=\sum_{k=1}^{N}\lambda_k m_k+\frac{\lambda_k}{\mu_k}v_k+\sum_{k=1}^{N}h_k E[I_k] \qquad (1)$$

Formula (1) denote the total production cost under the non-postponed production mode, where the first three items denote the manufacture cost, holding cost of processing product and the inventory cost respectively.

For easy analysis, we will further assume the condition based on the general instance, as following: in the random time zone, the probability of customer demand for the certain customized product is dependent on the length of time zone, but it independent of the terminal of time zone, and if the time zone is small enough, the probability that the number of customer demand for the certain customized product is more than twice can be ignored. Besides, in each independent time zone, the demand of product is independent. All the customized products are belonged to the same product family K , and thus each customized products are similar and have the same production time, so we can assume that λ_k , μ_k , m_k , v_k and $E[I_k]$ can be equal λ , μ , m , v and $E[I]$ respectively.

We can adopt the M/M/1model to compute the production cost under non-postponed production.

The production intension is: $\rho=\frac{\lambda}{\mu}$

The expected inventory amount is:

$$E[I]=s-\rho(1-\rho^s)$$

The total production cost is:

$$Z=Dm+\frac{\lambda}{\mu}v+Nh[s-\rho(1-\rho^s)]$$

3. The model under postponed production mode

3.1 Denotation

The manufacture will face the uncertain market demand and produce N kinds of customized products belonged to the same product family by postponed production. The investment cost that the manufacture invest in the equipment so as to implement postponed production is $F(r)$. Each kind of product need time T to be finished. The production time can be divided into two parts: before CODP, the pushed production process needs time T_1 , and after CODP, the pulled production process needs time T_2 . The customer demand of the K th customized product is a random variable whose average value is $\lambda_{2,k}$, and the variant value is $\sigma_{2,k}^2$. The average demand amount of total customized products is $D=\sum_{k=1}^{N}\lambda_{2,k}$, where λ_1 is the unit reaching ratio of the standard semi-finished products, μ_1 is the expected demand of the standard semi-finished product, $m_1(r)$ is the unit average production cost of the standard semi-finished product, $v_1(r)$ is the unit holding cost of semi-finished inventory, $h_1(r)$ is the average inventory cost and $E[I_1](r)$ is the expected inventory amount of semi-finished inventory. After CODP, $\lambda_{2,k}$ is the unit demand reaching ratio of the K th customized product, $\mu_{2,k}$ is the expected demand of the K th product, $m_{2,k}(r)$ is the unit average production cost, $v_{2,k}(r)$ is the unit holding cost of K th product inventory, $E[T_{2,k}](r)$ is the expected production time of the K th product and $w_{2,k}$ is the unit customer waiting cost of the K th customized product.

3.2 Assumption

①Whether to implement the postponed production will not influence the transportation time, so we will not consider the transportation cost. The location of CODP can be denoted as the ratio of production time of standard product to the total production time, i.e. $r(0 \le r \le 1)$.
② The manufacture will start production according to customized customer order, and once the customer ordered, it can't be hauled off, and thus, it can't influence the profit of manufacture, so the shortage cost will be not considered.

3.3 Model

Under postponed production mode, the manufacture should consider the total cost in a production period including manufacture cost, holding cost of processing product, the inventory cost, customer waiting cost, etc. The model of production cost under postponed production is as following:

$$\min Z(r) = F(r) + \sum_{k=1}^{N} \lambda_{2,k} m_1(r) + \sum_{k=1}^{N} \lambda_{2,k} m_{2,k}(r)$$

$$+ h_1(r) E[I_1](r) + \sum_{k=1}^{N} \lambda_{2,k} w_{2,k} ET_{2,k}(r) \qquad (2)$$

s.t. $$0 \le r \le 1 \qquad (3)$$

$$0 < \frac{\lambda_1(r)}{\mu_1(r)} < 1 \text{ , } 0 < \frac{\lambda_{2,k}(r)}{\mu_{2,k}(r)} < 1 \qquad (4)$$

Formula (2) denote the total production cost under the postponed production mode, where the first three items denote the investment cost, manufacture cost respectively. The fourth and fifth item is the holding cost of processing product before and after CODP. The sixth item is the inventory cost at CODP. The seventh item is the customer waiting cost afforded by manufacture so as to keep customer to purchase. Constraint (3) denotes the location of CODP can be at random stage in the whole production process. Constraint (4) denotes the production intension of manufacture is bigger than the reaching ratio of customer demand, which indicate that the overstock condition will not be appeared.

3.4 The extended model of M/M/1

For easy analysis, we will further assume the condition based on the general instance, as following: in the random time zone, the probability of customer demand for the certain customized product is dependent on the length of time zone, but it independent of the terminal of time zone, and if the time zone is small enough, the probability that the number of customer demand for the certain customized product is more than twice can be ignored. Besides, in each independent time zone, the demand of product is independent. All the customized products are belonged to the same product family K , and thus each customized products are similar and have the same production time, so we can assume that $\lambda_{2,k}$, $\mu_{2,k}$, $m_{2,k}$, $v_{2,k}(r)$, $E[T_{2,k}](r)$ can be equal λ_2 , μ_2 , $m_2(r)$, $v_2(r)$ and $E[T_2](r)$ respectively. With the location of CODP moving to the end of production process, more process will be standard and modular, and the manufacture center finishing more standard semi-finished products need be more flexible, leading the increasing investment cost. We can assume $F(r)$ is the simple increasing function of r . The customized products of product family are not the high additional-value product, and the incremental process is a continuous and even process of production time, i.e. $m_1(r)$, $h_1(r)$ are the simple increasing functions of r , and $m_2(r)$ is the simple decreasing functions of r .

According to the assumption of extended model, the manufacture intension of standard semi-finished product is $\rho_1(r)$ before CODP, and the production intension of customized product is $\rho_2(r)$.

$$\rho_1 = \frac{\lambda_1}{\mu_1} = rDT \text{ , } \rho_2 = \frac{\lambda_2}{\mu_2} = (1-r)DT$$

We can obtain the efficiency index before and after CODP according to the research of Buzacott and Shanthikumar (1993). The expected inventory amount of standard semi-finished products and the expected production time of customized products can be denoted as following:

$$E[I_1] = s_1 - \frac{\rho_1(1-\rho_1^{s_1})}{1-\rho_1} = s_1 - \frac{rDT-(rDT)^{s_1+1}}{1-rDT}$$

$$E[T_2] = \frac{1}{\mu_2-\lambda_2} = \frac{NT(1-r)}{1-DT(1-r)}$$

In the formula, s_1 is the safety factor relative to the inventory before CODP, so the cost model of manufacture can be transformed as:

$$\begin{aligned}&\min Z(r) = F(r) + Dm_1(r) + Dm_2(r) + rDTv_1(r)\\&+N(1-r)DTv_2(r)+\\&h_1(r)[s_1 - \frac{rDT-(rDT)^{s_1+1}}{1-rDT}] + \frac{DwN(1-r)T}{1-D(1-r)T}\end{aligned} \quad (5)$$

s.t. $0 \le r \le 1$ (6)

$$0 < DT < 1 \quad (7)$$

$$v_1(r) < h_1(r) < v_2(r) \quad (8)$$

$$\frac{\partial F(r)}{\partial r} > 0, \frac{\partial m_1(r)}{\partial r} > 0, \frac{\partial h_1(r)}{\partial r} > 0 \quad (9)$$

$$\frac{\partial m_2(r)}{\partial r} < 0 \quad (10)$$

$$s_1 \in Z^+ \quad (11)$$

Formula (5) denote the total production cost under the postponed production mode, where the first three items denote the investment cost, manufacture cost before and after CODP respectively. The fourth and fifth item is the holding cost of processing product before and after CODP. The sixth item is the inventory cost at CODP. The seventh item is the customer waiting cost afforded by manufacture so as to keep customer to purchase. Constraint (6) denotes the location of CODP can be at random stage in the whole production process. Constraint (7) denotes the production intension of manufacture is bigger than the reaching ratio of customer demand, which indicate that the overstock condition will not be appeared. Constraint (8) means the unit holding cost of processing product before CODP is smaller than the unit holding cost of standard semi-finished product after CODP, and thus it is smaller than the unit holding cost of processing product after CODP and unit inventory cost of standard semi-finished product at the same time. The three items of constraint (9) that the investment cost, value and inventory cost of standard semi-finished product before CODP is increasing with increasing r. Constraint (10) denote the production time of customization after CODP is shortening gradually and incremental value of customization after CODP is decreasing gradually at the same time. Constraint (11) indicates the safety stock is an integer.

4. Simulation and computation

For further research, we will apply MATLAB software to simulate the two models under postponed production mode and non-postponed production, so as to compare the cost in the two different production modes. By referring to the data in research of Rietze (2006), the parameter value of production cost is as following:
$F(r) = 10 + r$, $m_1(r) = r$, $m_2(r) = 2 - 0.6r$, $D = 50$, $T = 0.018$, $v_1(r) = 0.5r$, $v_2(r) = 0.75r$, $h_1(r) = 0.6r$, $s_1 = 2$, $N = 50$, $w = 0.05$ and $h = 0.6$.

4.1 The simulation expression under non-postponed production

The unit inventory cost is 0.6, unit manufacturing cost and unit holding cost of processing product are all equal to that in $r = 0$ after CODP of postponed production mode. According to the assumption, the probability that the number of customer demand for the certain customized product is more than twice can be ignored, and thus the safety stock of each product under non-postponed production is $s = 1$. The production intension is equal to that in $r = 0$ after CODP of postponed production mode, i.e. $\rho = \frac{\lambda}{\mu} = DT$ and $E[I] = s - DT[1-(DT)^s]$. So the simulation expression under non-postponed production is

$$\begin{aligned}Z &= Dm + DTv + Nh[s - DT + (DT)^{s+1}]\\&= 50\times 2 + 0.9\times 0.75 + 50\times 0.6(1-0.9+0.9^2)\\&= 127.975\end{aligned}$$

4.2 The simulation expression under postponed production

$$\begin{aligned}&Z(r) = 10 + r + 50r + 50(2-0.6r) + 0.9r\times 0.5r +\\&45\times 0.75(1-r) + 0.6r[3 - \frac{0.9r-(0.9r)^4}{1-0.9r}] + \frac{2.25(1-r)}{1-0.9(1-r)}\end{aligned}$$

The simulation result on the compact of CODP location on the objective function is shown in fig.1, and the abscissa represents the position of the CODP, and when the position of CODP is more right means the CODP location is more inclined to the end of the production process. The vertical axis represents the manufacturing cost, and when the position of the CODP is higher, the total production cost is bigger. In the simulation figure, the production cost is at the lowest position at $r = 0.13$.

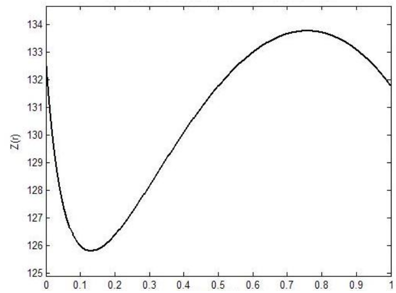


Fig.1 The effects of CODP on total production cost

4.3 Analysis of production cost in two different production modes

(1) Under non-postponed production mode, the total production cost $Z = 127.975$.
(2) Under postponed production mode, the total production cost is 132.5 at $r = 0$, which is bigger than that of non-postponed production mode. As CODP is moving to the end of production process, the production cost is decreasing and it will reach the lowest position at $r = 0.13$, $z = 125.799$, and later, because the rate of cost reduction before CODP is bigger than the rate of cost increasing after CODP, leading to the increasing production cost of the manufacturer. After the lowest point, when the CODP is continuing to move to the end of the production process, the customer waiting cost will continue to reduce gradually and the compact on the total production cost is gradually increasing, so the total production cost will reach a maximum value, and after that, it will decrease.

Reference to the above data, from the perspective of the total cost of the manufacturer, the implementation of the postponed production does not necessarily bring lower cost for manufacturer. If Manufacturer can't choose the right location of CODP, then the postponed production can't bring the cost advantage, or they can't achieve the overall optimization of resources.

4.4 The factor analysis of CODP location

(1) Customer waiting cost

$$Z(r) = 10 + r + 50r + 50(2 - 0.6r) + 0.9r \times 0.5r +$$
$$45 \times 0.75(1-r) + 0.6r[3 - \frac{0.9r - (0.9)^4}{1 - 0.9r}] + \frac{45w(1-r)}{1 - 0.9(1-r)}$$

In the above formula, the optimal position of CODP is $r^* = \arg\min\{Z(r) : 0 \le r \le 1\}$, and we can get the simulation result in table 1.

In table 1, when customer waiting cost increase, the objective function will also increase, but the trend of increase is continuously weakened, while optimal location of CODP is closer to the end of the production process. With customer waiting cost is rising, the compact of customer waiting cost on the total production cost is more obvious. The manufacturer will prefer to moving the CODP to closer to the end of the production process, and thus to shorten customer waiting time and reduce the growth rate of the total cost.

(2) The customized product variety
The simulation expression of product variety N on the optimal location of CODP is

$$Z(r) = 10 + r + 50r + 50(2 - 0.6r) + 0.9r \times 0.5r +$$
$$0.9 \times 0.75 \times N(1-r) + 0.6r[3 - \frac{0.9r - (0.9)^4}{1 - 0.9r}]$$
$$+ \frac{0.045 \times N(1-r)}{1 - 0.9(1-r)}$$

In the formula, the optimal location of CODP is $r^* = \arg\min\{Z(r) : 0 \le r \le 1\}$, and we can obtain the simulation result in table 2.

In table 2, when the product variety is increasing, the objective function will increase, but magnitude of increase in the objective function will decrease gradually, and the optimal position of CODP will be closer to the end of production process with the increasing of product variety. It indicates that when the product variety increase, the manufacture will increase the customized products to meet the more various customer demand, but for the uncertain market, the postponed production can reduce the uncertain risk of demand forecast and control the cost increase within the lower fluctuation range by scale economy of standard product.

5. Conclusion

This paper set the basic model and extended M/M/1 queuing model under non-postponed production and postponed production mode respectively. By solving the model and compare the total production cost, we found CODP location will bring the obvious influence on the production cost of manufacturer, and if the manufacturer can't choose the right CODP location in postponed production mode, then the production cost will exceed the cost of non-postponed production mode. In the simulation of customer waiting cost and product variety on the total production cost and optimal CODP location, we found that when the customer waiting cost and product variety increase, the total production cost will increase too, but the increase ratio will be controlled by moving CODP to the end of production process in postponed production. So when the manufacturers decide to apply postponed production mode, they should find out the right optimal CODP location, and consider the effects of each factors on the CODP location and total production cost, such as customer waiting cost, customized product variety, and so on.

Table1: The influence of customer waiting cost on optimal CODP

w	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
r^*	0	0.0318	0.068	0.1002	0.1302	0.1588	0.1867	0.2143	0.2485
Z	114.50	118.54	121.49	123.84	125.80	127.47	128.93	130.22	131.59

Table 2: The influence of product variety on optimal CODP

N	10	20	30	40	50	60	70	80	90
r^*	0.0269	0.0685	0.0952	0.1147	0.1302	0.1429	0.1537	0.1631	0.1714
Z	114.32	117.61	120.49	123.20	125.80	128.38	130.83	133.30	135.73

References

[1] J. C. P. Su, Y. L. Chang, and Ferguson Mark, "Evaluation of postponement structure to accommodate mass customization" [J], Journal of Operations Management, 23(3/4), 2005, 305-318.

[2] Z. L. Yang, and Y. M. Tian, "A study of partial backorder policy under random demand "[J], Journal of Systems & Management, 18, 6, 2009, 686-691.

[3] X. F. Shao, and J. H. Ji, "Research on impact of postponement strategies on inventory cost in mass customization" [J], System Engineering Theory Methodology Application, 13, 5, 2004, 385-389.

[4] X. J. Xu, and W. L. Zhou, "Simulation Study on Virtual Inventory Decision Making" [J], Industrial Engineering Journal, 12, 5, 2009, 76-81.

[5] Q. Jin, and Z. Zhang, "A visualized e-kanban system for small batch JIT production" [J], Industrial Engineering Journal, 14, 3, 2011, 137-145.

[6] H. L. Lee, and C. Tang, "Modeling the costs and benefits of delayed product differentiation" [J], Management Science, 43, 1, 1997, 40-53.

[7] Z. S. Hua, Flexible manufacturing systems and flexible supply chain—Modeling optimization [M], Beijing: Science Press, 2007.

[8] H. R. I. Van, "Logistics and virtual integration: postponement, outsourcing and the f1ow of information"[J], International Journa1 of Physical Distribution & Logistics Management, 28, 7, 1998, 508-523.

[9] L. Krajewski, J. C. Wei, and L. L.Tang, "Responding to schedule changes in build-to-order supply chains"[J], Journal of Operations Management, 23, 5, 2005, 452-469.

[10] K. Rao, B. Dan, and Y. Liu, "Research on cost optimization model of manufacturer implementation postponement strategy in mass customization"[J], Computer Integrated Manufacturing Systems, 13, 9, 2007, 1665-1671.

[11] Y. Y. Huang, and S. J. Li, "Suitable application situations of different postponements approaches: Standardization vs. Modularization"[J], Journal of Manufacturing Systems, 27, 2, 2008, 111-122.

[12] K. Rao, B. Dan, and Y. Liu, "Cost optimization model of supply chain implementation postponement strategy in mass customization"[J], Computer Integrated Manufacturing Systems, 15, 2, 2009, 287-291.

[13] D. D. Li, and J. X. Wang, "Customer order decoupling point position for mass customization"[J], Journal of Qingdao University (Natural Science Edition), 23, 2, 2010, 89-92.

[14] S. M. Rietze, "Case Studies of Postponement in the Supply Chain"[D], Cambridge: Massachusetts Institute of Technology. 2006.

[15]Mohamed Ramzi Bouzid, Naoufel Kraiem, "An Empirical Evaluation of a Reuse Based Approach for ERP Customization", International Journal of Computer Science Issues, Vol. 9, No. 3, 2012, pp. 17-26.

[16]Kompan Chomsamutr, Somkiat Jongprasithporn, "Optimization Parameters of tool life Model Using the Taguchi Approach and Response Surface Methodology", International Journal of Computer Science Issues, Vol. 9, No. 1, 2012, pp. 120-125.

Yanhong Qin Chongqing Province, China. Work as associate professor in Management School of Chongqing Jiaotong University, and research interests on mass customization and supply chain management.

Danping Xiong Jiangxi Province, China. Graduate student in Management School of Chongqing Jiaotong University, and research interests on mass customization and postponement strategy.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

The Implementation of Hard-Disk Protection Method Based on Disk Filter Driver

Yangwu Liu [1], Xisan Wei[1]

[1] School of Mathematics and Computer Science, Hubei University of Arts and Science
Xinagyan 441053, Hubei Province, China

Abstract

In this paper, it analyzes the working mechanism for the hard disk protection based on the current disk filter driver. The core ideology of the mechanism is to redirect the I/O request packet to the disk operation and to restore data automatically after restart the system in the case of transparent operating system to its upper storey. Also, it specifically analyzes the safety trouble by new virus, like the emerging virus, which can penetrate through the filter driver and results the failure of disk protection, and realizes a method to prevent this penetration. With monitoring the IRP in the kernel layer of operating system, it can identify the requests which have the virus read and write features and prevent it from sending requests to real disk. These methods enhance the protection function of operating system and improve the security of the hard disk protection system.

Keywords: *Filter driver, Hard-disk, Restore, Pass-through, IRP.*

1. Introduction

At present, the hard disk protection (restore) technique is normally making use of Microsoft's system architecture to detect and filter the operations of reading and writing disk in the system so as to achieve the disk data protection.

However, the emerging viruses, such as robot dog which can penetrate the filter drive and lead the failure of disk protection. Thus, it is necessary that the supervision for the IRP in the kernel layer of operating system is enforced and the method is found out using the request with feature reading and writing virus to prevent the request to real disk so as to prevent this kind of penetration and enhance the security of the restoring functions.

2. Principle of Disk Filter Drive

Microsoft's Windows management for the equipment in the kernel is hierarchical, at the same time, the access method to any device for users will be sent to the driver of target device object as final IRP I/O request packet.
The user writes and installs the drive which can be load by the system with certain specifications for the system to load in specific so that the operation request of user's layer for some equipment go through these drivers layer by layer according to certain sequence, each layer of the driver can be selected to modify, download and discard these requests.

In general, a disk filter driver is established making use of a disk filter equipment and attach to disk volume equipment, i.e. Device\Hard-disk\Volumes, so that the system can monitor the reading, writing and redirection for data on the disk.

The advantage of hard drive restore technology lie in its device-independent because filter driver is lie in disk drive and the user layer, it allows us neither consider the disk capacity in current computer system nor consider whether the disk is IDE or SCSI because these work will be completed by lower-level disk drive. The system will change reading and writing operation of the files from the user-level into IRP and deliver to the kernel disk object. And then, the kernel sent IRP one by one according to driver order hanged on the equipment. As it is completed by the driver, we can do slightly complex calculations in the driver to achieve the following performance: the storage area for user is smaller than the protected areas, the storage area can be discontinuous and the validity of the storage area is specified, etc.

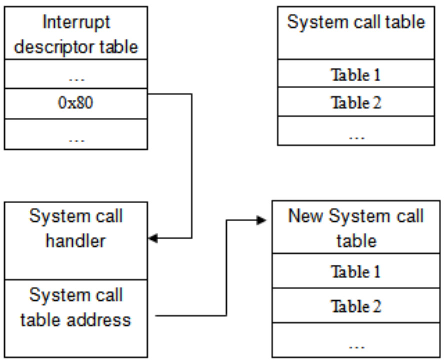


Fig. 1 Redirect the system call table.

3. Design of Hard-Disk Restore

After loading the file system in accordance with the normal operation, the disk drive is loaded in next step, here disk drive is hanged in the event of protect partition. Between loading the file system and the disk driver, the system interrupts the operation and joins its filter driver, and then, disk driver is hanged and the system is start up.
There are two kinds of methods to realize: Backup restore and Mapping restore.

3.1 Backup restore

Backup restore is the most common methods of restore. As shown in Figure 2, assuming that the users want to protect the data of 10 sectors, the data will be backed up to another area, we assume that the region is sector 100, so when the user carries on the operation again, the sector 10 is actually operated. If the sector 10 is destroyed, the data in sector 100 will be to re-copy to a buffer sector 10 in the recovery process.

The general steps of the backup restore can be listed as follows.

Step 1: Judgment whether it needs to be protected.

Step 2: Record the section information which is copied. For example: Establishes an index form record the relations of sector 10 and sector 100.

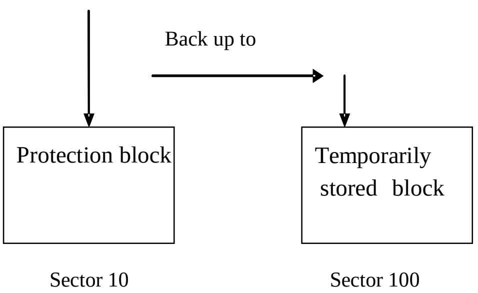


Fig. 2: Backup restore

Advantages of backup restore: simple and safe.
Shortcoming of backup restore: It takes too much space. Because a sector can store 512 bytes, but it should be used 8 bytes when it records the backup relationship of index table, so a lot of disk space will be occupied.
Solution of backup restore: Create a valid data bit map, which can reduce the index table.

3.2 Mapping Restore

Similarly, if the number of sectors which need to be protected is 10, these sectors will be protected. When the user implement the writing operation, Superficially, the user still operates the sector 10, in fact, the data has been written to the other sectors, such as the sector 100, then the user operates sector 10 again, it will be transferred to the sector 100. If the user wants to restore data, it restores directly the data in sector 10.

- A reading operation for user is shown in Figure 4. When the data is loaded, first of all, the user reads the protection bitmap and checks the module. If the module is protected, he will look into the map to search the module of data storage.

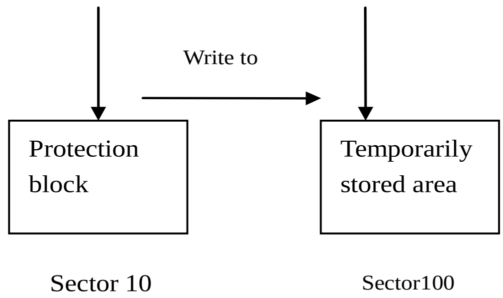


Fig. 3: Mappings restore.

- Meanwhile, in order to prevent the application procedure from accessing directly the physical disk devices, that is, \Device\Hard-disks\DRX, bypassing the filter driver of disk volume, a disk filter drive is commonly re-created and attached to the physical disk device as to filter the reading and writing on real physical disk device.

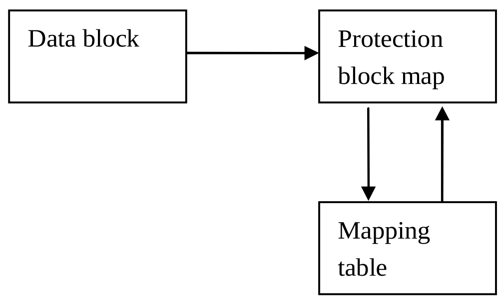


Fig. 4: Read operation.

4. Realization

Microsoft's own restore system is based mainly on the filtering of the file system. By installing a file system filter driver sr. sys with Srrstr.dll, Srsvc.dll and Srclient.dll, filtering of system will monitor the systems and some change of application file. And then, it will use the copy on write technology to automatically create restore points which is easy to be identified. The System Restore will lapse if it skips the file system of operation protection of reading and writing.

Based on the protective system of the disk filter drive just implement filtration treatment for the reading and writing of subarea and disk, it will not truly revise the real data on the physical disk. After restarting the system, all data in buffer is discarded, the system returns the original state.

4.1 Realization process of filter drive

We know that the stack of windows management drive equipment uses the Attached-Device domain in Device-Object. For example, after an IRP of disk request is send to disk equipment of Disk.sys, it will be then forwarded to the port equipment of bus Atapi.sys. There are actual relations: Attached Device domain of Atapi.sys equipment (such as kDevice\Ide\IdeDevicePOTOLO-X) = Disk.sys equipment (such as \Device\ harddisk0\DR0).

However, Windows system does not rely on this domain to decide which equipment to send for the IRP. In most cases, the drive saves the object which is attached to the next level drive equipment in the device extension of device object which is created by itself. For example, classpnp.sys is the driver of processing routines of Device\Harddisk0\DR0, it saves the DeviceObject of kDevice\Ide\IdeDevicePOTOLO-X of lower structure in the equipment Device extension (DeviceObject-> DeviceExtension), and then, the Device-Object is used when which needs to transmit IRP down. Because the DeviceObject is stored in the custom structure, the user won't know the IRP will be sent to which lower equipment without understanding the structure of the detection tool. Only if the pointer of DeviceObject in the structure is replaced, we can easily hijack the IRP to our driver and make filter work in it. This change isn't known by the lower and upper driver.

The I/O manager is an extensive structure; the user can expand the function of I/O subsystem by developing filter driver program. The I/O manager support layered driver model, the process of each I/O request packet (IRP) will go through each layer driver till one of the layered driver responses and completes the request. The I/O manager will create the driver object and transfer to the device driver as a parameter at the entrance of the driver when Windows I/O manager is loaded device driver. The driver will use this pointer which defines the routine and register to the I/O manager, so that the I/O manager can call these routines in proper time. The driver object is stored in a DRIVER_OBJECT structure.

Type	Size
DriverObject	
Flags	
DriverStart	
DriverSize	
DriverSection	
DriverExtension	
DriverName	
HardWareDatabase	
FastIODispatch	
DriverInit	
DriverStartIo	
DriverUnload	
MajorFunction	

Fig. 5: DeviceObject of the structure.

The driver always needs to create device objects to serve for Win32 applications or other drivers. The device object describes physical or logical equipment. And then, the same driver object can create many device objects.

Type	Size
ReferenceCount	
DriverObject	
NextDevice	
AttachedDevice	
CurrentIrp	
Timer	
Flags	
Characteristics	
. . . .	
DeviceExtention	
DeviceType	
StackSize	
. . . . .	
AlignmentRequirment	
....	

Fig. 6: DriverObject of the structure.

We create false DriverObject and DeviceObject (we don't use the system standard function IoCreateDevice, but allocate memory by ExAllocatePool), and then determines the relationship between the upper and lower equipments of disk drive through the AttachedDevice domain. The user fills some important domains to false DeviceObject and maintains consistent with the original lower DeviceObject, such as StackSize. the DriverObject domain will be filled with false DriverObject, at the same time, the MajorFunction of DriverObject filled with the processing function will search the device extension of the upper equipment, find the field position saved the lower device and fill it with false DeviceObject, as a result, Hijack is completed finally. Next, all IRP sent to the bus driver by disk equipment driver will go through our processing function, so long as we care about the content through the MajorFunction filter, for example, implement write operation to protected disk or modify the real number of sectors etc.. And then, we can achieve the disk protection (reduction) operation by calling the processing function of the original lower equipment.

4.2 Maintaining the multithreading and multi-CPU

Firstly, we should consider the multithreading and multi-CPU, there sometimes may be some reading and writing IRP for the disk which arrive to drive interior at the same time. In view of the exclusive events for transferring the memory block, we might perform queuing process to IRP. The system creates driver object, meanwhile to assign the DriverStartIo routine. In addition, IRP_MJ_WRITE and IRP_MJ_READ should be lined up because the similarity for IRP, thus, we can assign both routines as the same one:

```
    ……
    DriverObject->DriverStartIo = DFIoStart;
        DriverObject->MajorFunction[IRP_MJ_READ] = DFReadWrite;
        DriverObject->MajorFunction[IRP_MJ_WRITE] = DFReadWrite;
    ……
```

The system delivers the received IRP to the StartIO formation in the IRP_MJ_WRITE and IRP_MJ_READ routine：

```
    NTSTATUS DFReadWrite(IN PDEVICE_OBJECT DeviceObject, IN PIRP Irp)
      {
    IoMarkIrpPending(Irp);
    IoStartPacket(DeviceObject, Irp, 0, NULL);
    return STATUS_PENDING;
      }
```

In the StartIo routine, the system process IO and apply for the next one with IoStartNextPacket function：

```
    void DFIoStart(IN PDEVICE_OBJECT DeviceObject, IN PIRP Irp)
      {
        ……
    IoStartNextPacket(DeviceObject, TRUE);
        ……
      }.
```

5. File System Filter Pass-Through Technology

At present, the hard disk restore card (Recovery Card) normally uses the disk filter driver to process data recovery and restore. Because it works in the core layer, it is hard to be broke its stability. With more and more virus such as robot dog which can penetrate the protection card, it isn't suitable once but for all the protections. In the future, the lower level such as the monitor of the directly port I/O should be improved.

5.1 The principle of penetrate filter drive

The key realizing the restoring function of restoring software based on filter driver or interrupt-redirection is to intercept I/O operation somewhere during the process of disk I/O operation.

The application program of user mode will call API when it needs to do all kinds of disk operations, and then, API will enter kernel mode and be executed by I/O manager which will create an IRP and deliver it to various kernel mode drivers to process in turn. A device level interruption will be taken place in Kernel mode driver during the process of IRP. The interrupt handler will operate hardware to finish the operation of IRP by calling the related functions in hardware abstraction layer and send the result back to application program of user mode reversely after the operation is finished.

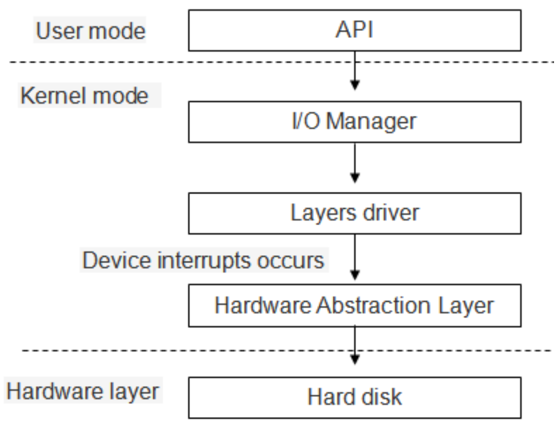


Fig. 7: Disk I/O operations of the basic processes

Due to restoring software adds the software driver in each layer in Figure 8, it must be processed so as to intercept the disk I/O operations before which is executed, As shown in Figure 7, the restoring software based on the interrupt redirection at the position of device level interruption will modify the address of interrupt handler and intercept the disk I/O operation.

In order to penetrate the restoring software, the key of penetrating design is to ensure that the disk operation which user expected won't be intercepted by the two kinds of restoring software before which is delivered to the hardware process. An idea way is that the I/O operation is directly delivered to hardware to process. For example, the robot dog virus is just designed on this way. The I/O operation process loaded virus is shown in Figure 9.

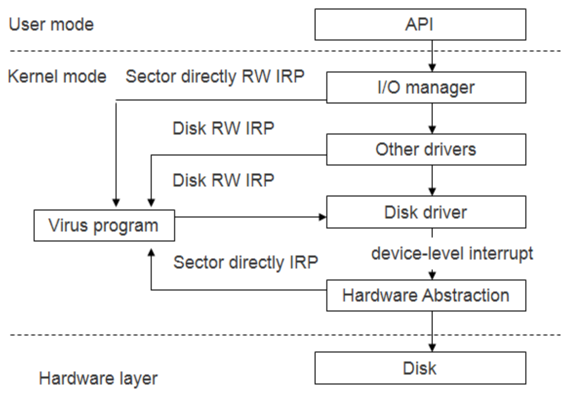


Fig. 8: The disk I/O operations processes in the virus program is loaded

As shown in Figure 8, the read-write IRP in sector is namely I/O operation not to be restored by user. This kind of IRP created by I/O manager will be processed by penetrate software directly, since it bypasses the original driver of all layers in the operating system, it can penetrate the restoring software based on filter driver. At the same time, the penetrate software calls the hardware abstraction layer function during the process of conducting this kind of IRP and interruption in the device level will not happen, so it can penetrate the restoring software based on interrupt-based redirection.

5.2 Implementation of Pass-through

Step 1: Breakthrough the limitation of driver

There are many ways to breakthrough the restriction of read-write driver. Here, we realize it by setting the IOPM mode. Changing IOPM in kernel mode driver needs two unpublished calling functions:Ke386 IoSetAccessProcess and Ke386SetIoAccessMa. The function of Ke386IoSetAccessProcess is to set of IoAccessMap for some process, the function of Ke386SetIoAccessMap is to set a number IOPM for some IoAccessMap.

Step 2: Sector directly RW

The penetrate software can call the functions of hardware abstraction layer to operate the I/O port and realize the read-write of sector directly after breakthrough the restriction for the read-write of Driver in operating system. For read-write operation of disk, you can adopt the custom's I/O control code mode except common read-write operations. In order to realize the read-write in sector directly, three I/O control codes must be defined in the penetrate software. They are IOCTL_DETECT, IOCTL_READ_SECTOR and IOCTL_ WRITE_SECTOR, their functions are getting disk information, reading sector directly and writing sector directly respectively.

Step 3: Intercept the read-write operation of disk

The core intercepting the read-write disk operation is to save and modify the dispatch function's address of disk driver. After modifying the dispatch function's address of the disk driver, the IRP sent to the disk driver originally is still sent to the disk driver, during the accomplishment of the IRP, it will call the modified dispatch function instead of the original dispatch function of the disk driver. The dispatch function's address of main function code related to the read-write disk operation includes IRP_MJ_SHUTDOWN,IRP_MJ_FLUSH_BUFFERS, IRP_MJ_CREATE,IRP_MJ_CLOSE,IRP_MJ_READ, IRP_MJ_WRITE, IRP_MJ_DEVICE_CONTROL and IRP_MJ_INTERNAL_DEVICE_ CONTROL.

Step 4: Distribute dispatch function

The dispatch function of IRP_MJ_SHUTDOWN, IRP_MJ_FLUSH_BUFFERS and IRP_MJ_INTERNAL_DEVICE_CONTROL is ShutFlush; the dispatch function of IRP_MJ_ CREATE and IRP_MJ_CLOSE is CreateClose; the dispatch function of IRP_MJ_READ and IRP_MJ_ WRITE is ReadWrite; the dispatch function of IRP_MJ_DEVICE_CONTROL is DeviceControl.

Initialize the disk reading and writing operation queue Function prototype of initialized queue as follows:

```
VOID   ExInitializeNPagedLookasideList(
    IN PNPAGED_LOOKASIDE_LIST   Lookaside,
    IN   PALLOCATE_FUNCTION            Allocate
OPTIONAL,
    IN PFREE_FUNCTION   Free   OPTIONAL,
    IN ULONG   Flags,
    IN SIZE_T   Size,
    IN ULONG   Tag,
    IN USHORT   Depth
```

);

Among them, Look aside is defined a queue, allocate is a distributed node function, Free is a released node function, Flags and Depth are retention values which are always zero, Size is the size of node, tag is a sign of the queue.

5.3 The implementation of pass-through

1) To be sure that save the changed objects

- Get the disk quantity

We can realize the function through calling the function of IoGetConFigureurationInformation.

- Get the Device object of current disk partition one.

"\ Device, HarddiskX \ Partition0" represents the symbolic links of the first partition of the current disk, the X value ranges from zero to disk quantity. We can get the file object and the device object which linked to this symbol by calling the function of IoGetDeviceObjectPointer.

- Get the current disk partition Numbers

Calling IoBuildDeviceIoControlRequest function and creating an IRP which has a main function code IRP_MJ_DEVICE_ CONTROL and a vice function code DISK_GET_DRIVE_LAYOUT. Then, IoCallDriver function is called to finish this IRP. This IRP will return the current disk partition numbers after it executed successfully. The prototype of IoBuildDeviceIoControlRequest function is as follows:

```
PIRP   IoBuildDeviceIoControlRequest(
    IN ULONG   IoControlCode,
    IN PDEVICE_OBJECT   DeviceObject,
    IN PVOID   InputBuffer   OPTIONAL,
    IN ULONG   InputBufferLength,
    OUT PVOID   OutputBuffer   OPTIONAL,
    IN ULONG   OutputBufferLength,
    IN BOOLEAN   InternalDeviceIoControl,
    IN PKEVENT   Event,
    OUT PIO_STATUS_BLOCK   IoStatusBlock
    );
```

- Verify the effectiveness of the device object

Calling NtOpenFile function and opening a device object next partition of current disk. If it successes, there will prove that the device object is effective, and then close the device object and realize the saving and changing operations. Loop operation will complete the saving and changing operations for all disk partitions.

2) Realize the saving and changing operations

In order to realize these functions, we need to define two data structures for saving the information of disk drive object and disk device object. The data structure is defined as below:

```
typedef struct _major_table
{
  PDRIVER_OBJECT    DrvObj;
  PDRIVER_DISPATCH    ul[28];
struct _major_table *next;
}MajorTable,*PMajorTable;
```

This structure above is used to save driver object and all dispatch functions' addresses. The DrvObj is to save driver object and ul[28] is to save all dispatch functions' addresses which are 28.

```
typedef struct _dev_table
{
struct _major_table * Pmajtab;
  PDEVICE_OBJECT    devObj;
  ULONG    num;
struct _dev_table *next;
}DevTable,*PDevTable;
```

This structure is to save device objects. Pmajtab points to the address saved by the driver object of this device object, devObj is to save device object, num is to indicate the disk number of the device object and partition numbers, the high 16 bits of the num is to save the disk number and the low 16 bits of the num is to save the partition number. All the disk partition information will be linked by the field such as 'next'.

Through above-mentioned data structures, i.e. major table and dev_table, it will save all disk partition information of diver object and device object in the form of linked list. After saving each partition, it will change the dispatch function address of the partition driver object into the function address of the initial distributed dispatch function.

5.4 The queue management of disk R/W operation

Allocate memory in queue. Realize it by calling ExAllocateFromPagedLookasideList function.

Save the information related to IRP. The data structure defined in the penetrate software represents the related information of IRP, the data structure is as follows:

```
typedef struct _context
{
PIO_COMPLETION_ROUTINE oldCplRot;
  PVOID   oldContext;
  ULONG   control;
ULONG    majFun;
} Context,*PContext;
```

OldCplRot is to save the routine finished by the IRP, OldContext is to save the Context field of the IRP, Control is to save the field of Control and majFun is to save the main function code of the IRP.

Set the complete routine for IRP. It can be defined by itself or it can be the dispatch function of the saved disk driver. We can Call IoSetCompletionRoutine function to realize, the prototype of the function is defined as below:

```
VOID   IoSetCompletionRoutine(
    IN PIRP   Irp,
    IN PIO_COMPLETION_ROUTINE CompletionRoutine,
    IN PVOID   Context,
    IN BOOLEAN   InvokeOnSuccess,
    IN BOOLEAN   InvokeOnError,
    IN BOOLEAN   InvokeOnCancel
    );
```

IRP represents the I/O request which need to be processed, CompletionRoutine is the set completion routine, and Context is the related information saved by the last step. When InvokeOnSuccess, InvokeOnError and InvokeOnCancel are true, they represent respectively the success, failure and cancel returned by IRP calling the completion routines.

Release allocated memory in queue. It can be realized by calling ExFreeToPagedLookasideList function. The queue management of read-write disk operation defined four dispatch functions in initialization.

5.5 Sector to R/W

The penetrate software realize read-write operation in sector directly by designing I/O control code and using the mode of driver PIO RW. The penetrate software needs to define an additional I/O control code to enquiry if the current disk driver can be uninstalled except to define the I/O control code of the sector directly RW, IOCTL_DETECT, IOCTL_WRITE_SECTOR and IOCTL_READ_SECTOR. The defined additional I/O control code is IOCTL_ UNLOADQUERY.

The main function of the I/O control code is to enquiry uninstall. It confirms that all disk operation is finished and avoids the loss of disk RW operations by judging the status and times of RW.

5.6 Prevention

After the virus drive loaded successfully, it needs to penetrate reduction and sends the IRP of sector RW to the disk device object directly bypassing the reduction disk filter drive. If the reduction system can modify the information of the related disk device object in the windows kernel before robot virus drive starts, it can hide the true disk device objects' pointers and make all other programs and device objects' pointers point to the pointer of disk filter drive device of reduction system, The IRP sent by robot virus will all go through the reduction disk filter drive.

Modify all information of the bottom disk device objects in system, make the robot dog virus can't get the true information of the bottom device object no matter what layers or what levels. The pointers of the disk device object which the robot dog virus got all point to the pointer of the reduction system disk filter driver's device object, and achieve the purpose to defense the robot virus to penetrate it.

Here, we use the method of Inline Hook IofCallDriver to intercept the IRP sent by robot virus and know what information of the bottom disk device object it wants to access. We modify the information in advance and the robot virus can't get the true information of the bottom disk device objects.

6. TEST EFFECT

The program stays in the ROM chip booted ahead of the system after it was realized, make reduction system can establish a disk volume equipment smoothly and filter files through the disk equipment filter drivers. It is loaded during computer starts and protects documents not to be amended from the kernel layer.

Table 1: Test environment

Item	Version
OS	Microsoft Windows XP (Service Pack 3)
CPU	QuadCore Intel Core 2 Quad Q8200, 2333 MHz
BIOS	AMI 2TKT00AUS 11/19/07
Motherboard ID	64-0100-009999-00101111-050509-Eaglelake$5HKT18AG_08.00.15
Motherboard chipset	Intel Eaglelake G41
IDA Pro	V 4.7
Cbrom	V 2.20

Installing device drivers on Windows XP is extremely simple. The entire process consists of only two steps:

Step 1: Copy the necessary files to the system：
These files include the driver executable image (.sys) file, as well as any other files that are required by the driver.

Step 2: Create the necessary Registry entries:

These entries indicate when the driver is to be started, and also store any driver specific or device specific information that the driver may need during its initialization process. Each driver in Windows XP must have its own key in the Registry, named with the driver's name, under the HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet \Services key.

Testing steps:

Step 1: Load hard disk protection drive on a normal Windows system and execute samples program of robot dog virus. Restart the computer then you can see system has been written Trojan virus by the virus.

Step 2: Write the realized program to the hard disk protection card on another uninfected computer by the virus, start it in advance before the computer starts and load it.

Step 3: Execute the sample program of robot dog virus. Restart the computer.

You will see the system is not infected by the virus when you restart the computer after having executed the sample of robot dog virus. It proves that adding prevention penetrate function to computer is effective. After many times tests, it shows that the method can avoid attacks by robot dog virus and some other type variant virus.

7. SUMMARY AND THE FUTURE WORK

The protection technology of hard disk based on filter driver realizes the restore or backup function by adding restoring software to kernel module of OS in the form of driver. The driver makes all access to disks through the "filter" of restoring software firstly, and then submits them to disk driver of operating system to be processed. This can filter the writing operations to hard disk and achieve the purpose of data reduction. However, the new emerging virus such as "robot dog" can penetrate disk filter drivers and execute read-write operation of hard disk directly. In this paper, we analyzed the methods of pass-through protection and figured out some methods on how to prevent this kind of virus in hard disk protection system. The experimental results proved that our method can prevent the varieties of robot virus and other viruses effectively.

The protection (restore) software needs to conduct data filtering protection in bus level so as to prevent some penetration techniques which aims at hard disk protection. The best way is to set the I/O access ports breakpoints by means of debug register and to intercept reading and writing operations to the ports of hard disk drive controller to fight piercing attack. Monitoring I/O operation of user mode directly is also an available method to prevent the restoring penetration under the ring3. In addition, user mode applications allowed to access I/O is a very dangerous operation; the protection software for hard disk should pay attention to intercept the privilege escalation of progress.

Acknowledgments

This work was supported by the scientific research program No.2010YA024 from Xiangyan University.

References

[1] Peisert S, Bishop M, Karin S, et al. "Analysis of computer intrusions using sequences of function calls" , IEEE Transactions on Dependable and Secure Computing, 2007, Vol. 4, No. 2, pp. 137-150.
[2] Dalton C, Gephardt C and Brown R, "Preventing hypervisor-based rootkits with trusted execution tech-enology", Network Security, No. 11, 2008, pp. 7-12.
[3] Levine John, Gizzard Julian, Bowen, Henry . "Detecting and categorizing kernel-level rootkits to aid future detection" , IEEE Security and Privacy, Vol.4, 2006, pp. 27-32.
[4] Chunxia Zhang, Peng Ji, "Application of WDM device driver in the power-supply controlling system of HIRFL-CSR", Computer Engineering, Vol. 20， No. 2, 2005, pp. 189-191.
[5] Ling Li, "Driver-developing model of virtual position sensor mapped by mouse", Journal of Zhejiang University (Engineering Science), Vol. 40，No. 8，2006，pp. 1344-1347.
[6] Tan X B, Xi H S, "Hidden semi Markov model for anomaly detection" , Applied Mathematics and Computation, Vol. 205, No. 2, pp. 562-567, 2008.
[7] Jinqian Liang, Xiaohong Guan, "A virtual disk environment for providing file system recovery" , Computers & Security, Vol.25, No. 8, pp. 589~599, 2006

Yangwu Liu was born in 1972, is a Chinese living in Xiangyang City, Hubei Province. He graduated from Wuhan University of Technology in 2009 with education in computer science and technology. He is now the director of the Basic Science Lab under the School of Math and Computer of Hubei University of Arts and Science. He has been involved in academic development, scientific research and talents cultivation. He is good at building and maintenance of LAN, maintenance of computer hardware and lab management.

Xisan Wei was born in Xiang yang of Hubei province on February, ninth 1976. He received his engineering degree in computer science and his master degree in computer technology professional from Wuhan University in 2010 . Since 2004 he is Technician at the School of Mathematics and Computer Information Systems and Software Lab, Hubei University of Arts and Science. Currently, he has the qualifications of the project management division.

Design for a 5-DOF Cable-Driven Anthropomorphic Arm

Weihua Su[1], Xudong Ren[2], Fu Niu[2] and Zhiyuan Fu[3]

[1] **School of Mechanical Engineering, Tianjin University, Tianjin, 300072, PR China**

[2] **Institute of Medical Equipment, Academy of Military Medical Sciences, Tianjin, 300161, PR China**

[3] **Research Institute of South Sichuan Gas Field, South-West China Petroleum Branch Co., Luzhou, 646001, PR China**

Abstract

In this paper, a motion control approach for a 5-DOF cable-driven manipulator is designed, and the mechanical structure design of this anthropomorphic-arm is introduced. For the 5-DOF manipulator, a hybrid algorithm is proposed to make the trajectory tracing of the manipulator in task space with high accuracy; the goals of the first phase of the robot arm have been meet. Although the method of motion control is limited in the current state, it serves as a strong foundation on which to test the performance and interface of the electronic components. The coupling cable lengths among the different joint modules are analyzed in detail.

Keywords: *5-DOF, Design, Anthropomorphic Arm.*

1. Introduction

Recently, a research effort has been made over the past few years to model and control flexible structures and, in particular, flexible arms. Many papers on the control of flexible arms with fixed or variable payloads at the tip have been reported. A 5-DOF anthropomorphic-arm is a kind of typical redundant robot [1]. However, the actuators of a conventional robot are all installed on the joints, which not only increase the inertia and weight of robot, but also reduce its load capacity. Meanwhile, this type of design is also unsuitable for high speed motion and rapid response of a robot. Therefore, Cable-driven method brings a new idea to overcome the defect from conventional robot for its motion performance [2-4].

The regional and international competitions began to appear during 1950s that gave people a lot of fun, playing rules, rank are not resolved yet. The game is one that anyone can take participate in, regardless of age, sex, weight or experience, so that arm wrestling robots for entertainment are easily to be popularized. Several arm wrestling robots have been developed in recent years. In the 1990s, Taihei built an arm wrestling robot, whose driving mechanism is electric drive. In 2002, a system that enables players to arm wrestling via a force-position I/O device was built by Kunihiko Kano, their main topic is force testing, but the operation is complex and test results are low accuracy, so it did not have wide application. After two years, Kekuk University in Korea developed an intelligent arm wrestling robot [5] for senior, which is used to help them keep physical and mental health in order to lessen social welfare cost.

Parallel mechanisms have been widely examined in terms of its theory and applications. Its main attractions are its low structural weight, very high rigidity, and very high accuracy in positioning. However, its major drawback is its limited workspace. Cable-driven mechanisms have also been examined in terms of its theory and applications. These mechanisms have very low structure weight because the actuators are not moved and can be located away from the mechanism. The cable-driven concept has been applied to both serial and parallel mechanisms, however, for the cable-driven serial mechanisms, it encounters difficulties in guiding the cables around the joints and while for cable-driven parallel mechanisms, it still encounters the drawback of limited workspace [6, 7].

This paper details the design and development of the arm and hand assembly within the first phase. The arm and hand, henceforth referred to as arm, are designed to meet the following requirements. First, it must have the ability to grasp an object and place it in a different location. Second, it must be similar in scale to that of a human arm and be able to replicate similar motions. The final design should be made with standard components, such that it could be easily reproduced and mirrored to create left and right versions. Finally, the arm should be easily mounted to the mobile base [8-10].

This project involves using an existing head and neck and modifying an existing mobile base. The arm, however, is designed and built from scratch. For this reason, the majority of work on the arm in the first phase revolves around its mechanical design and construction [11].

2. Mechanism Design

A 5-DOF robot prototype, as shown in Fig. l, similar to human arm, the 5-axis motion of the robot includes three horizontal axis and a vertical axis of rotary motion, plus a hundred and eighty degree rotation of the pinch. The robot can carry out self-motion around the shoulder-wrist line; hence it fits for flexible operation under complex environment. To realize the force-closure of the manipulator, the redundant force is obligatory, because cables can only be pulled unilaterally. In order to realize an n-DOF motion, at least n+1 cable is needed as actuating elements [12, 13]. To use six cables to drive sphere joint, this paper proposes an approach to drive sphere joint by four cables.

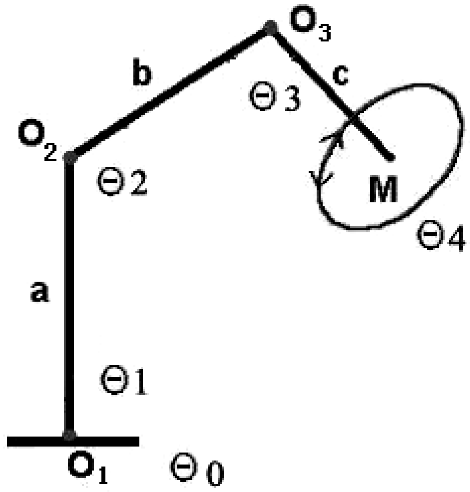


Fig. 1 5-DOF Robot.

By using two cables, the elbow joint realizes 1-DOF rotation. As shown in Fig. 2, to eliminate the motion influence from the elbow joint to the wrist joint, the oriented pulley should be installed on the rotation center of the elbow joint. Meanwhile, the actuating cables should pass along the axial core of the rotary wheel. Thus the cables will only have curving change, whereas the cable length will not be affected by elbow joint.

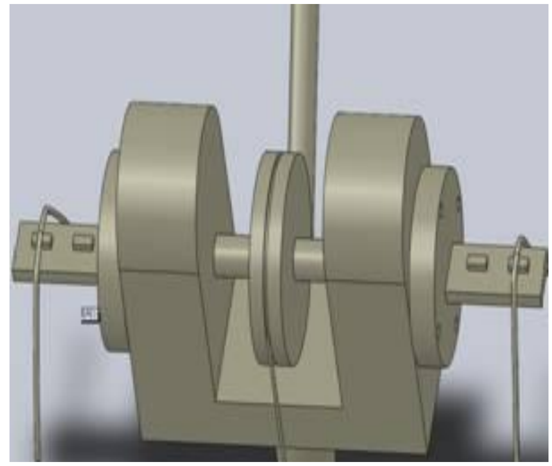


Fig. 2 Elbow joint driven by cables.

3. Solution of Joint Angles

It is difficult to directly solve the cable lengths from the pose of the end-efiactor, thus we propose an approach to find joint angles of the manipulator first, and then to find the cable lengths through analyzing the geometry relationship between cable lengths and joint angles.

According to the workspace analysis, the motion ranges of joint angles q_1, q_2, q_3 have influences to each other; joint angles q_4 and q_5 have such influences also. In order to improve fault-tolerance ability of the manipulator, here we propose an approach to avoid such situation through joint rate redistribution.

Assuming the dimensions of task space and joint space are m and n respectively, here we also define that the number of locked joints for failures is n_f, the number of joints with joint rate saturations is n_s, and the number of joints with normal work ability is n_g respectively. The joint rate redistribution can be included: According to the practical working ability of every joint, individual joint rate can be redistributed to ensure the trajectory of the end-effector being unchanged. Normally the velocity of end-effector is described by the hybrid velocity of a rigid body, as expressed below:

$$V_n^h = \begin{pmatrix} p_n \\ w_n^s \end{pmatrix} \quad (1)$$

Where $P_n' \in R$ represents the instantaneous linear velocity, and w_n^s represents the instantaneous angular velocity, both are viewed in the base frame. So, the relationship between the velocity of the end-effetor and joint rates can be expressed as:

$$V_n^h = j_n^h(q)q' \tag{2}$$

Where $j_n^h = [\begin{matrix} I & p_n \\ 0 & I \end{matrix}]$, $J_n^s \in R^{4\times n}$, J_n^h represents the hybrid Jocobian matrix of the manipulator, $J_n^s \in R^{4\times n}$ is termed as the spatial Jocobian of the manipulator, where matrix $J_n^s(q) = [(\frac{\partial T_{0,n}}{\partial_{q_1}} T_{0,n}{}^{-1})^V \ldots (\frac{\partial T_{0,n}}{\partial_{q_n}} T_{0,n}{}^{-1})^V]$ so that the joint rates can be obtained according to Eq.(2):

$$q' = (J_w^h)^+ V_n^h + K_{mom}(I - (J_n^h)^+ J_n^h)[\nabla H(q)] \tag{3}$$

Where K_{mom} is the coefficient of homogeneous solution, $\nabla H(q)$ is the gradient vector of kinematic performance criterion H(q), $(J_w^h)^+$ is the weighted pseudoinverse of the Jacobian matrix, it can be given by

$$(\tilde{J}_n^h)^+ = w^{-1}(J_n^h)^t (J_n^h w^{-1}(J_n^h)^t)^{-1} \tag{4}$$

Where $W \in R^{n\times n}$ is a symmetric positive-definite matrix, and is referred to as a weighted matrix .Usually it is a diagonal matrix:

$$W=\mathrm{diag}\,[w_1, w_2,\ldots, w_n] \tag{5}$$

with its ith entry given as:

$$W_i = 1 + \left|\frac{\partial H(q)}{\partial q_i}\right|, \text{ i=1, 2, 4….n} \tag{6}$$

From Eq. (4), the contribution from the ith joint to the pose of end-effector is relevant to the ith column of J_n^h only. If the ith joint is locked with rate $q_1^f=0$ because of failure, r the ith joint rate is commanded to run with rate that is higher than its saturation rate q_1^s, in these cases, the contribution difference, $J_i^h(q_1 - q_1^s - q_1^f)$, should be ensure the end-effector to move compensated by other joints to along predetermined trajectory.

If there are f failure joints, and s saturation joints, thus the contribution of n_g normal joints to the end-effector should be revised as following

$$\tilde{V}_n^h = V_n^h - \sum_{j=n_{f1}}^{nf_f + n_{st}} J_j^k q_j'^s \tag{7}$$

Where j=n_{f1}, n_{f2}, n_{f3},…, n_{ff} refers to the index of the ith failure joint, while n_{si} (i=1, 2, 3,…, s) refers to the index of the ith saturation joint.

For calculating simplicity, assuming only the ith joint is locked because of reaching its joint angle limit, thus:

$$\tilde{q}' = [q_1', q_2', q_3', \ldots, q_5'] \tag{8}$$

$$\tilde{J}_n^h = [J_1^h, J_2^h, \ldots, J_5^h] \tag{9}$$

Then we have

$$V_n^h = \tilde{J}_n^h \tilde{q}' \tag{10}$$

$$\tilde{q}' = (\tilde{J}_w)^+ V_n^h + K_{mom}(I - (\tilde{J}_n^h)^+ \tilde{J}_n^h)[\nabla \tilde{H}(q)] \tag{11}$$

Where

$$(\tilde{J}_w^h)^+ = \tilde{W}^{-1}(\tilde{J}_n^h)^T (\tilde{J}_n^h \tilde{W}^{-1}(\tilde{J}_n^h)^T)^{-1} \tag{12}$$

$$\tilde{W} = diag[w_1, w_2, \ldots, w_{i-1}, w_{i+1}, \ldots w_5] \tag{13}$$

Where $I \in R^{n\times n}$ is the identity matrix, $\nabla \tilde{H}(q)$ is the remainder parts of ∇H after its ith entry is extracted.

Consider t_0 as the interpolation period, G_k as the pose of the end-effector for kth interpolation, and q_i (i=1, 2…5) as the individual joint angle. If the pose of the end-effector is G_{k+1} for (k+1)th interpolation, will be changed as:

$$G_{k+1} = e^{\tilde{s}_1 q_1} e^{\tilde{s}_2 q_2} e^{\tilde{s}_3 q_3} e^{\tilde{s}_4 q_4} e^{\tilde{s}_5 q_5} G_k \tag{14}$$

The process to find the inverse solution of the given pose G_{k+1} can be shown as following:

Firstly, solving joint angel q_i for special joint

After finding out the joint rate q_i^k for the kth interpolation through Eq. (3) the joint angle for the(k+1)th interpolation can be obtained shown as following:

$$q_3^{k+1} = q_3^k + q_3^k t_0 \tag{15}$$

Fix the joint q_i to convert the manipulator from redundant to non-redundant to solve other joint angles.

Secondly, solving q_1 and q_2 with Eq. (2).

Since q_i have been solved relating to q_1 and q_2. Take a point P_1 on the axis s_1, we have $e^{s_1 q_1} P_1 = P_1$, then q_1, q_2 can be solved via Eq. (2).

Finally, solving q_4, q_5 by Eq. (1) and Eq. (2).

Since q_1, q_2, q_3 are known, let:

$$G_{k+1}{}' = e^{\tilde{s}_1 q_1} e^{\tilde{s}_2 q_2} e^{\tilde{s}_3 q_3} G_{k+1} G_k^{-1} \tag{16}$$

So that we can rearrange Eq. (14).

$$e^{\tilde{s}_4 q_4} e^{\tilde{s}_5 q_5} = G_{k+1}{}' \tag{17}$$

Take a point p_5 that is on the axis of s_5 but not on the axess s_3, s_4, For $e^{s_5 q_5} P_5 = P_5$ right multiply p_5 on both sides of the Eq. (17), then:

$$e^{\tilde{s}_3 q_3} e^{\tilde{s}_4 q_4} P_5 = G_{k+1}{}' P_5 \tag{18}$$

Take a point P, on the axis s_3, we can get q_3, q_4 via Eq. (2). Further q_5 can be obtained via Eq. (1).

Then, all joint angles, from q_1 to q_5, have been determined. Totally there are eight group possible solutions that come from multiple solutions. Usually, we can choose a group solution with minimum norm of joint rates to make the inverse solution is in the same trajectory group with its initial solution.

4. Coupling influences Among Joints

Because the actuating motor of each joint is installed on the base, every cable that actuates the corresponding joint should reach the base along the anthropomorphic-arm. Hence, the lengths of cables passing the shoulder joint will also be changed when the shoulder joint moves. The coupling influences from shoulder to the wrist and the elbow joint should be compensated to ensure the cable lengths of elbow and wrist correct.

As presented in this paper, the cables drawn from the elbow and wrist joints pass the moving platform P_1, P_2, P_3, P_4 of the shoulder join; P_5, P_6, P_7, P_8, P_9, P_{10} is respectively the connecting points for the six cables to pass moving platform of shoulder joint; B_5, B_6, B_7, B_8, B_9, B_{10} is respectively the connecting points for the six cables to pass fixed platform of shoulder joint. They are all distributed symmetrically.
Since the orientation matrix of the shoulder joint, ${}^A R_p$ can be solved from the paper, so that the cable length change, because of the coupling from the shoulder to the elbow and wrist joint, can be obtained by

$$\Delta l_1' = [\![{}^B B_i - {}^B R_p \cdot {}^P P_i]\!] l_i' \quad \text{(i=5, 6,…, 10)} \tag{19}$$

where ${}^P P_i$ and ${}^B B_i$ is the coordinates of point P_i in the frame P and the point B_i in frame B respectively.

5 .Drive motor and ancillary equipment selection

The low-power servo motor in drive motor will reduce the weight of the drive system. Using the direct control of the micro motor, meet the basic control requirements, at the same time, Design and implementation are relatively simple. Manipulator rod parameters are shown in Fig. 3. In order to facilitate the presentation, deputy campaign numbered from 0-3, Adjacent the spacing of the motor rotary shaft is in turn defined as i0 to i3, in which i3 including the execution unit.

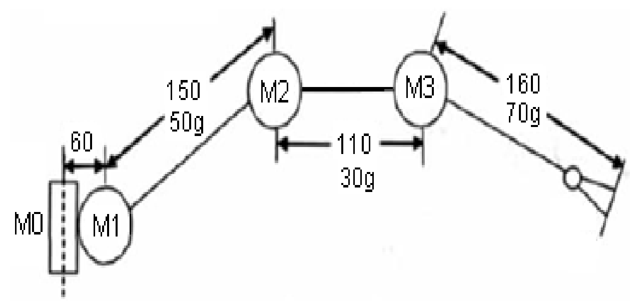


Fig. 3. Schematic manipulator rod parameters

Based on the manipulator rod Parameters, calculating the joint maximum drive torque required, then selecting the appropriate gearbox torque and efficiency. Ultimately determine the appropriate drive motor, and choose the matching code disc.

6. Conclusions

A 5-DOF Cable-Driven Anthropomorphic Arm is designed in this paper, and the mechanism design for this cable driven anthropomorphic-arm is analyzed. It proposes a hybrid inverse kinematic algorithm to combine both gradient project method and Paden-Kahan Subproblem. Therefore the optimization in redundancy solution can be realized with high trajectory tracing accuracy. Overall, the goals of the first phase of the robot arm have been meet. Although the method of motion control is limited in the current state, it serves as a strong foundation on which to test the performance and interface of the electronic components. The forward kinematic equations have been developed and the process has been well documented for future research with this robot.

Acknowledgments

The author acknowledges the support of the natural science foundation of Tianjin, P. R. China (No.10JCZDJC25100).

References

[1] X.J. Liu, J.S. Wang, F. Gao, "Design of a Serial-Parallel 5-DOF Redundant Anthropomorphic Arm", Chinese Journal of Mechanical Engineering, 2011, Vol. 13, pp.101-104.
[2] J.W. Jeong, S.H., Kim, Y.K., Kwak, "Kinematics and Workspace Analysis of a Parallel Wire Mechanism for Measuring a Robot Pose", Mechanism and Machine Theory, 2006, Vol. 34, pp.825-841.
[3] G.L. Yang, W. Lin, S.K. Mustafa, "Kinematic Design of a 5-DOF Cable-Driven Humanoid Arm: a Solution-in-nature Approach", Proc. of IEEE/ASME Inter. Conf. on Advanced Intelligent Mechatronics, 2012, Vol. 2, pp. 444-449.
[4] A.B. Alp, S.K. Agrawal, "Cable suspended robots: Design, planning and control", Proc. of IEEE Conf. on Robotics and Automation, 2012, Vol. 3, pp. 4275-4280.
[5] M. Yamamoto, N. Yanai, A. Mohri, "Inverse Dynamic and Control of Crane-Type Manipulator", Proc. of IEEE/RSJ Inter. Conf. on R. & S., 1999, Vol. 3, pp. 1228-1233.
[6] S.R. Oh, S.K. Agrawal, "Cable-Suspended Planar Parallel Robots with Redundant Cables: Controllers with Positive Cable Tensions", Proc. of IEEE Inter. Conf. on R. & A., 2010, Vol. 3, pp. 3023-3028.
[7] R.V. Dubey, J.A. Euler, S.M. Babcock, "An Efficient Gradient Projection Optimization Scheme for a Seven-Degree-Freedom Redundant Robot with Spherical Wrist", Proc. of IEEE Inter. Conf. on R. &A., 2009, Vol.1, pp. 28-36.
[8] W.H. Chen, I.M. Chen, W.K. Lim, G.L. Yang, "Cartesian Coordinate Control for Redundant Modular Robots", Inter. Conf. on System, Man, and Cybernetics, 2009, Vol. 2, pp. 3253-3258.
[9] L.Y. Li, "On the Control of Redundant Robots, PhD Thesis, Beijing University of Aeronautics and Astronautics", Proc. of IEEE Conf. on Robotics and Automation, 2011, Vol. 3, pp. 23-28.
[10] X.K. Qian, "Mechanism design of BUAA-RR 7-DOF robot", Journal of Beijing University of Aeronautics and Astronautics, 2011, Vol. 2, pp.354-357.
[11] D.H. Zhou, "On the Mechanism of Redundant Robots, PhD Thesis, Beijing University of Aeronautics and Astronautics", Proc. of IEEE Conf. on Robotics and Automation, 2012, Vol. 3, pp. 3023-3028.
[12] N.G. Ullrich, G.D. Lieto, F. Salsedo, "Design and Optimization of a Purely Rotational 3 DOF Haptic Device", IEEE Inter. Workshop on Robot and Human Interactive Communication, 2001, Vol. 1, pp. 100-105.
[13] Tamilselvi, Mercy shalinie, Hariharasudan, “Optimal Path Selection for Mobile Robot Navigation Using Genetic Algorithm”, International Journal of Computer Science Issues, Vol. 8, No. 4, 2011, pp. 433-440.

Weihua Su received his B.S. Degree in 2000 from Shijiazhuang Ordnance Engineering College, and the M.S. Degree in 2003 from Tianjin Institute of Medical Equipment, both in mechanical Engineering. He is currently a Ph.D. candidate in Tianjin University, his research interests focus on the rescue and evacuation robot.

Xudong Ren received his B.S. Degree in 2000 from Shijiazhuang Ordnance Engineering College, and the M.S. Degree in 2003 from Tianjin Institute of Medical Equipment, both in mechanical Engineering. He is currently a researcher of Tianjin Institute of Medical Equipment, his research interests focus on the field medical equipments.

Fu Niu is currently a researcher of Tianjin Institute of Medical Equipment, his research interests focus on the field medical equipments.

Zhiyuan Fu received his B.S. Degree in 1994 from Petroleum University of China, both in mechanical Engineering. His research interests focus on the oil and gas production.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Non-fragile H_∞ Filtering for a Class of Nonlinear Sampled-data System with Long Time-Delay

Shi-gang Wang

School of Mechanical & Electronic Engineering
Heilongjiang University
Harbin, 150080, P. R. China

Abstract

This paper studies the problem of non-fragile filtering for a class of nonlinear sampled-data system with long time-delay. The nonlinearity is assumed to satisfy global Lipschitz conditions, and the filter to be designed includes multiplicative gain variation, which results from filter implementations. Furthermore, to long time-delay, the direct distribution method is applied to design non-fragile filter that assures asymptotic stability and satisfy a prescribed performance index for the filtering error system. The proposed algorithm is given in terms of linear matrix inequality, whose feasibility and effectiveness has been shown by a numerical example.
Keywords: *Nonlinear Sampled-data System, Long Time-delay, Non-fragile H_∞ filtering, Uncertainties.*

1. Introduction

Sampled-data system extensively exists in lots of industrial processes, such as welding industry, aeronautics and astronautics (Tian 2012), chemical industry, etc. (Chen 1995), which is characterized by a continuous control plant and discrete controller. Due to uncertainties and time-delay (Xian 2012) frequently appearing in sampled-data system, which makes the system instable and its performance deteriorated. Therefore, robust control and robust filtering have gradually become hot topics of control field and signal processing (Wu 2001; Wu 2002; Theodor 1994; Xie 1991; Xie 1996; Li 1997). However, above-mentioned results are based on the accurate feedback controllers. In fact, because of the existence of the accuracy problem parameter drift, and other factors, it is shown that relatively small perturbation of the controller parameters might destabilize the closed-loop system, even lead to the performance degradation. Thus, it is necessary to design a controller which can tolerate some level of controller parameter variables. This is known as the non-fragile control problem. To data, this problem of non-fragile control and fitering has been widely investigated by many researchers, (Reinaldo 2001, Wang 2011; Wang 2011, Mahmoud 2004, Yang 2008, Che 2008, Al-Doori 2012).

On the other hand, H_∞ theory has been important effect on stability analysis and state estimation. In recent years, researches has made lots of works in H_∞ theory, such as (Mahmoud 1998, Zhang 2006).

This paper deals with non-fragile H_∞ filtering problem for nonlinear sampled-data systems whose time-delay is longer than a sampling period. The uncertain parameters are assumed to belong to a given norm-bounded type. A methodology for the design of a full-order stable linear filter that assures asymptotic stability and a prescribed H_∞ performance for the filtering error system, irrespective of the uncertainty and long time delay, is developed by solving a set of LMIs.

2. Problem Statement and Preliminaries

Consider the plant of sampled-data system

$$\begin{cases} x(t) = A_0 x(t) + A_1 x(t-\tau) + B_0 \omega(t) + f(x,u,t) \\ y(t) = C_0 x(t) + D_0 \omega(t) \\ z(t) = L_0 x(t) \\ x(t) = x_0 \,,\, t \in [-\tau, 0] \end{cases} \tag{1}$$

where $x(t) \in \mathbf{R}^n$ is the state vector, and $y(t) \in \mathbf{R}^r$ is the measured output, $z(t) \in \mathbf{R}^l$ is the signal to be estimated, $\omega(t) \in \mathbf{R}^p$ is the external disturbance input that belongs to $L_2[0,\infty]$, A_0, A_1, B_0, C_0, D_0 and L_0 are the constant matrices of appropriate dimensions. x_0 is the initial state vector. τ is time delay, which is uncertain and assumed to be evaluated between two adjacent sampling periods, namely, $(m-1)h \le \tau \le mh$, where $m \ge 1$ is a known constant.

Discretizing system (1) in one period, we can obtain the discrete state equation of the sampled-data system

$$\begin{cases} x(k+1) = G_0 x(k) + G_1 x(k-m+1) + \\ \qquad G_2 x(k-m) + H_0 \omega(k) + \bar{f}(x,u,t) \\ y(k) = C_0 x(k) + D_0 \omega(k) \\ z(k) = L_0 x(k) \\ x(k) = x_0, \quad k \le 0 \end{cases} \tag{2}$$

where

$$G_0 = e^{A_0 h},\ G_1 = \int_0^{mh-\tau} e^{A_0 t} dt A_1$$
$$G_2 = \int_{mh-\tau}^{h} e^{A_0 t} dt A_1,\ H_0 = \int_0^h e^{A_0 t} dt B_0$$
$$\bar{f}(x,u,t) = \int_0^h e^{At} dt\ f(x,u,t)$$

Since time-delay τ is uncertain, G_1 and G_2 are uncertain matrices. Let

$$A_0 = L \operatorname{diag}\{\lambda_1, \cdots, \lambda_n\} L^{-1}$$

where L is a $n \times n$ nonsingular matrix, $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of matrix A_0, here, assuming that $\lambda_1, \cdots, \lambda_n$ are unequal to 0, then

$$\begin{aligned} G_1(\tau) &= \bar{G}_1 + DF(\tau)E \\ G_2(\tau) &= \bar{G}_2 - DF(\tau)E \end{aligned} \tag{3}$$

where

$$\bar{G}_1 = -L \operatorname{diag}\{1/\lambda_1, \cdots, 1/\lambda_n\} E$$
$$\bar{G}_2 = L \operatorname{diag}\{e^{\lambda_1 h}/\lambda_1, \cdots, e^{\lambda_n h}/\lambda_n\} E$$
$$D = L \operatorname{diag}\{e^{\lambda_1 \beta_1}/\lambda_1, \cdots, e^{\lambda_n \beta_n}/\lambda_n\},\ E = L^{-1} A_1$$
$$F(\tau) = \operatorname{diag}\{e^{\lambda_1 (mh-\tau-\beta_1)}, \cdots, e^{\lambda_n (mh-\tau-\beta_n)}\}$$

Selection of make $e^{\lambda_i (mh-\tau-\beta_i)} \le I$. Thus it is clear that

$$F^{\mathrm{T}}(\tau) F(\tau) \le I \tag{4}$$

Remark 1 If A_0 can't be transformed into a diagonal matrix or it has $j\ (0 \le j \le n)$ eigenvalues equal to 0, a similar result can be obtained, but $\bar{G}_1$, $\bar{G}_2$, D, $F(\tau)$, E, should be changed correspondingly.

The aim of this section is to design a full-order, linear, time-invariant asymptotically stable filter for system (2), The state-space realization of the filter has the form

$$\begin{cases} \hat{x}(k+1) = A_f \hat{x}(k) + B_f y(k) \\ \hat{z}(k) = C_f \hat{x}(k) \end{cases} \tag{5}$$

where $A_f \in \mathbf{R}^{n \times n}$, $B_f \in \mathbf{R}^{n \times r}$, $C_f \in \mathbf{R}^{l \times n}$ are filter parameters to be determined,

$$\begin{aligned} A_f &= A_{f1}(I + \Delta_1) \\ B_f &= B_{f1}(I + \Delta_2) \\ C_f &= C_{f1}(I + \Delta_3) \end{aligned} \tag{6}$$

and $\Delta_1, \Delta_2, \Delta_3$ represent filter parameter uncertainties to satisfy

$$\begin{bmatrix} \Delta_1 & \Delta_2 & \Delta_3 \end{bmatrix} = DF(\tau) \begin{bmatrix} E_2 & E_3 & E_3 \end{bmatrix} \tag{7}$$

In order to be convenient to solve the following linear matrix inequality, the corresponding linear transformation is used.

letting $\tilde{x}(k) = M \hat{x}(k)$

and then $\hat{x}(k) = M^{-1} \tilde{x}(k)$

filter is transformed to

$$\begin{cases} \tilde{x}(k+1) = M A_f M^{-1} \tilde{x}(k) + M B_f y(k) \\ \tilde{z}(k) = C_f M^{-1} \tilde{x}(k) \end{cases} \tag{8}$$

Denote

$$\xi(k) = \begin{bmatrix} x(k) \\ \tilde{x}(k) \end{bmatrix},\ e(k) = z(k) - \tilde{z}(k)$$

Then, fitering error system

$$\begin{cases} \xi(k+1) = \tilde{G}_0 \xi(k) + \tilde{G}_{m-1} \xi(k-m+1) + \\ \qquad \tilde{G}_m \xi(k-m) + \tilde{H}_0 \omega(k) + \hat{f}(x,u,k) \\ e(k) = z(k) - \tilde{z}(k) = \tilde{L}_0 \xi(k) \end{cases} \tag{9}$$

where

$$\tilde{G}_0 = \begin{bmatrix} G_0 & 0 \\ M B_f C_0 & M A_f M^{-1} \end{bmatrix},\ \tilde{G}_{m-1} = \begin{bmatrix} G_{m-1} & 0 \\ 0 & 0 \end{bmatrix}$$
$$\tilde{G}_m = \begin{bmatrix} G_m & 0 \\ 0 & 0 \end{bmatrix},\ \tilde{H}_0 = \begin{bmatrix} H_0 \\ M B_f D_0 \end{bmatrix}$$
$$\hat{f}(x,u,k) = \begin{bmatrix} \bar{f}(x,u,k) \\ 0 \end{bmatrix},\ \tilde{L}_0 = \begin{bmatrix} L_0 & -C_f M^{-1} \end{bmatrix}$$

The non-fragile H_∞ filtering problem in this section is stated as follows:
Given scalars $\gamma > 0$, find a full-order linear time-invariant, asymptotically stable filter with a state-space realization of the form (8) for system (2), such that,

(1) Filtering error system (9) with $\omega(k) = 0$ is asymptotically stable;
(2) H_∞ performance index satisfies

$$\|e(k)\|_2 < \gamma \|\omega(k)\|_2 \tag{10}$$

is guaranteed under zero-initial conditions for all non-zero $\omega(k) \in \boldsymbol{l}_2[0, +\infty]$。

Assumption 1 $\hat{f}(x,u,k)$ satisfies the quadratic inequality is the domain of continuity, namely

$$\begin{aligned} \hat{f}^T(x,u,k) \hat{f}(x,u,k) \le\ & \xi^T(k) M_{11} \xi(k) + \\ & \xi^T(k-m+1) M_{22} \xi(k-m+1) + \\ & \xi^T(k-m+1) M_{33} \xi(k-m+1) \end{aligned} \tag{11}$$

Lemma 1[18] For given matrices $Q = Q^{\mathrm{T}}$, H and E, with appropriate dimension

$$Q + HF(k)E + E^T F^T(k) H^T < 0$$

holds for all $F(k)$ satisfying $F^T(k)F(k) \le I$ if and only if there exists $\varepsilon > 0$, such that

$$Q + \varepsilon HH^T + \varepsilon^{-1}E^T E < 0$$

3. Main Results

Theorem 1 Consider the filtering error system (9). For given scalar $\gamma > 0$, the system is asymptotically stable and (11) is satisfied under zero-initial conditions for all non-zero $\omega(k) \in \boldsymbol{l}_2[0, +\infty]$, if there exist matrices S, R_i $(i = 1, \cdots, 5)$ such that the following matrix inequalities hold:

$$\begin{bmatrix} \Omega & * & * & * & * & * & * & * & * \\ \varepsilon_1\theta_1^{\mathrm{T}} & -\varepsilon_1 I & * & * & * & * & * & * & * \\ \theta_2 & 0 & -\varepsilon_1 I & * & * & * & * & * & * \\ \varepsilon_2\theta_3^{\mathrm{T}} & 0 & 0 & -\varepsilon_2 I & * & * & * & * & * \\ \theta_4 & 0 & 0 & 0 & -\varepsilon_2 I & * & * & * & * \\ \varepsilon_3\theta_5^{\mathrm{T}} & 0 & 0 & 0 & 0 & -\varepsilon_3 I & * & * & * \\ \theta_6 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_3 I & * & * \\ \varepsilon_4\theta_7^{\mathrm{T}} & 0 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_4 I & * \\ \theta_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_4 I \end{bmatrix} <0 \quad (12)$$

where

$\Omega =$

$$\begin{bmatrix} \Upsilon_1 & * & * & * & * & * & * & * & * & * \\ \Upsilon_1 & \Upsilon_2 & * & * & * & * & * & * & * & * \\ 0 & 0 & \Upsilon_3 & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \Upsilon_4 & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & -\gamma^2 I & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & -I & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & -I & * & * & * \\ SG_0 & SG_0 & SG_{m-1} & SG_m & SH_0 & I & S & -S & * & * \\ \Upsilon_5 & \Upsilon_6 & R_0G_{m-1} & R_0G_m & \Upsilon_7 & I & R_0 & -I & -R_0 & * \\ \Upsilon_8 & L_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -I \end{bmatrix}$$

$\Upsilon_1 = -S + R_1 + R_2 + R_3$

$\Upsilon_2 = -R_0 + R_1 + R_2 + R_3$

$\Upsilon_3 = -R_1 + R_4$

$\Upsilon_4 = -R_2 + R_5$

$\Upsilon_5 = R_0G_0 + \hat{B}_f C_0 + \hat{A}_f$

$\Upsilon_6 = R_0G_0 + \hat{B}_f C_0$

$\Upsilon_7 = R_0H_0 + \hat{B}_f D_0$

$\Upsilon_8 = L_0 - \hat{C}_f$

$\theta_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & D^T S^T & D^T R_0^T & 0 \end{bmatrix}$

$\theta_2 = \begin{bmatrix} 0 & 0 & E_1 & -E_1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

$\theta_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & D^T\hat{B}_f^T & 0 \end{bmatrix}$

$\theta_4 = \begin{bmatrix} E_3C_0 & E_3C_0 & 0 & 0 & E_3D_0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

$\theta_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & D^T\hat{A}_f^T & 0 \end{bmatrix}$

$\theta_6 = \begin{bmatrix} E_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

$\theta_7 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & D^T\hat{C}_f^T \end{bmatrix}$

$\theta_8 = \begin{bmatrix} -E_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

then filter paramenter is solved by

$$A_{f1} = (S-R)^{-1}\hat{A}_f,\ B_{f1} = (S-R)^{-1}\hat{B}_f,\ C_{f1} = \hat{C}_f$$

Proof. Choose a Lyapunov functional candidate to be

$$V(\xi(k)) = \xi^{\mathrm{T}}(k)P_1\xi(k) + \sum_{j=k-m+1}^{k-1} \xi^{\mathrm{T}}(k)Q_1\xi(k) + \sum_{j=k-m}^{k-1} \xi^{\mathrm{T}}(k)Q_2\xi(k) \quad (13)$$

Denote

$\tilde{\xi}(k) =$

$$\begin{bmatrix} \xi^{\mathrm{T}}(k) & \xi^{\mathrm{T}}(k-m+1) & \xi^{\mathrm{T}}(k-m) & \omega^{\mathrm{T}}(k) & \hat{f}^{\mathrm{T}}(x,u,k) \end{bmatrix}^{\mathrm{T}}$$

and then

$\Delta V = V(k+1) - V(k)$

$$= \tilde{\xi}^{\mathrm{T}}(k) \begin{bmatrix} -P_1 + Q_1 + Q_2 & * & * & * & * \\ 0 & -Q_1 & * & * & * \\ 0 & 0 & -Q_2 & * & * \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tilde{\xi}(k) +$$

$$\xi^{\mathrm{T}}(k) \begin{bmatrix} \tilde{G}_0^{\mathrm{T}} \\ \tilde{G}_{m-1}^{\mathrm{T}} \\ \tilde{G}_m^{\mathrm{T}} \\ \tilde{H}_0^{\mathrm{T}} \\ I \end{bmatrix} P_1 \begin{bmatrix} \tilde{G}_0 & \tilde{G}_{m-1} & \tilde{G}_m & \tilde{H}_0 & I \end{bmatrix} \xi(k)$$

By (11) and Schur complements lemma, we can attain

$$\begin{bmatrix} -P_1 + Q_1 + Q_2 + M_{11} & * & * & * & * & * & * \\ 0 & -Q_1 + M_{22} & * & * & * & * & * \\ 0 & 0 & -Q_2 + M_{33} & * & * & * & * \\ 0 & 0 & 0 & -\gamma^2 I & * & * & * \\ 0 & 0 & 0 & 0 & -I & * & * \\ P_1\tilde{G}_0 & P_1\tilde{G}_{m-1} & P_1\tilde{G}_m & P_1\tilde{H}_0 & P_1 & -P_1 & * \\ \tilde{L}_0 & 0 & 0 & 0 & 0 & 0 & -I \end{bmatrix} < 0 \quad (14)$$

In order to solve conveniently, let

$$P_1 = \begin{bmatrix} R & X_{12} \\ X_{12}^{\mathrm{T}} & X_{22} \end{bmatrix}, P_1^{-1} = \begin{bmatrix} S^{-1} & Y_{12} \\ Y_{12}^{\mathrm{T}} & Y_{22} \end{bmatrix}$$

where $R > 0, S > 0$, and assume matrices X_{12} and Y_{12} has full rank

Define

$$T_1 = \begin{bmatrix} S^{-1} & I \\ Y_{12}^{\mathrm{T}} & 0 \end{bmatrix}, T_2 = \begin{bmatrix} I & R \\ 0 & X_{12}^{\mathrm{T}} \end{bmatrix}$$

choose congruent transformation matrix

$$\mathrm{diag}\{T_1, \quad I, \quad I, \quad I, \quad I, \quad T_1, \quad I\}$$

pre- and post- multiplying both sides of inequality (14), one obtains the following inequality,

$$\begin{bmatrix} \phi_1 & * & * & * & * & * & * & * & * & * \\ \phi_2 & \phi_3 & * & * & * & * & * & * & * & * \\ 0 & 0 & \phi_4 & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \phi_5 & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & -\gamma^2 I & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & -I & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & -I & * & * & * \\ G_0 S^{-1} & G_0 & G_{m-1} & G_m & H_0 & S^{-1} & I & -S^{-1} & * & * \\ \phi_6 & \phi_7 & R_0 G_{m-1} & R_0 G_m & \phi_8 & I & R_0 & -I & -R_0 & * \\ \phi_9 & L_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -I \end{bmatrix} < 0 \quad (15)$$

where

$\phi_1 = -S^{-1} + S^{-1} R_1 S^{-1} + S^{-1} R_2 S^{-1} + S^{-1} R_3 S^{-1}$

$\phi_2 = -I + R_1 S^{-1} + R_2 S^{-1} + R_3 S^{-1}$

$\phi_3 = -R_0 + R_1 + R_2 + R_3$

$\phi_4 = -R_1 + R_4$

$\phi_5 = -R_2 + R_5$

$\phi_6 = R_0 G_0 S^{-1} + X_{12} M B_f C_0 S^{-1} + X_{12} M A_f M^{-1} Y_{12}^T$

$\phi_7 = R_0 G_0 + X_{12} M B_f C_0$

$\phi_8 = R_0 H_0 + X_{12} M B_f D_0$

$\phi_9 = L_0 S^{-1} - C_f M^{-1} Y_{12}^T$

choose congruent transformation matrix

$$\mathrm{diag}\{S, \quad I, \quad I, \quad I, \quad I, \quad I, \quad I, \quad S, \quad I, \quad I\}$$

pre- and post- multiplying both sides of inequality (15), the following inequality is obtain by,

$$\begin{bmatrix} \phi_{10} & * & * & * & * & * & * & * & * & * \\ \phi_{10} & \phi_{11} & * & * & * & * & * & * & * & * \\ 0 & 0 & \phi_{12} & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \phi_{13} & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & -\gamma^2 I & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & -I & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & -I & * & * & * \\ SG_0 & SG_0 & SG_{m-1} & SG_m & SH_0 & I & S & -S & * & * \\ \phi_{14} & \phi_{15} & R_0 G_{m-1} & R_0 G_m & \phi_{16} & I & R_0 & -I & -R_0 & * \\ \phi_{17} & L_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -I \end{bmatrix} < 0 \quad (16)$$

where

$\phi_{10} = -S + R_1 + R_2 + R_3$

$\phi_{11} = -R_0 + R_1 + R_2 + R_3$

$\phi_{12} = -R_1 + R_4$

$\phi_{13} = -R_2 + R_5$

$\phi_{14} = R_0 G_0 + (S - R) B_f C_0 + (S - R) A_f$

$\phi_{15} = R_0 G_0 + (S - R) B_f C_0$

$\phi_{16} = R_0 H_0 + (S - R) B_f D_0$

$\phi_{17} = L_0 - C_f$

Separate uncertainties from (15), we can get

$$\Omega + \theta_1 F(\tau)\theta_2 + \theta_2^{\mathrm{T}} F(\tau)\theta_1^{\mathrm{T}} + \theta_3 F(\tau)\theta_4 + \theta_4^{\mathrm{T}} F(\tau)\theta_3^{\mathrm{T}} + \theta_5 F(\tau)\theta_6 + \theta_6^{\mathrm{T}} F(\tau)\theta_5^{\mathrm{T}} + \theta_7 F(\tau)\theta_8 + \theta_8^{\mathrm{T}} F(\tau)\theta_7^{\mathrm{T}} < 0$$

and lemma 1 is applied to get

$$\Omega + \varepsilon_1 \theta_1 \theta_1^{\mathrm{T}} + \varepsilon_1^{-1} \theta_2^{\mathrm{T}} \theta_2 + \varepsilon_2 \theta_3 \theta_3^{\mathrm{T}} + \varepsilon_2^{-1} \theta_4^{\mathrm{T}} \theta_4 + \varepsilon_3 \theta_5 \theta_5^{\mathrm{T}} + \varepsilon_3^{-1} \theta_6^{\mathrm{T}} \theta_6 + \varepsilon_3 \theta_7 \theta_7^{\mathrm{T}} + \varepsilon_3^{-1} \theta_8^{\mathrm{T}} \theta_8 < 0$$

Finally, (11) is attained.

4. Numerical Simulation

We consider the system (1) with

$$A_0 = \begin{bmatrix} 0.5 & -3.5 \\ -1.2 & 0.6 \end{bmatrix}, A_1 = \begin{bmatrix} -1 & 0.5 \\ 0 & 0.1 \end{bmatrix}, B_0 = \begin{bmatrix} -5.5 \\ 1 \end{bmatrix}$$

$$C_0 = [-3 \quad 0.2], D_0 = 0.5, L_0 = [-2 \quad 1]$$

Letting $h = 0.1, m = 2$, and discretizing system (1), a new state equation is attained with corresponding parameter

$$G_0 = \begin{bmatrix} 1.0735 & -0.3724 \\ -0.1277 & 1.0841 \end{bmatrix}, H_0 = \begin{bmatrix} -0.5862 \\ 0.1381 \end{bmatrix}$$

$$G_1 = \begin{bmatrix} -0.1033 & 0.0498 \\ 0.0062 & 0.0073 \end{bmatrix}, G_2 = \begin{bmatrix} 0.1011 & -0.2337 \\ 0.6279 & -0.3188 \end{bmatrix}$$

$$D = \begin{bmatrix} 0.4982 & 0.4277 \\ 0.2847 & -0.2566 \end{bmatrix}, E = \begin{bmatrix} 0.5899 & -0.3933 \\ -0.5689 & 0.1849 \end{bmatrix}$$

Apply MATLAB LMI Toolbox to solve, and then, filter parameter is given, as follows.

$$A_f = \begin{bmatrix} 1.8721 & -1.136 \\ -1.1152 & -0.0325 \end{bmatrix}$$

$$B_f = \begin{bmatrix} 2.2305 & -1.4230 \end{bmatrix}^{\mathrm{T}}$$

$$C_f = \begin{bmatrix} -1.2175 & 0.6215 \end{bmatrix}$$

By numerical experiment, fitering effect of non-fragile H_∞ filter is better than regular filter obviously.

5. Conclusions

The problem of the non-fragile H_∞ filtering for a class of nonlinear sampled-data system with long time-delay has been investigated. Using direct distribution approach, dimension of state variables is deceased, a novel filter is established, and sufficient condition for the H_∞ performance index of the combined system is derived via linear matrix inequality (LMI). Finally, a simulation example is presented to show the validity and advantages of the proposed method.

Acknowledgments

This work is partially supported by Science and Technology Research Foundation of Heilongjiang Education Department (12521437).

References

[1] T. Chen and B. Francis, "Optimal Sampled-data Control System". New York: Springer, 1995.
[2] F. Tian, W. J. Chai, C. Y. Wang, X. P. Sun, " Design and Implementation of Flight Visual Simulation System". International Journal of Computer Science Issues, vol. 9, issue 5, no. 3, Semtemper 2012, pp. 1-7.
[3] J. M. Xing, W. Shao, " Adaptive Iterative Learning Control for a Class of Linear Time-varying Systems". International Journal of Computer Science Issues, vol. 9, issue 5, no. 1, September 2012, pp. 61-66.
[4] J. F. Wu, Q. Wang and S. B. Chen, "Robust Control of a Class of Sampled-data Systems". Control Theory & Applications, vol. 18, no. s1, August 2001, pp. 99-102.
[5] J. F. Wu, Q. Wang and S. B. Chen, "Robust Control of a Class of Sampled-data Systems with Structured Uncertainty". Control and Decision, vol. 17, no. 5, September 2002, pp. 567-570.
[6] Y. Theodor, U. Shaked, and C. E. de souza, "A Game Theory Approach to Robust Discrete-time H_∞ Estimation". IEEE Trans. Signal Processing, vol. 42, no. 6, June 1994, pp. 1486-1495.
[7] L. Xie, C. E. de Souza, and M. Fu, " H_∞ Estimation for Discrete-time Linear Uncertain Systems". Int. J. Robust Nonlinear Contr., vol. 1, no. 2, April 1991, pp. 111-123.
[8] L. Xie, C. E. de Souza, and Y. Wang, " Robust Filtering for a Class of Discrete-time Uncertain Nonlinear Systems: An H_∞ Approach" . Int. J. Robust Nonlinear Contr., vol. 6, no. 4, April 1996, pp. 297- 312.
[9] H. Li and M. Fu, "A Linear Matrix Inequality Approach to Robust H_∞ Filtering". IEEE Trans. Signal Processing, vol. 45, no. 9, September 1997, pp. 2338-2350.
[10] S. G. Wang, J. F. Wu, "Observer-based Non-fragile H_∞ Control for a Class of Uncertain Time-delay Sampled-data Systems". System Engineering and Electronics, vol. 33, no. 6, June 2011, pp. 1352- 1357.
[11] S. G. Wang, "Non-fragile H-infinity Filtering for a Class of Sampled-data System with Long Time-delay". ICIC Express Letters, Part B: Applca- tions, vol. 2, no. 6, December 2011, pp. 1447-1452.
[12] M. S. Mahmoud, "Resilient Linear Filtering of Uncertain Systems". Automatica, vol. 40, no. 10, October 2004, pp. 1797-1802.
[13] G. H. Yang, W. W. Che, "Non- fragile H_∞ Filter Design for Linear Continuous-time Systems". Automatica, vol, 44, no. 11, November 2008, pp. 2849-2856.
[14] W. W. Che, G. H. Yang, "Non- fragile H_∞ Filter Design for Discrete-time Systems with Finite Word Length Consideration". Acta Automatica Sinica, vol. 34, no. 8, August 2008, pp. 886-892.
[15] M. H. Al-Doori, R. Badlishah Ahmad, A. Yahya, M. R. Arshad, " FPGA Design and Implementation of Multi-Filtering Techniques Using Flag-Bit and Flicker Clock". International Journal of Computer Science Issues, vol. 9, issue4, no. 2, July 2012, pp. 39-48.
[16] S. Mahmoud, "Passive Control Synthesis for Uncertain Time-delay System". Proceeding of the 37th IEEE Conference on Decision and Control, Junary,1998, pp. 4139-4143.
[17] P. Zhang, Y. M. Fu, G. R. Duan, "Design of Robust Passive Filter for Linear Uncertain Descriptor Time-delay Systems". Control and Decision, vol. 21, no.11, November 2006, pp. 1275-1279.
[18] B. Barmish, "Necessary and Sufficient Conditions for Quadratic Stabilizability of an Uncertain System", Journal of Optimization Theory and Applications, vol. 46, no. 4, August 1985, pp. 399-408.

Trust Based Infererence Violation Detection Scheme Using Acut Model

Mr. K. Karthikeyan[1], Dr. T. Ravichandran[2]

[1] Research Scholar, Department of Computer Science, Karpagam University, Coimbatore, Tamilnadu-641021, India,

[2] Principal, Hindusthan Institute of Technology, Coimbatore, Tamilnadu-641032, India

Abstract

The inference problem is the major problem in securing the sensitive data in the secured database. In the inference problem the user tries to access the sensitive data from the secured database by giving series of queries with the non-sensitive fields. Inference problems affect the securities in the database and peculiarly in the Multilevel secure databases in which the data and users are differentiated into different levels. Our approach provides the control over the inference problems in the single as well as multi level database. We are provided with a trust based system which results in more efficiency when compare to the other techniques.

Keywords: *Inference Problem, Secured Database, Multilevel database, Trusted Servers*

1. Introduction

From early 80s, the inference problem in the multilevel database had been put forward for the research. It can be described that if the information is secured with the higher level security then user can retrieve the secured data with the data with lower security level, this phenomenon which we are calling as "inference problem". For example, consider an enterprise is maintaining the project details of every staff in a higher level security, but the meeting held and the persons who were attending the meeting is placed in a lower security levels. Then a user can retrieve the projects and the company associated by giving the query with the persons attended the meetings.

The inferred result from the secured database may produce some harmful results if the inferred user knowledge is not authorized to use the secured database, later only the database manager came to know that some inference problem occur in the secured database.

The response taken for the inference problem can be made easy by finding the computation between the accessing feasibility and response time. But there are some complications that we have to identify. In the secured databases there occurs a conflict while

granting a query result for a user, if multi users are allowed to responds for a

particular query then we should made the security alerts for the servers and the queries are analyzed before responding to the queries. This phenomenon is called collision resistance. This process identifies the inference problem but it requires a scheme for inference detection system among the users and their queries.

We are focusing only in the inference problem occurring in the multilevel secured database. In such database the data and users are differentiated with different security levels. Our system authenticates every user before accessing the database, if the users are not authorized means they cannot access the database. This is the first level of security in our approach.

The following approaches are needed to prevent the inference problems, the first approach is to control the unwanted authentications of the users in the secured database, the second approach is to design the database so that with the inference control. The third process is to create a "Trust Model" which acts as the advisor for the secured database. Since the inference problem is very complex and a single layered control does not provide more security effectively, hence we go for this integrated approach to overcome the inference problem.

2. Related Works

The former works posted many researches related to the Inference problems, one of the proposal were key based schema [3]. In this process, the initial process was to generate set of key pairs; the key generation process can be done by using any of the key generation algorithms.

Each key was associated with the objects. The number of key pairs depended on the length of the inference channels. The key pairs are denoted by the letter 'Ks', in this approach two types of key sets are represented in which the first approach is used by the database systems in which the user did not need to hold the keys. In the other approach, every user needs to have the secret key. In the initial phase, all objects in the inference channel are linked with the key or key sets.

If a key is accessed for an object by a query algorithm, other queries have to use the same key to access the object, this process can be achieved by rejecting the associations between the objects and keys. To do so, this process provides the easy access and fast query processing. But different keys cannot be utilized to access the same objects.

Another approach is proposed in the database security technology [5]. A knowledge based inference control machine had been proposed to detect the inference strategies. Users can use this inference controller to monitor for the unauthorized inferences and simultaneously secure the database from violations. To make the inference controller more efficient, it should have knowledge about the various inference strategies. A controller should determine the sensitive data or objects. Finally a controller should detect or prevent the security violation of inference attacks.

A knowledge based inference violation detector is proposed named XINCON (eXpert INference CONtroller) which has the capability to detect unauthorized access of the users. It uses some logical detection and analogical reasoning concepts to find out the inference attacks. The major advantage of XINCON is the user interface components, knowledge management, and truth maintenance system. XINCON uses security constraints to find out the sensitive levels of the data and objects.

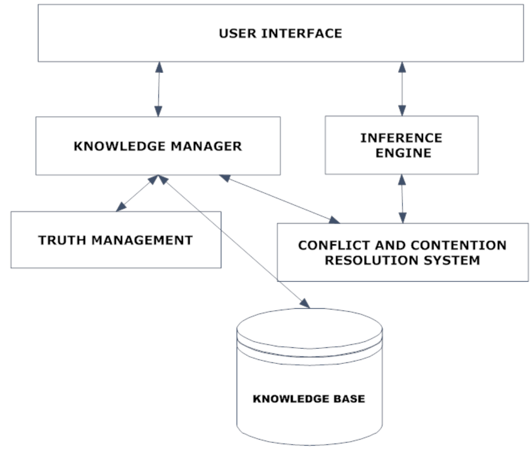


Fig. 1 Represents the XINCON framework

Data and knowledge are distinguished into different security levels. Security constrains plays vital role in assigning the security levels for data and knowledge. We describe the various constrains in the followings;

• Constrains that separate the database, relation or attribute.

• Constrains that separate the part of the database based on the value of some data. These processes are called as content-based constraints.

• Constrains which splits any part of the database based on the real world events. These processes are called as value-based constraints.

• Constrains that classify based on the information based on already processed. These are called aggregate constraints.

In this paper [3] a system with a modern intrusion-tolerant technique had been proposed in distributed database system security model. In the former secured distributed database systems lie on the precautions are limited to the malicious attacks, but there is no action taken for the intrusion tolerance mechanism, isolate access and recovery from the intrusion in the secured database. Our proposed approach maintains the integrity and availability of the data even when the intrusion attacks may occur. In this system a threshold secret share schema is used to ensure the data from the servers with intrusion on it.

Distributed database are vastly used in many industrial applications and research applications, but the security concerns in the distributed database is longer milestones to achieve it. More researches have been put forward for the information security. DDBS have been integrated with many attack overcome techniques like secure audit, message tracing, firewalls and IDS systems. But some time the preventive actions fail to find the malicious attacks in attack methods and frequently attack occurs. So a new approach with Intrusion tolerance has been proposed to find even a single rare intrusion attacks in the database.

This system uses some triggers tolerance mechanism which secures the intrusion and provides the normal service to the users. The intrusion tolerance mechanism is presented with the distributed database security model to provide the distributed database applications with integrity and data availability.

A distributed database is written of databases put in physically classified schemes, co-ordinated by communicating networks, and dealt by disseminated database management system. While conventional database security techniques frequently go wrong to conduct with vicious attacks or intrusions an intrusion-tolerant database system can discover invasions, isolate attacks, assess and repair the damage cause by attacks or intrusions, and keep confidential data safe. The ITDDB model is designed for these goals and has four main subsystems:

• The Proxy Sever subsystem, which will receive and filter users' requests and communicate with other sites in DDBS;

• The Intrusion Detection subsystem, which acts as an intrusion tolerance trigger to the whole system;

• The Assessment and Repair subsystem, which will assess and repair the damage caused by attacks or intrusions in a timely manner;

• The Isolation subsystem, which will isolate suspicious users (transactions) when the intrusion detection subsystem gives an alert.

Knowledge management [4] improves the values by finding the possessions and resources for the management efficiency. But we are lacking in the security for the knowledge management. Every management should protect their important possessions. So authenticated individuals only will be granted to perform operations in the organizations. In this paper the security of the knowledge management will be done based on the confidentiality and trust. But mainly the access control will be made based on the trust management and higher level of privacy will be achieved.

Knowledge management is the combination of many technologies like data mining, multimedia and www. So the knowledge management’s manager cannot be thinking about the security. The possessions they are having higher security risks. Trade secrets are high confidential one, so that the competitors should not access it. So we need an access control mechanism, credential mechanism to secure the intellectual possessions. We have introduced a process which is having higher secured operations. The security of the knowledge management architecture will be build around the intranet. Trust management is the key role in the secure knowledge and it increases the additional security features of the security.

This process states that the trust of the individual will be taken into the account and based on the trust knowledge sharing will be allotted. In our process we are focusing on corporate, so every corporate needs to posses their trust values and based on the trust values knowledge will be shared otherwise negotiation process will be carrying on. In our trust negotiation models enough knowledge management tools will be used efficiently. Trust management is a vital role in the secured technique for knowledge management.

Many recent researches [5] in the multi secure database management system focus on the centralized security systems in the distributed database. Due to the demand of the higher performance and higher availability, database systems are migrated from the centralized approach to the distributed approach. Since there occurs lot of problems in the distributed architectures, like concurrency, security causes major problems in the distributed architectures of the database. So concurrency control has to be integrated with the distributed architecture of the databases. In this approach we propose several concurrency algorithms for the centralized multilevel secure database management systems. To make the distributed architecture more secure we present a virtual model and we analyze the performance of them with concurrency control in a distributed database systems.

In this paper, we propose two security levels: high and low. A main pertain in multilevel surety is information leakage. Leakage can occur in two ways: directly through intelligence such as interpretation an information item or circuitously through a cover channel. Cover channels are ways not normally meant for data flow. In multilevel secure information low security level deal can be retarded or discharged by a high security level deal due to partook data access. Direct leakage can be controlled by the required control schemes but for the covert channels is requiring more modification in the former concurrency technology with two phase locking and time stamp ordering.

Experimental result shows that the result obtains from our system shows the effective result when compare with the former method.

The distributed database architecture [6] is having many successful designs on it, the designs help to improve many criteria like scalability, accessibility and flexibility for many data. Developing such a efficient distributed database requires security issues which is to be solved for the integrity and access control of the systems. Distributed database has some security issues, from the former techniques have proposed many technologies but it is still under research only. These databases are majorly affected by the denial of service and information modifications, etc., some important security requirements for the database management system such as access control, reliability, integrity and recovery should be achieved.

In our process, we provide a user interface has been built for the user request which will be under the control of the distributed Transaction Manager. A customized user interface provides the entry for the user requests and gives the query results to the users. The parameters that we have mentioned as the issues for the security will be providing solutions in the user interface for the query submissions. The transaction manager structures the view of the data before giving to the users and also it validates the output of the query submitted.

We propose a method [7] which is based on the different stages of security process and access control rights to the users. In the process a stages of security for the database have been proposed and different types of security tags also proposed for the objects and data. The major concept behind this is the creation of rules for the tags and tag tables. Access control can be obtained by changing the user's query by the security tag table in distributed database. Different stage security refers that the objects are splited into different groups with the specified security level, every groups has its own security levels and the objects in the similar groups have the same security levels, the data across the objects with multi security levels must meet the security access control policies. In the concept we have proposed Bell-LaPadula model shows the considerable way for the information security system and redefines the system requirements for the data security for the data. This concept forms the basis for every different stage of database security systems.

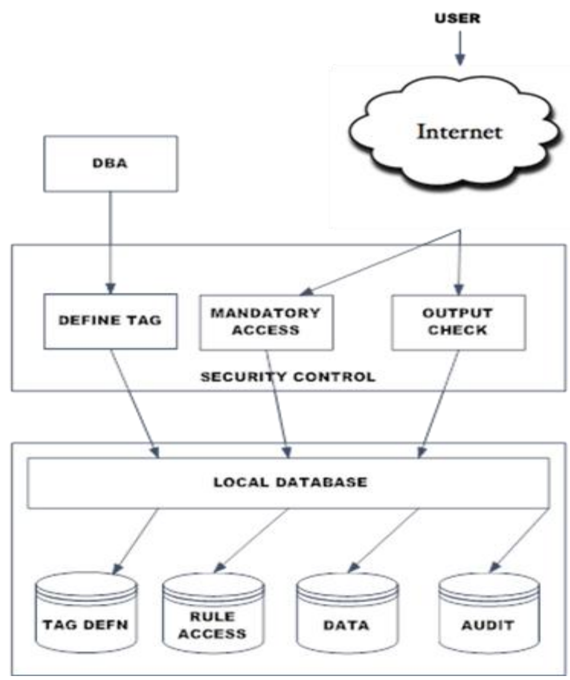


Fig. 2 represents the framework for the multilevel database security

Define Tag: In this module the complete security tag definition and object definition for the database administration, in this process additionally the integrity and consistency checks have been processed.

Access Control: In this module, the security access of the query statements given by the users has been achieved. The process of this module is for confirming the security tags then according to the system security policies it gives the privilege to the users submitted queries from the local and global.

Output Validation: In this module, the query results for the user submitted queries will be validated before giving the result to the users.

3. Proposed System

3.1 ACUT Model

In this model we propose a Access Control Model named **ACUT (ACCESS CONTROL USING TRUST)** which grants the access for accessing the data using trust value calculated from the user logs or history or behavior. In the previous paper we have seen that how a user behavior can be analyzed with additionally we state that based on the behavior we are calculating the trust value. The trust value is the one which decides whether the user will be given access to use the database or their request will be declined initially. The overall diagrammatic flow has been stated as follow;

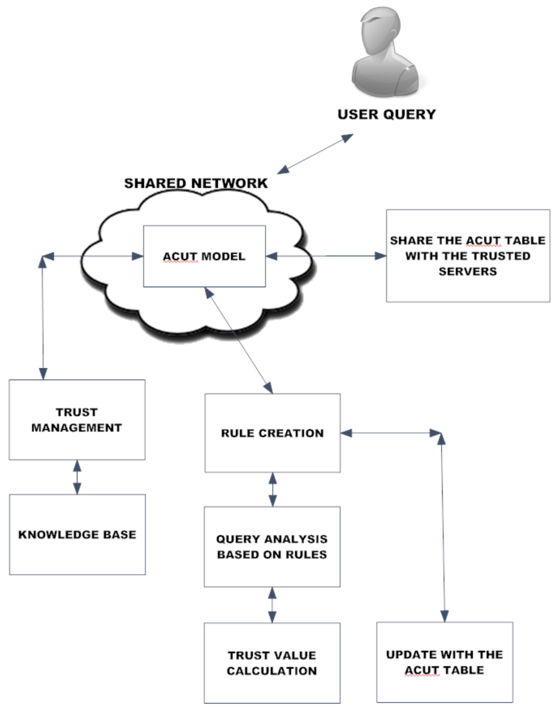


Fig. 3 represents the working procedure of ACUT model

In this model we use ACUT model process, so that we use shared network to update the behavior and trust value of the users to share the knowledge with the trusted servers. In the shared network we are having our ACUT model framework; it is having three processes in it.

- Trust Management
- Rule Creation
- Share the ACUT Table.

3.1.1 Trust Management

In the Trust Management process, we are going to manage the Trust based model with it. The trust based model is having the back end data as the knowledge base data. In the previous paper we have done based on IVDS which has the working procedure based on the prior knowledge threshold limitations.

In our Trust Management process we use the prior history of the user's behavior and previous query access and results. These backlogs help us to manage the trust for every user. The Trust Management block helps the ACUT model to grant the access for every user in the access control policy.

3.1.2 Rule Creation

In the Rule Creation process we have the query analysis phase. The query submitted by the users will be analyzed here. The rules for the data or objects which is to be posed as the query result will be determined here.

The sensitive data will not be posted for the user submitted query but how we are going to achieve that, for that we need to create rules. These rules are having the results as binary values and if the sensitive data are requested from the query means the rules result the '0' value and the query will be denied if the rule result will be '0'.

3.1.3 Share the ACUT Table

In our process we maintain an ACUT table which acts as the backlog for every user. In the ACUT table the user details and their respective trust values. The trust values will be given to the ACUT process for the access provisions.

In the access provision based on the ACUT value of the users, he will be given result for the query, so both the knowledge base and the behavior based will be carrying on in our process.

4. Experimental Result

We made experiments with the number of users and their queries submitted. Based on the query analysis we have analyzed what type of data they try to access from their queries. Based on the query they have submitted we will be assigning the trust value for them.

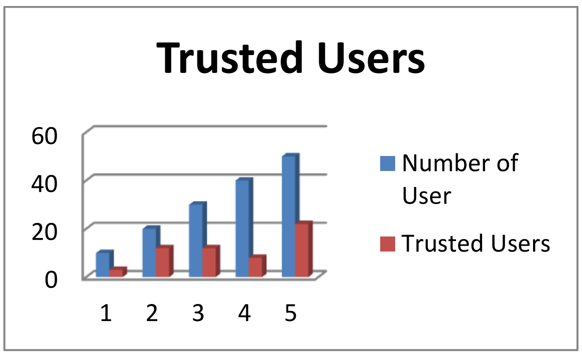


Fig. 4 represents the graph between the number of users and trusted users.

In the graph depicts that we are having the trust management and the trust management analyzes every user based on the prior knowledge and based on the knowledge we come to a conclusion that the trusted users from the whole users who had submitted the queries.

In the second experiment we have analyze the queries submitted by the users. Our query analysis process analyzes the query for inidiviudal users and based on the data they have obtain from their queries we have calculate the query analysis value and their result will be displayed.

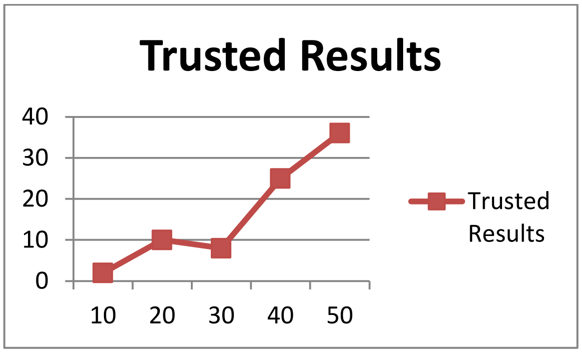


Fig. 5 represents the Trusted result for the queries given by the queries.

The above graph shows how the trust query results are depicted. In the graph we calucate number of queries have been submitted to the query analysis phase and based on the query analysis phase we have calculated the trust values from the query analysis phase.

5. Conclusion

The inference problem is the major problem in securing the sensitive data in the secured database. In the inference problem the user tries to access the sensitive data from the secured database by giving series of queries with the non-sensitive fields. To overcome that we have introduced a process namely ACUT model which analyses the query as well as the user who is giving the queries. Based on the trust values obtain from the backlogs as well as the behaviors we can avoid the inference problems and also the experimental results show that our approach works well for knowledge base as well as behavior based using trust values.

6. Reference

[1] Gu-Ping Zheng and Lu-Feng Xu, “Distributed Database System Security Model of Power Enterprise Based on Intrusion Tolerance Technology” 2006 International Conference on Power System Technology.
[2] K. Pradeep and Z. Mohammad, "DIDMA: A Distributed Intrusion Detection System Using Mobile Agents," in Proc. Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks, SNPDISAWN 2005, Sixth International Conf., pp. 238-245.
[3] “Distributed Database System Security Model of Power Enterprise Based on Intrusion Tolerance Technology “ by Gu-Ping Zheng and Lu-Feng Xu 1-4244-0111-9/06/$20.00c02006 IEEE.
[4] “Secure Knowledge Management: Confidentiality, Trust, and Privacy” by Elisa Bertino, Fellow, IEEE, Latifur R. Khan, Ravi Sandhu, Fellow, IEEE, and Bhavani Thuraisingham, Fellow, IEEE 1083-4427 © 2006 IEEE.
[5] “Performance Evaluation of Secure Concurrency Control Algorithm for Multilevel Secure Distributed Database Systems” by Navdeep Kaur, Rajwinder Singh, A.K.Sarje, Manoj Misra 0-7695-2315-3/07 IEEE.
[6] “On Distributed Database Security Aspects” by Zakaria Suliman Zubi 978-1-4244-3757-3/09/ ©2009 IEEE
[7] “Access Control Method Based on Multi-level Security Tag for Distributed Database System” by Ying-Guang Sun 978-l-61284-088-8/ll ©2011 IEEE
[8] Jiang Wen-bin,Zhang Ren-jin,Zhang Fang-xia.Analysis of Security Strategy on DDBS[J].Computer Knowledge and Technology.2009, vol.5,no.4, pp.769-776.

First Author: **MR.K.KARTHIKEAYAN** is pursuing Ph.D in Database Security. He received the B.Sc degree from Kamaraj University, Tamilnadu, India in 1994, and M.C.A degree from Bharathidasan University, Tamilnadu, India in 1998 and M.Phil from Manonmaniam Sundaranar University, Tamilnadu, India in 2004. He is currently working as Assistant Professor in Department of Computer Application, Anna university of Technology, Madurai. Before Anna University of Technology he was working as a Lecturer in Sengunthar Arts & Science College, Tirunchengodu, Tamilnadu, India.

Second Author: Professor **Dr.T.Ravichandran** received the BE degree from Bharathiar University, Tamilnadu, India and ME degrees from Madurai Kamaraj Universtiy, Tamilnadu, India in 1994 and 1997 respectively, and PhD degree from the Periyar University, Salem, India in 2007. He is currently the **Principal** of Hindustan Institute of Technology, Professor Ravichandran has been a professor and Vice Principal in Vellalar College of Engineering and Technology, Erode, Tamilnadu, India. His research interest includes theory and practical issues of building distributed systems, internet computing and security, mobile computing, performance evaluation, and fault tolerant computing, Professor Ravichandrann is a member of the IEEE, CSI and ISTE. Professor Ravichandran has published more than 80 papers in referred International journals and refereed international conferences proceedings.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Analysis and Implementation Fuzzy Multi-Attribute Decision Making SAW Method for Selection of High Achieving Students in Faculty Level

Widayanti-Deni[1], Oka-Sudana[2] and Arya-Sasmita[3]

[1]Department of Information Technology, Udayana University
Bali, 80361, Indonesia

[2]Department of Information Technology, Udayana University
Bali, 80361, Indonesia

[3]Department of Information Technology, Udayana University
Bali, 80361, Indonesia

Abstract

This paper we provide an overview of the analysis and implementation a method of decision-making system for high achieving students selection. The method used is Fuzzy Multi-Attribute Decision Making (FMADM) Simple Additive weighting (SAW). FMADM SAW itself is a method of decision-making that uses a simple weighting system. The use of this method is expected to help and provide the best decision in the selection of high achieving students in the faculty level.
Keywords: *Criteria, FMADM, SAW, weight, high achieving students.*

1. Introduction

Decision-making in the selection of high achieving students requires data processing time given the number of students are owned by a college and not a little to be completely in accordance with the conditions set. The process of selecting high achieving students must be precise, accurate and quality to be able to achieve the expected outcome is to get students achievement appropriate to their criteria or standards set. But the process of selecting and processing data of participants sometimes still use manual techniques that will require a longer time in both data processing and transmission of results.

Along with current technological developments that can be addressed with the establishment of a system that can help the Faculty of Engineering, Udayana University in the decision-making by using a Decision Support System (DSS) in which this system provides the facility to do an analysis so that each process decisions made based on existing criteria. The method used in this research is to use models of Fuzzy Multi-Attribute Decision Making (FMADM), Simple Additive Weighting Method (SAW). The research is expected to provide convenience and efficiency of data processing in the selection of high achieving students in faculty level.

2. Methodology

Overview of Fuzzy SAW method the selection high achieving students in faculty level can be seen in Fig 1.

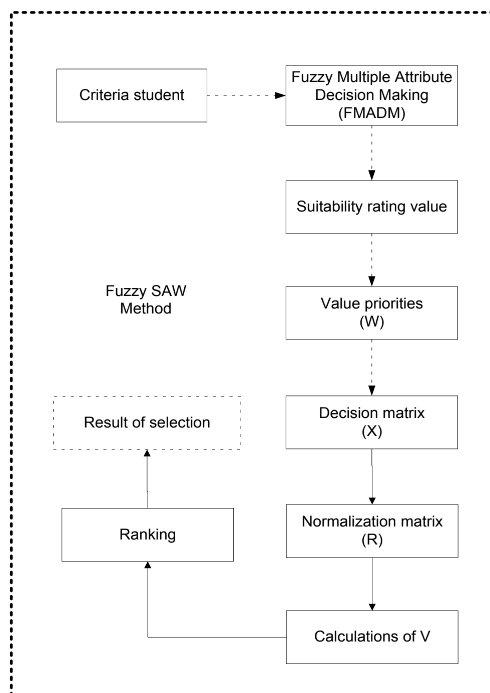


Fig. 1 Overview data flow of the selection of high achieving students using FMADM SAW Method

2.1 Fuzzy Multi-Attribute Decision Making (FMADM)

Fuzzy Multi-Attribute Decision Making is a method used to find the optimal alternative from a number of alternatives to certain criteria. FMADM is the core of determining the value of the weights for each attribute, followed by a ranking process that will select the alternative that has been given. Basically, there are three approaches to find the weights of attributes, namely the approach of subjective, objective approach and the approach to the integration between the subjective and objective. Each approach has advantages and disadvantages. In the subjective approach, the weights are determined based on the subjectivity of decision-makers par, so that some of the factors in ranking the alternatives can be determined independently. While the objective approach, the weights are calculated mathematically that ignoring the subjectivity of the decision makers [4].

There are several methods that can be used to solve the problem FMADM namely [3]:
1. Simple Additive Weighting (SAW)
2. Weighted Product (WP)
3. ELECTRE
4. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)
5. Analytic Hierarchy Process (AHP)

2.2 Simple Additive Weighting (SAW) Method

Churchman and Ackoff (1945) first utilized the SAW method to with a portfolio selection problem. The SAW method is probably the best known and widely used method for multiple attribute decision making MADM. Because of its simplicity, SAW is the most popular method in MADM problems [2].

SAW method also known as the term is often weighted summation method. The basic concept of SAW method is to find a weighted sum of rating the performance of each alternative on all attributes. SAW method requires a process of normalizing the decision matrix (X) to a scale that can be compared with all the rating of the alternatives.

$$\boldsymbol{r_{ij}} = \begin{cases} \dfrac{x_{ij}}{Max_i\, x_{ij}} \\ \\ \dfrac{Min_i x_{ij}}{x_{ij}} \end{cases} \qquad (1)$$

$\frac{x_{ij}}{\text{Max}_i\, x_{ij}}$ is used if the attribute/criteria is benefit

$\frac{\text{Min}_i\, x_{ij}}{x_{ij}}$ is used if the attribute/criteria is cost

Where r_{ij} is the normalized performance ratings of alternatives Ai on attributes C_j, i = 1,2, ..., m and j = 1,2, ..., n. Preference value for each alternative (V_i) is given as:

$$V_i = \sum_{j=1}^{n} w_j r_{ij} \qquad (2)$$

A larger V_i value indicates that the alternative A_i more selected.

2.2.1 Step Completion SAW Method

This research uses a model FMADM SAW method. The steps are [3]:
Step 1: Determining the criteria that will be used as a reference in decision-making, namely C_i.
Step 2: Determine the suitability rating of each alternative on each criterion.
Step 3: Making decisions based on criteria matrix (C_i).
Step 4: Normalized matrix equations based tailored to the type attribute (attribute or attributes benefit costs) to obtain normalized matrix R.
Step 5: The final results obtained from the ranking the sum of normalized R matrix multiplication with the weight vector in order to obtain the greatest value is selected as the best alternative (A_i) as a solution.

3. Requirement Analysis for FMADM SAW Method

The criteria taken into consideration in the selection of high achieving students as in Tables 1.

Table 1: Criteria selection of high achieving students

No	Criteria
1	Grade Point Average / GPA
2	TOEFL Score
3	Number of paper that ever made
4	Number of seminars / workshops have been followed – Department level – Faculty level – University level – National level – International level
5	Number of committees have been followed – Department level – Faculty level – University level

6	Number of award-owned – Department level – Faculty level – University level – District level – Province level – National level – International level

High achieving student criteria listed above each have six criteria where there are some criteria that have high levels of valuation. Each level is the criteria that have to be multiplied first by their respective weights in Table 2 for total points of each of these criteria.

Table 2: Level

Level	Weight
Department	1
Faculty	2
University	3
District	4
Province	5
National	6
International	7

Determine the criteria that will be used as a reference in decision-making, namely C_i. Selection of high achieving students in the faculty level using six criteria listed in Table 3. Selection refers to the ability of students in the academic field as seen from the GPA, English language skills (TOEFL), papers, and awards earned. While the non-academic assessment based on active students in campus activities such as committees and seminars / workshops.

Table 3: Criteria of high achieving students (Ci).

Criteria	Name of criteria
C1	Grade Point Average/GPA
C2	TOEFL Score
C3	Number of paper that ever made
C4	Point of seminars/workshops have been followed
C5	Point of committees have been followed
C6	Point of award-owned

Each criterion will be defined as a priority assessment weighting (W) selection high achieving students. The weight consists of five fuzzy numbers such as Very Low (VL), Low (L), Medium (M), High (H) and Very High (VH).

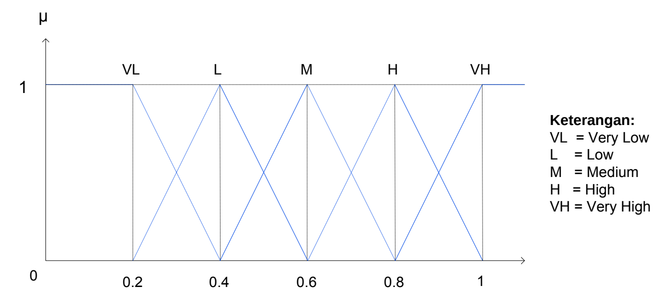


Fig. 2 Overview of fuzzy linguistic variables of the weight for each criterion

The weights of each criterion for the selection of high achieving student are as follows:

Table 4: Priority assessment weighting (W)

Criteria	Fuzzy
C1	Very High
C2	High
C3	Medium
C4	Very Low
C5	Low
C6	High

Criteria C1 and C3 consists of four fuzzy numbers are Very Low (VL), Low (L), Medium (M), and High (H).

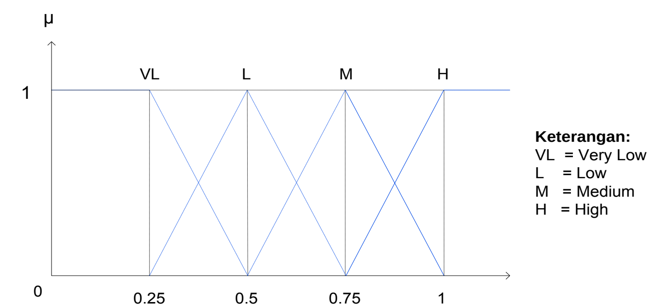


Fig. 3 Overview of fuzzy linguistic variables for criteria GPA (C1) and number of paper (C3)

Table 5: Fuzzy linguistic variables and their correspondent fuzzy number for GPA (C1)

GPA (C1)	Linguistic terms	Linguistic values
GPA < 2,51	Very Low	0.25
GPA 2,51-3,00	Low	0.5
GPA 3,01-3,50	Medium	0.75
GPA > 3,50	High	1

Table 6: Fuzzy linguistic variables and their correspondent fuzzy number for Number of paper (C3)

Number of paper (C3)	Linguistic terms	Linguistic values
Number of paper < 3	Very Low	0,25

Number of paper 3-5	Low	0.5
Number of paper 6-10	Medium	0.75
Number of paper > 10	High	1

Criterion C2 consists of five fuzzy numbers are Very Low (VL), Low (L), Medium (M), High (H) and Very High (VH).

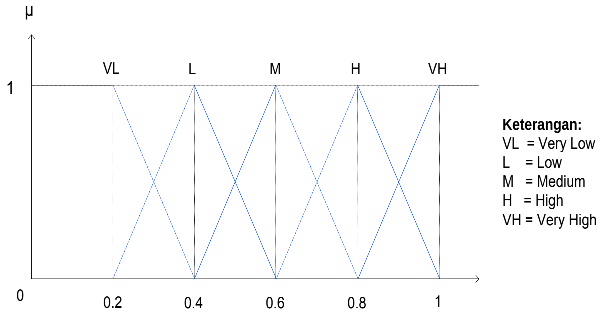


Fig. 4 Overview of fuzzy linguistic variables for criteria TOEFL Score (C2)

Table 7: Fuzzy linguistic variables and their correspondent fuzzy number for TOEFL Score (C2)

TOEFL Score (C2)	Linguistic terms	Linguistic values
TOEFL Score < 311	Very Low	0,2
TOEFL Score 311-420	Low	0.4
TOEFL Score 421-480	Medium	0.6
TOEFL Score 481-520	High	0.8
TOEFL Score > 520	Very High	1

Criterion C4, C5 and C6 consists of three fuzzy numbers are Low (L), Medium (M), and High (H).

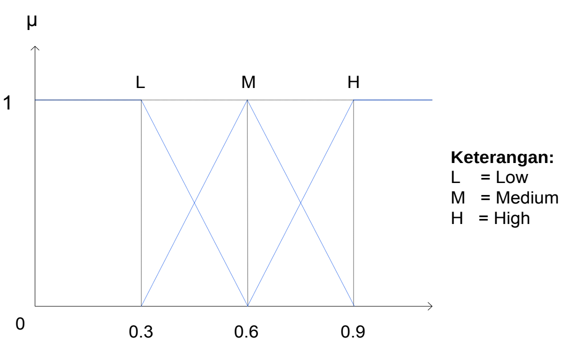


Fig. 5 Overview of fuzzy linguistic variables for criteria point of seminars/workshops (C4), point of committees (C5), and point of award-owned (C6)

Table 8: Fuzzy linguistic variables and their correspondent fuzzy number for Point of seminars/workshops (C4)

Point of seminars/workshops (C4)	Linguistic terms	Linguistic values
Point of seminars/workshops < 31	Low	0.3
Point of seminars/workshops 31-60	Medium	0.6
Point of seminars/workshops > 60	High	0.9

Table 9: Fuzzy linguistic variables and their correspondent fuzzy number for Point of committees (C5)

Point of committees (C5)	Linguistic terms	Linguistic values
Point of committees < 21	Low	0.3
Point of committees 21-40	Medium	0.6
Point of committees > 40	High	0.9

Table 10: Fuzzy linguistic variables and their correspondent fuzzy number for Point of award-owned (C6)

Point of award-owned (C6)	Linguistic terms	Linguistic values
Point of award-owned < 21	Low	0.3
Point of award-owned 21-40	Medium	0.6
Point of award-owned > 40	High	0.9

4. Implementation and Results

Referring to high achieving students data tables described above may be formed suitability rating each alternative on each criterion.

Table 11: Suitability rating

Alternative	Criteria					
	C1	C2	C3	C4	C5	C6
A1	1	0.6	0.5	0.9	0.6	0.9
A2	0.75	0.8	0.75	0.6	0.6	0.3
A3	0.75	0.8	0.5	0.6	0.6	0.3
A4	0.75	0.4	0.5	0.9	0.6	0.3
A5	1	0.4	0.75	0.9	0.6	0.6

Decision matrix (X):

$$X = \begin{bmatrix} 1 & 0.6 & 0.5 & 0.9 & 0.6 & 0.9 \\ 0.75 & 0.8 & 0.75 & 0.6 & 0.6 & 0.3 \\ 0.75 & 0.8 & 0.5 & 0.6 & 0.6 & 0.3 \\ 0.75 & 0.4 & 0.5 & 0.9 & 0.6 & 0.3 \\ 1 & 0.4 & 0.75 & 0.9 & 0.6 & 0.6 \end{bmatrix}$$

Normalized matrix (R):
The example of normalization matrix calculations:

$$r_{11} = \frac{1}{\mathrm{Max}\{1; 0.75; 0.75; 0.75; 1\}} = \frac{1}{1} = 1$$

then the results obtained normalization matrix (R):

$$R = \begin{bmatrix} 1 & 0.75 & 0.67 & 1 & 1 & 1 \\ 0.75 & 1 & 1 & 0.67 & 1 & 0.33 \\ 0.75 & 1 & 0.67 & 0.67 & 1 & 0.33 \\ 0.75 & 0.5 & 0.67 & 1 & 1 & 0.33 \\ 1 & 0.5 & 1 & 1 & 1 & 0.67 \end{bmatrix}$$

Weights:

W =
[Very high; High; Medium; Very low; Low; High]

$$W = [1;\ \ 0.8;\ \ 0.6;\ \ 0.2;\ \ 0.4;\ \ 0.8]$$

The example of V calculations for the ranking:

$$V1 = (1)(1) + (0.8)(0.75) + (0.6)(0.67) + (0.2)(1) + (0.4)(1) + (0.8)(1)$$

$$= 1 + 0.536 + 0.402 + 0.2 + 0.4 + 0.8$$

$$= 3.338$$

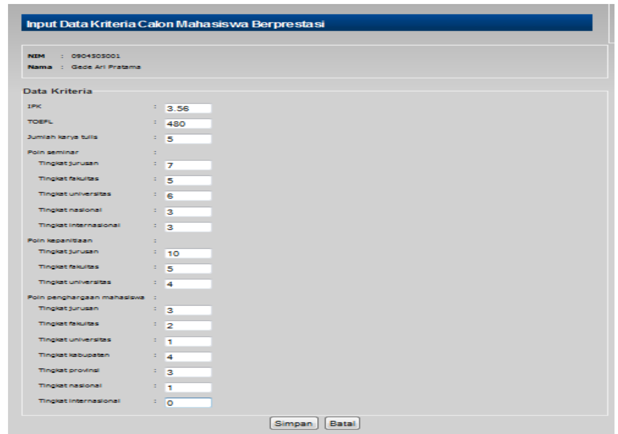
Based on calculations using those calculations, the greatest value is in V1 so that A1 could feasibly be used as the best alternative in selecting of high achieving students. Here are the final results in tabular form.

Table 12: Assessment result

Rank	Alternative	Value
1	A1	3,338
3	A2	2,948
4	A3	2,886
5	A4	2,416
2	A5	3,136

4.1 Implementation in System

In the Fig. 6 is a display system for inputting form high achieving students criteria values where some criteria had levels as described in Table 1.


Input Data Kriteria Calon Mahasiswa Berprestasi

NIM : 0904505001
Nama : Gede Ari Pratama

Data Kriteria

IPK	:	3.56
TOEFL	:	480
Jumlah karya tulis	:	5
Poin seminar	:	
Tingkat jurusan	:	7
Tingkat fakultas	:	5
Tingkat universitas	:	6
Tingkat nasional	:	3
Tingkat internasional	:	3
Poin kepanitiaan	:	
Tingkat jurusan	:	10
Tingkat fakultas	:	5
Tingkat universitas	:	4
Poin penghargaan mahasiswa	:	
Tingkat jurusan	:	3
Tingkat fakultas	:	2
Tingkat universitas	:	1
Tingkat kabupaten	:	4
Tingkat provinsi	:	3
Tingkat nasional	:	1
Tingkat internasional	:	0

Simpan Batal

Fig. 6 Overview input criteria form.

The user fills in the form with the value of each criterion and the system will process these values in accordance with the calculation method FMADM SAW. Here's the view from each of the outcome of the calculation on Fig. 7, Fig. 8, and Fig. 9.

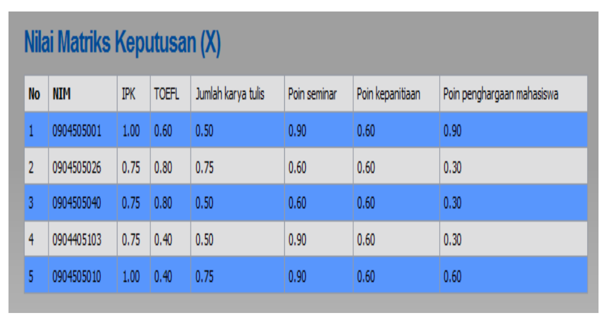

Penilaian Mahasiswa Berprestasi

No	NIM	Nama Lengkap	IPK	TOEFL	Jumlah karya tulis	Poin seminar	Poin kepanitiaan	Poin penghargaan mahasiswa
1	0904505001	Gede Ari Pratama	High	Medium	Low	High	Medium	High
2	0904505026	Putu Maharani	Medium	High	Medium	Medium	Medium	Low
3	0904505040	Surya kencana	Medium	High	Low	Medium	Medium	Low
4	0904405103	Made Bayu Saputra	Medium	Low	Low	High	Medium	Low
5	0904505010	Ketut Dewiyanti	High	Low	Medium	High	Medium	Medium

Nilai Rating Kecocokan

No	NIM	Nama Lengkap	IPK	TOEFL	Jumlah karya tulis	Poin seminar	Poin kepanitiaan	Poin penghargaan mahasiswa
1	0904505001	Gede Ari Pratama	1.00	0.60	0.50	0.90	0.60	0.90
2	0904505026	Putu Maharani	0.75	0.80	0.75	0.60	0.60	0.30
3	0904505040	Surya kencana	0.75	0.80	0.50	0.60	0.60	0.30
4	0904405103	Made Bayu Saputra	0.75	0.40	0.50	0.90	0.60	0.30
5	0904505010	Ketut Dewiyanti	1.00	0.40	0.75	0.90	0.60	0.60

Fig. 7 Overview data of suitability rating value.


Nilai Matriks Keputusan (X)

No	NIM	IPK	TOEFL	Jumlah karya tulis	Poin seminar	Poin kepanitiaan	Poin penghargaan mahasiswa
1	0904505001	1.00	0.60	0.50	0.90	0.60	0.90
2	0904505026	0.75	0.80	0.75	0.60	0.60	0.30
3	0904505040	0.75	0.80	0.50	0.60	0.60	0.30
4	0904405103	0.75	0.40	0.50	0.90	0.60	0.30
5	0904505010	1.00	0.40	0.75	0.90	0.60	0.60

Fig. 8 Overview data of matrix X.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Nilai Matriks Ternormalisasi (R)

No	NIM	IPK	TOEFL	Jumlah karya tulis	Poin seminar	Poin kepanitiaan	Poin penghargaan mahasiswa
1	0904505001	1.00	0.75	0.67	1.00	1.00	1.00
2	0904505026	0.75	1.00	1.00	0.67	1.00	0.33
3	0904505040	0.75	1.00	0.67	0.67	1.00	0.33
4	0904405103	0.75	0.50	0.67	1.00	1.00	0.33
5	0904505010	1.00	0.50	1.00	1.00	1.00	0.67

Fig. 9 Overview data of matrix R.

After processing system to search each value V will then be shown the results of the value of V for each of the attributes that have been sorted by the largest value of V as shown in Fig. 9. Thus it can be determined that students with student ID (NIM) 0904505001 selected as high achieving students.

Hasil Perankingan Pemilihan Mahasiswa Berprestasi

Rank	NIM	Nama	Nilai
1	0904505001	Gede Ari Pratama	3.4
2	0904505010	Ketut Dewiyanti	3.13333333333
3	0904505026	Putu Maharani	2.95
4	0904505040	Surya kencana	2.75
5	0904405103	Made Bayu Saputra	2.41666666667

Fig. 10 Overview of assessment result selection of high achieving students.

5. CONCLUSION

Based on the research that has been done, it can be concluded that the FMADM SAW method can be used in the selection process of high achieving students. The selection results obtained in the form of ranking the final value of the participant. Although using a simple weighting calculations, FMADM SAW method can provide the best decision in the decision process.

Acknowledgments

Our thank goes to Department of Information Technology Udayana University, Bali, who has helped organize this research.

References

[1] Dikti, "Pedoman Mahasiswa Berprestasi Program Sarjana", Jakarta, 2011

[2] Huang. Jeng. Jih, "Multiple Attribute Decision Making Methods and Applications", Chapman and Hall/CRC, 2011

[3] Kusumadewi. Sri, Hartati. S, Harjoko. A, and Wardoyo. R, "Fuzzy Multi-Attribute Decision Making (FUZZY MADM)", Yogyakarta: Graha Ilmu, 2006

[4] Kusumadewi. Sri, "Diktat Kuliah Kecerdasan Buatan, Jurusan Teknik Informatika", Fakultas Teknologi Industri Universitas Islam Indonesia, 2007

[5] S. Wibowo. Henry, Amalia Riska, M. Fadlun. Andi, and Arivanty. Kurniawan, "Sistem Pendukung Keputusan Untuk Menentukan Penerima Beasiswa Bank BRI Menggunakan FMADM (Studi Kasus: Mahasiswa Fakultas Teknologi Industry Universitas Islam Indonesia)", National Conference: Seminar Nasional Aplikasi Teknologi Informasi (SNATI), 2009

[6] Modarres, M., Sadi-Nezhad. "Fuzzy Simple Additive Weighting Method By Preference Ratio", Intelligent Automation and Soft Computing, Vol. 11, No. 4, 2005, pp. 235-244

[7] Rumaisa. Fitrah and Nurafianti. Tanti, "Sistem Pendukung Keputusan Kelulusan Ujian Saringan Masuk Jalur PMDK Berdasarkan Nilai Rata-Rata Matematika dan Bahasa Inggris", National Conference: Design and Application of Technology, 2010

[8] Putra. Apriansyah and Hardiyanti. Yunika. Dinna, "Penentuan Penerima Beasiswa dengan Menggunakan Fuzzy MADM". National Conference: Seminar Nasional Informatika (semnasIF). 2011

AUTHORS PROFILE

Made Widayanti Deni is a student at Faculty of Engineering, Department of Information Technology, Udayana University. Her research interest on decision support system for selection of high achieving students to get her bachelor degree.

A.A Oka Kompiang Sudana. S.Kom, MT received his S.Kom degree in Informatics Engineering from Institut Teknologi Sepuluh Nopember University in 1997 and his MT. degree in Informatics and Computer System from Gajah Mada University in 2001. He was Technical Manager in PT. INFOS Teknologi Indonesia (Software Developer) during April 2008-Sept 2008, Information Technology Leader-Human Resources and General Affair Division at PT. JAS Catering International Airport Ngurah Rai Bali during April 2003-July 2006, Person in Charge of Technological and Professional Skills Development Sector Project (TPSDP)-Asian Development Bank (ADB) Loan, Batch II in Electrical Engineering Study Program during 2002-2006, and now he is lecturer at Magisterial Program of Electrical Engineering Department of Udayana University, lecturer at Electrical Engineering Department (major in Computer System and Informatics) of Udayana University, lecturer at Information Technology Department of Udayana University, and member of Development Project Team of Academic Management Information System and Networking Implementation of Udayana University. His research experiences are in Analysis and Design of Information System and Biometric Identification and Recognition.

I Gusti Made Arya Sasmita, ST, MT received his ST. degree in Electrical Engineering from Udayana University in 1997 and received his MT. degree from Gajah Mada University in 2003. Now he is lecturer at Electrical Engineering Department of Udayana University and Information Technology Department of Udayana University.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

The System of Demand Analysis for URAV Based on Offense and Defense

Guodong Jin[1,2], Libin Lu[1], Liangxian Gu[2], Juan Liang[1], Xiaofei Zhu[1]

**[1] Xi'an research institute of high technology, Faculty 907
Xian 710025, China**

**[2] College of Astronautics, Northwestern polytechnic University
Xian 710071, China**

Abstract

Because of its advantage of zero casualties Unmanned Reconnaissance Aerial Vehicle (URAV) plays an important role in battlefield monitoring and information acquiring, thus caught great attention of the world. URAV is developed fast in our country, yet no scientific assessment methods was ever given due to different fight needs of armed forces. Considering demand of the missile artillery about the martial information, the model of information requirement of combat force, reconnaissance ability of URAV, survivability of URAV, and task reliability of URAV were constructed respectively. Synthesizing mathematic models above, the model of developing demand was constructed about URAV equipment. It simulated and calculated some URAV equipment developing scales, and explores a way of settling the problem of URAV equipment developing demand.

Keywords: *URAV; demand; survivability; reliability*

1. Introduction

At the present time, URAV is developed greatly in our country, but because all services and arms are different in combat requirement, there is no scientific way to evaluate the developing scope of URAV and troops. This paper aims at the Second Artillery combat requirement, and tries to find a new way to solve four urgent problems in URAV developing.

(1) How to improve the URAV operational efficiency?

(2) How to improve the URAV survivability?

(3) How to improve the URAV task reliability?

(4) How to rationally define the URAV developing scope?

2. Headings and Footnotes Construct the model of developing scale of URAV

The model of developing scale is as follows:

(1) Information of the missile artillery in wartime and reconnaissance ability of URAV is quantified. The paper has realized the unification of the quantification by building connections between target numbers and reconnaissance ability of UAV, with the result of laying the foundations for analyzing and calculating the number of UAV.

(2) The model of reconnaissance ability of URAV is constructed. The paper points out some methods to improve URAV's penetration ability by supplementing and innovating analytical methods of military airplane survival ability.

(3) The model of reconnaissance ability of URAV is constructed

Synthesize mathematic models above, the model of developing scale of URAV is constructed. Figure 1 shows a structure diagram of the model of developing scale of URAV.

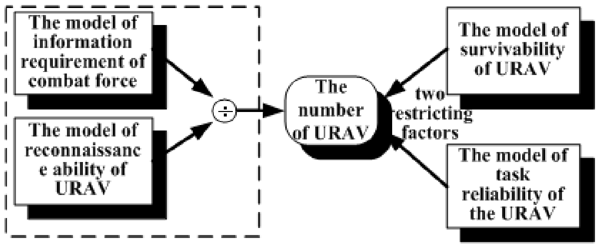


Figure 1 structure diagram of the model of developing scale of URAV

We get the number of URAV needed of one campaign unit in one campaign day:

$$N = \frac{M}{M' \times P_{S/E} \times P_R} \tag{1}$$

"M" is the number of targets which could be attacked by a campaign unit in a day, "M'" is the number of targets that the URAV can scout, "$P_{S/E}$" is the survival probability of

URAV, and " P_R " is the task reliability probability of URAV.

Because of its destruction when breaking through the enemy defense and the reliability of URAV system, new supply is needed every day. Based on the tasks of the Second Artillery Force in the battle, the missile attack time is estimated "T", and campaign units is supposed "K", so the loss of the URAV during one battle is:

$$K \times T \times (2 - P_{S/E} - P_R)$$

The number of the URAV needed is:

$$N' = K \times N + K \times T \times (2 - P_{S/E} - P_R) \qquad (2)$$

3. The model of method and blue print about subsystem

3.1 Construct the model of information requirement of combat force

The core of the model is quantifying information requirement for the missile artillery in wartime. The missile artillery needs information about the characteristic of targets and effects of striking in wartime. How to quantify and calculate the information in mathematical model? The paper connects the information with targets and quantifies them to target numbers in wartime. Thus, it could complete the quantifying calculation of model.

The model of information requirement is as follows:

(1) Take a fighting unit as a template, analyze the average number of the missiles can be launched in a day and the number of the missiles which can break the defense and attack targets , then let the number be "m".

(2) Analyze the type of targets and the number of missiles to destroy every type of targets. Define the number of type of targets to be "i", and the number of missiles to destroy the target " n_i ".

(3) Analyze the method of attack .As the launch of missile has two modes: salvo and dartling, the difference of attacked effects and penetration and the number of missiles to expend is analyzed .Correct the necessary number of missiles to destroy the target and define it as " n_i' ".

As is analyzed above, information demand of the missile artillery in wartime can be quantified by the number of targets which could be attacked by a campaign unit in a day .For the convenience, suppose every campaign unit can attack one kind of target in a day:

$$M = \frac{m}{n'_i} \qquad (3)$$

3.2 Construct the model of reconnaissance ability of URAV

The core of the model is quantifying reconnaissance ability of URAV. In this paper, according to the characteristics of URAV of the missile artillery, the reconnaissance ability of URAV is quantified by the number of targets which can be scouted in a day.

The model is as follows:

(1) Define the number of targets which can be scouted on the condition of a safe out and home course to be "l".

(2) Reckon the number of sorties which each flight could take in a day.

It is the mid-altitude long-range of URAV that is discussed in the paper .To reckon the number of targets which can be scouted in each sortie, time of an out and home course (" t_1 ") , time spent on scouting after arrive at the termini (" t_2 "), the prepare time from the launch position(" t_3 "), time spent on withdrawing the equipment from launch position(" t_4 "), time of next launch after normally callback (" t_5 ") , and time of time-lapsed maintaining (" t_6 ")——the average time for each flight, are required.

Then the sorties receivable in a day is:

$$n = \frac{t}{t_1 + t_2 + t_3 + t_4 + t_5 + t_6}$$

In this formula, " n" refers to the sorties per day, and " t" means the overall time in a day.

Since URAV needs to scout the target in advance and evaluate the effects of attack, there have to be at least two flights to make sure that the mission can be fulfilled. So the number of targets that the URAV can scout should be:

$$M' = \frac{l \times n}{2} \qquad (4)$$

So the paper has realized the unification of the quantification by building connections between target numbers and reconnaissance ability of URAV, with the

ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

result of laying the foundations for analyzing and calculating the number of URAV.

In order to get the information about the number of targets that troops attack in a day, the number of URAV required to fly to the targets can be known due to formula (3) and (4):

$$N = \frac{M}{M'} \qquad (5)$$

3.3 Construct the model of survivability of URAV

3.3.1 Define scheme of model

URAV and manned planes are similar in appearance and threats they may face in a battle. In the paper, the model survival ability of URAV is constructed by using the approaches of analyzing the survival ability of military airplane and also making some necessary complements according to the characteristics of the URAV. Figure 2 shows a structure diagram of the model of survivability of URAV.

" P_K " is the destroying probability of URAV. It is made up of allergy (P_H) and vulnerability ($P_{K/H}$). P_H multiplied by $P_{K/H}$ is P_K :

$$P_K = P_H \cdot P_{K/H} = P_d \cdot P_{H/d} \cdot P_{K/H}$$

" P_d " is the probability of detecting a URAV by radar; $P_{H/d}$ is the hitting probability?

The survival probability of URAV is showed :

$$P_{S/E} = 1 - P_K$$

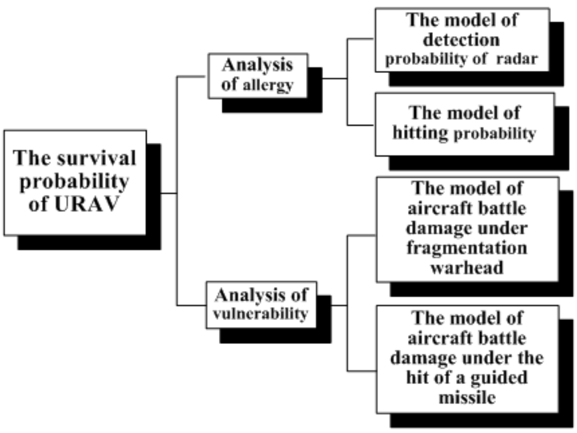


Figure 2 structure diagram of the model of survivability of URAV

3.3.2 Algorithm of the survival probability of URAV

Calculating steps about the survival probability of URAV against air defense firepower are:

(1) According to URAV and early warning radar's technical parameter, the utmost distance of detecting a URAV by radar is calculated;

(2) According to flight speed of URAV, the time that URAV spend on fling over enemy territory of air defense is calculated;

(3) According to the performance of enemy anti-aircraft weapons, the strength of enemy air defense firepower is evaluated;

(4) The $P_{H/d}$ and $P_{K/H}$ is calculated about different anti-aircraft weapons;

(5) The survival probability of URAV is calculated by the formula of $P_{S/E} = 1 - P_K$.

3.3.3 Simulative calculation and analysis

(1)When RCS is reduced, detecting distance will be reduced obviously. It is showed on table 1;

Table 1 H=5000m, different RCS corresponding radar's detecting distance

RCS (m^2)	1.0	0.5	0.1	0.01
detecting distance (km)	202	103	56	34

(2)RCS=1m^2,When the URAV flights on the different high, radar's detecting distance is showed on table 2.

Table 2 RCS=1m2, radar's detecting distance on the different high

H(m)	5000	1000	500	100
detecting distance (km)	200	144	87	75

According to setting URAV and early warning radar's technical parameter**,** the survival probability of URAV is simulated. The result is showed on Figure 3**.**

To advance penetration capability about URAV, Some conclusion can be acquired by analysis and calculation. It is as follows:

(1) It is important for RCS to influence detection probability of radar. If RCS is reduced effectively , detection probability of radar can be reduced. The way above markedly advances survivability of URAV, but it cost much.

(2) The mobile low altitude flight can advance survivability of URAV. When URAV flight on flat landform, some way are adopted, such as low altitude flight, continual changing flight path and pose. Clutter noise can keep out detecting by radar on the land or sea. When URAV flight on a mountainous area, the mobile flight is adopted to control URAV's attitude and pose by hand.

(3) To advance survivability of URAV, some measure such as reducing the area of vulnerable parts and standby important parts can be done, which will make weight of UAV increased, so it is given an overall consideration.

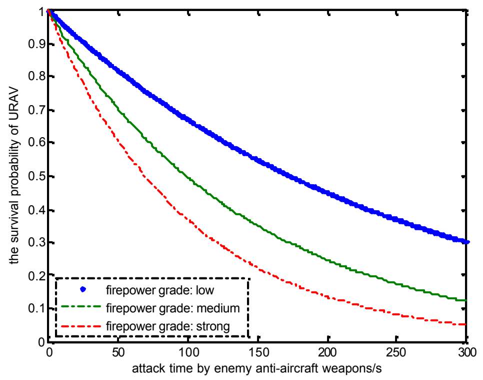


Figure 3 the survival probability of URAV

3.4 Construct the model of task reliability of URAV

3.4.1 The traditional reliable model of system

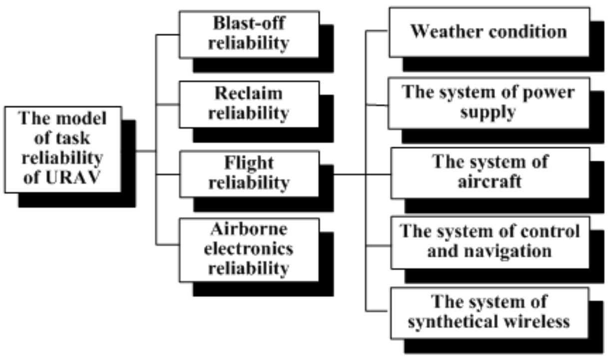


Figure 4 structure diagram of the model of task reliability of URAV.

When the URAV performs tasks, the course from blast-off to reclaim is evaluated. Capacity that URAV finishes task is analyzed, and it calculates reliability probability. Reliability of the system of URAV is divided into four subsystems: blast-off system, reclaim system, flight system, airborne electronics system. Figure 4 shows a structure diagram of the model of task reliability of URAV.

By analysis above, Four subsystems are series arrangement, So Reliability probability of the system of URAV is product representation.

3.4.2 The integrated reliable evaluation model

The model is based on the fault that the URAV cannot be repaired when carrying out tasks, and analyzes reliability of each function unit. According to fuzzy mathematics and malfunction analysis, the integrated reliable evaluation model is constructed. Figure 5 shows fault tree about the system of UARV.

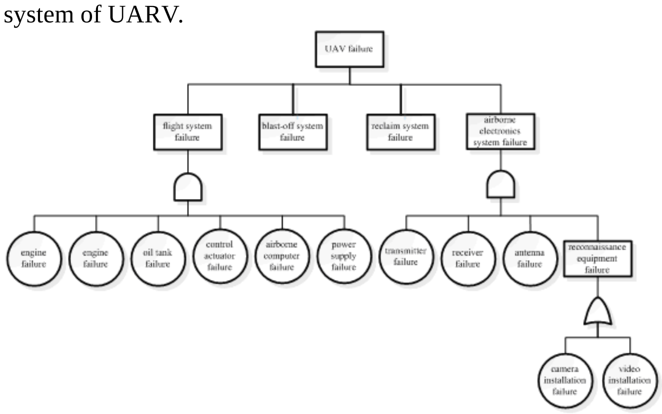


Figure 5 fault tree about the system of UARV

It is supposed the field $S_s = \{s_0, s_1, \cdots, s_k\}$. A and B show that system is in normal state and fault state respectively. A and B is fuzzy event of S_s ,and can show:

$$A = \frac{m_A(s_0)}{s_0} + \frac{m_A(s_1)}{s_1} + \cdots + \frac{m_A(s_k)}{s_k}$$

$$B = \frac{m_B(s_0)}{s_0} + \frac{m_B(s_1)}{s_1} + \cdots + \frac{m_B(s_k)}{s_k}$$

A and B is fuzzy normal state and fuzzy fault state respectively, then the aggregative indicator about task reliability (task confidence level) is :

$$R_A = P(A) = E(m_A) = \sum_{j=0}^{k} m_A(s_j) P(s_j)$$

$P(s_j)$ is each function state probability, $m_A(s_j)$ is membership function, namely weight or importance.

3.4.3 Calculation and validation

Based on some UAV, reliability is calculated.

The result of first model is: 0.8598

The result of second model is: 0.8598

Two models have similar result, it is shows that models are correct, and reliability of the URAV is comparatively high.

4. Simulate and analyze the system of developing scale of URAV

A simulation of the system is carried out with the environment of MATLAB/SIMULINK? Figure 6 shows a structure diagram of the simulation of developing scale of URAV.

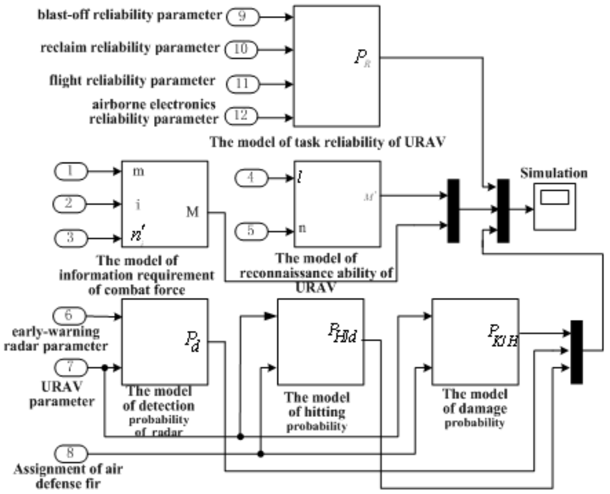


Figure 6 scheme of the simulation of developing scale of URAV.

Suppose enemy air defenses firepower is divided into three kind of Assignment, there are low, medium, and strong. The time of simulation "T" is the time of a campaign. The result of simulation is the number of URAV that is demanded in three kind of Assignment of air defenses fir. It is showed on figure 7.

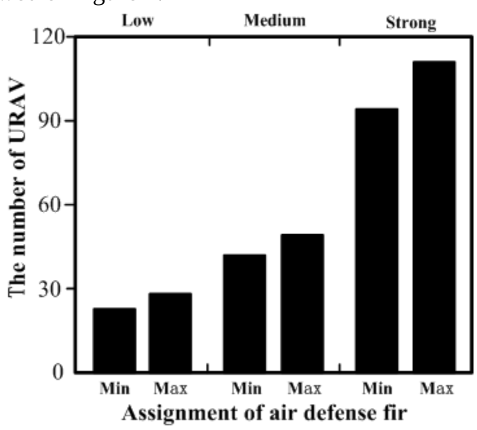


Figure 7 the number of URAV

The result shows that the modeling method and the train-of-thought are both correct.

5. Conclusion

The model of System of developing Scale for URAV is construed and simulated, that offers a scientific evaluating way for URAV equipment developing demand. To advance veracity of the simulation, the result of each sub-model is validated in feasibility and accuracy.

The model offers useful accessorial decision -making information for command, but it is comparative idealization. There are not enough for some complicated factor that is considered, such as military expenditure spending, maintenance capacity, tactical application of URAV, that influence end product. The model need be deepened and expanded to make the result press close to fact by more investigation.

References

[1] Wen Zhonghui. Tactical Missiles Systems Analysis. Beijing, National Defenses Industry Press, 2000, 6: 115~117 , 135~137
[2] Skolnik M I. Readar Handbook. Beijing, Publishing House of Electronics Industry, 2004,11: 5~8, 18~66
[3] Zhang Kao, Ma Dongli. Military Aircraft Survivability and Stealth Design. Beijing: National Defense Industry Press, , 2002 , 6: 19~53
[4] Guo Xiaohui, Song Biefeng, Wang Xu .Evaluation of Survivability to an Aircraft with Electronic Countermeasures System. Systems Engineering--Theory & Practice. 2005, 6: 72~75
[5] Gao Hongjian, Song Bifeng, Li Zhanke. Detection time and range research during the aircraft-detection system engagement. Systems Engineering - Theory & Practice , 2003, 10: 104 ~105.
[6] Wei Yong. Research on an Optimized Method for Torpedo Launcher. Equipment Environmental engineering , 2006, 5: 63~65.
[7] Lu Yanxiao, Zheng Pengzhou. Reliability design and analysis. Beijing, National Defense Industry Press, 1995, 9: 170~178

Guodong Jin was Born in 1979, He is now Ph.D student in Northwestern Polytechnic University. He received B.S. and M.S. from the Second Artillery Engineering Institute in 2002 and 2005 respectively, and then became a teacher there. He did research work in application and simulation of UAV.

A Sub-block Based Image Retrieval Using Modified Integrated Region Matching

E R Vimina[1] and K Poulose Jacob[2]

[1] **Department of Computer Science, Rajagiri College of Social Sciences**
Kochi, Kerala, India

[2] **Department of Computer Science, Cochin University of Science and Technology**
Kochi, Kerala, India

Abstract

This paper proposes a content based image retrieval (CBIR) system using the local colour and texture features of selected image sub-blocks and global colour and shape features of the image. The image sub-blocks are roughly identified by segmenting the image into partitions of different configuration, finding the edge density in each partition using edge thresholding followed by morphological dilation. The colour and texture features of the identified regions are computed from the histograms of the quantized HSV colour space and Gray Level Co- occurrence Matrix (GLCM) respectively. The colour and texture feature vectors is computed for each region. The shape features are computed from the Edge Histogram Descriptor (EHD). A modified Integrated Region Matching (IRM) algorithm is used for finding the minimum distance between the sub-blocks of the query and target image. Experimental results show that the proposed method provides better retrieving result than retrieval using some of the existing methods.

Keywords: *CBIR, Colour histogram, Edge histogram descriptor, Euclidean distance, GLCM, IRM similarity.*

1. Introduction

Content Based Image Retrieval (CBIR) has become an important area of research with the ever increasing demand and use of digital images in various fields such as medicine, engineering, sciences, digital photography etc. Unlike the traditional method of text-based image retrieval in which the image search is based on textual description associated with the images, CBIR systems retrieve images based on the content of the image such as colour, texture, shape or any other information that can be automatically extracted from the image itself and using it as a criterion to retrieve content related images from the database. The retrieved images are then ranked according to the relevance between the query image and images in the database in proportion to a similarity measure calculated from the features [1][2][3].

2. Related Work

Of the many variants of CBIR systems, query-by-example (QBE) is the most widely supported method. Here the user formulates the query by giving an example image. The features of this query image will be extracted and compared with the pre-extracted features of the images in the database and the most similar images will be returned to the user. Most of the early CBIR systems rely on global features of the query image to retrieve similar images [4][5][6][15]. But they more often fail either due to the lack of higher-level knowledge about what exactly was of interest to the user in the query image or due to the fact that global features cannot sufficiently capture the important properties of individual objects. Recently, much research has focused on region-based techniques [2][3][7][16][19][31]. Such systems can be classified into two types, the ROI defined by the user or ROI identified by machine learning methods. In the first type the user can randomly select the region of the image based on his or her need and search for similar images [16][31]. Although this method captures meaningful object regions, sometimes it is a tedious and boring task for the user. The second type either subdivide the image into fixed blocks [19][20][21] or partition the image into different meaningful regions using segmentation algorithms [2][3][7][23]. Performance of segmentation based methods depends highly on the quality of the segmentation as the average features of all pixels in a segment are often used as the features of that segment. Small areas of incorrect segmentation might make the representation very different from that of the real object. Incorrect segmentation may also affect the shape features. Also accurate segmentation is still a challenging problem and the computational load of segmentation method is heavier. For the fixed block segmentation methods the computational cost is less and also provides satisfactory results comparable with that of the pixel-wise segmentation methods even if the objects are not segmented correctly. Some other CBIR systems [16] [30] extract salient points (also known as interest points) [28] [29], which are locations in an image where there is a significant variation with respect to a chosen image feature. In salient point based methods, feature vector is created for each salient point and the selection of the number of salient points is very important. These representations enable a retrieval method to have a representation of different local regions of the image, and thus these images can be searched based on their local characteristics.

3. Proposed Method

In the proposed method fixed block segmentation is used. The images are divided into different sized blocks for feature extraction. Feature vectors are extracted from selected grids of different configurations (3x3 grid, horizontal and vertical grids, central block and the entire image) (Fig.1). Unlike some block based image retrieval systems that uses all the sub-blocks for feature extraction and similarity measurement, our system uses selected blocks only reducing the computational time and cost.

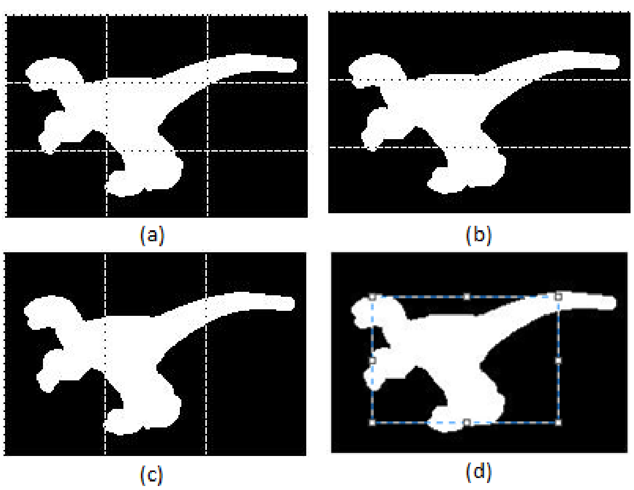


Fig. 1. Different image configurations for feature extraction

3.1 Attention Center and Central Block Extraction

To find the attention center of an image, the first step is to find the salient regions. In an image all regions may not be important or perceptually salient. When an image is mapped into the appropriate feature space salient regions will stand out from the rest of the data and can more easily be identified. To identify the salient regions the images are initially cropped by 20 pixels in the horizontal and vertical direction from the border in-order to avoid the effect of unwanted edges in the border regions. The resultant image is then converted to gray scale and blurred with Gaussian filter to discard noise. The canny edge filter is used for extracting the prominent edges. Center of mass (centroid) of the resultant image is found and is termed as attention center.

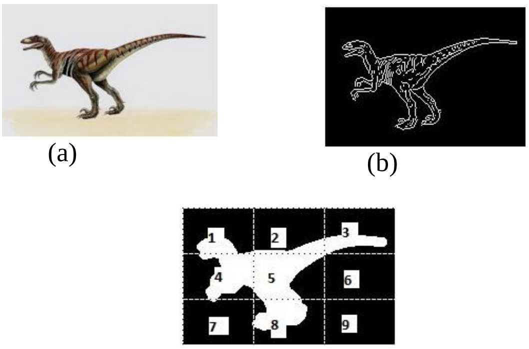
Fig. 2. Original image (Left) and the edge image marked with attention center (Right)

The rectangular region around the attention center with dimensions half the size of the original image is taken as the center block.

3.2 Sub-block Selection

To identify the salient sub-blocks /object regions in (a), (b) and (c) of fig. 1, first the grayscale image is computed for each image and edge map is detected using Sobel edge filter with a threshold value of τ (τ <1 so that the edges are boosted). The gaps in the edge map are bridged by dilating it with 'line' structuring element, that consists of three 'on' pixels in a row, in the 0, 45, 90 and 135 directions. The holes in the resultant image are then filled to get the approximate location of the objects. The objects are identified correctly if the background is uniform.

A sub-block is selected for further processing, feature extraction and is identified as region of interest (ROI) if τ'% of the sub-block is part of the object region. Ie, if the number of white pixels in that sub-block is τ'% of the sub-block with maximum white pixel density, it is identified as a region of interest. For example, for the 3x3 partitioned image in Fig.3, regions 1, 3, 4, 5, and 8 are the ROIs. Only these sub-blocks take part in further computations for



(a) (b) (c)

Fig. 3. (a)Original image (b)Edge map after sobel edge filtering
(c) Edge map after edge thresholding and morphological dilation

calculating the similarity along with the global colour and shape features of the entire image [26]. The horizontal and vertical ROIs are also identified in the same manner

4. Feature Extraction

The colour and texture features of the selected sub-blocks are extracted for similarity computation between the query and the candidate images in the database. Global colour and shape features are also computed for this purpose.

4.1 Colour

Colour features are extracted using the histograms of HSV colour space. For this purpose, the HSV colour space is quantized into 18 bins of Hue, 3 bins of Saturation and 3 bins of Value. The histogram of each of these channels are extracted resulting in a 24 dimensional colour feature vector that is normalized in the range of [0,1]. For each image both global and local colour features are extracted.

4.2 Texture

Texture features are extracted using the Gray Level Co-occurrence Matrix (GLCM). It is a matrix showing how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j. It is defined by P(i,j| d,ϴ), which expresses the probability of the couple of pixels at ϴ direction and d interval. Once the GLCM is created various features can be computed from it. The most commonly used features are contrast, energy, entropy, correlation and homogeneity. We have taken d=1 and ϴ = 0^o, 45^o, 90^o and 135^o for computing the texture features. Contrast, energy, correlation and homogeneity are taken in all the four directions and entropy of the whole block is separately calculated as it gave better retrieving results. Thus 17 texture feature vectors are calculated for each sub-block.

4.3 Shape

Shape feature provide important semantic information due to human's ability to recognize objects through their shape. However, this information can only be extracted by means of a segmentation similar to the one that the human visual system implements which is still a challenging problem. Here Edge Histogram Descriptor (EHD) is used for shape feature extraction[13][14]. It represents the local edge distribution of the image by dividing image space into 4× 4 sub-images and representing the local distribution of each sub-image by a histogram. For this, edges in the sub-images are categorized into five types; vertical, horizontal, 45-degree diagonal, 135-degree diagonal and non-directional edges (Fig.4). The edge histogram for the sub-images are computed resulting in a shape feature vector of size 80.

a) vertical b) horizontal c)45 degree d) 135 degree e)non-directional

Fig 4 Five types of edges in the Edge Histogram Descriptor

5. Similarity Computation

The L_2 norm or Euclidean distance measure is used for computing the distance between the images. It is given by the formula,

$$d_{(I1,I2)} = \sqrt{(F_{I1} - F_{I2})^2} \quad (1)$$

Where, f_{I1} and f_{I2} are the feature vectors of images I_1 and I_2.

5.1 Minimum distance between images

For computing the minimum distance between the regions of the images a modified Integrated Region Matching algorithm [3] is used. The IRM algorithm allows one region in an image to be matched with several regions of another image. In the proposed algorithm, for each ROI in the query image, the colour and texture features are computed and is compared with each ROIs of the target images (Fig.5). Assume that image I_1 has m ROIs represented by $R_1 = \{r_1, r_2, \ldots\ldots, r_m\}$ and I_2 has n ROIs represented by $R_2 = \{ r'_1, r'_2, \ldots\ldots r'_n\}$. Let the distance between r_i and r'_j be d(ri,r'j) denoted as di,j. Every region

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

ri of R1 is compared with every region r_j of R_2. This results in 'n' comparisons for a single region in R_1 and n distance measures. These distances are stored in ascending order in an array and the minimum distance (d[1]) only is taken for the final computation of the distance D; the distance between I_1 and I_2. Every d[1] of the 'm' distances is then multiplied with the minimum significance of the corresponding regions. Finally out of the m × n distances m distances are added to get the distance D. Using this method if image I_1 is compared with itself, D will be equal to zero indicating perfect match.

The significance matrix S_1 and S_2 of image I_1 and I_2 respectively consist of the white pixel density in each identified region. Ie, if I_1 has m regions and I_2 has n regions,

$S_1=[s_{11'},s_{12'}\ldots.s_{1m'}]$ (2)

$S_2=[s_{21'},s_{22'},\ldots.s_{2n'}]$ (3)

Where, $s_{1i'}$ and $s_{2i'}$ are the white pixel density in each identified region of I_1 and I_2. Also, S_1 and S_2 are normalized so that $\Sigma S_1=0$ and $\Sigma S_2=0$.

The algorithm is summarized as follows:

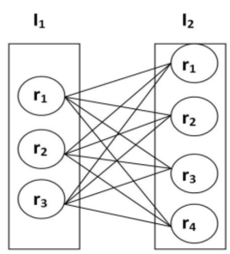


Fig. 5 m regions of I_1 are compared with n regions of I_2

Input:R_1, R_2; the ROIs of I_1 and I_2
S_1, S_2; significance of selected regions of I_1 and I_2
Output:*D*, minimum distance between regions of I_1 and I_2
Begin
for each region in the query image I_1, i=1 to m **do**
for each region in the target image I_2, j=1 to n **do**
compute distance $d[j]=d_{i,j;}$
end
Sort distance array 'd' in ascending order;
if ($\Sigma S_1>0$ and $\Sigma S_2>0$)
$s'_{i'j'}$ = minimum ($s_{i'}$, $s_{j'}$);
$D=D + d[1] \times s'_{i',j'}$;
$s_{i'}=s_i' - s'_{i'j'}$;
$s_{j'}=s_j' - s'_{i'j'}$;
else
$D=D + d[1]$;
end if
end for
end begin

Here 'd' is the array containing the distances between the r_i of R_1 with the n regions of R_2. If d[1] is the minimum distance in the array; the region pair being i of R_1 and j of R_2, then $s_{i'}$ is the significance of region i in S_1, and $s_{j'}$ is the significance of region j in S_2 and $s'_{i'j'}$ is the minimum significance among the two.

In some cases ΣS_1 or ΣS_2 or both will become zero before all the m regions of the query image I_1 is considered for the similarity calculation. In such cases d[1] of the uncounted regions is taken for similarity computation.

The minimum distance between the horizontal and vertical blocks are also computed in a similar manner and is denoted as D_h and D_v respectively. The final distance between I_1 and I_2 is given by

$$D'=D+D_h+D_v+D_g + d_{central\ block\_colour\ \_texture\_feature} \quad (4)$$

Where, $D_g = d_{global\_colour\_feature} + d_{global\_shape\_feature}$; $d_{global\_colour\_feature}$ and $d_{global\_shape\ \_feature}$ being the Euclidean distance between the global colour and shape feature vectors of I_1 and I_2 and $d_{central\ block\_colour\ \_texture\_feature}$ is the distance between the feature vectors of the central blocks of I_1 and I_2 .

6. Experimental Results and Discussions

The Wang's image database [9] of 1000 images, which is considered to be one of the benchmark databases for CBIR, consisting of 10 categories is used for evaluating the performance of the proposed method. Each category contains 100 images. A retrieved image is considered to be correct if and only if it is in the same category as the query. For each query, a preselected number of images are retrieved which are illustrated and listed in the ascending order of the distance between the query and the retrieved images. The results of the proposed method is compared with that of [10], [11] and [27] in terms of average precision. Precision (P) of retrieved results is given by

$$P(k)=n_k/k \quad (5)$$

Where, k is the number of retrieved images, n_k is the number of relevant images in the retrieved images. The average precision of the images belonging to the q^{th} category A_q is given by

$$\overline{P}_q = \sum_{k \in A_q} P(I_k) \Big/ \left|A_q\right|, \; q = 1,2,\ldots.10. \quad (6)$$

The final average precision is

$$\overline{P} = \sum_{q=1}^{10} \overline{P}_q / 10 \tag{7}$$

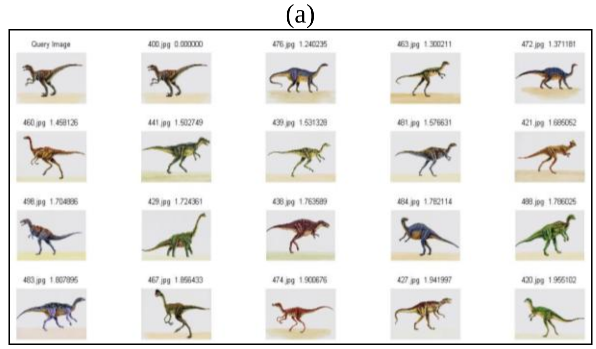
Table.1. shows the average precision of the retrieved images for different categories when k=20 for different methods. It is seen that for most of the categories the proposed method provides better or comparable results with that of the other methods. For a few categories like 'Beaches', 'Buildings' and 'Mountains' the performance of the proposed method is lower than that of some of the compared methods because of the similarity of the background of the images. For the categories 'Dinosaur' and 'Flowers' the average precision when k=20 is very high. This means that for images with single object the proposed algorithm works better than the compared algorithms.

Table.1 % Average Precision (K=20) of retrieved images using different methods

Category	*% Average precision of retrieved images for k=20*			
	Jhanwar et al[11]	*Hung and Dai's [10]*	*CTDCIRS [27]*	*Proposed method*
Africa	45.25	42.40	56.20	71.52
Beaches	39.75	44.55	53.60	43.60
Buildings	37.35	41.05	61.00	53.55
Bus	74.10	85.15	89.30	85.30
Dinosaur	91.45	58.65	98.40	99.55
Elephant	30.40	42.55	57.80	59.10
Flowers	85.15	89.75	89.90	90.95
Horse	56.80	58.90	78.00	92.40
Mountains	29.25	28.5	51.20	38.35
Food	36.95	42.65	69.40	72.40
Average	**52.64**	**53.24**	**70.48**	**70.67**

Fig.6 depicts the top 19 retrieved images for two sample query image using proposed method. In each set, on top left corner is the query image and the retrieved images are listed according to their distance with the query image. On top of every retrieved image the image identification number and its distance from the query image is shown. It should be noted that in each set of retrieval, the first image retrieved is the query image itself with distance zero. This emphasizes the fact that if the query image itself is present in the searching database, it will be retrieved first indicating perfect match.

(a)



(b)

Fig. 6 Top 19 retrieved images for the two sample query image. For both the results the image in the top left corner is the query image and the retrieved images are listed according to their distance from the query image.

6. Conclusion and future work

A content based image retrieval system using the colour and texture features of selected sub-blocks and global colour and shape features of the image is proposed. The colour features are extracted from the histograms of the quantized HSV color space, texture features from GLCM and shape features from EHD. A modified IRM algorithm is used for computing the minimum distance between the selected sub-blocks of the query image and the candidate images in the database. Unlike the most sub-block based methods that involves all the sub-blocks of the query image to be compared with that of the candidate images, our system involves only selected sub-blocks for similarity measurement, thus reducing the number of comparisons and computational cost. Experimental results also show that the proposed method provides better retrieving result than some of the existing methods. Future work aims at the selection of sub-blocks based on their saliency in the image to improve the retrieval precision. Also the proposed method has to be tested on various databases to test the robustness.

References

[1] Jia Li and James Z. Wang, "Real-time computerized annotation of pictures", Proceedings of the 14th annual ACM international conference on Multimedia, (2006), pp. 911-920.

[2] Y. Chen, J. Z. Wang, and R. Krovetz, "CLUE: Cluster-based retrieval of images by unsupervised learning", IEEE Transactions on Image Processing, (2005), 14(8):1187-1201

[3] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive Integrated Matching for Picture Libraries", IEEE Transactions on Pattern Analysis and Machine Intelligence, (2001), vol. 23, no. 9, pp. 947-963.

[4] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom et al. "Query by Image and Video Content: The QBIC System", IEEE Computer, (1995) vol. 28, pp. 23-32

[5] A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Content-based Manipulation of Image Databases,", Proc. SPIE Storage and Retrieval for Image and Video Databases II, SanJose, CA (1994), pp. 34–47.

[6] M. Stricker, and M. Orengo, "Similarity of Color Images," in Proc. SPIE Storage and Retrieval for Image and Video Databases, (1995) pp. 381-392.

[7] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, and J. Malik, "Blobworld: A System for Region-Based Image Indexing and Retrieval", Proc. Visual Information Systems, (1999) pp. 509-516.

[8] Haralick, R.M., K. Shanmugan, and I. Dinstein, "Textural Features for Image Classification", IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-3, (1973), pp. 610-621.

[9] http://wang.ist.psu.edu/docs/related/

[10] P.W. Huang, S.K. Dai, "Image retrieval by texture similarity", Pattern Recognition, 36 (3) (2003), pp.665–679.

[11] N. Jhanwar, S. Chaudhuri, G. Seetharaman, B. Zavidoviqu, "Content based image retrieval using motif co-occurrence matrix", Image and Vision Computing, (2004), 22 (14) pp. 1211-1220.

[12] X.C. He and N.H.C. Yung, "Curvature Scale Space Corner Detector with Adaptive Threshold and Dynamic Region of Support", Proceedings of the 17th International Conference on Pattern Recognition, (2004) 2:791-794.

[13] Overview of the MPEG-7 standard, (2001) December. ISO/IEC/TC/SC29/WG11 N3914

[14] B. S. Manjunath, Philippe Salembier, Thomas Sikora, "Introduction to MPEG-7", JOHN WILLEY & SONS, LTD , (2002) pp. 183-184

[15] Subrahmanyam Murala, Anil Balaji Gonde, R. P. Maheshwari, "Color and Texture Features for Image Indexing and Retrieval", IEEE International Advance Computing Conference, (2009), , March, Patiala, India, pp.1411-1416,

[16] Banerjee Minakshi, Kundu Malay K, Maji Pradipta. "Content-based image retrieval using visually significant point features", Fuzzy Sets & Systems, (2009), 160(23):3323-3341.

[17] M.J. Swain, D.H. Ballard, "Color indexing", International Journal of Computer Vision (1991), pp. 11-32.

[18] Hafner, J., Sawhney, H.S., Wquitz, W, "Efficient color histogram indexing for quadratic form distance functions", IEEE Transactions on Pattern Analysis and Machine Intelligence (1995), 17 (7), 729–736.

[19] MJ Hsiao, J P Huang, t Tsai, T W Chiang, "An Efficient and Flexible Matching Strategy for Content-based Image Retrieval", Life Science Journal, Vol 7, No. 1, pp. 99–106 (2010).

[20] Q. Tian, Y. Wu, and T. S. Huang, "Combine user defined region-of-interest and spatial layout for image retrieval", International Conference on Image Processing, (2000), Vol.3, pp. 746 – 749.

[21] P. Howarth and S.Ruger, "Robust texture features for still-image retrieval", IEEE. Proceedings of Visual Image Signal Processing, (2005), Vol. 152, No. 6.

[22] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, (2008), vol. 40, pp. 1-60.

[23] W T Su, J C Chen, J J Lien, "Region Based Image Retrieval System with Heuristic Pre-clustering relevance feedback", Expert systems with Applications (2010), Vol.37, pp. 4984-4998

[24] B.S.Manjunath and W.Y.Ma , "Texture Features for Browsing and Retrieval of Image Data", IEEE transactions on PAAMI, (1996) Vol 18, No. 8. pp. 837 - 842

[25] L. M Kaplan, "Fast texture database retrieval using extended fractal features", SPIE 3312, SRIVD (1998), VI:162-173.

[26] Vimina E R, Poulose Jacob K, "Image Retrieval using Local and Global Properties of Image Regions with Relevance Feedback", Proceedings of the International Conference on Advances in Computing, Communications and informatics – ICACCI'12, August 3-5, 2012, Chennai, T Nadu, India, pp. 683-689.

[27] M. B. Rao, B. P. Rao, and A. Govardhan, "CTDCIRS: Content based Image Retrieval System based on Dominant Color and Texture Features", International Journal of Computer Applications, Vol. 18–No.6, pp-0975-8887, (2011).

[28] David G. Lowe, "Distinctive image features from scale-invariant keypoints", International journal of computer vision, vol.60, pp. 91-110, (2004).

[29] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)", Computer Vision and Image Understanding, vol. 110, pp. 346 - 359, (2008).

[30] Hiremath P.S, Jagadeesh Pujari. "Content Based Image Retrieval using Color Boosted Salient Points and Shape features of an image". International Journal of Image Processing, 2(1): 10-17, (2008).

[31] Zhuozheng Wang, Kebin Jia , Pengyu Liu, "A Novel Image Retrieval Algorithm Based on ROI by Using SIFT Feature Matching", International Conference on Multimedia and Information Technology,2008, pp. 338-341.

[32] Ma Y F, Zhang H J, "Contrast based image attention analysis by using fuzzy growing", 11th ACM international conference on multimedia, pp 734-781, 2003

E R Vimina, Department of Computer Science, Rajagiri College of Social Sciences, India received B.Tech in Electrical and Electronics Engineering from Mahatma Gandhi University, India and ME degree in Computer Science and Engineering from Bharathiar University, India. She is currently pursuing doctoral research in Image Retrieval at Cochin University of Science and Technology, India. Her research interests include Image Retrieval and Artificial intelligence.

Dr. K. Paulose Jacob, Professor of Computer Science at Cochin University of Science and Technology since 1994, is the Director of the School of Computer Science Studies and Dean of the Faculty of Engineering. Dr. Jacob has been teaching at the Cochin University since 1980. A National Merit Scholar all through, he is a graduate in Electrical Engineering and postgraduate in Digital Electronics. He obtained Ph D in Computer Engineering for his work in Multi-Microprocessor Applications. His other research

interests are Information Systems Engineering, Intelligent Architectures and Networks. He has more than 50 research publications to his credit, and has presented research papers in several International Conferences in Europe, USA, UK and other countries. He is a Permanent Professional Member of the ACM and a Life Member of the Computer Society of India.

A Novel Parallel Computing Method for Computational Fluid Dynamics

Bai Yuguang[1,2,*], Wu Guoqiang[1,2] and Zhang Yuguang[2]

[1] **State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology Dalian, Liaoning 116024, China**

[2] **School of Aeronautics and Astronautics, Dalian University of Technology Dalian, Liaoning 116024, China**

Abstract

A novel parallel computing method for computational fluid dynamics research in engineering is presented. The message passing language PVM was employed to develop a convenient parallel environment with effective computing resources. It makes three dimensional numerical simulations of fluid-structure interaction problems feasible for most engineers. Wind-induced vibration problems of blunt structures with three dimensional turbulent flow simulations are implemented on the proposed clusters built with the present parallel computing method and get a direct benefit in engineering.

Keywords: *Parallel Computing Method; PVM Code; Computational Fluid Dynamics; Three Dimensional numerical Simulation*

1. Introduction

Fluid-structure interaction (FSI) is a challenging engineering problem which has been studied extensively through wind tunnel and real life tests. Over the past decade much computational work has been undertaken in the area of computational fluid dynamics (CFD) because advances in computer power and algorithms have made this approach increasingly feasible [1]. But for some blunt structures existing widely over the world, e.g. long-span bridges, two dimensional (2D) numerical simulations cannot be accurate enough, while three dimensional ones can give accurate predictions and direct visual evaluations [2].

For 3D CFD simulations, even the most advanced personal computer (PC) over the present world cannot undertake the burden of the computational demands. So parallel computing clusters are necessary. Until recently, these clusters are mainly from professional suppliers and rather expensive. For most engineers, they cannot afford these huge hardware budgets. For the past few years, PC cluster is a cost effective solution for people to have computing power of supercomputers. Besides the budget issue, another reason for the popularity of the PC cluster is that people can deploy the system as what they want. This is the benefits of open standard and open source software.

Some studies have given encouraging advices for parallel computing method. Khan *et al*. [3] presented the classification of load balancing conditions for parallel and distributed systems and discuss the use of his method. Nourah *et al.* [4] proposed the Multiprocessor scheduling when the parallel genetic algorithm was used, and he discussed task scheduling which is a most challenging problem in the parallel computing. Synchronous master to slave algorithm outperforms the sequential algorithm in case of complex and high number of generations problem was also proposed. An iterative decoding of generalized parallel concatenated block codes using cyclic permutations was proposed by Allouch *et al.* [5], which has significant benefit for numeric simulation in engineering. The effects of various parameters component codes such as inter-leaver size, block size, and the number of iterations were evaluated. The work can provide important samples for parallel applications. Olmedo *et al*. [6] proposed an approach of the point to point processing of digital images using parallel computing. It can provide an effective way for us to build smooth communication among master computers and slave computers.

A novel parallel computing method based on free operating system and software to connect common PCs together for 3D CFD numerical simulations in engineering is presented. And the advantages of it are introduced. After several clusters have been built successfully with this method, 3D wind-induced vibrations of long-span bridge structures are simulated on these clusters with detailed introductions of parallel computing including flow field, pressure contours.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

2. Parallel Computing Method

2.1 Hardware and Software Demands

The demands of hardware for the present parallel computing method are fast Ethernet switch devices and common PCs, each of which has a PXE network interface card. PCs are connected to the switch and they can be transformed into a working group.

The demands of software are free GNU/Linux operating systems such as Centos. We developed the parallel code and run it in our PC cluster with Linux system. The message passing language PVM [7] (i.e. Parallel Virtual Machine) was employed to develop a parallel environment because it is convenient for people to grasp.

2.2 Installation of the Parallel Computing Clusters

One cluster consists of a server computer and many client computers. For 3D CFD simulations, many computers with same hardware were used, each of which has two cores in one Inter or AMD CPU. A roadmap for installation steps is given by Figure 1 and the last three steps are only for CFD. Actual clusters built with the present method are pictorial in Figure 2.

2.3 Advantages of the Proposed Parallel Method

The advantages of the present parallel computing method can be summarized as:

• The installation is simple. The present method uses PXE/etherboot, NFS, and NIS to provide services to client computers so that it is not necessary to install GNU/Linux on the client hard drives individually. Once the server is ready, the client computers can boot via PXE/etherboot (diskless).

• Save on hardware, budget and maintenance fees. Clusters built with the present method can achieve the same computing power as professional clusters with much lower budget than that of the professional ones. The hardware is easy to buy and the maintenance fee can be controlled to a lower standard than that of the professional clusters.

• Good expansibility. New computers can be connect to the server easily as client computers and can not influence other client ones.

So the presented parallel computing method has great potential to be applied in engineering simulations. Many engineers can afford the budget of this kind of clusters and use them simply to give better prediction of complicated engineering problems.

2.4 System Services for CFD Simulations

There are several absolutely necessary system services for CFD parallel computing. They are XINETD, RSH and RLOGIN. Many kinds of commercial software for CFD parallel computing need these services which are not started by default. They must be started before parallel computing is implemented on the clusters.

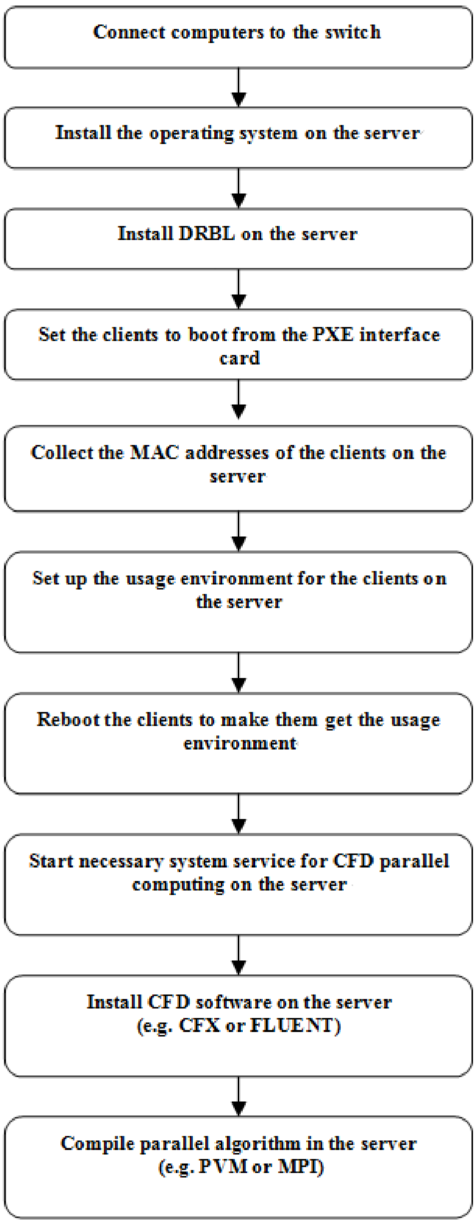


Fig. 1 Installation steps.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

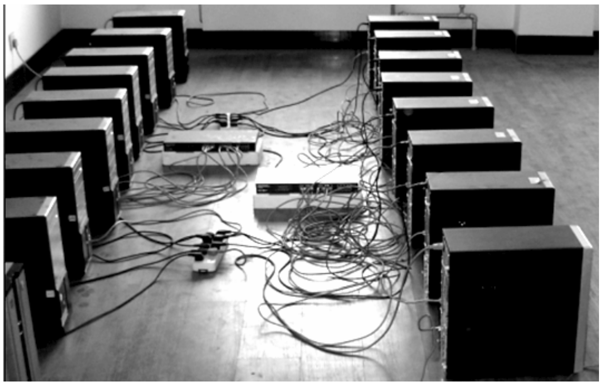


Fig. 2 Actual parallel clusters built with the proposed method

3. Application in the Simulation with CFD

3.1 Numerical Algorithm for FSI

The Navier-Stokes (NS) equations and the associated continuity equation are used with: a modification to account for turbulence; a moving mesh; and the Smagorinsky eddy viscosity model [2].
The governing equations for the FSI come from both the corresponding flow and structural analyses. All these equations can be treated as time and space dependent and need to be discretised before solution methods can be applied. The discretised incremental Navier–Stokes and structural equations can be expressed by

$$\mathbf{N}(\mathbf{a}, \mathbf{b}) = 0 , \quad \mathbf{S}(\mathbf{b}) = \mathbf{f}(\mathbf{a}) \qquad (1)$$

where **a** and **b** are the field vectors consisting of the unknowns at the time step n+1 currently being solved for; $\mathbf{a} = \langle \mathbf{u}^{n+1}, \mathbf{p}^{n+1} \rangle$ contains the variables from the fluid domain; $\mathbf{b} = \langle \mathrm{h}^{n+1}, \dot{\mathrm{h}}^{n+1} \rangle$ contains the variables from the structural domain, where h is the vector of the decoupled generalized displacements. The field variables at the previous time step n are assumed to be known, and are not reflected in Eq. (1). The two equation sets are fully coupled.

Eq. (1) can be solved by the block-iterative method [2]. By using any available solvers for each of the two parts of (1) in conjunction with the block-Gauss–Seidel iterative algorithm, the NS equations are solved first for **a** and then for **b**. The iteration scheme which results can be written as

$$\mathbf{N}(\mathbf{a}^{(i+1)}, \mathbf{b}^{(i)}) = 0 \qquad (2)$$

$$\mathbf{S}(\mathbf{b}^{(i+1)}) = \mathbf{f}(\mathbf{a}^{(i+1)}) \qquad (3)$$

where i is the iteration counter, and it converges linearly.

For generality, (2) and (3) are both treated as nonlinear. Therefore linearization methods like the Newton–Raphson method or the Picard (fixed point) method must be used [2]. For the latter the linearization iteration is

$$\mathbf{a}^{(i+1)}_{(j+1)} = \mathbf{F}(\mathbf{a}^{(i+1)}_{(j)}, \mathbf{b}^{(i)}), \quad \mathbf{b}^{(i+1)}_{(j+1)} = \mathbf{G}(\mathbf{a}^{(i+1)}, \mathbf{b}^{(i+1)}_{(j)}) \qquad (4)$$

where j is the linearization iteration counter. The two layers of iteration i and j in (4) address both the field coupling and the non-linearity. With global convergence checked at every time step, the solution obtained should be identical to that given by the direct coupled solution to (1). Note too that the two iterations can be mixed and when merged, the equivalent iteration is

$$\mathbf{a}^{(k+1)} = \mathbf{F}(\mathbf{a}^{(k)}, \mathbf{b}^{(k)}), \quad \mathbf{b}^{(k+1)} = \mathbf{G}(\mathbf{a}^{(k+1)}, \mathbf{b}^{(k)}) \qquad (5)$$

where k is the merged iteration counter. Solvers in the form of (5) can be found in existing CFD and structural analysis software.

Figure 3 plots the flowchart of this process.

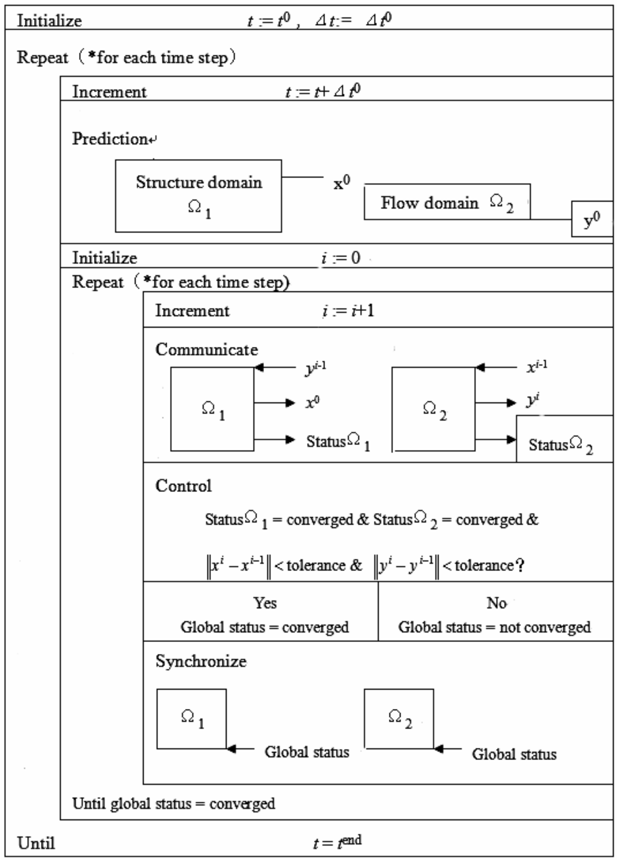


Fig. 3 Numerical analysis flowchart of the FSI problem

3.2 3D CFD Modelling

Blunt body structures like long-span bridges with different aerodynamic stability were simulated with 3D CFD turbulent modelling here. The geometrical features of them are shown in Figure 4. Figure 5 shows a 3D CFD modelling example using structure G1. The amounts of the mesh number for the five bridges are summarized in Table 1. It can be found that 3D CFD has much larger amounts of meshes than 2D CFD. The burden of 3D CFD computing must be undertaken at least by a cluster with 64 CPUs and corresponding memory. The present parallel computing for 3D CFD simulations used a cluster with 64 CPUs (i.e. including 128 CPU cores) and 128GB memory which was built with 64 PCs.

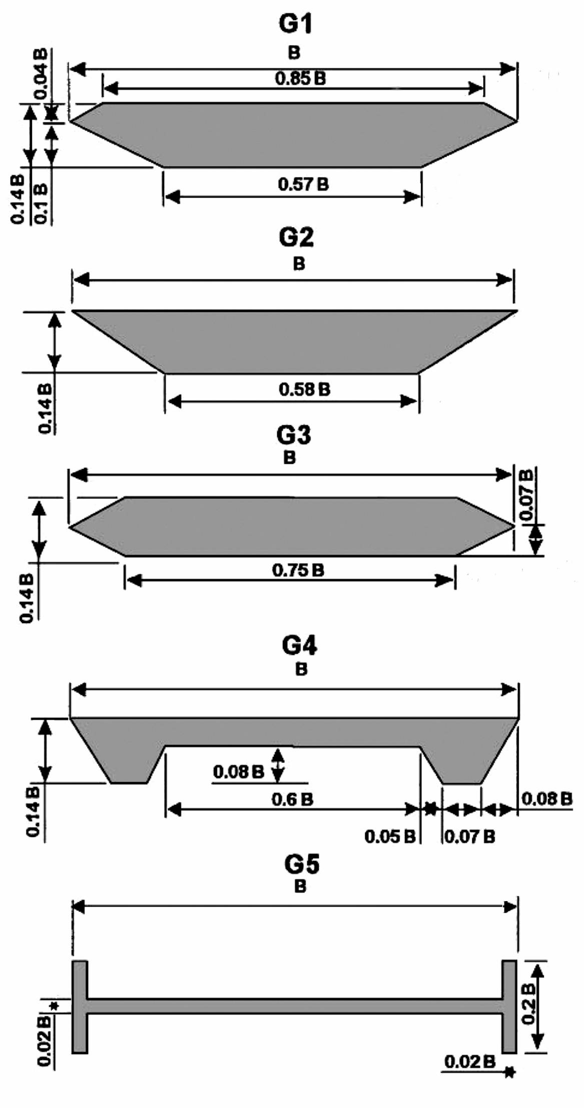


Fig. 4 The five long-span bridge deck sections G1 to G5 used.

Table 1: Mesh numbers of structures with 2D or 3D CFD simulations

	G1	G2	G3	G4	G5
2D CFD	59k	57k	67k	65k	73k
3D CFD	3183k	3093k	3696k	3602k	4090k

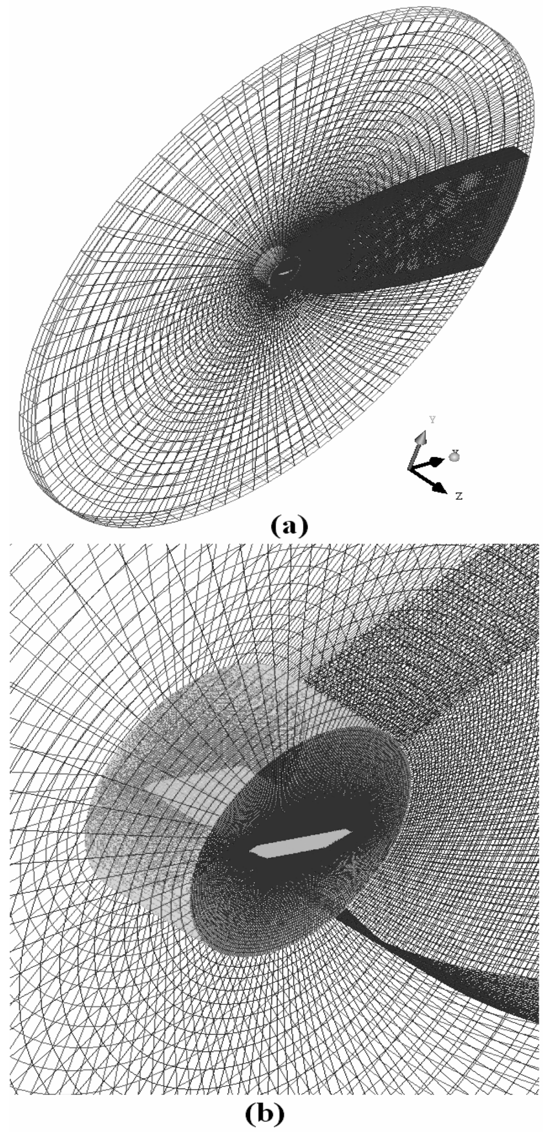


Fig. 5 3D meshes used for bridge G1: (a) overview; (b) enlargement of the inner region

3.3 Parallel Computing Process of 3D CFD

During the 3D simulation, the fluid domain is divided into 128 blocks. The block parallel code can be proposed by Allouch *et al.* [5]. The computation for each of the blocks is delegated to one of 128 processors, with one processor used as the master and the rest as the slaves.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

The flux conservation between blocks is managed by the master processor. Solutions for both the fluid flow and the structural analyses according to (5) are synchronized at each coefficient iteration loop. The time for calculating structural response is much shorter than that for the fluid analysis and to reduce the time for communicating the structural response, the structural analysis is carried out on both the master and the slaves. During the iteration, the master and the slave nodes perform the mesh deformation for their own blocks using the calculated structural responses. The structural and fluid analyses here have been closely coupled.

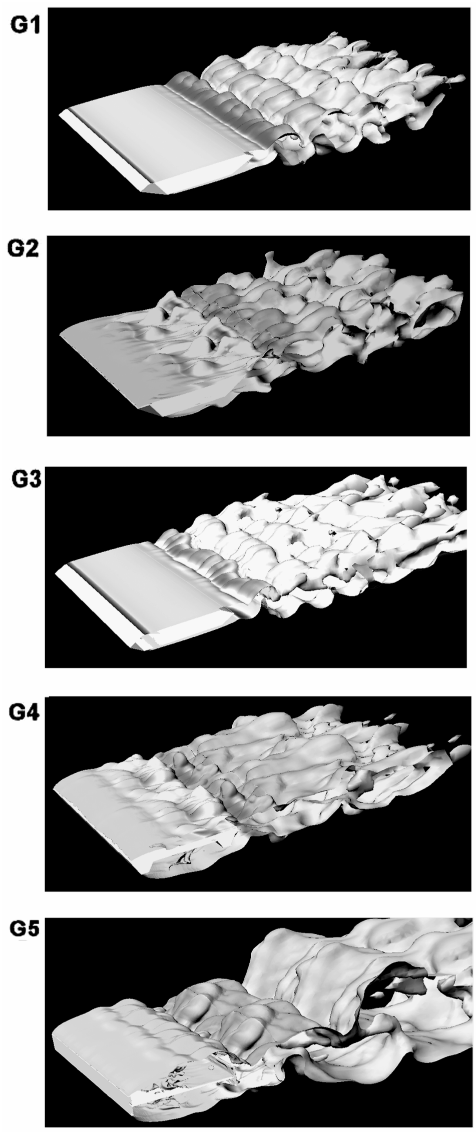


Fig. 6 Pictures of the 3D wake flow for the five bridge sections G1 to G5.

4. Results and Discussion

4.1 Simulation Results

The features of the flow field at different parts of the long-span deck cross section can be shown visually with 3D CFD simulations through the present parallel computing, e.g. the 3D wake flows for the five deck cross sections are shown pictorially in Figure 6. Hence it can be seen that structure G1 has the best aerodynamic stability, while structure G5 has the worst. Such visualization of wake through 3D CFD simulations is of direct benefit for aerodynamic analysis of structures and engineers can make their design and manufacture much more convenient with low budget.

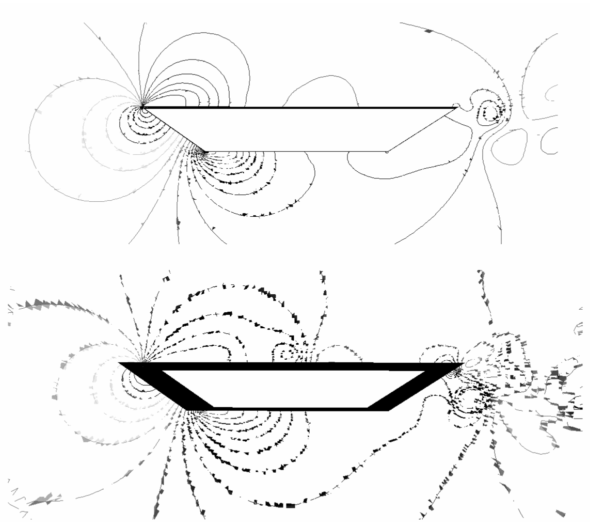


Fig. 7 Pressure contours of structure G2

Figure 7 plotted the pressure contours of structure G2 (i.e. the upper part from two-dimensional simulation and the lower part from three-dimensional simulations). It can be found that pressure distribution from three-dimensional simulations are quite different from those from two-dimensional simulations. From the contour of the lower half of Figure 7, it can be found that around structure G2, the flow reattachment occurs at the middle of the upper deck, and a large vortex region can be found at the rear part, where the structure has a slant angle.

4.2 Parallel Computing Efficiency

```
Total CPU time: 3.684E+04 seconds
          or: (        0:        10:        13:     57.516 )
              (     Days:     Hours:   Minutes:    Seconds )

Total wall clock time: 4.252E+04 seconds
          or: (        0:        11:        48:     44.000 )
              (     Days:     Hours:   Minutes:    Seconds )
```

Fig. 8 The efficiency of the present parallel computing method.

The efficiency of the present parallel computing method can be calculated from Figure 8, in which CPU time means actual computing time and wall time means total time containing CPU time and communication time. It is 86.6% and it is sufficiently good for 3D CFD simulations.

4. Conclusions

A novel parallel computing method with free operating system and software is presented. It can be simply applied to common computers and has significant benefits in terms of saving on hardware, budget and maintenance fees. The parallel computing efficiency is also satisfactory. 3D CFD simulations with the present parallel method are practiced with five long-span bridge structures. The visualizations of the flows around these blunt structures can provide direct help to engineers for aerodynamic analysis of complicated structures. Also, clusters built with the present parallel computing method can be applied to computational analysis in many other parts of engineering without huge budget.

Acknowledgments

The work was financially supported by a Foundation for the Author of National Excellent Doctoral Dissertation of PR China (Funding Code 2002030) and China Postdoctoral Science Foundation funded project (Funding Code 2012M520618)

References

[1] J. D. Anderson, Computational fluid dynamics, New York: Mc Graw Hill, 2005.

[2] Y. G. Bai, D. K. Sun, and J. H. Lin, “Three dimensional numerical simulations of long-span bridge aerodynamics using block-iterative coupling and DES,” Computers & Fluids, Vol. 39, 2010, pp. 1549-1561.

[3] Z. Khan, S. Singh, J. Alam, and S. Saxena, "Classification of load balancing conditions for parallel and distributed systems", International Journal of Computer Science Issues, Vol 8, No. 55(1), 2011, pp. 411-419.

[4] A. A. Nourah, and A. Abdullatif, "Multiprocessor scheduling using parallel genetic algorithm", International Journal of Computer Science Issues, Vol. 9, No. 44(3), 2012, pp. 260-264.

[5] H. Allouch, I. Chana, and M. Belkasmi, "Iterative decoding of generalized parallel concatenated block codes using cyclic permutation", International Journal of Computer Science Issues, Vol. 9, No. 55(1), 2012, pp. 191-198.

[6] E. Olmedo, J. de la Calleja, A. Benitez, and M. Medina Auxilio1, "Point to point processing of digital images using parallel computing", International Journal of Computer Science Issues, Vol. 9, No. 33(3), 2012, pp.1-10.

[7] Z. Roberto, P. Mario, and I. Angelo, "Enabling PVM to exploit the SCTP protocol", Journal of Parallel and Distributed Computing, Vol. 66, No. 11, 2006, pp. 1472-1479.

First Author: Yuguang Bai is a Lecturer in Dalian University of Technology, P.R.China. He received a Ph.D. degree in 2011 from Faculty of Vehicle Engineering and Mechanics in Dalian University of Technology (DUT), Dalian, P.R.China, also received his BE degrees in 2004 from Faculty of Vehicle Engineering and Mechanics in DUT. His research interests are in area of multi-physics modeling and simulation, high performance computing, and computational fluid dynamics.

Second Author: Guoqiang Wu is a Lecturer in Dalian University of Technology, P.R.China. He received a Ph.D. degree in 2009 from Astronautics School in Harbin Institute of Technology (HIT), Harbin, P.R.China, also received his BE and ME degrees in 2001 and 2004 from Solid Mechanics in HIT. His research interests are in area of multidisciplinary modeling and simulation, high performance computing, micro-satellite communication, and channel decoding.

Third Author: Yuguang Zhang is a master student in Dalian University of Technology, P.R.China. He received his BE degrees in 2010 from Department of Engineering Mechanics in Liaoning Technical University. His research interests are in area of multi-physics modeling and simulation, high performance computing, and computational fluid dynamics.

The Activity Based Costing method opportunity to assess and master the aircraft maintenance service cost for Third Party: a case study

Driss Rchid[1], Otmane Bouksour[1] and Zitouni Beidouri[1]

[1] Laboratoire de Mécanique Productique & Génie Industriel (LMPGI) ; Université Hassan II Ain Chock, Ecole Supérieure de Technologie, Km 7 Route El Jadida, Casablanca, Maroc

Abstract

This work illustrates how to calculate accurate aircraft maintenance costs using Activity Based Costing (ABC). Comparing to traditional cost calculation, the proposed method is useful for aircrafts maintenance activity integrated to Airline Company and who wants to diversify its revenue by selling some available maintenance slots to third party clients for a competitive price. This paper takes base maintenance check process of a Moroccan airline aircrafts maintenance activity as a case study. The actual method used to calculate aircraft maintenance check cost can't not be applied to third party client pricing system. The ABC method shows substantial benefits if it is successfully implemented. Data collection through interviews and questionnaires are the main sources to this study.

Keywords: *Activity Based Costing, Aircraft Maintenance, Costs.*

Introduction

The global air traffic has undergone many changes in recent years. On the one hand, its development has been extremely rapid since it was multiplied by thirty in almost thirty years (82 million passengers in 1982 against nearly 2.7 billion in regular lines in 2011[1] .On the other hand, market liberalization, embedded in a context of economic globalization, increases in fuel prices as well as the ongoing global financial crisis that started in 2008 [2], has had an impact on airlines, with increased competition and a prices softening. With this increased competitiveness and war price, airlines are forced to reduce their costs, including maintenance cost .Typically it ranges from approximately 10 to 20 % of Airplane-Related Operating Costs (AROC) [3-5]. In this context, the compression of these costs is a priority for the survival of these companies.

For maintenance organizations, knowledge of aircraft maintenance cost is an essential information to the services pricing provided to airlines, which are increasingly demanding the maintenance of their fleets, to the best conditions of safety, quality and economy. However, most organizations are still using the same traditional cost accounting systems developed decades ago, while most financial departments (70%) recognize ABC [1] as a performance measure method. If this method is pragmatically developed and methodology implemented, it will provide quick and visible results, especially in crisis times[6].

The aim of this paper is to contribute to the design of a new model of cost calculation according to the ABC approach, dedicated primarily to aircrafts maintenance activity integrated to an Airline Company, through a case study conducted within a maintenance centre of a major Moroccan airline.

2. Activities Based Costing system interest

In the eighties in the U.S.A, an enthusiasm was growing for a new approach in cost accounting called Activity Based Costing method (ABC). Pioneered by Cooper, Kaplan and Johnson [7-14] and presented as a method adapted to the new economic environment. Its ambition is to provide a framework for overcoming the difficulties encountered in Traditional Cost Accounting (TCA). In today's globally competitive market, ABC is a powerful tool that provides reliable and accurate cost information. It enables manufacturing and services organizations to improve their competitiveness and make better decisions based on an improved understanding of their product cost behaviour.

The method intends to guide costs with maximum visibility and remove up all the elements that may give a distorted picture of resources consumption cost as commonly observed in TCA[15]. Traditional cost systems known to distort the cost information by using traditional overhead allocation methods .It involves collecting indirect costs from departments and then allocates them to products or services. In contrast, ABC collects indirect costs of activities and assigns activity costs to cost object

[1] Study conducted by IDC (International Data Corporation) in July 2006 after a series of 200 interviews.

[16]. The assignment of costs through ABC occurs in two stages: cost objects (i.e., products, services, customers) consume activities and activities consume resource costs [16,17]. This means, opposing the traditional view of expenses direction (**Fig.1**).

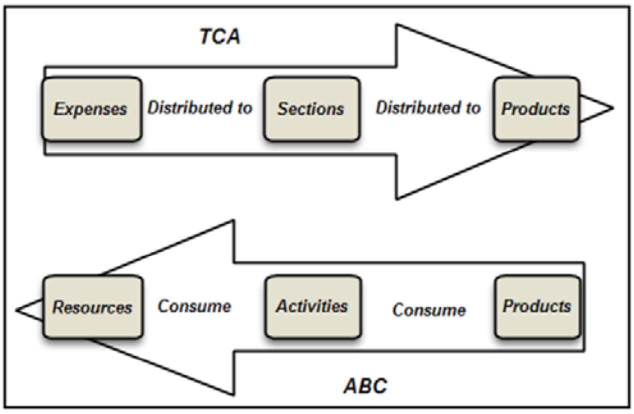


Fig. 1 Resource consumption ABC vs. TCA

In the first stage (**Fig.2)**, costs assigned to cost pools within an activity centre, based on a cost driver. There is no equivalent step in a traditional costing approach. In the second stage, costs allocated from the cost pools to cost objects based on these objects' consumption of the activities. This stage is similar to TCA except that the traditional approach uses a single volume-based cost driver to allocate overhead costs to cost objects without consideration for non-volume-related characteristics [18] . After seeing the activities costing system value, we will see in the next section, the actual method and its limits to calculate aircraft maintenance cost.

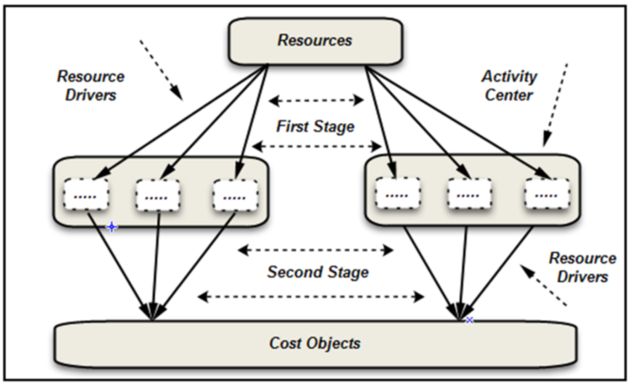


Fig. 2 Cost assignment procedure in ABC[16]

3. The Moroccan Airline Maintenance Centre

The Moroccan Airline Maintenance Centre (MAMC) is a technical department of a major Moroccan airline. Its main mission is to maintain the company's fleet in airworthiness state and provide, as possible, maintenance work to third parties. MAMC cost accounting practices performed according to the complete cost method, primary for annual budget determination. The allocating key used for overhead cost is ***Flight Time*** Company's airline fleet.

3.1 Analysis of the actual aircraft maintenance cost calculating

Through an example, we will illustrate the actual methodology used to calculate the aircraft maintenance.

Table 1: Aircrafts consumed charges

Costs types	*Annual Costs*	*Boeing 737-400*	*Boeing 737-700*	*Boeing 737-800*	*Total*
Direct Costs (KMAD)		**71800**	**48100**	**217900**	**337800**
	Base maintenance Consumable Material	3500	2000	10000	15500
	Heavy maintenance Consumable Material	40000	30000	100000	170000
	Engine maintenance	4000	3000	5500	12500
	Aircraft equipments maintenance package	4000	3000	2000	9000
	Direct labour	20000	10000	100000	130000
	Aeronautical material rent	300	100	400	800
Overhead Costs (KMAD)					**222000**
	Engineering and aircraft documentation				7000
	Indirect labour				60000
	Depreciation				80000
	General expenses				75000
Total Cost (KMAD)					**559800**

Assume that the company's fleet consists of three aircrafts types. These aircrafts receive distributed expenses (**Table 1**). Note that at this level the overhead costs are not distributed. After annual flight time accounting for each aircraft type is done, a distribution coefficient is determined and used as overhead costs allocation key among aircrafts types (**Table 2**). The overhead costs distribution is as follow (**Table 3**). All expenses are then, divided by the accumulated flying time to get the full cost per flight hour (**Table 4**).

Table 2: Allocation key

Aircraft Types	***Mean annual flight Hours (1)***	***Coefficient% (1)/(2)***
Boeing 737-400	30000	**20**
Boeing 737-700	50000	**33**
Boeing 737-800	70000	**47**
Total Annual Flight Hours (2)	**150000**	**100**

Table 3: Aircrafts consumed charges after allocation

Costs types	***Annual Costs***	***Boeing 737-400***	***Boeing 737-700***	***Boeing 737-800***	***Total***
Direct Costs (KMAD)		**71800**	**48100**	**217900**	**337800**
	Base maintenance Consumable Material	3500	2000	10000	15500
	Heavy maintenance Consumable Material	40000	30000	100000	170000
	Engine maintenance	4000	3000	5500	12500
	Aircraft equipments maintenance package	4000	3000	2000	9000
	Direct labour	20000	10000	100000	130000
	Aeronautical material rent	300	100	400	800
Overhead Costs (KMAD)		**44400**	**73992,6**	**103607,4**	**222000**
	Engineering and aircraft documentation	1400	2333,1	3266,9	7000
	Indirect labour	12000	19998	28002	60000
	Depreciation	16000	26664	37336	80000
	General expenses	15000	24997,5	35002,5	75000
Total Cost (KMAD)		**116200**	**122092,6**	**321507,4**	**559800**

Table 4: Cost per flight hour

Annual Costs	***Boeing 737-400***	***Boeing 737-700***	***Boeing 737-800***	***Total***
Direct Costs (KMAD)	71800	48100	217900	**337800**
Overhead Costs (KMAD)	44400	73992,6	103607,4	**222000**
Total Cost (KMAD)	116200	122092,6	321507,4	**559800**
Cost per Flight Hours (KMAD)	**3,87**	**2,44**	**4,59**	**10,91**

Thus, the Boeing 737-800 maintenance cost is 45 900 MAD per flight hour. So for a technical check type A that is done at 400 flight hours interval, it will cost using this method 45 900 x 400 = 18360000 MAD.

3.2 Limits

The cost system, pursued in (MAMC), is based on a narrow management control view. The primary responsibility is to prepare budgets and daily commitments monitoring to ensure that the (MAMC) complies with headquarter financial department given budget. The actual cost system established is based on existing tools to budget the most important positions. These tools come essentially from the general accounting, which is analytical monitoring of aircraft heavy maintenance. Regarding the aircraft base maintenance, the direct cost is determined by type of aircraft is based on the historical and technical support contracts, in order to have the budget for this activity. Therefore, the cost monitoring system is established primary to meet budgetary needs.

Thus, the missing of complete cost system per aircraft maintenance check makes it very difficult to assess, on one hand, the competitiveness of (MAMC) at two levels: The direct costs system through aircraft type does not allow having a benchmarking to study the possibility to outsource expensive maintenance work compared to industry costs. Also The (MAMC) is pursuing a development policy for maintenance work to third parties.

However, the current system does not determine the correct price to charge customers. On the other hand, the use of a single key (Flight Time) for overhead costs allocation without direct causal link with the real consumption of resources, induces a subsidizing effect of aircraft check costs between the different aircrafts types.

4. ABC implementation process

The implementation of ABC method implies to define the process of resources consumption. This will require the following steps [19]:

- Define cost object,
- Define activities that contribute to the achievement of these cost objects,
- Define drivers associated with each activity that measure resources consumption,
- Define relationship between cost object, activities and resources.

4.1 First Step: Define Cost Object

The (MAMC) maintains aircrafts and their equipments. Thus, there are two main processes: Aircraft maintenance and Equipment maintenance. The engineering and industrial logistics functions support these two main processes. Due to huge resources required for modelling all the centre's activities, we limited our focus to the process of aircraft maintenance. The importance of aircraft maintenance activity, which is the primary mission of this centre, motivated this choice.

The various services offered by the aircraft maintenance activity are:

- Aircrafts periodical technical checks (Pre-flight, daily, Weekly, A, C, and D checks).
- Service Bulletins (SB) and the Airworthiness Directive (AD) application for aircrafts;
- Aircrafts systems or structure modifications according to standards approved by the manufacturers and official services (DGAC, FAA, EASA)[1]
- Structural repairs in accordance with approved standards
- Aircrafts engines removal and installation.

Our study (cost object) will focus on ***the standard technical check (Type A)*** which forms the basis of aircraft maintenance. The engineering support intervention is, in general, at non-standard services (application of SB, AD...) therefore, it is excluded from our study field. After determining the cost object, comes the activities research that contribute to the achievement of the latter.

4.2 Second Step: Define the activities that contribute to the achievement of the cost object.

The different organization activities definition is an important step in the method. It is indeed, cutting the production organization to activities that will define the resources process consumption by services. Three steps can be set for the completion of a technical check (**Fig.3**):

- The technical check preparation;
- Work execution.
- Updating information system.

Once we well assimilated our study domain, we targeted personnel who were involved in aircraft maintenance process. We also, developed questionnaires in addition to what we collected through the procedures organizing the tasks and roles of each actor in the process. These data led to a list of activities called the initial list (Example **Table 5**).

[1] DGAC=Direction Générale d'Aviation Civil (Morocco), FAA=Federal Aviation Administration (USA), EASA= European Aviation Safety Agency (Europe).

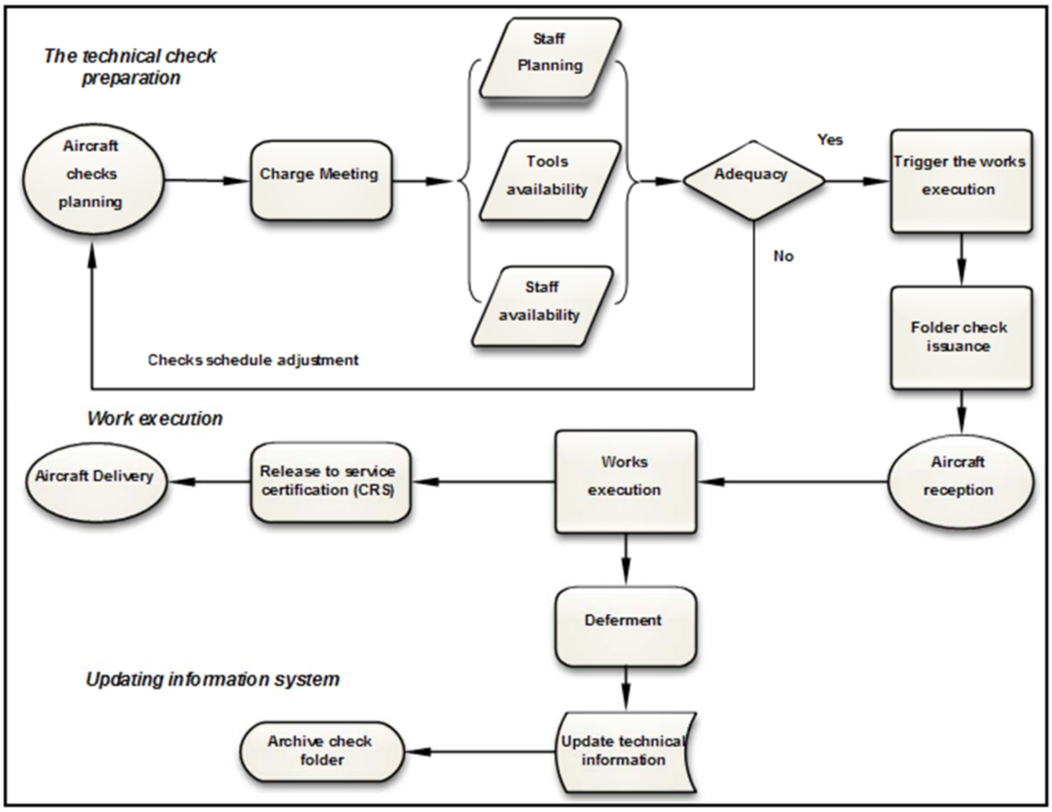


Fig. 3 Aircraft check process

Table 5: Initial list example

Activities	*Tasks*	*N*
Tools preparation	Manage industrial tools and resources (Docks, Stepladders, etc.).	44
	Manage the audit operations, periodic maintenance and calibration of tools.	45
	Manage tools Stores.	46
	Distribute tools.	47
	Monitor and ensure the terms of stores tools entry.	48
	Ensure the storage conditions and ensure the physical removal of resources and equipments in storage limit.	49
	Control the movement between stores tools.	50
	Ensure the provision of tools stores.	51
Aircraft maintenance	Perform pre-flight and Transit operations.	52
	Carry out daily, Weekly, A, C, and D checks.	53
	Achieve the planned interventions.	54
	Troubleshoot aircraft in operation.	55
	Carry out cabin configuration change for special flights.	56
	Wash the aircraft exterior.	57
Aircraft cabin maintenance	Carry out maintenance of seats, carpets, plastic rugs and curtains.	58
	Performs aircraft toilet servicing.	59
	Install and remove the stretchers and provide oxygen to patients.	60
	Perform cabin version change.	61
Aircraft structure maintenance	Repair the structure of aircraft in operation and in checks.	62
	Overhaul and repair removed aircraft structural elements in the workshop.	63
	Carry out aircraft welding and painting.	64

We follow an analysis phase. The goal is to make a synthesis and reduce the number of activities by eliminations and consolidations. The elimination criterion, is Pareto law (20/80) which is usually checked and it allows to select only the most significant activities in view of their relative importance [20]. We grouped some activities from the initial list because they involve the same notion of activity and use identical ways at the distribution of the checks. Unlike, we eliminated others because they represent too small cost that will have little impact on the profitability of services. Hence, we carried out the ***final list of activities***.

Once the initial list of activities defined, several interesting documents can serve as a guide including the summary of activity (**Table 6**). For each activity, and through questionnaires and interviews with operational personnel, we have defined each activity, knowing what triggered it (the supplier), who executed the next activity (the customer) and finally the cost driver. After establishing the list of activities that contribute to the achievement of cost object and the summary sheet for each activity we deduced the cost drivers.

Table 6: Control the aircraft activity summary example

Activity name	***Control the aircraft***
Definition	Control maintenance tasks such as run up, boroscope, adjustments and tests.
	Assist production entities.
Suppliers	Engineering.
Clients	Production.
	Exploitation.
Cost driver	Number of inspections carried out.

4.3 Third Step: Define the drivers associated with each activity that will measure resources consumption

The cost drivers' choice is critical, because they will determine the activity overall cost and the share to be assigned to a particular service. This determination is made using analysis cause / effect methods as Ishikawa Fishbone diagram and Pareto Flowcharts. After achieving the first two steps, the analysis reveals the following information (**Table 7**, **Part A**).

Table 7: Cost drivers

	(A)	**(B)**	**(C)**	**(D)**
Activities	***Selected Cost drivers***	***Annual Volume (1)***	***Cost (MAD) (2)***	***Unit Cost (2) / (1)(MAD)***
Aeronautical purchase	Number of control	10000	432000	**43.20**
	Number of Consumable references	20000	144000	**7.20**
	Number of equipments	5000	600000	**120.00**
	Number of parts supplied	10000	1500000	**150.00**
Production	Workforce Time (aircraft)	300000	34000000	**113.33**
Support	Number of aircraft scheduled	500	700000	**1 400.00**
	Number of work Entries (launching)	3400	1400000	**411.76**
	Number Material Items prepared	6000	1200000	**200.00**
	Number of references identified	5000	1800000	**360.00**
	Workforce Time (tools)	15400	7000000	**454.55**
	Number of job cards made	1000	1600000	**1 600.00**
	Number of documents sequenced	180000	2500000	**13.89**
	Number of inspections carried out	12000	7500000	**625.00**
	Number of work entries (monitoring)	8000	800000	**100.00**

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

4.4 Last Step: Define the relationship between cost object, activities and resources

The main difficulty encountered in the (MAMC) financial department, is the lack of cost accounting on which we can rely to evaluate the activities. However, the budget control service was useful in this evaluation. Two types of costs will affect the products or services cost price:

- Direct costs allocation: consumable materials.
- Overhead costs allocation: activities.

The allocation of aircraft equipment consumables, repairable and revisable to the technical check can be obtained through the Technical Information System, where every aircraft equipment use is charged to the concerned check (A, C or D) and to the concerned aircraft (Boeing 737-400,700 or 800). The second phase of services cost calculation consists of allocating the cost of activities to the cost objects. The method presupposes that all information is available for the definition of various relationships. For our cost object ***the standard technical check (Type A)***, the procedure follows two sub steps: The first one is to calculate the activities' cost driver quantities consumed by cost object during one year (**Table 7**, **part B**). The second sub step follows the same logic as the previous one, but at this time, it is at activities level. The aim is to determine the resources consumed by different activities (**Table 7**, **part C**). Finally, the unitary cost of each cost driver is calculated (**Table 7, part D**). At the end, the Boeing737-700/800 ***Aircraft standard technical check (Type A)*** cost-calculating table can stand as follows (**Table 8**):

Table 8**:** Technical Check (type A) cost for aircraft Boeing 737-700/800

Activities	***Cost Nature***	***Measure Unit***	***Used Quantities***	***Unit Cost***	***Total Cost (MAD)***
(Direct costs)	Aeronautical consumable materials				**10 000.00**
(Direct costs)	Overhauled repairable materials (maintenance costs only)				**950 000.00**
Production (Direct costs)	Maintain the aircraft	Workforce time	200	113.33	**22 666.67**
Aeronautical purchase (Overhead costs)	Establish orders	The order	50	43.20	**2 160.00**
	Receive Orders	The order	40	43.20	**1 728.00**
	Manage Consumables	The reference	40	7.20	**288.00**
	Managing Service and Repair	Equipment	10	120.00	**1 200.00**
	Parts store	The part	40	150.00	**6 000.00**
Support (Overhead costs)	Develop the program of visits	The plane	1	1 400.00	**1 400.00**
	Launch works	Work entry	20	411.76	**8 235.29**
	Prepare the materials	The item	40	200.00	**8 000.00**
		Reference	20	360.00	**7 200.00**
	Prepare tools	Time workforce	8	454.55	**3 636.36**
	Prepare job cards	Job card	4	1 600.00	**6 400.00**
	Plan and monitor production	Working paper	70	13.89	**972.22**
	Control the aircraft	Control	5	625.00	**3 125.00**
	Enter works	Work entry	40	100.00	**4 000.00**
Total cost					**1037011.55**

The (**Table 8)** presents the (MAMC) expenses account as a list of valued activities. This new reading will lead naturally to resource allocation analysis against the objectives of the organization. Thus, the application of activity-based costing leads to the determination of overhead costs, resources consumer, on which we focus our attention. For example, to determine the accurate ***check A*** cost of Boeing 737-800 for a client, we can easily see the cost articulation of activities and remove those who aren't used to perform this check. For this case we can remove the activities "Develop the program of visits and Prepare job cards" cost and know the charge for every additional aeronautical purchase order requested. With this data, the manager gains a greater overall understanding of the expenses and can compete by having a reliable information system on the costs of services offered by his organization.

5. Conclusion

Reliable cost information is a powerful tool in increasing organization's profitability and competitiveness. Traditional cost accounting methods, which use a single key, such as flight time, for overhead costs allocation without direct causal link with the real resources consumption, can induce a subsidizing effect between various organization's services offered. This paper offers a new method for calculating costs according to the ABC approach for aircraft maintenance activity and helps the decision makers to determine the price of each service based on the allocation of overheads. This study takes a Moroccan Airline Maintenance Centre as a case study.

References

[1] IATA,2012. Strong traffic growth in 2011 reflects improved global economic climate. Available at. http://www.icao.int/Newsroom/News%20Doc%202012/PIO.28.11.EN.pdf

[2] Yu-Hern, Chang and Ei-Chi, Shao,"Operating cost control strategies for airlines", African Journal of Business Management, Vol. 5 (26), No.,2011,pp.10396-10409.

[3] Berrittella, Maria, La Franca, Luigi, and Zito, Pietro,"An analytic hierarchy process for ranking operating costs of low cost and full service airlines", Journal of Air Transport Management, Vol. 15, No.,2009,pp.249-255.

[4] Boeing Commercial airplanes,2001. Reduced maintenance costs for the 737-600/-700/-800/-900 family of airplane. Available at. http://www.boeing.com/commercial/aeromagazine/aero_15/costs.html

[5] O'Connell, John F,"The rise of the Arabian Gulf carriers: An insight into the business model of Emirates Airline", Journal of Air Transport Management, Vol. 17, No.,2011,pp.339-346.

[6] Antonio, abissad,2009. The ABC/M method: activities management for cost pilot. Available at. http://www.journaldunet.com/solutions/expert/36263/la-methode-abc-m---gestion-par-les-activites-pour-un-pilotage-des-couts.shtml

[7] Cooper, Robin,"The rise of activity-based costing-Part 1: What is an activity-based cost system?", Journal of Cost Management, Vol. 2, No.2,1988,pp.45-54.

[8] Cooper, Robin,"The rise of activity-based costing-Part 2: When do I need an activity-based cost system?", Journal of Cost Management, Vol. 2, No.3,1988,pp.41-48.

[9] Cooper, Robin and Kaplan, Robert S,"How cost accounting distorts product costs", Management Accounting, Vol. 69, No.10,1988,pp.20-27.

[10] Cooper, Robin and Kaplan, Robert S,"Measure costs right: Make the right decisions", Harvard Business Review, Vol. 66, No.5,1988,pp.96-103.

[11] Cooper, Robin,"The rise of activity-based costing-Part 3: How many cost drivers do you need and how do you select them?", Journal of Cost Management, Vol. 2, No.4,1989,pp.34-46.

[12] Cooper, Robin,"The rise of activity-based costing-Part 4: What do activity-based cost systems look like?", Journal of Cost Management, Vol. 3, No.1,1989,pp.38-49.

[13] Johnson, H Thomas, and Kaplan, Robert S, Relevance lost: The rise and fall of management accounting, Boston : Harvard Business School Press, 1987.

[14] Kaplan, R. S, and Cooper, R, Cost and effect: Using integrated cost systems to drive profitability and performance, Boston : 1998.

[15] Mahmood, Shafiee, Golriz, Amooee, and Yaghoub, Farjami,"Developing an Activity-Based Costing Approach to Maximize the Efficiency of Customer Relationship Management Projects", International Journal of Computer Science Issues, Vol. 9, No.3,2012,pp.221-229.

[16] Tsai, Wen Hsien and Kuo, Lopin,"Operating costs and capacity in the airline industry", Journal of Air Transport Management, Vol. 10, No.4,2004,pp.269-275.

[17] Tsai, Wen Hsien and Hsu, Jui Ling,"Corporate social responsibility programs choice and costs assessment in the airline industry- A hybrid model", Journal of Air Transport Management, Vol. 14, No.4,2008,pp.188-196.

[18] Roztocki Narcyz, Porter, J David, Thomas, Robin M, and Needy, Kim LaScola,"A Procedure for smooth implementation of Activity-Based Costing in small companies", Engineering Management Journal, Vol. 16, No.4,2004,pp.19-27.

[19] Cargenello, Charles Emmanuelle, Morard, Bernard, and Trahand, Jacques, Management accounting: Cost, activity, distribution, Grenoble : 2000.

[20] Lorino, Philippe, Performance methods and practice , the piloting guide, Paris : 1997.

Influence of parameters variation on a shielded line submitted at a punctual injection current at any position along its length

Kaoutar Senhaji Rhazi[1], Mohamed Najmouddine[2]

[1] **ENSEM's Lab of electrical engineering Km 8, road El Jadida B.P 8118 Oasis Casablanca, Morocco,**

[2] **ENSEM's Lab of electrical engineering Km 8, road El Jadida B.P 8118 Oasis Casablanca, Morocco,**

Abstract

Firstly, I note that this draft article is an article continuation already published in your journal in September 2012; n°IJCSI-2012-9-5-3873. This is the reason why you will find repeated diagrams and explanations in order to introduce new work. In reference to articles previously published, one in your journal and two others articles already published in the Annals of telecommunications journal, which we will quote a summary in the following paragraph again, this paper provides a study of the factors influence on a shielded line submitted at a punctual injection current at any position along its length. In the previous article, published in your journal, we have studied the influence of four of them. This study helps to study the influence of four other parameters. Using simulated line for geometric dimensions respecting the line theory assumptions, we try to analyze the sensitivity to changes in various parameters.

Keywords: *transmission line, electromagnetic disturbance, modeling, punctual injection current, parametric sensitivity*.

Summary

1. Introduction

The design of low-level circuits involved in control systems in the large industrial sites has been for many years, discussions subject of the choice of cable types to use and especially the mode of grounding their shields. Indeed, the cables that connect low electronic components are particularly sensitive to various sources of electromagnetic interference. For more clarification on the this article's finality, we can refer to articles previously published, by the same team in your journal entitled "study of parametric sensitivity of the general model of transmission line" in September 2012; number IJCSI-2012-9-5-3873; and in the annals of telecommunications, respectively entitled "Approached expressions at the low and high frequencies of the induced tensions of common-mode at the end of a shielded line submitted to a punctual injection of current" in March 2005 (n ° 3/4) and "Determination of the disturbance at the ends of a shielded line subject to a punctual excitation current", in November 2002 (n°11/12). Transfer impedance is the quantity which allows the quantitative assessment of the quality of shielding a cable. We also know that the modeling project of a shielded line subject to current disruption in any position of its length, involves several factors. Some of these factors have a major influence on the assessment of parasitic tensions, others are probably less important, so it is necessary to determine the effect on the result for each of them, to conduct a study on parametric sensitivity.

The digital model validated by the previous experiences will allow us to make this study and assess the influence obtained by the variation of each parameter characterizing a line with two strictly coaxial conductors above a perfectly conducting plane of ground. We are still interested in the case of a screen characterized by a poor electromagnetic protection that may form for example a metallic ribbon helically wrapped. This time, we will explore additional parameters in addition to those discussed in the previous article published in your journal.

1.1. Summary of previously published articles

Two articles mentioned above were published in the “telecommunications’ annals” journal. For the article published in November 2002 (n°11/12), we were able to determine the perturbation induced at the ends of each coaxial line on a wide band of frequencies. The source of disruption is a punctual direct current injection on the shielding of the cable in any position. For that, we applied the theory of state variables, on the base of approximation quasi-TEM, to model the coaxial line over a plan of ground with a finite conductivity. This allowed us to provide expressions of tensions and currents at the two ends of the line under their analytic forms for any load conditions. Then, an experimental verification was made to validate the developed code of calculation. Concerning the article published in March 2005 (n°3/4), we were able to determine approached expressions at the low and high frequencies of the induced tensions of common-mode at the end of a shielded line submitted to a punctual current injection, for the case of a helical ribbon. This article also discusses the correction function, the relationship between the exact value and the approximated one of the disturbance and the limit of validity to the high frequencies approached expressions established at low frequencies. For the Article published in your journal in September 2012 (n°IJCSI-2012-9-5-3873°, the transfer impedance is the size that allows quantitative assessment of the quality of a shielded cable. The numerical model validated by various previous studies to perform this study and evaluate the influence obtained by the variation of each parameter characterizing a line with two coaxial conductors strictly above a perfectly conducting ground plane. In this paper (published in September 2012), we studied the influence on the module Z_{tg}(transfer impedance overall), for four parameters:

-The length ℓ_2 of the line portion, located between the point of injection and the left end of the line
-The total coaxial line length $\mathcal{L}$ of the coaxial line
-The Transfer resistance R_T of outside driver of the coaxial line
-The Transfer inductor L_T from outside driver of the coaxial line

In this paper, we will study the influence of other parameters, namely:

-The inner radius “a” of the shield of the coaxial line
-The height h_R of the coaxial line
-Resistance Z_R the inner conductor of the coaxial line
-Relative permittivity ε_{r1} of the dielectric within the coaxial line
-Relative permittivity ε_{r3} of the dielectric outside the coaxial line

1.2. Geometrical characteristics and configurations of excitation of the model of line of transmission

Always in reference to the articles previously published, the type of shielded transmission line that we consider as the basic circuit includes two coaxial conductors (length $\mathcal{L}$) above a perfectly conducting plane of ground. This condition on the conductivity of the plane of ground is usually required to not further complicate propagation conditions that can occur at high frequencies. The transverse dimensions and the position of the disturbance are represented on figure 1.

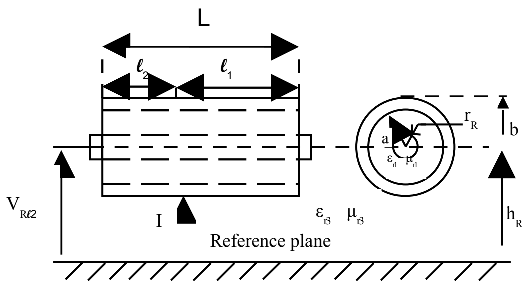


Fig. 1 Geometric configuration of the transmission line studied

Always remember that:

ε_{r1} and μ_{r1} are respectively the permittivity and permeability of the inner dielectric environment, and ε_{r3} and μ_{r3} are the corresponding values of the outer dielectric. We consider that μ_{r1} and μ_{r3} are those of the vacuum and the dielectric environment outside the cable is the open air. $V_{R\ell 2}$ is the induced voltage of common mode on the internal conductor to the left end of the line, and I is the total disruptive current. $\mathcal{L}$ is the total length of the line, h_R is the height at which the line is from the ground; a and b are respectively the inside and outside rays of the external conductor and r_R is the radius of the inner conductor. ℓ_1 and ℓ_2 are the portions of the line between which the disruptive power is.

The values of the parameters presented in the table below (already given in the previous article, March 2005) are chosen arbitrarily but comply with the basic theory of transmission lines assumptions.

Table 1: reference values

Reference values used for lineic parametric calculation	
Values for which we study the sensitivity	Values kept constant
L =100 m ℓ_2 = 5 m a = 4 mm b-a = e= 0.2 mm h_R =50 mm Z_R = 0.01 Ω/m ε_{r1} = 2 ε_{r3} = 1	r_R = 0.7 mm μ_{r1} = 1 μ_{r3} = 1

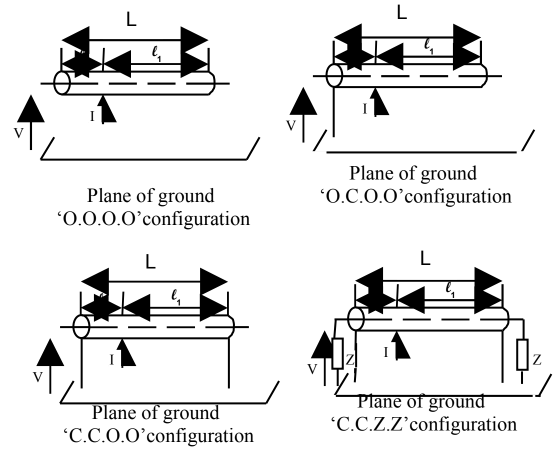


Fig. 2 Configurations of excitation considered in the simulations

In this study, they represent the reference's values that correspond to a reference's curve related to a definite load configuration.

To simplify this study, we made the assumption that the resistance of the conductor is a constant. We are interested in the case of a shielding that strongly promotes the penetration of the electromagnetic field (helical ribbon).

As in the previously published article (March 2005), simulations concern the variations according to the frequency of the Z_{tg} module, that is to say the ratio of the induced voltage of common mode on the inner conductor in the left end of the line to total disruptive current. The left end of the line is supposed to be the end where the disturbance-sensitive instrumentation is.

Three configurations of excitation characterized by conditions at the ends often encountered in practice, will allow us to simulate a transmission line shielded connecting an equipment to another. Loads at the ends of the internal conductor simulate input and output impedance of the equipment. Additional configuration for which shielding is open at both line's ends, is used for comparison and may represent the case of involuntary defect in connections between ground and shielding.
Excitation configurations are represented in figure 2 below.

To highlight the effect produced by each parameter on the result of Z_{tg}, we propose to vary successively each parameter while keeping, the others equal to their reference's values. The results reproduced on all of the plates are calculated between 1 kHz and 10 MHz by setting two additional values around the reference value to each parameter. For more convenience in the analysis of the results, we assembled the effect of a parameter for the four configurations on the same plate. In addition, for a good precision, we increased the number of points calculated by decade of frequency to 150 points, which is not without consequences on the calculation time of each simulation. Furthermore, we retained the four-letter code to designate each configuration. These four letters represent the successive order of load impedances at both line's ends starting with the shielding and from right to left. The letter 'O' corresponds to an open circuit, 'C' to a short circuit and 'Z' to the characteristic impedance of the line formed by two coaxial conductors.

1.3 Case of a row with a cable whose shielding is likened to a helical ribbon above a plane of ground

Always in reference to previously published articles and in order to quantitatively represent the phenomenon of diffusion through the screen of a cable, we consider a poor quality screen made by a helical ribbon and whose transfer impedance Z_T is characterized by the following approximated expression:

$Z_T = R_T + j\omega L_T$ (1)

Like the reference value, the resistance value in continuous current R_T is chosen equal to 10mΩ/m and the inductor L_T is taken equal to 16nH/m. Such values are arbitrary and should not be interpreted as obtained from the geometric dimensions from the line; in other words, they are independent from the diameter and the thickness of the screen.

1.3.1. Parametric sensitivity

Figures 3 to 7 give below, provide the calculated values of Z_{tg} module by successively varying each parameter for the four load configurations.

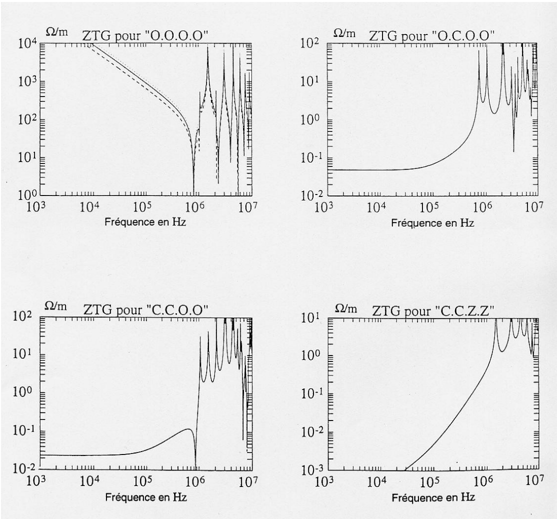


Fig. 3 Influence, on the Z_{tg} modulus, of the radius a of the coaxial line shield
(_____ :a=4 mm),(------ : a= 10 mm) et (….. a= 2 mm)

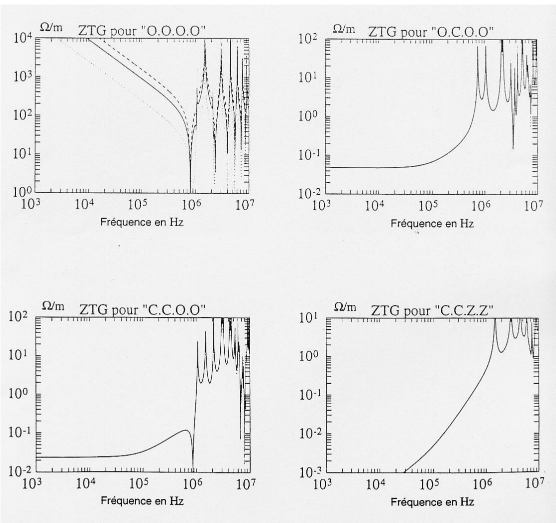


Fig. 4 Influence, on the Z_{tg} modulus of the coaxial line height h_R
(_____ : h_R=5cm), (----- :h_R = 50 cm) et (…… h_R=0,5 cm)

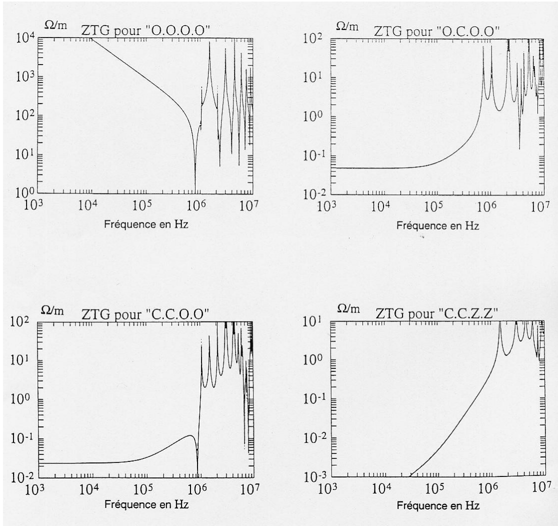


Fig. 5 Influence, on the Z_{tg} modulus, of Z_R resistance of the coaxial line inner conductor
(_____ : Z_R=10mΩ/m), (----- :Z_R=100mΩ/m) et (….. : Z_R= 1mΩ/m)

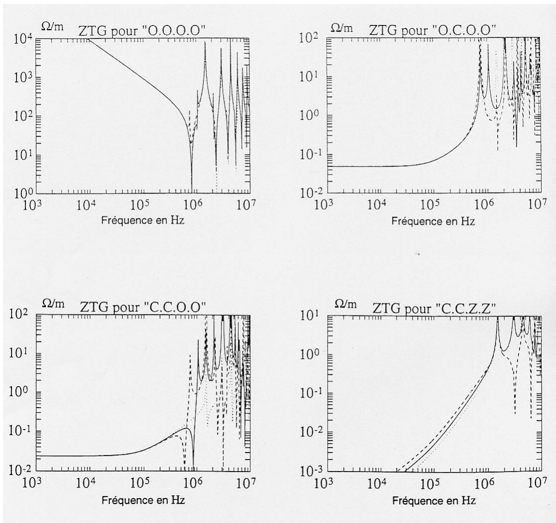


Fig. 6 Influence, on the Z_{tg} modulus, of the dielectric relative permittivity ε_{r1} inside the coaxial line
(____ : ε_{r1}=2), (---- : ε_{r1}=4) et (…. :ε_{r1}=1)

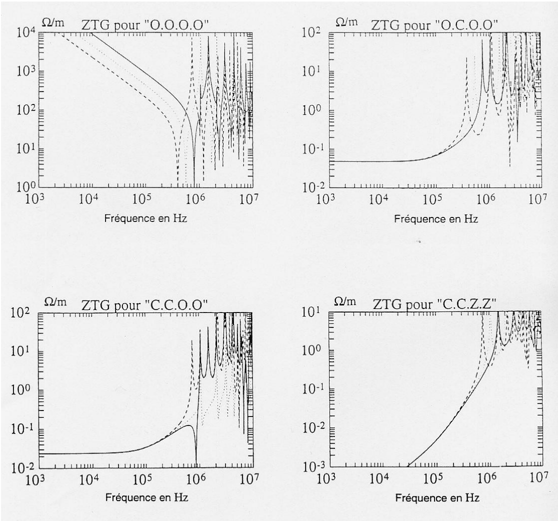


Fig. 7 Influence, on the Z_{tg} modulus, of the dielectric relative permittivity ε_{r3} outside the coaxial line
(____ : ε_{r3}=1), (---- : ε_{r3}=4) et (…. :ε_{r3}=2)

1.2.2. Discussion

The results presented in Figure 3 (variation of the space between the inner conductor and the shield) show that in this case where the transfer impedance is defined as an independent function of this space, this parameter only affects the configuration in which the shield is open circuited at both ends. It is obvious that for this configuration where the coupling is essentially capacitive type, the Z_{tg} module is determined by the line admittance formed by the shield and the ground plane and therefore increases when the admittance decreases. The same is true for the line height variation (Fig. 4), but especially when at least one end of the shield is grounded the height parameter does not affect the result.

On the other hand, respecting the realistic values for the resistance of the inner conductor, Figure 5 shows that this resistance has no significant effect on the calculation results because of its negligible value compared to the terminal impedances of common mode considered in this study. However, it gives rise to an attenuation constant that can absorb some resonance peaks at higher frequencies must take place where the skin effect which obviously neglected in this study.

The relative permittivity influence of the inside dielectric medium ε_{r1} cable on the simulation results obtained for the four configurations shown clearly in Figure 6.

We note that on low frequencies, the interference voltages amplitudes are independent of the value assigned to this parameter except for the configuration "CCZZ" where impedance characteristic of the line Z = Z_{CR} itself is theoretically defined as a function of the constant, that is to say, the load impedances are different for different values of ε_{r1}.

Indeed, one of the features of particular interest of the transmission lines at high frequencies is the characteristic impedance. A coaxial line for using an insulating material which the relative permittivity is ε_{r1} and the permeability is that of the vacuum, this expression reduces to the expression:

$$Z_{CR} = \sqrt{\frac{L_R}{C_R}} = L_R \frac{v_0}{\sqrt{\varepsilon_{r1}}}$$

(2)

Where L_R, C_R and v_0 are respectively the inductance per unit length, capacitance per unit length and the speed of light.

The same remarks apply to the curves of Figure 7 for the medium dielectric permittivity that surrounds the line.
We also note that ε_{r3} does not affect the low frequencies in the configuration "CCZZ" since in this case ε_{r1} being constant, the Z_{CR} is being too.

These results highlight the relative permittivity effect of dielectric media on the evolution versus frequency of Z_{tg} calculated modules.

Thus, the propagation modes can be considered as quasi-TEM and dielectric media can be assimilated in the open air at low frequencies. By cons, for high frequencies, from solving coupling problems, signal waveforms are very attached to velocities and therefore the values assigned to the dielectric media relative permittivity.

These results confirm the interest of using structures better electromagnetic immunity shielded to protect lines against low external disturbances.

These results show once again that when the cable shield is in the air at both ends, its ability to reduce disturbance is almost zero and we find almost the same common mode voltage in the total absence of shielding.

2. Conclusion

The numerical model validity allowed for a study of parametric sensitivity. In this study, we presented and discussed the numerical simulations obtained results. It was made clear that the weak coupling assumption is still effective. It has also emphasized the need to properly ground the two ends of the cable shielding, in particular the right equipment sensitive to disturbance.

It is also clear from this study that, apart from cases where the shield is in the air at both ends and the position of the resonance peaks, only the transfer impedance, line length and frequency decisive.

Was determined separately influence on the level of interference voltages, some parameters involved in the modeling of a shielded line above a perfectly conducting ground plane.

Among the main parameters, the transfer impedance is of particular interest since we see her appear essential that clearly explains the function screen shielded transmission lines.

3. Bibliography

[1] K. Senhaji Rhazi , M.Najmouddine, “study of parametric sensitivity of general model of transmission line”. IJCSI – September 2012 Issue (Volume 9, Issue 5):IJCSI-2012-9-5-3873
[2] K. Senhaji Rhazi , M.Najmouddine, J. Hoeffelmann, , A. Lamine , "Approximate expressions at low and high frequencies of induced voltages of common mode at the ends of an armored submissive line at a punctual injection of current". Ann. Telecommun.60, n ° 3-4, pp454-471, March April 2005
[3] M.Najmouddine, K. Senhaji Rhazi J. Hoeffelmann, “Determination of disturbances induced at the ends of a line reinforced subjected to point excitation current ", Ann. Telecommun. , 57, n ° 11-12, pp.1091 - 1123, November-December 2002"
[4] SHELKUNOFFE.F (SA) "The electromagnetic theory of coaxial transmission lines and cylindrical shields," Bell System Tech. Day., Pp. 533-579, October 1934.
[5] BLECH (P.), DIJAMATOVIC (Y) and Ianoz (Mr.) "Measurement of the impedance transfer in pulsed mode," Tech. MES, Bulletin ASE / UCS 78, no.. 9, pp. 488-492, May 1987.

4. Biography

I am Mrs. Senhaji Rhazi Kaoutar, an assistant professor at the superior school of technology (EST) in Casablanca, Morocco. I wish to inform you that I'm preparing my habilitation (in the field of electromagnetic compatibility). I completed the national thesis in July 2006. I got the research preparation certificate (CPR) in telecommunications from EMI (Mohammadia School of Engineers Rabat Morocco December 97). I am an engineer of the State owened (EMI) in Rabat, Morocco (in 1991). I have already published two articles in the annals of telecommunications' journal respectively entitled: "Determination of the disturbance at the ends of a shielded line submitted to a punctual current excitation" in the November-December 2002 number and "expressions approached at the low and high frequencies of the voltage induced in the common-mode at a shielded line's ends submitted to a punctual injection of current" in the March / April 2005 number. M. Najmouddine is my supervising professor in the superior normal school of electricity and mechanics (ENSEM Casa Morocco)

Study on Stability of Metal Mine Overlying Strata for Artificial Pillar Support

Kang Zhao[1], Zhongqun Guo[1], Xiaodong Zhao[1], Zeming Zhou[1] and Xiaojun Wang [2]

[1] School of Architectural and Surveying Engineering,Jiangxi University of Science and Technology,Ganzhou 341000,PR China

[2] School of Resources and Environment Engineering, Jiangxi University of Science and Technology,Ganzhou 341000,PR China

Abstract

In this paper, artificial pillar in one gold mine is designed and the stability of overlying strata at its supporting goaf is investigated. From the perspective of stress distribution, the results show that the stress which applied to roof overlying strata and artificial pillar cannot reach damage degree of instability, and the devised artificial pillar can meet the requirement of safety production. However, the stability monitoring at some key positions on the surrounding rock of goaf should be strengthened. It is found from the displacement changes that the larger the goaf span is, the greater the subsidence of roof overlying strata is. In addition, the mining in the next mid section has a great influence on the shifting of roof overlying strata at the goaf of the above mid section.

Keywords: *Artificial pillar; Overlying strata; Metal mine; Stability*

1. Introduction

The pillar is an important, economic and effective way for maintaining stability of surrounding rock and roof control at the goaf in the underground mining support[1-3]. The pillar in the open-stope method is divided into primary ore pillar and artificial pillar. The primary ore pillar is the mineral rock left directly in the underground in ore mining, and should not be mined and reclaimed. The artificial pillar is mostly applied in rare and precious metal mine [4-8]. As mineral ores are of high-grade and valuable, the filling material constructed artificially in certain dimensions is used to replace the mined ore so that it can change the surrounding rock stress at the goaf, support overlying strata and prevent surrounding rock deformation. With ore resources being exhausted, the artificial pillar will be gradually adopted for mining in many ordinary mines in order to fully recover mineral resources and enhance economic performance. Thus, it is of practical significance to systematically study the stability of artificial pillar in supporting overlying strata.

As the artificial pillar in supporting overlying strata is affected by complex factors such as filling material proportioning, design strength and filling construction technology, its supporting effect is inferior to that of the primary ore pillar.

Although some researchers had undertaken studies on supporting role and performance of artificial pillars in underground mining in metal mines and had achieved some research results [9-14], most of researches are focused on construction technology and on-site application of artificial pillar and relatively less research are reported to concentrate on damage dynamic process and displacement changes of overlying strata under the artificial pillar supporting conditions.

2. Model Buiding

The research was done in one gold mine in China. This mine's surface was level in dip angle and +160 meters above sea level, and its ore body was 5 meters thick on average, 90 meters long, and 500 meters deep. It is mined with the open-stope method. In order to maintain production safety and to recover more metal ore, artificial pillar was used to replace primary ore one so as to guarantee stability of surrounding rock at the goaf. Seven mining mid sections were designed for mining, namely: -390m, -410m, -430m, -460m, -500m, -540m and -580m. Currently -390m and -410m mid sections have been mined out, the -430m mid section is actually mined, and the -460m mid section is being developed. As the main shaft of the mine is within the moving belt, it is very important to study on movement pattern of overlying strata in -430m and -460m mid sections with artificial pillar supporting. This will directly affect normal operation and safety of the main shaft. For this reason, based on practical engineering requirements, we devised a model with a size set at 600m×600m×400m, its specific parameters of physical mechanics are shown in Table 1. It has 124614 nodes and 117096 units. The burial depth for the model bottom is 790 meters. The stability of overlying strata under artificial pillar support is mainly studied for -430m and -460m mid sections, and the dimensions of room and pillars are seen in Table 2. Grid is divided with a gradual change mode, and the densest grids are at the center. The

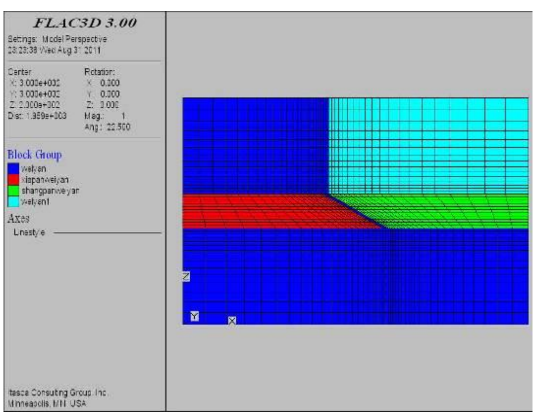
grids are distributed uniformly at the key research area. The model adopts displacement boundary condition: Rolling support
is used around themodel($u_x = 0, u_y = 0$), its bottom is fixed ($u_x = 0$、$u_y = 0$、$u_z = 0$)， and gravity stress for overlying strata at the top boundary $\sigma_{zz} = -10.92Mpa$. With respect to tectonic stress, the horizontal stress is 1.25 times of vertical stress in inclination of ore body($\sigma_x = 1.25\sigma_z$), while the horizontal stress is 0.75 time of vertical stress along ore body ($\sigma_x = 0.75\sigma_z$). The Mohr-Coulomb strain softening principle is applied in calculation. The model is shown as Fig. 1.

Table 1 mechanical parameters of rock mass

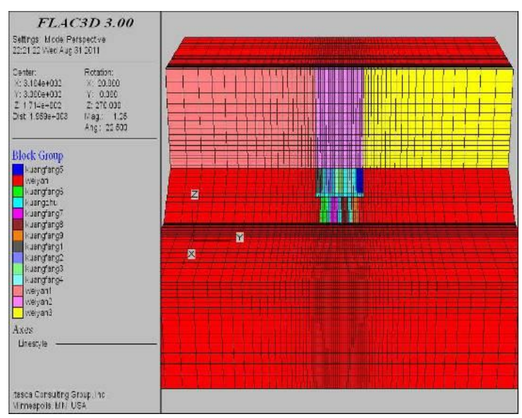
Constituent	*Density (kN/m^3)*	*Elastic modulus (MPa)*	*Cohesion (MPa)*	*Friction angle(°)*	*Poisson's ratio*	*Tensile strength(MPa)*
Surrounding rock	28	60000	15	45	0.2	7.5
Ore body	27.1	65000	15	42	0.19	7.5
Artificial pillar	21	230	0.171	35	0.25	0.01

Table 2 Parameter design for ore block structure of -430m and -460m mid sections

Mid section	*Ore block structure parameter*			
	Number of pillars per section	Pillar width (m)	Number of rooms per section	Room width (m)
-430m	4	6.6	5	11.7
-460m	3	5.8	4	11.9



(a) Division of grids for the model



(b) Cutaway view for the model

Fig. 1 Model diagram of overlying strata with artificial pillar support

3. Results and Anlysis

3.1 Distribution and changes of stress for overlying strata

It is known from the maximum main stresses in Figure 2 that stresses are centralized mainly at the boundary of the mining zone, especially in the surrounding rock area outside of the room at two ends. The centralized stress is 61.03MPa at the lower part of surrounding rock outside of the room of the -460m mid section, the compressive strength is around 75MPa for the surrounding rock. Though it is less than the standard surrounding rock compressive strength, these areas should be monitored in stability and the effective reinforcement measures should be taken. The load-carrying capacity for the artificial pillar itself is also less than compressive strength, the artificial pillar is in a relatively stable state. It is known from minimum main stress in Figure 3 that tensile stress occurs mainly at the top of the room in -430m mid section and in the middle of the baseboard at mining areas of two mid sections. The closer to the mid section at the goaf the room is, the larger tensile stress its top (bottom) board receives, and the bigger the high stress area with a maximum tensile stress at 3.54MPa, which is far less than the tensile strength 7.5MPa of the surrounding rock itself. So the surrounding rock will not be damaged. In addition, it is known from the figure that there is no tensile stress in artificial pillar, which ensures the artificial pillar is in a stable supporting state. It is known from distribution of plastic areas at the goaf roof (Fig. 4) that only sporadic plastic areas exist in the stope after -430m and -460m mid sections are mined out. Artificial pillar itself has no plastic breakdown area, few plastic areas occur in artificial pillar only during process of mining. But artificial pillar gradually enters in a stable stress state with mining, which indicates the structural parameters for devised artificial pillar can be used to effectively prevent plastic area forming and damaging of overlying strata at the goaf. Through analysis of force diagram and distribution diagram of plastic area of overlying strata and artificial pillar itself at goaf, it is known that their compressive stress or tensile stress is less than damage strength in instability and failure strength, which doesn't result in rupture and instability of overlying strata and failure of artificial pillar itself. However, the areas where there is the larger stress at two sides of surrounding rock at the goaf should be monitored in stability, and the reinforcement measure should be taken as necessary.

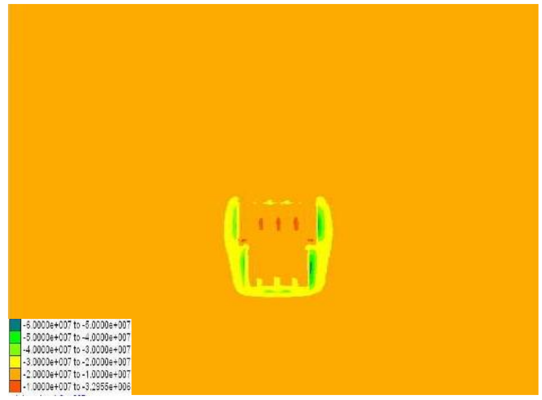


Fig. 2 Contour map of maximum principal stress

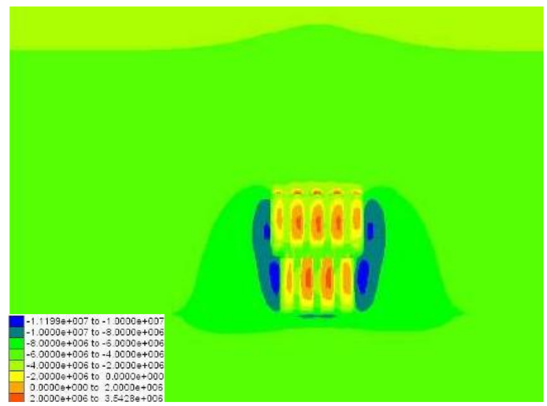


Fig. 3 Contour map of minimum principal stress

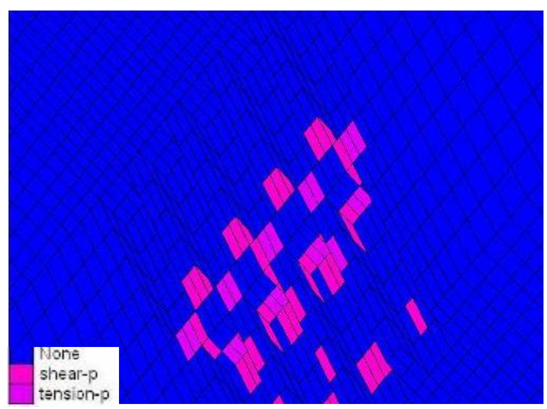


Fig. 4 Plastic zone of the room roof at the goaf

3.2Change of displacement for overlying strata

The change of displacement for overlying strata at the goaf is an important parameter reflecting its stability. Therefore, No. 1-5 displacement monitoring points are set up in the middle of room roof at the goaf of the -430m mid section, and No. 6-9 displacement points are set up in the middle of room roof at -460m mid section (Fig. 5). After the -430m mid section is mined out, the displacement changes of room roof at the goaf of the section are shown in Fig. 6. It is known from Fig. 6 that the room roof in the middle of the entire goaf, i.e. No. 3 monitoring point, presented the largest subsidence of 6.75mm. There is a small subsidence, around 4.24mm for No. 1 and No. 5 monitoring points located room roof of the boundary at two sides of the goaf. This result is consistent with distribution of maximum tension stress analyzed aforementioned. Because rock materials is characterized with compressive resistance rather than tensile resistance, the tensile stress is centralized at the same area where the overlying strata will ruptures easily and results in great displacement of overlying strata.

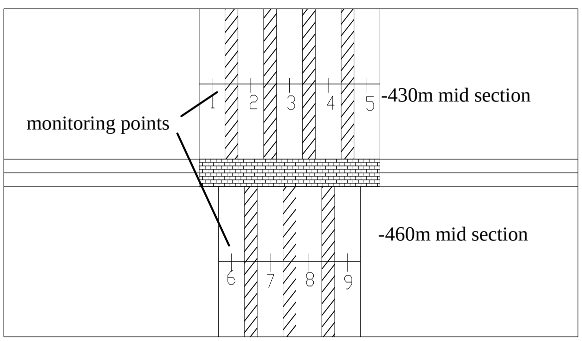


Fig. 5 Distribution of displacement monitoring points of the room roof at the goaf in different mid sections

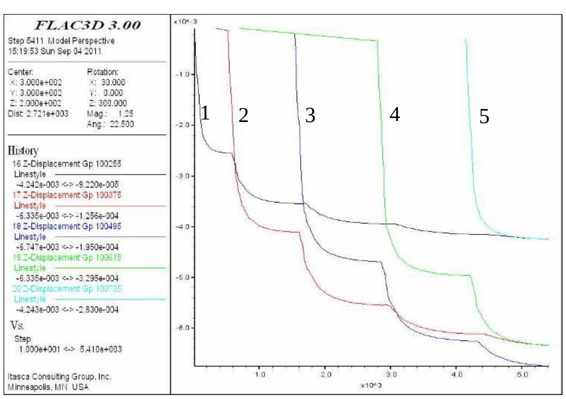


Fig. 6 Displacement change of the goaf roof at No. 1-5 points after the -430m mid section is mined out

After the -460m mid section is mined out, there is a greater subsidence of room roof at Points No. 7 and 8, i.e. 7.76mm or so (Fig. 7), which is about 1mm higher than the maximum subsidence of 6.75mm at room roof after -430m mid section is mined up. This result is caused by actions of gravity stress and tectonic stress.

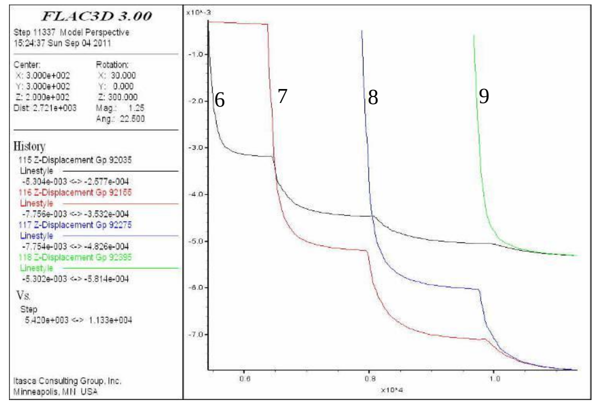


Fig. 7 Displacement change of the goaf roof at No. 6-9 points after the -460m mid section is mined out

3.3 Interaction of displacement changes for the mined overlying strata at different mid sections

After the -460m mid section is mined out, the maximum subsidence of the room roof comes to 8.02mm at No. 3 monitoring point in the -430m mid section (Fig. 8). It is known from Fig. 9 that the displacement of room roof at No. 1-5 monitoring points after -460m mid section is mined up is much larger than the displacement at the corresponding points in -430m mid section after the section is mined up. The subsidence of room roof reaches greatest value after -460m mid section is mined up. when -430m mid section is mined up ,No. 3 monitoring point's subsidence is 6.75mm, while its subsidence goes up to 8.02mm and its displacement increases by 18.8% after the -460m mid section is mined up. In comparison to the subsidence at different monitoring points of foregoing different mid sections, it is found that the maximum subsidence of room roof at -460m mid section is less than that of the -430m mid section. This suggests that the bigger the mining mid section span is, the greater the subsidence of roof at the goaf is, and more unstable the overlying strata are.

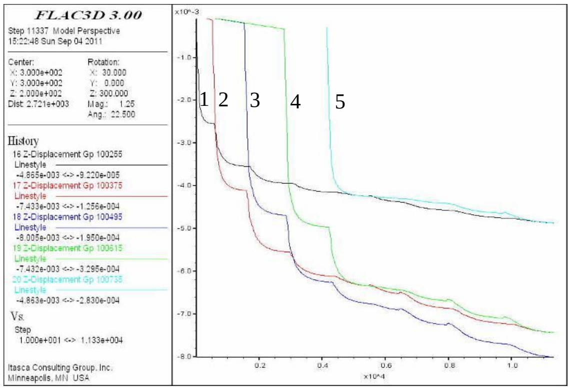


Fig. 8 Displacement change of the goaf roof at No. 1-5 points after the -460m mid section is mined out

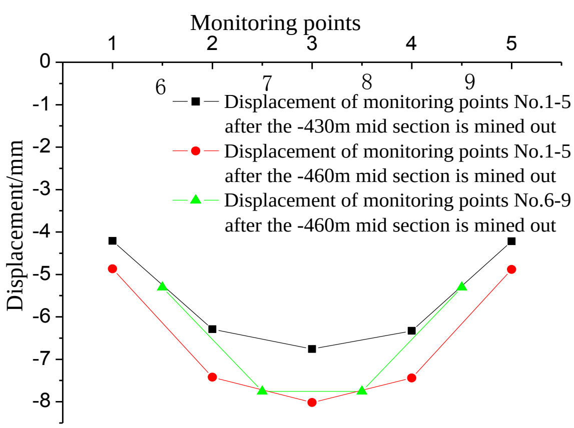


Fig. 9 Comparison of displacements of roof monitoring points at the goaf after different mid sections are mined out

Taking all the calculations into consideration, the mining out of -460mm mid section will have a greater influence on the subsidence of No. 1-5 monitoring points than the mining out of -430m mid section, and their displacement of monitoring points increases by 15.6--18.8%.

4. Conclusions

(1) It is shown by the stress distribution research that the strata will not reach instability state when maximum compressive stress of 61.03Mpa(Compressive strength for surrounding rock is around 75MPa) or maximum tensile stress of 3.54MPa (Tensile strength is 7.5MPa) is applied to roof overlying strata. However, the stability monitoring at some key position of surrounding rock should be strengthened and corresponding reinforcement measures should be taken if necessary. From analysis of stress strength and plastic area distribution, the designed artificial pillar is stable itself and can bear pressure forces from overlying strata and meet requirements of safety in production.

(2) It is found from displacement changes that, attributed to physical and mechanical properties of rock materials, the bigger the goaf span is, the greater tensile stress it bears, and the greater the subsidence of roof overlying strata is in the mining disturbing process.

(3) When two adjacent mid sections with different elevations are mined, the mining for next mid section has a great influence on the shifting of roof overlying strata at the goaf in the above mid section. The mining for the -460m mid section causes 8.02mm subsidence at No. 3 monitoring point, which increases by 18.8% at No. 3 monitoring point, compared with the mining of the -430m mid section.

Acknowledgments

The authors gratefully acknowledge the National Natural Science Foundation of China (Grant 51064010) ,Jiangxi Province Science Foundation for Youths(GJJ12361)and Foundation of Jiangxi University of Science and Technology(jxxj11005) .

References

[1] A. Newman, E. Rubio, R. Caro, A. Weintraub, K. Eurek, "A review of operations research in mine planning", Interfaces 2010, Vol. 3, pp.222-245.
[2] D.R. Tesarik, J.B. Seymour, T.R. Yanske, "Long-term stabilityof a backfilled room-and-pillar test section at the Buick Mine,Missouri, USA", International Journal of Rock Mechanics and Mining Sciences, 2009, Vol. 7, pp.1182-1196.
[3] M. Najafi, S.E. Jalali, A. R. Yarahmadi Bafghi, F. Sereski, "Prediction of the confidence interval for stability analysis of chain pillars in coal mines", Safety Science, 2011, Vol. 5, pp. 651-657.
[4] Kh.I Aglyukov, "Mining of the protective pillars using a packed fill",Journal of Mining Science, 2004, Vol.3, pp.292-297.
[5] M. Monjezi, M. H. Seyed, M. Khandelwal, "Superiority of neural networks for pillar stress prediction in bord and pillar method", Arabian Journal of Geosciences, 2011, Vol.5-6, pp.845-853.
[6] O.V. Ovcharenko, I.I. Ainbinder, K.Yu. Shilin, N.P. Kramskov, "Geomechanical Substantiation of the Parameters for Underground Mining of “Mir” Kimberlite Pipe" ,Journal of Mining Science, 2002, Vol.6, pp.528-533.
[7] A. Mortazavia, F.P. Hassanib, M. Shabania, "A numerical investigation of rock pillar failure mechanism in underground openings",Computers and Geotechnics, 2009, Vol.5 pp.691-697.
[8] J. Deng, L. Bian, "Investigation and characterization of mining subsidence in Kaiyang Phosphorus Mine", Journal of

Central South University of Technology, 2007, Vol.3, pp. 413-417.
[9] A. S. Tawadrous, P. D. Katsabanis, "Prediction of surface crown pillar stability using artificial neural networks", Int. Journal for Numerical and Analytical Methods in Geomechanics, 2007, Vol. 7, pp. 917-931.
[10] Chen Q.F., Zhou K.P., Action mechanism of low-grade backfill on stability of mining environment structure. Rock and Soil Mechanics, 2010, Vol. 9, pp. 2811-2816.
[11] X.J. Wang, X. Feng, T.B. Yang, K. Zhao, K. Zhao, "Reasonable width calculation and analysis of artificial pillar in deep mining", Journal of Mining & Safety Engineering, 2012, Vol. 1, pp.54-59.
[12] J. P. Loui, P. R. Sheorey, "Estimation of non-effective width for different panel shapes in room and pillar extraction", International Journal of Rock Mechanics and Mining Sciences, 2002, Vol. 1, pp.95-99.
[13] J. Deng, Z.Q. Yue, L.G. Tham, H.H. Zhu, "Pillar design by combining finite element methods, neural networks and reliability: a case study of the Feng Huangshan copper mine, China", International Journal of Rock Mechanics and Mining Sciences, 2003, Vol. 4, pp.585-599.
[14] J. Zhou, X.B. Li, X.Z. Shi, W. Wei, B.B. Wu, "Predicting pillar stability for underground mine using Fisher discriminant analysis and SVM methods",Transactions of Nonferrous Metals Society of China, 2011, Vol.21, pp.2734-2743.

Kang Zhao is currently a lecturer in the Mechanics Institute in Jiangxi University of Science and Technology. He received the B.S. Degree from Henan Polytechnic University, in 2003 and the M.S. Degree from Jiangxi university of science and technology in 2006, both majoring in Mining Engineering. He received PH. D degree in Engineering Mechanics from Beijng university of science and technology in 2011. He is now hosting or taking part in the National Natural Science Foundation of China, the Natural Science Foundation of Jiangxi Province, and Jiangxi Province Science Foundation for Youths etc. He has published more than 20 science papers, including 7 ones indexed by EI. His research interests are engineering rock monitoring and stability assessment, numerical methods for geotechnical engineering and computer applications.

Zhongqun Guo received his M.S. Degree in 2012 from Jiangxi university of science and technology, both in Geotechnical Engineering. His research interests focus on stability of engineering rock.

Xiaodong Zhao and Zeming Zhou are both undergraduates in the Geotechnical Engineering of Jiangxi university of science and technology.

Xiaojun Wang is now a Ph.D and lecturer in the Mining Engineering Institute in jiangxi University of Science and Technology.He received his B.S. Degree in 2003 from Henan Polytechnic University, both in Mining Engineering. His research interests focus on the mine safety.

Perform wordcount Map-Reduce Job in Single Node Apache Hadoop cluster and compress data using Lempel-Ziv-Oberhumer (LZO) algorithm

Nandan Mirajkar[1], Sandeep Bhujbal[2] , Aaradhana Deshmukh[3]

[1]Department of Advanced Software and Computing Technologies
IGNOU – I[2]IT Centre of Excellence for Advanced Education and Research
Pune, Maharashtra 411 057, India

[2]Department of Advanced Software and Computing Technologies
IGNOU – I[2]IT Centre of Excellence for Advanced Education and Research
Pune, Maharashtra 411 057, India

[3]Department of Computer Engineering
Smt. Kashibai Navale College of Engineering. Vadgaon (Bk), Off Sinhgad Road.
Pune, Maharashtra 411 041, India

Abstract

Applications like Yahoo, Facebook, Twitter have huge data which has to be stored and retrieved as per client access. This huge data storage requires huge database leading to increase in physical storage and becomes complex for analysis required in business growth. This storage capacity can be reduced and distributed processing of huge data can be done using Apache Hadoop which uses Map-reduce algorithm and combines the repeating data so that entire data is stored in reduced format. The paper describes performing a wordcount Map-Reduce Job in Single Node Apache Hadoop cluster and compress data using Lempel-Ziv-Oberhumer (LZO) algorithm.

Keywords: *Hadoop, Map-reduce, Hadoop Distributed file system HDFS, HBase, LZO*

1.0 Introduction

Hadoop was created by Doug Cutting an employee at Yahoo and Michael J. Cafarella. It was originally developed to support distribution for the Nutch search engine project [12]. Hadoop was inspired by papers published by Google regarding its approach in handling an avalanche of data, and became a standard for storing, processing and analyzing hundreds of terabytes, and even petabytes of data. Hadoop's breakthrough advantages mean that businesses and organizations can now find value in data that was recently considered useless [11]. Hadoop enables a computing solution that is: Cost effective – Due to massive parallel computing approach by Hadoop, there is decrease in the cost per terabyte of storage. Fault tolerant – When a node is missed or a fault arises the system navigates work to another location of the data and continues processing. Flexible – Hadoop is schema-less, and can accept any type of data,structured or not, from any number of sources. Data from multiple sources can be joined and aggregated in arbitrary ways enabling deeper analyses than any one system can provide. Scalable – New nodes can be added as required and added without changing data formats [13].

Compression reduces number of bytes read from or written to HDFS. Compression enhances efficiency of network bandwidth and disk space. HBase is used when need arises for random, realtime read/write access to Big Data [19]. HBase comes with only Gzip compression, compression by GZip is not as fast as Lempel-Ziv-Oberhumer (LZO) compression. For maximum performance LZO is used but HBase cannot ship with LZO because of the licensing issues hence LZO installation is done post-HBase installation [18]. The section 2 of this paper depicts information related to Apache Hadoop , Hadoop distributed file system (HDFS) and its architecture, Map-reduce technique. Section 3 gives details about installation of single node Hadoop cluster on Ubuntu 10.04 server edition open source operating system. Section 3 also gives results of wordcount example using Hadoop which is a Map-reduce job. Section 4 describes installation of Hbase and its usage in Hadoop cluster. Section 5 gives an example of compression of data in Hadoop using Lempel-Ziv-Oberhumer (LZO) algorithm.

2.0 Hadoop

2.1 Apache Hadoop

The framework Apache Hadoop is used for distributed processing of huge data sets known as "Big Data" across clusters of computers using a simple programming model [2][1]. Hadoop permits an application to map, group, and reduce data across a distributed cloud of machines so that applications can process huge data [1]. It can scale up to large number of machines as required for the job; each machine will provide local computation and storage. Apache Hadoop software library itself detects and handles any failures at application layer [2].

2.2 Hadoop Distributed File System - HDFS

A distributed user-level filesystem HDFS—Hadoop Distributed File System written in Java [15] stores huge files across machines in a large cluster. Hadoop DFS stores each file as a sequence of blocks, all blocks in a file except the last block are the same size typically 64 MB [14][15]. Blocks belonging to a file are replicated for fault tolerance. The block size and replication factor are configurable per file. An application can specify the number of replicas of a file. The replication factor can be specified at file creation time and can be changed later. Files in HDFS are "write once" and have strictly one writer at any time [15][16].

2.3 Architecture

HDFS comprises of interconnected clusters of nodes where files and directories reside. An HDFS cluster consists of a single node, known as a Name-Node that manages the file system namespace and regulates client access to files. In addition, Data-Nodes store data as blocks within files [27].

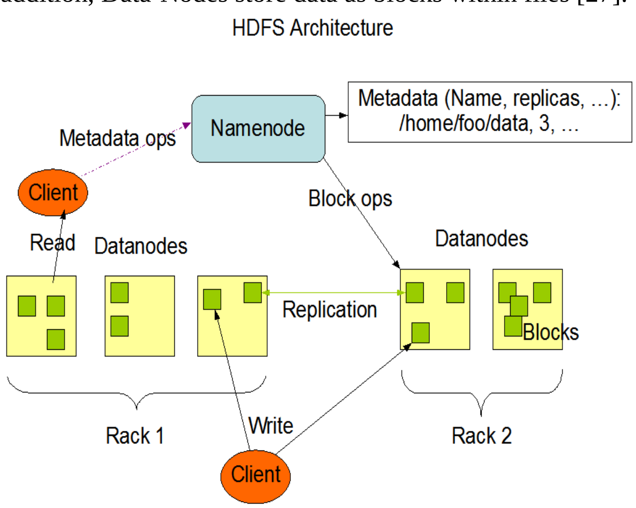


Fig 1: HDFS Architecture [17]

2.4 Name-Node

The Name-Node executes file system namespace operations like opening, closing, and renaming files and directories. The Name-Node does not store HDFS data itself, but rather maintains a mapping between HDFS file name, a list of blocks in the file, and the Data Node on which those blocks are stored. The Name-Node makes all decisions regarding replication of blocks [15][17].

2.5 Secondary Name-Node

HDFS includes a Secondary Name-Node, there is a misconception that secondary Name-Node comes into action after Primary Name-Node (i.e Name-Node) fails. The fact is Secondary Name-Node is continuously connected with Primary Name-Node and takes snapshots of Name-Node's memory structures. These snapshots can be used to recover the failed Name-Node and recent memory structure [12].

2.6 Data-Node

A Data-Node stores data in the Hadoop File System. A functional filesystem has more than one Data-Node, with data replicated across them. On startup, a Data-Node connects to the Name-Node; spinning until that service comes up. It then responds to requests from the Name-Node for filesystem operations. Client applications can talk directly to a Data-Node, once the Name-Node has provided the location of the data [22].

2.7 Job-Tracker

Job-Tracker keeps track of which Map-Reduce jobs are executing, schedules individual Maps, Reduces or intermediate merging operations to specific machines, monitors the success and failures of these individual Tasks, and works to complete the entire batch job. The Job-Tracker is a point of failure for the Hadoop Map-Reduce service. If it goes down, all running jobs are halted [20].

2.8 Task-Tracker

A Task-Tracker is a node in the Hadoop cluster that accepts tasks such as Map, Reduce and Shuffle operations from a Job-Tracker. Task-Tracker is set up with set of slots which depicts the number of tasks it can accept. The Task-Tracker spawns a separate JVM processes to do the actual work. The Task-Tracker supervises these spawned processes, capturing the output and exit codes. When the process finishes, successfully or not, the task tracker notifies the Job-Tracker. The Task-Trackers also transmit heartbeat messages to the Job-Tracker, usually every few minutes, to reassure the Job-Tracker that it is still alive. These messages also inform the Job-Tracker of the number of available slots, so the Job-Tracker can stay up to date with where in the cluster work can be assigned [20].

2.9 Map-reduce

Map-reduce is a Parallel Programming approach used for extracting and analyzing information from unstructured Big Data storage [8]. In Map-reduce their is a map function that processes a key/value pair to generate a set of intermediate key/value pairs, and a reduce function which will immix all intermediate values associated with the same intermediate key[7].

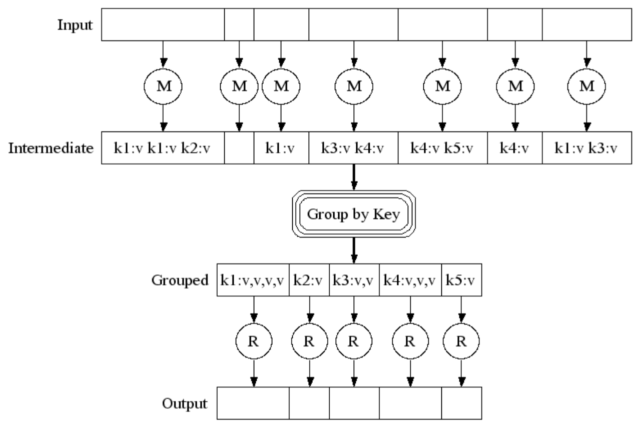


Fig 2: Map-reduce structure [28]

Mapping (M) is done on input data to get intermediate key/value pairs as shown in Figure 2 then this intermediate data is grouped by key e.g all values v with key k1 in one group , all values v with key k2 etc. This grouped data is reduced to give following output i.e. {k1, 4} {k2, 1} {k3, 2} {k4, 3} and {k5, 1}

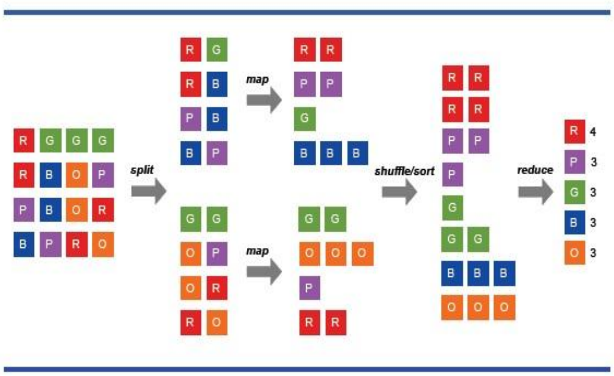


Fig 3: Map reduce example [21]

Above is example of color data on which map-reduce is performed. The 16 blocks are split into two sets with 8 blocks in each and it is mapped to arrange blocks with respect to colors where R is red, P is pink, G is green, B is blue and O is orange. The first set contains 2 red, 2 pink, 1 green, 3 blue and second set contains 2 green, 3 orange, 1 pink, 2 red. These 2 sets are shuffled or sorted to get a single set and is reduced to give following output {R, 4} {P, 3} {G, 3} {B, 3} {O, 3}.

Following is a code for map-reduce: The mapper read-s input records and produces <word, 1> as intermediate pairs. After shuffling, intermediate counts associated with the same word are passed to a reducer, which adds the counts together to produce the sum [8].

```
map(String key, String value)
    {
       for(each word w in value)
     {
    EmitIntermediate(w, 1);
     }
   }
```

In above case key: document name and value: document contents.

```
reduce(String key, Iterator values)
  {
     int sum = 0;
     for(each v in values)
    {
     sum += v;
    }
    Emit(key, sum);
  }
```

In above case key: a word and values: a list of counts.

3.0 Installation of Hadoop

3.1 Install Ubuntu 10.04 server

1.Insert Ubuntu 10.04 server edition and select "Install ubuntu server"
2.Select "configure network manually"
IP address: 192.168.0.216
Netmask: 255.255.255.0
Gateway: 192.168.0.1
Nameserver: 202.138.xx.x
3. Hostname give as 'n1'
4. Fullname for the new user: nandan
5. Software selection: select only "Open SSH server"
6. Install grub boot loader to master boot loader:yes
7. Complete installation and reboot the system
8. Now update the system using following command
sudo apt-get update
9. If desktop is required following command can be used
sudo apt-get install ubuntu-desktop

3.2 Install SUN-Java

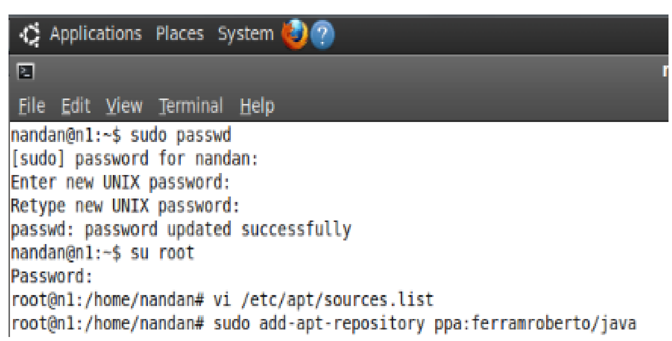


Fig 4: Screenshot of password update and installation of PPA

Login as user “nandan” and change the password of root. For Ubuntu 10.04 LTS, the sun-java6 packages have been dropped from the Multiverse section of the Ubuntu archive. It is recommended to use openjdk-6 instead. If one cannot switch from the proprietary Sun JDK/JRE to OpenJDK, install sun-java6 packages from the Canonical Partner Repository [29]. Any of the repositories can be added or modified by directly editing the files in /etc/apt/sources.list [30] so login as root and open sources.list file and add following lines in it at the end,

deb http://archive.canonical.com/ lucid partner

The JDK has tools needed for developing and testing programs written in the Java programming language and running on the Java Platform [5]. Sun-java6-jdk has been removed from the default Ubuntu 10.04 hence repositories are required, add a PPA (Personal Package Archives) repository then update the source list and install java6-jdk [3].Personal Package Archives (PPA) allows to upload Ubuntu source packages to be built and published as an apt repository by Launchpad [4]. Accept the operating system distributor license for Java during installation.

root@n1:/home/nandan# sudo apt-get update
root@n1:/home/nandan# sudo apt-get install sun-java6-jdk
root@n1:/home/nandan# java -version

3.3 Create Hadoop group and user

Add a group hadoop and user hadoop in same group, choose default for user information for hadoop [31]

root@n1:/home/nandan# addgroup hadoop
root@n1:/home/nandan# adduser --ingroup hadoop hadoop

Now Configuring sudo to allow users in the group "hadoop" to run commands as root [35].

root@n1:/home/nandan# visudo
root   ALL=(ALL) ALL
hadoop ALL=(ALL) ALL

To manage nodes Hadoop requires SSH access, remote machines plus local machine if one want’s to use Hadoop on it. For single node setup configure SSH access to localhost for the hadoop user [6].
root@n1:/home/nandan# su - hadoop
hadoop@n1:~$ ssh-keygen -t rsa -P ""

The above command will create an RSA key pair with an empty password. Empty password is not recommended, but in this case it is needed to unlock the key so that one doesn’t require entering the passphrase every time Hadoop interacts with its nodes [6].

Enable SSH access to local machine with this newly created key.

hadoop@n1:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys

Test the SSH setup by connecting to local machine with the hadoop user here local machine’s host key fingerprint is saved to hadoop user's known_hosts file [6]. Then exit localhost.

hadoop@n1:~$ ssh localhost
hadoop@n1:~$ exit
logout
Connection to localhost closed.

3.4 Install Hadoop software

Download Hadoop from the Apache Download Mirrors and extract the contents of hadoop in /usr/local. Change the owner of all files to hadoop user and hadoop group using *chown* command [6].

nandan@n1:~$ su - root
Password: *****
root@n1:~# cd /usr/local/
root@n1:/usr/local#wget
http://apache.communilink.net/hadoop/core/hadoop-0.20.2/hadoop-0.20.2.tar.gz
root@n1:/usr/local# tar -xvf hadoop-0.20.2.tar.gz
root@n1:/usr/local# chown -R hadoop:hadoop hadoop-0.20.2
root@n1:/usr/local# ln -s hadoop-0.20.2/ hadoop
ln command lets a file/directory on disk be accessed with more than one file/directory name, hadoop can be used instead of hadoop-0.20.2/

Remove the tar file after extraction
root@n1:/usr/local# rm -rf hadoop-0.20.2.tar.gz

3.5 Configure the Hadoop

One problem with IPv6 on Ubuntu is that using 0.0.0.0 for the various networking-related Hadoop configuration options will result in Hadoop binding to the IPv6 addresses of Ubuntu box. One can disable IPv6 only for Hadoop by adding the lines of IPv4 shown in figure to *conf/hadoop-env.sh*. The environment variable that has to be configured for Hadoop is JAVA_HOME. Open *conf/hadoop-env.sh* set the JAVA_HOME environment variable to the Sun JDK/JRE 6 directory

hadoop@n1:/usr/local/hadoop$ vi conf/hadoop-env.sh

export JAVA_HOME=/usr/lib/jvm/java-6-sun
export HADOOP_OPTS="-Djava.net.preferIPv4Stack=true"

As of Hadoop 0.20.x and 1.x, the configuration settings previously found in hadoop-site.xml were moved to core-site.xml (hadoop.tmp.dir, fs.default.name), mapred-site.xml (mapred.job.tracker) and hdfs-site.xml (dfs.replication).

Configure the directory where Hadoop will store its data files, the network ports it listens to, etc. The hadoop.tmp.dir variable can be changed to the directory of own choice. Here the directory is /home/hadoop/cloud. Hadoop's default configurations use hadoop.tmp.dir as the base temporary directory both for the local file system and HDFS [6].If required localhost can be replaced with n1 in following xml files.

hadoop@n1:/usr/local/hadoop$ mkdir ~/cloud
hadoop@n1:/usr/local/hadoop$ vi conf/core-site.xml

```
<configuration>
<property>
<name>hadoop.tmp.dir</name>
<value>/home/hadoop/cloud/hadoop-${user.name}</value>
</property>
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
<description>The name of the default file system.
</description>
</property>
</configuration>
```

hadoop@n1:/usr/local/hadoop$ vi conf/mapred-site.xml

```
<configuration>
<property>
<name>mapred.job.tracker</name>
<value>localhost:9001</value>
<description>The host and port that the MapReduce job tracker runs  at.  If 'local', then jobs are run in-process as a single map and reduce task.
</description>
</property>
</configuration>
```

hadoop@n1:/usr/local/hadoop$ vi conf/hdfs-site.xml

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
<description>Default block replication.
The actual number of replications can be specified when the file is created. The default is used if replication is not specified in create time.
</description>
</property>
</configuration>
```

3.6 The Word-count Map-reduce Job

As it is single node Hadoop cluster with master and slave on same machine, n1 should be mentioned in master and slave files.

hadoop@n1:/usr/local/hadoop$ vi conf/masters
n1
hadoop@n1:/usr/local/hadoop$ vi conf/slaves
n1

The first step in starting up Hadoop installation is formatting the Hadoop filesystem, which is implemented on top of the local filesystems of cluster. This has to be carried out when first time Hadoop installation is done. Do not format a running Hadoop filesystem, this will cause all data to be erased

hadoop@n1:/usr/local/hadoop$ bin/hadoop namenode –format

Now run following command which will startup a Namenode, Datanode, Jobtracker and a Tasktracker on the machine.

hadoop@n1:/usr/local/hadoop$ bin/start-all.sh

Now run the Java process status tool jps to list all processes in Hadoop.
hadoop@n1:/usr/local/hadoop$ jps
5621 JobTracker
5782 TaskTracker
5861 Jps
5545 SecondaryNameNode
5372 DataNode
5200 NameNode

'bin/hadoop dfsadmin' command supports a few HDFS administration related operations.
-report : reports basic stats of HDFS[32]

hadoop@n1:/usr/local/hadoop$ bin/hadoop dfsadmin –report

Configured Capacity: 302827593728 (282.03 GB)
Present Capacity: 283720749071 (264.24 GB)
DFS Remaining: 283720724480 (264.24 GB)
DFS Used: 24591 (24.01 KB)
DFS Used%: 0%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0

\---

Datanodes available: 1 (1 total, 0 dead)
Name: 192.168.0.216:50010
Decommission Status : Normal
Configured Capacity: 302827593728 (282.03 GB)
DFS Used: 24591 (24.01 KB)
Non DFS Used: 19106844657 (17.79 GB)
DFS Remaining: 283720724480(264.24 GB)
DFS Used%: 0%
DFS Remaining%: 93.69%
Last contact: Thu Nov 08 10:09:23 IST 2012

hadoop@n1:/usr/local/hadoop$ bin/hadoop dfs -mkdir datain

lsr[33] Usage: hadoop fs -lsr <args>
Recursive version of ls. Similar to Unix ls -R.
ls -R : lists directory tree recursively.

hadoop@n1:/usr/local/hadoop$ bin/hadoop dfs -lsr
drwxr-xr-x - hadoop supergroup 0 2012-11-08 10:12 /user/hadoop/datain

hadoop@n1:/usr/local/hadoop$ vi data1.txt

Full virtualization provides a complete simulation of underlying computer hardware, enabling software to run without any modification. Because it helps maximize the use and flexibility of computing resources, multiple operating systems can run simultaneously on the same hardware, full virtualization is considered a key technology for cloud computing. For cloud computing systems, full virtualization can increase operational efficiency because it can optimize computer workloads and adjust the number of servers in use to match demand, thereby conserving energy and information technology resources

Above data is taken from [36]
hadoop@n1:/usr/local/hadoop$ vi data2.txt
data2.txt also contains same data1.txt data.

```
hadoop@n1:/usr/local/hadoop$ vi data1.txt
hadoop@n1:/usr/local/hadoop$ vi data2.txt
hadoop@n1:/usr/local/hadoop$ bin/hadoop dfs -put data1.txt datain/
hadoop@n1:/usr/local/hadoop$ bin/hadoop dfs -put data2.txt datain/
hadoop@n1:/usr/local/hadoop$ bin/hadoop jar hadoop-0.20.2-examples.jar wordcount datain dataout
12/11/08 10:27:47 INFO input.FileInputFormat: Total input paths to process : 2
12/11/08 10:27:47 INFO mapred.JobClient: Running job: job_201211081007_0001
12/11/08 10:27:48 INFO mapred.JobClient:  map 0% reduce 0%
12/11/08 10:27:55 INFO mapred.JobClient:  map 100% reduce 0%
12/11/08 10:28:07 INFO mapred.JobClient:  map 100% reduce 100%
12/11/08 10:28:09 INFO mapred.JobClient: Job complete: job_201211081007_0001
12/11/08 10:28:09 INFO mapred.JobClient: Counters: 17
12/11/08 10:28:09 INFO mapred.JobClient:   Job Counters
12/11/08 10:28:09 INFO mapred.JobClient:     Launched reduce tasks=1
12/11/08 10:28:09 INFO mapred.JobClient:     Launched map tasks=2
12/11/08 10:28:09 INFO mapred.JobClient:     Data-local map tasks=2
12/11/08 10:28:09 INFO mapred.JobClient:   FileSystemCounters
12/11/08 10:28:09 INFO mapred.JobClient:     FILE_BYTES_READ=1686
12/11/08 10:28:09 INFO mapred.JobClient:     HDFS_BYTES_READ=1204
12/11/08 10:28:09 INFO mapred.JobClient:     FILE_BYTES_WRITTEN=3442
12/11/08 10:28:09 INFO mapred.JobClient:     HDFS_BYTES_WRITTEN=596
12/11/08 10:28:09 INFO mapred.JobClient:   Map-Reduce Framework
12/11/08 10:28:09 INFO mapred.JobClient:     Reduce input groups=61
12/11/08 10:28:09 INFO mapred.JobClient:     Combine output records=122
12/11/08 10:28:09 INFO mapred.JobClient:     Map input records=22
12/11/08 10:28:09 INFO mapred.JobClient:     Reduce shuffle bytes=1692
12/11/08 10:28:09 INFO mapred.JobClient:     Reduce output records=61
12/11/08 10:28:09 INFO mapred.JobClient:     Spilled Records=244
12/11/08 10:28:09 INFO mapred.JobClient:     Map output bytes=1858
12/11/08 10:28:09 INFO mapred.JobClient:     Combine input records=164
12/11/08 10:28:09 INFO mapred.JobClient:     Map output records=164
12/11/08 10:28:09 INFO mapred.JobClient:     Reduce input records=122
hadoop@n1:/usr/local/hadoop$
```

Fig 5: Screenshot of Word-count Map-reduce Job

hadoop@n1:/usr/local/hadoop$ bin/hadoop dfs -cat /user/hadoop/output1/part-r-00000

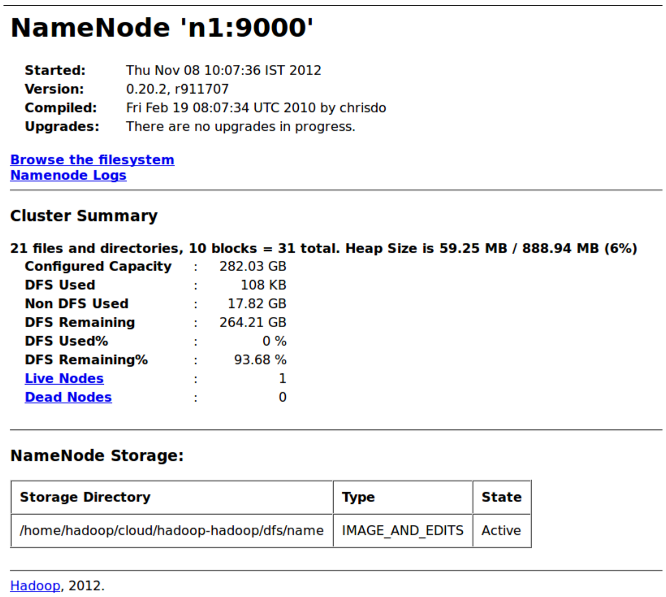
```
Applications  Places  System
hadoop@n1: /usr/
File  Edit  View  Terminal  Help
hadoop@n1:/usr/local/hadoop$ bin/hadoop dfs -cat /user/hadoop/dataout/part-r-00000
Because 2
For     2
Full    2
a       4
adjust  2
and     6
any     2
because 2
can     6
cloud   4
complete        2
computer        4
computing       4
computing.      2
conserving      2
considered      2
demand, 2
efficiency      2
enabling        2
energy  2
flexibility     2
for     2
full    4
hardware,       4
helps   2
in      2
increase        2
information     2
is      2
it      4
key     2
match   2
maximize        2
modification.   2
multiple        2
number  2
of      6
on      2
operating       2
operational     2
optimize        2
provides        2
resources       2
resources,      2
run     4
same    2
servers 2
simulation      2
simultaneously  2
software        2
systems 2
systems,        2
technology      4
the     6
thereby 2
to      4
underlying      2
use     4
virtualization  6
without 2
workloads       2
hadoop@n1:/usr/local/hadoop$
```

Fig 6: Screenshot of output of Wordcount Map-reduce Job

NameNode Web Interface (HDFS layer)

The name node web UI depicts a cluster summary including information about total/remaining capacity, live and dead nodes. Additionally, it allows user to browse the HDFS namespace and view the contents of its files in the web browser. It also gives access to the local machine's Hadoop log files [6].

By default, it's available at http://localhost:50070/.


NameNode 'n1:9000'

Started: Thu Nov 08 10:07:36 IST 2012
Version: 0.20.2, r911707
Compiled: Fri Feb 19 08:07:34 UTC 2010 by chrisdo
Upgrades: There are no upgrades in progress.

Browse the filesystem
Namenode Logs

Cluster Summary

21 files and directories, 10 blocks = 31 total. Heap Size is 59.25 MB / 888.94 MB (6%)

Configured Capacity	:	282.03 GB
DFS Used	:	108 KB
Non DFS Used	:	17.82 GB
DFS Remaining	:	264.21 GB
DFS Used%	:	0 %
DFS Remaining%	:	93.68 %
Live Nodes	:	1
Dead Nodes	:	0

NameNode Storage:

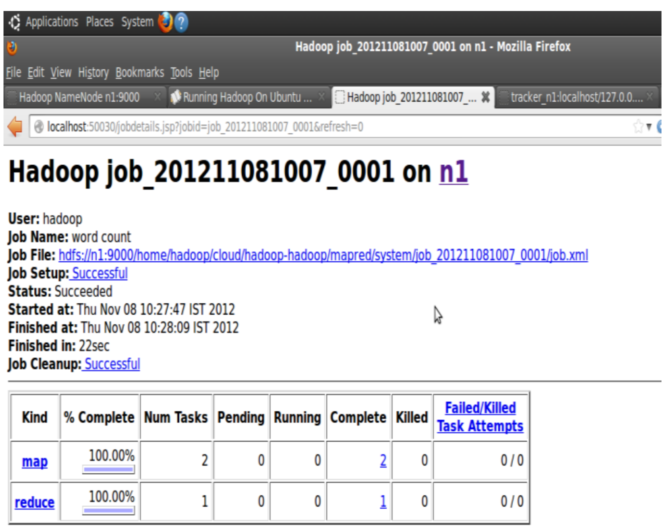
Storage Directory	Type	State
/home/hadoop/cloud/hadoop-hadoop/dfs/name	IMAGE_AND_EDITS	Active

Hadoop, 2012.

Fig 7: Screenshot of NameNode

JobTracker Web Interface (MapReduce layer)

The job tracker web UI provides information about general job statistics of the Hadoop cluster, running/completed/killed jobs and a job history log file. It also gives access to the local machine's Hadoop log files (the machine on which the web UI is running on)[6].
By default, it's available at http://localhost:50030/.


Hadoop job_201211081007_0001 on n1

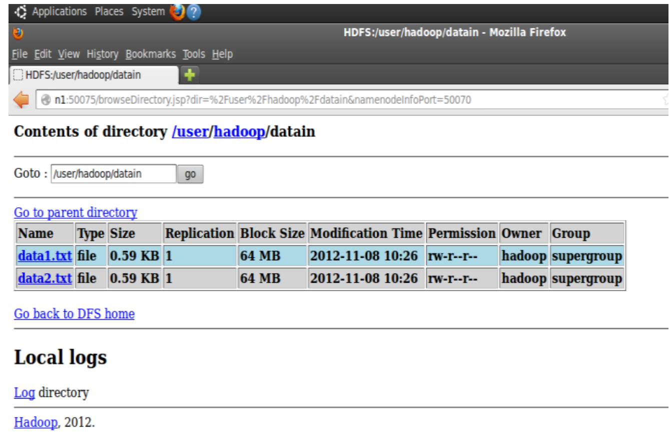
User: hadoop
Job Name: word count
Job File: hdfs://n1:9000/home/hadoop/cloud/hadoop-hadoop/mapred/system/job_201211081007_0001/job.xml
Job Setup: Successful
Status: Succeeded
Started at: Thu Nov 08 10:27:47 IST 2012
Finished at: Thu Nov 08 10:28:09 IST 2012
Finished in: 22sec
Job Cleanup: Successful

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	2	0	0	2	0	0 / 0
reduce	100.00%	1	0	0	1	0	0 / 0

Fig 8: Screenshot of JobTracker

TaskTracker Web Interface (MapReduce layer)

The task tracker web UI depicts running and non-running tasks. It also gives access to the local machine's Hadoop log files [6].
By default, it's available at http://localhost:50060/.


Contents of directory /user/hadoop/datain

Goto : /user/hadoop/datain go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
data1.txt	file	0.59 KB	1	64 MB	2012-11-08 10:26	rw-r--r--	hadoop	supergroup
data2.txt	file	0.59 KB	1	64 MB	2012-11-08 10:26	rw-r--r--	hadoop	supergroup

Go back to DFS home

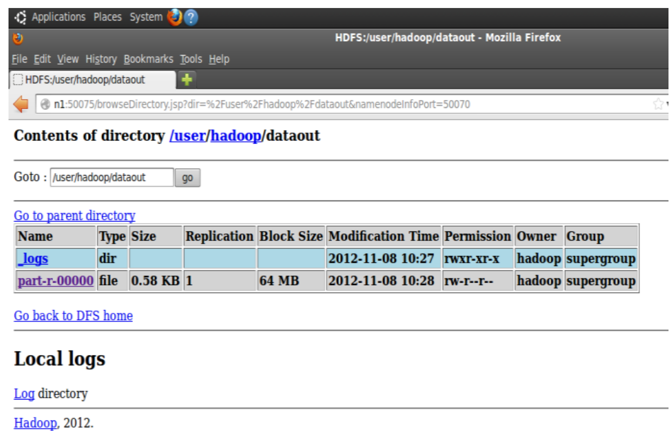
Local logs

Log directory

Hadoop, 2012.

Fig 9: Screenshot of text files in Hadoop

The two data text files are reduced to one text file saving 64 MB of block size using Hadoop. Combined two files contained 164 words which is map-reduced to one file of 61 records.


Contents of directory /user/hadoop/dataout

Goto : /user/hadoop/dataout go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
_logs	dir				2012-11-08 10:27	rwxr-xr-x	hadoop	supergroup
part-r-00000	file	0.58 KB	1	64 MB	2012-11-08 10:28	rw-r--r--	hadoop	supergroup

Go back to DFS home

Local logs

Log directory

Hadoop, 2012.

Fig 10: Screenshot of output file after Map-reduce

4.0 HBase Installation

Download hbase-0.94.2.tar.gz or higher version of HBase from Apache Download Mirrors. Extract it into a folder and change conf/hbase-site.xml file, set hbase.rootdir which is the directory where HBase writes data to, and hbase.zookeeper.property.dataDir, the directory ZooKeeper writes its data too [25]. ZooKeeper is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services [24].

```
<property>
        <name>hbase.rootdir</name>
        <value>hdfs://master:9000/hbase</value>
    </property>
    <property>
        <name>hbase.zookeeper.property.dataDir</name>
        <value>/home/nandan/usr/local/hadoop/hbase/hbase-0.90.5/tmp</value>
    </property>
```

Fig 11: Screenshot of hbase-site.xml

Now start the HBase using following command

$ ./bin/start-hbase.sh

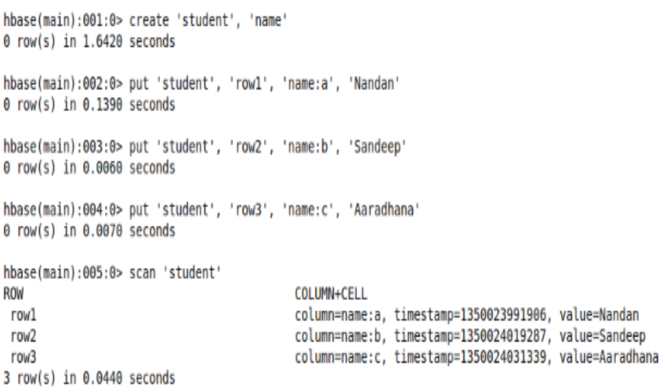
```
hadoop@n1:/usr/local/hadoop/hbase/hbase-0.90.5$ bin/start-hbase.sh
localhost: starting zookeeper, logging to /usr/local/hadoop/hbase/hbase-0.90.5/bin/../logs/hbase-hadoop-zookeeper-n1.out
starting master, logging to /usr/local/hadoop/hbase/hbase-0.90.5/bin/../logs/hbase-hadoop-master-n1.out
n1: starting regionserver, logging to /usr/local/hadoop/hbase/hbase-0.90.5/bin/../logs/hbase-hadoop-regionserver-n1.out
hadoop@n1:/usr/local/hadoop/hbase/hbase-0.90.5$ jps
5402 Jps
2218 JobTracker
1954 DataNode
5121 HQuorumPeer
2382 TaskTracker
5171 HMaster
5342 HRegionServer
1790 NameNode
2123 SecondaryNameNode
hadoop@n1:/usr/local/hadoop/hbase/hbase-0.90.5$
```

Fig 12: Screenshot of starting HBase

Now connect to running Hbase via shell as follows

hadoop@n1:/usr/local/hadoop/hbase/hbase-0.90.5$ ./bin/hbase shell

Using shell commands a sample table is created in Hbase named ‘student’ which has an attribute ‘name’. In this table compression is not specified.



```
hbase(main):001:0> create 'student', 'name'
0 row(s) in 1.6420 seconds

hbase(main):002:0> put 'student', 'row1', 'name:a', 'Nandan'
0 row(s) in 0.1390 seconds

hbase(main):003:0> put 'student', 'row2', 'name:b', 'Sandeep'
0 row(s) in 0.0060 seconds

hbase(main):004:0> put 'student', 'row3', 'name:c', 'Aaradhana'
0 row(s) in 0.0070 seconds

hbase(main):005:0> scan 'student'
ROW                          COLUMN+CELL
 row1                        column=name:a, timestamp=1350023991906, value=Nandan
 row2                        column=name:b, timestamp=1350024019287, value=Sandeep
 row3                        column=name:c, timestamp=1350024031339, value=Aaradhana
3 row(s) in 0.0440 seconds
```

Fig 13: Screenshot of running HBase and Shell Commands

Using exit command one can come out of running Hbase.
To stop HBase use following command

$ ./bin/stop-hbase.sh

5.0 LZO Compression

In order to allow compression in HBase itself a compression software has to be installed in Hadoop. In this experiment LZO compression is used to compress the data which is as follows.

We will use the hadoop-lzo library to add LZO compression support to HBase: [26]

1. Get the latest hadoop-lzo source from https://github.com/toddlipcon/hadoop-lzo.
2. Build the native and Java hadoop-lzo libraries from source. Depending on your OS, to build 64-bit binaries run the following commands:
$ export CFLAGS="-m64"
$ export CXXFLAGS="-m64"
$ cd hadoop-lzo
$ ant compile-native
$ ant jar
These commands will create the hadoop-lzo/build/native directory and the hadoop-lzo/build/hadoop-lzo-x.y.z.jar file. In order to build 32-bit binaries, just change the value of CFLAGS and CXXFLAGS to –m32.

3. Copy the built libraries to the $ hbase-0.94.2/lib and $ hbase-0.94.2/lib/native directories on node:

$ cp hadoop-lzo/build/hadoop-lzo-x.y.z.jar hbase-0.94.2/lib
$ mkdir hbase-0.94.2/lib/native/Linux-amd64-64
$ cp hadoop-lzo/build/native/Linux-amd64-64/lib/* hbase-0.94.2/lib/native/Linux-amd64-64

4. Add the configuration of hbase.regionserver.codecs to your hbase-site.xml file:

$ vi hbase-0.94.2/conf/hbase-site.xml

<property>
<name>hbase.regionserver.codecs</name>
<value>lzo,gz</value>
</property>

5. Sync the $ hbase-0.94.2/conf and $ hbase-0.94.2/lib directories across the cluster.
6. HBase ships with a tool to test whether compression is set up properly. Use this tool to test the LZO setup. If everything is configured accurately, we get the SUCCESS output:

```
root@n1:/usr/local/hadoop/hbase-0.94.2# ./bin/hbase org.apache.hadoop.hbase.util.CompressionTest file:///usr/local/hadoop/data2.txt lzo
12/11/27 16:20:46 INFO util.ChecksumType: org.apache.hadoop.util.PureJavaCrc32 not available.
12/11/27 16:20:46 INFO util.ChecksumType: Checksum can use java.util.zip.CRC32
12/11/27 16:20:46 INFO util.ChecksumType: org.apache.hadoop.util.PureJavaCrc32C not available.
12/11/27 16:20:46 DEBUG util.FSUtils: Creating file:file:/usr/local/hadoop/data2.txtwith permission:rwxrwxrwx
12/11/27 16:20:46 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
12/11/27 16:20:46 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library
12/11/27 16:20:46 INFO compress.CodecPool: Got brand-new compressor
12/11/27 16:20:46 DEBUG hfile.HFileWriterV2: Initialized with CacheConfig:disabled
12/11/27 16:20:46 INFO compress.CodecPool: Got brand-new decompressor
SUCCESS
root@n1:/usr/local/hadoop/hbase-0.94.2#
```

Fig 14: Screenshot of Success of Loading native-lzo/native gpl library

7. Test the configuration by creating a table with LZO compression.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org


```
hbase(main):001:0> create 'Faculty', {NAME=>'name', COMPRESSION=>'lzo'}
0 row(s) in 1.9500 seconds

hbase(main):002:0> put 'Faculty', 'row1', 'name:a', 'Nandan'
0 row(s) in 0.1010 seconds

hbase(main):003:0> put 'Faculty', 'row2', 'name:b', 'Sandeep'
0 row(s) in 0.0140 seconds

hbase(main):004:0> put 'Faculty', 'row3', 'name:c', 'Aaradhana'
0 row(s) in 0.0090 seconds

hbase(main):005:0> scan 'Faculty'
ROW                                   COLUMN+CELL
 row1                                 column=name:a, timestamp=1354014818509, value=Nandan
 row2                                 column=name:b, timestamp=1354014827813, value=Sandeep
 row3                                 column=name:c, timestamp=1354014836738, value=Aaradhana
3 row(s) in 0.0500 seconds
```

Fig 15: Screenshot of creating table in HBase with LZO compression

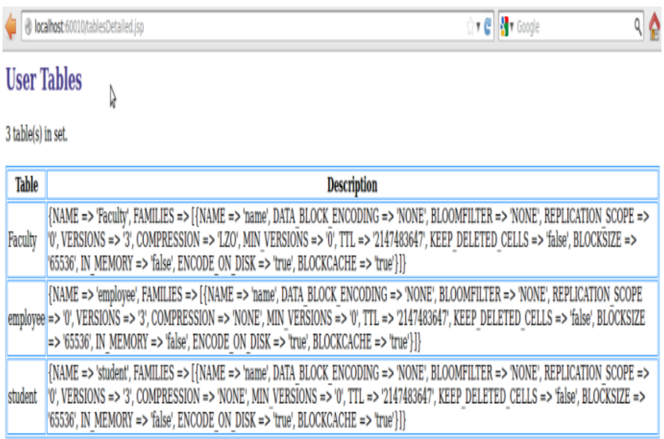


Table	Description
Faculty	{NAME => 'Faculty', FAMILIES => [{NAME => 'name', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'NONE', REPLICATION_SCOPE => '0', VERSIONS => '3', COMPRESSION => 'LZO', MIN_VERSIONS => '0', TTL => '2147483647', KEEP_DELETED_CELLS => 'false', BLOCKSIZE => '65536', IN_MEMORY => 'false', ENCODE_ON_DISK => 'true', BLOCKCACHE => 'true'}]}
employee	{NAME => 'employee', FAMILIES => [{NAME => 'name', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'NONE', REPLICATION_SCOPE => '0', VERSIONS => '3', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => '2147483647', KEEP_DELETED_CELLS => 'false', BLOCKSIZE => '65536', IN_MEMORY => 'false', ENCODE_ON_DISK => 'true', BLOCKCACHE => 'true'}]}
student	{NAME => 'student', FAMILIES => [{NAME => 'name', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'NONE', REPLICATION_SCOPE => '0', VERSIONS => '3', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => '2147483647', KEEP_DELETED_CELLS => 'false', BLOCKSIZE => '65536', IN_MEMORY => 'false', ENCODE_ON_DISK => 'true', BLOCKCACHE => 'true'}]}

Fig 16: Screenshot of Tables in HBase

By adding LZO compression support, HBase StoreFiles (HFiles) will use LZO compression on blocks as they are written. HBase uses the native LZO library to perform the compression, while the native library is loaded by HBase via the hadoop-lzo Java library that is built. In order to avoid starting a node with any codec missing or misinstalled, add LZO to the hbase.regionserver.codecs setting in the hbase-site.xml file. This setting will cause a failed startup of a region server if LZO is not installed properly. "Could not load native gpl library" is visible then there is an issue with the LZO installation. In order to fix it, the native LZO libraries are installed and the path is configured properly. A compression algorithm is specified on a per-column family basis. We create a table Faculty, with a single column family name, which uses LZO compression on it. Although it adds a read-time penalty as the data blocks probably is decompressed when reading, LZO is fast enough as a real-time compression library. We recommend using LZO as the default compression algorithm in production HBase [26].

6.0 Conclusion

The LZO compression format was designed considering speed as priority, it decompresses about twice as fast as gzip, meaning it's fast enough to keep up with hard drive read speeds. It doesn't compress quite as well as gzip— expect files that are on the order of 50% larger than their gzipped version. But that is still 20-50% of the size of the files without any compression at all, which means that IO-bound jobs complete the map phase about four times faster [34]. Following table is a typical example, starting with an 8.0 GB file containing some text-based log data:

Compression	File	Size (GB)	Compression Time (s)	Decompression Time (s)
None	some_logs	8.0	-	-
Gzip	some_logs.gz	1.3	241	72
LZO	some_logs.lzo	2.0	55	35

Fig 17: Comparison of different compression formats [34]

As per above table the LZO file is slightly larger than the corresponding gzip file, but both are much smaller than the original uncompressed file. Additionally, the LZO file compressed nearly five times faster, and decompressed over two times faster [34].
In this way paper illustrates the importance of Hadoop in current Big-Data world, power of Map-reduce algorithm and necessity of compression of data.

References

[1] http://adcalves.wordpress.com/2010/12/12/a-hadoop-primer/

[2] http://hadoop.apache.org/

[3] http://programminggems.wordpress.com/2012/02/20/cloud-foundry-install-error-no-candidate-version-available-for-sun-java6-bin/

[4] https://launchpad.net/ubuntu/+ppas

[5] http://packages.debian.org/squeeze/sun-java6-jdk

[6] http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/

[7] Map-Reduce: Simplified Data Processing on Large Clusters Jeffrey Dean and Sanjay Ghemawat jeff@google.com, sanjay@google.com Google, Inc.

[8] Oracle In-Database Hadoop: When MapReduce Meets RDBMS Xueyuan Su Computer Science Yale University New Haven, CT 06520 xueyuan.su@yale.edu Garret Swart Oracle Corporation Redwood Shores, CA 94065 garret.swart@oracle.com SIGMOD'12, May 20–24, 2012,

Scottsdale, Arizona, USA. Copyright 2012 ACM 978-1-4503-1247-9/12/05 ...$10.00.

[9] MapReduce Online Tyson Condie, Neil Conway, Peter Alvaro, Joseph M. Hellerstein UC Berkeley Khaled Elmeleegy, Russell Sears Yahoo! Research Published in: Proceeding NSDI'10 Proceedings of the 7th USENIX conference on Networked systems design and implementation Pages 21-21 USENIX Association Berkeley, CA, USA ©2010

[10] The Hadoop Distributed File System Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler Yahoo! Sunnyvale, California USA {Shv, Hairong, SRadia, Chansler}@Yahoo-Inc.com

[11] http://www.cloudera.com/content/cloudera/en/why-cloudera/hadoop-and-big-data.html

[12] http://en.wikipedia.org/wiki/Apache_Hadoop

[13] http://www-01.ibm.com/software/data/infosphere/hadoop/

[14] http://wiki.apache.org/hadoop/ProjectDescription

[15] The Hadoop Distributed Filesystem: Balancing Portability and Performance Jeffrey Shafer, Scott Rixner, and Alan L. Cox Rice University Houston, TX Performance Analysis of Systems & Software (ISPASS), 2010 IEEE International Symposium,28-30 March 2010

[16] http://hadoop.apache.org/docs/r0.18.0/hdfs_design.pdf

[17] http://hadoop.apache.org/docs/r0.16.4/hdfs_design.html

[18] http://wiki.apache.org/hadoop/UsingLzoCompression

[19] http://hbase.apache.org/

[20] http://wiki.apache.org/hadoop/MapReduce

[21] MapReduce and the Data Scientist Colin White, Copyright @ 2012 BI Research, All Rights Reserved.

[22] http://wiki.apache.org/hadoop/DataNode

[23] http://wiki.apache.org/hadoop/GettingStartedWithHadoop

[24] http://wiki.apache.org/hadoop/ZooKeeper

[25] http://hbase.apache.org/book/quickstart.html

[26] http://www.packtpub.com/article/hbase-basic-performance-tuning

[27] http://www.ibm.com/developerworks/library/wa-introhdfs/

[28] http://webpages.iust.ac.ir/hsalimi/Courses/89-90-1/CC/CloudComputing.html:- Cloud Programming Models.pptx By Khalili, Sh.

[29] http://www.ubuntugeek.com/sun-java-moved-to-the-partner-repository-in-ubuntu-10-04-lucid.html

[30] http://en.wikipedia.org/wiki/Debian

[31] http://www.youtube.com/user/kanoncloud

[32] http://hadoop.apache.org/docs/r0.17.1/hdfs_user_guide.html

[33] http://hadoop.apache.org/docs/r0.20.0/hdfs_shell.html

[34] http://blog.cloudera.com/blog/2009/11/hadoop-at-twitter-part-1-splittable-lzo-compression/

[35] http://fedorasolved.org/post-install-solutions/sudo

[36] ‘Implementation of Private Cloud using Eucalyptus and an open source operating system’ - Nandan Mirajkar, Mohan Barde, Harshal Kamble, Dr.Rahul Athale, Kumud Singh ‘International Journal of Computer Science Issues’ (IJCSI) Volume 9, Issue 3, 31st May 2012

First Author: **Nandan Nagarajappa Mirajkar** is pursuing **M.Tech** in Advanced Information Technology with specialization in Software Technologies from IGNOU – I^{2}IT Centre of Excellence for Advanced Education and Research, Pune, India. He is also **Teaching Assistant** in Advanced Software and Computing Technologies department. He has published one International Journal in IJCSI. He is IGNOU – I^{2}IT Centre of Excellence for Advanced Education and Research Academic Scholarship holder. He has pursued **B.E** Electronics and Telecommunications from University of Mumbai. His research interests include Cloud computing, Databases and Networking.

Second Author: **Sandeep Bhujbal** is **Sr. Research Associate** in Advanced Software and Computing Technologies department of IGNOU – I^{2}IT Centre of Excellence for Advanced Education and Research, Pune, India. He has pursued **M.C.S** from University of Pune. His research interests include Operating systems, Compiler construction, Programming languages and Cloud computing.

Third Author: **Prof. (Ms.) Aaradhana Arvind Deshmukh** is Asst.Professor in Dept. of Computer Engineering , Smt. Kashibai Navale Collge of Engineering, Pune. She is **pursuing PhD in Cloud Computing from Aalborg University, Denmark**. She is visiting Faculty in Advanced Software and Computing Technologies department of IGNOU – I^{2}IT Centre of Excellence for Advanced Education and Research, Pune, India. She obtained Masters [Computer Engineering] , A.M.I.E. Computer Engineering , B.E. (Computer Engineering) M.A. (Economics) from Pune University. She is having 10 years experience in Teaching Profession and 2 ½ years R & D experience in various institutes under Pune University. She has published 43 papers , 13 in International Journals like ACM, IJCSI, ICFCA, IJCA etc, 16 in International Conferences like IEEE etc., 9 in National Conferences, 4 in symposiums . She has received Gold Medal at International level Paper Presentation on "Neural Network" as well as one more for " UWB Technology based adhoc network", in International Conferences. She is recipient of ‘Distinguished Alumni Award’ in 2011 from Inst. Of Engineers [India] , Gunawant Nagrik Puraskar for the year 2004 – 2005, ‘Anushka Purskar’ from Pimpri Chinchwad Municipal Corporation, and also won many Firodiya awards. She has organized many 15 multidisciplinary Short Term Training Program, workshops, conferences on National and International Level. Her research interests include Cloud computing, Databases and Networking, security.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

Numerical Differentiation of 2D Functions by a Mollification Method Based on Legendre Expansion

Ou Xie[1], Zhenyu Zhao[1]

1 College of Science, Guangdong Ocean University, Zhanjiang, 524088, China

Abstract

In this paper, we consider numerical differentiation of bivariate functions when a set of noisy data is given. A mollification method based on spanned by Legendre polynomials is proposed and the mollification parameter is chosen by a discrepancy principle. The theoretical analyses show that the smoother the genuine solution, the higher the convergence rates of the numerical solution. To get a practical approach, we also derive corresponding results for Legendre-Gauss-Lobatto interpolation. Numerical examples are also given to show the efficiency of the method.

Keywords: *Ill-posed problem, Numerical differentiation, Legendre spectral method, Discrepancy principle.*

1. Introduction

Numerical differentiation is a problem of determining the derivatives of a function from its perturbed values on an interval or some scattered points. It arises from many scientific research projects and applications, e.g., the identification of the discontinuous points in an image process [1]; the problem of solving the Abel integral equation [2, 3]; the problem of determining the peaks in chemical spectroscopy [4] and some inverse problems in mathematical physics [5], etc. The main difficulty is that differentiation is an ill-posed problem, which means small errors in the measurement of a function may lead to large errors in its computed derivatives [5, 6]. Some computational methods have been suggested for one-dimensional case [5-9, 11, 12], but so far only a very few results on the high dimensional case have been reported [13-16] and most of these papers focus on the first order derivative. As far as we know, the literature on higher order differentiation in two dimensions is extremely poor. In [14], G. Nakamura, S. Z. Wang and Y. B. Wang proposed a method for constructing second order derivatives of 2D functions. They have present a convergence result for functions in $H^4(\Omega)$ and the convergence rate can not be improved even if the functions have a higher smoothness. Moreover, an additional boundary condition is needed for their method.

In the present paper, as an alternative way of dealing with numerical differentiation, we introduce a new mollification method. Mollification methods for the regularization of ill-posed problems have been studied and analysed in a number of publications, whereof we can only cite a short list [8, 9, 17-20]. Generally, the idea of mollification methods for an evaluation problem.

$$y = Ax$$

with perturbed data x^δ ,$\|x^\delta - x\| \le \delta$ (δ is is a known error level) consists of two stages:

- Take the mollification of x^δ :
$$x^\delta \to M_\alpha x^\delta .$$
- Take
$$y_\alpha^\delta = AM_\alpha x^\delta$$
as the approximation of y .

The key issues of mollification methods are construction of the mollification operator M_α and choice of the mollification parameter α . In this paper, we will construct the mollification operator by using subspace projection associated with Legendre polynomials. We also point out that the mollification parameter can be chosen by a general strategy----the discrepancy principle, which has been thoroughly studied [6,21,22]. The theoretical analysis shows that the smoother the genuine solution, the higher the convergence rate of the numerical solution. Moreover, the solution processes will be uniform in our method for the different order derivatives and the method is self-adaptive.

This paper is organized as follows. In section 2, we present some preliminary materials which will be used throughout the paper. The methods to construct approximate functions by Legendre expansion and Legendre-Gauss-Lobatto interpolation will be found in section 3 and 4. Some numerical results are given in section 5 to show the efficiency of the new methods.

2. Preliminaries

In this section, we present some preliminary materials which will be used throughout the paper. Let (x_1, x_2) and $\Omega = (-1,1)\times(-1,1)$ and denote by $L^2(\Omega)$ and $H^r(\Omega)$ the usual Lebesgue and Sobolev spaces and by $\|v\|$, $\|v\|_r$ their corresponding norm. Let N be the set of all non-negative integers. For any two tuples $\alpha = (\alpha_1, \alpha_2)$, $l = (l_1, l_2) \in N^2$, $|\alpha| = \alpha_1 + \alpha_2$, $|l|_\infty = \max(|l_1|, |l_2|)$. Throughout this paper, we denote by c a generic positive constant independent of any function. The Legendre polynomial of degree l is defined by

$$L_l(x) = L_{l_1}(x_1) L_{l_2}(x_2), \tag{1}$$

where

$$L_{l_q}(x_q) = \frac{(-1)^{l_q}}{2^{l_q} l_q!} \partial_{x_q}^{l_q} (1 - x_q^2)^{l_q}, q = 1, 2. \tag{2}$$

The set of Legendre polynomials is a L^2 -orthogonal system on Ω, i.e.,

$$\int_\Omega L_l(x) L_k(x) dx = \left(l_1 + \frac{1}{2}\right)^{-1} \left(l_2 + \frac{1}{2}\right)^{-1} \delta_{l,k} \tag{3}$$

where $\delta_{l,k}$ is the Kronecher symbol.

For any $v \in L^2(\Omega)$, we may write $v(x) = \sum_{|l|_\infty = 0}^{\infty} \hat{v} L_l(x)$, where

$$\hat{v}_l = \left(l_1 + \frac{1}{2}\right)\left(l_2 + \frac{1}{2}\right) \int_\Omega v(x) L_l(x) dx, |l| = 0, 1, \ldots. \tag{4}$$

We first recall some properties of the Legendre approximation. Let N be any positive integer and B_N be the set of all algebraic polynomials of degree at most N in each variable. We turn to the inverse inequalities in the space B_N.

Lemma 1[24] Let n be an non-negative integer and $|\alpha| = n$, then for any $\phi \in P_N$,

$$\|D^\alpha \phi\| \le c N^{2n} \|\phi\|. \tag{5}$$

Also for any $r \ge 0$,

$$\|\phi\|_r \le c N^{2r} \|\phi\|. \tag{6}$$

The L^2-orthogonal projection of a function $v \in L^2(\Omega)$ is

$$P_N v(x) = \sum_{|l|_\infty = 0}^{N} \hat{v}_l L_l(x) \cdot \tag{7}$$

Lemma 2 [24] If $r \ge 0$, then for any $v \in H^r(\Omega)$

$$\|v - P_N v\| \le c N^{-r} \|v\|_r . \tag{8}$$

We now turn to the discrete Legendre approximation. Let $x^{(j)} = (x_1^{(j_1)}, x_2^{(j_2)}),\ 0 \le j_q \le N,\ \{x_q^{(k)}\}_{k=0}^N$ are LGL points [24] and

$$\omega_{j_q} = \frac{2}{N(N+1)} \frac{1}{[L_N(x_q^{(j_q)})]^2}, 0 \le j_q \le N. \tag{9}$$

Let Ω_N be the set of all $x^{(j)}$. We can define discrete inner product in $C(\overline{\Omega})$ and its associated norm by

$$\langle u, v \rangle_{N,\omega} = \sum_{j_1=0}^{N} \sum_{j_2=0}^{N} u(x^{(j)}) v(x^{(j)}) \omega_{j_1} \omega_{j_2}$$

$$\|u\|_{N,\omega} = \langle u, u \rangle_{N,\omega}^{\frac{1}{2}} \tag{10}$$

We have

$$\langle L_l, L_k \rangle_{N,\omega} = \begin{cases} 0 & , \quad l \ne k \\ \dfrac{2}{N} & , \quad l = k. \end{cases} \tag{11}$$

Lemma 3 [24] $\quad \|u\| \le \|u\|_{N,\omega} \le \sqrt{3} \|u\|$ (12)

The Legendre interpolation $I_N v(x) \in B_N$ of a function $v \in C(\overline{\Omega})$ is defined by

$$I_N v(x^{(j)}) = v(x^{(j)}), \forall x^{(j)} \in \Omega_N. \tag{13}$$

Lemma 4 [24] If $v \in H^r(\Omega),\ r > \dfrac{1}{2},$ then

$$\|v - I_N v\| \le c N^{-r} \|v\|_r . \tag{14}$$

3.A mollification method by using Legendre expansion

We will discuss the following problem. Suppose that we know an approximate function $g^\delta \in L^2(\Omega)$ of $g \in H^r(\Omega)$ such that

$$\|g^\delta - g\| \le \delta, \|g^\delta\| \ge \tau\delta \tag{15}$$

where $\delta > 0$ is a given constant called the error level and $\tau > 1$. We want to approximate $D^\alpha g$ from g^δ. Our idea is to compute approximate derivatives with the following mollification method. At first, we mollify g^δ by using the projection operator

$$g^\delta \to f_{n,\delta} = P_n g^\delta = \sum_{l=0}^{n} \hat{g}_l^\delta L_l(x). \tag{16}$$

The mollification parameter n plays an essential role in the accuracy of these approximations. In this paper, we make use of a discrepancy principle to obtain an optimal an a posteriori chosen $n(\delta, g^\delta)$:

$$\left\|(I-P_n)g^\delta\right\| \le \tau\delta < \left\|(I-P_{n-1})g^\delta\right\|. \tag{17}$$

Then $D^\alpha f_{n,\delta}$ will be used as an approximation of $D^\alpha g$, we will prove a convergence estimate in the following.

Theorem 5 Suppose that $f_{n,\delta}$ is defined by (16) and (17), $g \in H^r(\Omega)$, then for any $|\alpha| \le \frac{r}{2}$, we have

$$\left\|D^\alpha f_{n,\delta} - D^\alpha g\right\| = O(\delta^{\frac{r-2|\alpha|}{r}}). \tag{18}$$

Proof: By Lemma 1 and 2

$$\begin{aligned}\left\| f_{n,\delta} - g\right\|_{r/2} &= \left\|P_n(g^\delta - g) + P_n g - g\right\|_{r/2} \\ &\le \left\|P_n(g^\delta - g)\right\|_{r/2} + \left\|P_n g - g\right\|_{r/2} \\ &\le cn^r + cn^{-r/2}\left\|g\right\|_{r/2}\end{aligned} \tag{19}$$

On the other hand

$$\begin{aligned}\left\|P_{n-1}g - g\right\| &= \left\|(P_{n-1}g^\delta - g^\delta) - (I-P_{n-1})(g-g^\delta)\right\| \\ &\ge \left\|P_{n-1}g^\delta - g^\delta\right\| - \left\|(I-P_{n-1})(g-g^\delta)\right\|.\end{aligned} \tag{20}$$

From (17), we have

$$\left\|P_{n-1}g^\delta - g\right\| \ge \tau\delta. \tag{21}$$

And

$$\left\|(I-P_{n-1})(g^\delta - g)\right\| \le \left\|g^\delta - g\right\| \le \delta. \tag{22}$$

Hence

$$\left\|P_{n-1}g - g\right\| \ge (\tau-1)\delta. \tag{23}$$

From Lemma 2, we can obtain

$$(\tau-1)\delta \le \left\|P_{n-1}g - g\right\| \le c(n-1)^{-r}\left\|g\right\|_r \tag{24}$$

Thus we can get

$$m \le \left(\frac{c\left\|g\right\|_r}{\tau-1}\right)^{\frac{1}{r}} \delta^{-\frac{1}{r}} + 1 \tag{25}$$

So by (19), there exists a constant M which does not depend on δ such that

$$\left\| f_{n,\delta} - g\right\|_{r/2} \le M. \tag{26}$$

Moreover,

$$\left\| f_{n,\delta} - g\right\| \le \left\| f_{n,\delta} - g^\delta\right\| + \left\|g^\delta - g\right\| \le (\tau+1)\delta. \tag{27}$$

The assertion of theorem will be obtained by interpolation [23].

$$\left\|u\right\|_\mu \le K\left\|u\right\|_{r/2}^{\frac{\mu}{r/2}} \left\|u\right\|^{\frac{r/2-\mu}{r/2}}, \mu \le r, \forall u \in H^{r/2}(\Omega). \tag{28}$$

4.A mollification method by using LGL interpolation

In this section, we derive corresponding results for pseudo-spectral approximations which are more convenient in actual computations. In practical, the perturbed data are usually given at nodes. Suppose that the data is given at points $x^{(j)} \in \Omega_N$, such that

$$|g^\delta(x^{(j)}) - g(x^{(j)})| \le \delta_1. \tag{29}$$

We can obtain the following Lemma

Lemma 6. Suppose that the perturbed data $g^\delta(x^{(j)})$ satisfies (29), then we have

$$\left\|I_N g^\delta - I_N g\right\|_{N,\omega} \le 2\delta_1 =: \hat{\delta}. \tag{30}$$

Proof:

$$\begin{aligned}\left\|I_N g^\delta - I_N g\right\|_{N,\omega} &= \left(\sum_{j_1=0}^{N}\sum_{j_2=0}^{N}(g^\delta(x^{(j)}) - g(x^{(j)}))^2 \omega_{j_1}\omega_{j_2}\right)^{1/2} \\ &\le \delta_1\left(\sum_{j_1=0}^{N}\sum_{j_2=0}^{N}\omega_{j_1}\omega_{j_2}\right)^{1/2} \\ &= 2\delta_1\end{aligned} \tag{31}$$

Using the LGL interpolation, we can give the approximate function as follows:

$$\mathrm{f}_\delta = \mathrm{f}_{n_1,\delta}(x) = P_{n_1} I_N g^\delta. \tag{32}$$

where n_1 is determined by the discrepancy principle

$$\left\|(I-P_{n_1})I_N g^\delta\right\|_N \le \tau\delta < \left\|(I-P_{n_1-1})I_N g^\delta\right\|_N \tag{33}$$

with $\tau > 1$.

We now state the main result of this section.

Theorem 7. Suppose that $\mathrm{f}_{n_1,\delta}$ is defined by (32) and (33), $g \in H^r(\Omega)$, then for any $|\alpha| \le \frac{r}{2}$, we have

$$\left\|D^\alpha \mathrm{f}_{n,\delta} - D^\alpha g\right\| = O(\hat{\delta}^{\frac{r-2|\alpha|}{r}} + N^{2|\alpha|-r}). \tag{34}$$

Proof: By Lemma 1 and 2

$$\begin{aligned}\left\|\mathrm{f}_{n_1,\delta} - g\right\|_{r/2} &= \left\|P_{n_1}(I_N g^\delta - I_N g) + P_{n_1} I_N g - I_N g + I_N g - g\right\|_{r/2} \\ &\le \left\|P_{n_1}(I_N g^\delta - I_N g)\right\|_{r/2} + \left\|P_{n_1} I_N g - I_N g\right\|_{r/2} + \left\|I_N g - g\right\|_{r/2} \\ &\le cn^r + cn^{-r/2}\left\|g\right\|_{r/2}\end{aligned} \tag{35}$$

From (33), we have

$$\left\|P_{n_1-1} I_N g^\delta - I_N g^\delta\right\|_{N,\omega} \ge \tau\hat{\delta} \tag{36}$$

and

$$\left\|(I-P_{n_1-1})(I_N g^\delta - I_N g)\right\|_{N,\omega} \le \hat{\delta}. \tag{37}$$

Hence

$$\left\|P_{n_1-1} I_N g - I_N g\right\|_{N,\omega} \ge (\tau-1)\hat{\delta}. \tag{38}$$

From Lemma 2 and 3, we can obtain

$$\begin{aligned}(\tau-1)\hat{\delta} &\le \left\|P_{n_1-1}I_N g - I_N g\right\|_{N,\omega} \\ &\le \sqrt{3}\left\|P_{n_1-1}I_N g - I_N g\right\| \\ &= \sqrt{3}\left\|P_{n_1-1}I_N g - P_{n_1} g + P_{n_1} g - g + g - I_N g\right\| \\ &\le 2\sqrt{3}\left\|I_N g - g\right\| + \left\|P_{n_1} g - g\right\| \\ &\le 2\sqrt{3}\left\|I_N g - g\right\| + \left\|P_{n_1} g - g\right\| \\ &\le 2\sqrt{3}cN^{-r}\left\|g\right\|_r + c(n_1-1)^{-r}\left\|g\right\|_r \end{aligned} \quad (39)$$

Thus we can get

$$n_1 \le \left(\frac{C\left\|g\right\|_r}{\tau-1}\right)^{1/r}\hat{\delta}^{-1/r}. \quad (40)$$

By (35), there exists a constant M_1 such that

$$\left\|f_{n_1,\delta} - g\right\|_{r/2} \le M_1. \quad (41)$$

Moreover,

$$\begin{aligned}\left\|f_{n_1,\delta} - g\right\| &\le \left\|f_{n_1,\delta} - I_N g^{\delta}\right\| + \left\|I_N g^{\delta} - I_N g\right\| + \left\|I_N g - g\right\| \\ &\le \left\|f_{n_1,\delta} - I_N g^{\delta}\right\|_{N,\omega} + \left\|I_N g^{\delta} - I_N g\right\|_{N,\omega} + cN^{-r}\left\|g\right\| \\ &\le (\tau+1)\hat{\delta} + cN^{-r}\left\|g\right\| \end{aligned} \quad (42)$$

The assertion of theorem will be obtained by interpolation inequality [23]

$$\left\|u\right\|_{\mu} \le K\left\|u\right\|_{r/2}^{\frac{\mu}{r/2}}\left\|u\right\|^{\frac{r/2-\mu}{r/2}}, \mu \le r, \forall u \in H^{r/2}(\Omega).$$

(43)

5. Numerical examples

We provide numerical examples in this section. Let $g(x)$ be a function with two variables given by

$$g(x) = \cos(\pi x_1)\sin(\pi x_2), x \in \Omega.$$

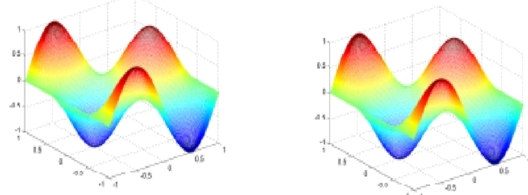


Figure 1: Functions g and its approximation

The discretization knots are $x^{(j)} \in \Omega_N$. The perturbed discrete data are given by

$$g^{\delta}(x^{(j)}) = g(x^{(j)}) + \varepsilon_j, |\varepsilon_j| \le \overline{\delta},$$

where $\{\varepsilon_j\}_{j=1}^{(N+1)^2}$ are generated by Function $2\times(\mathrm{rand}(N+1)-1)\times\overline{\delta}$ in Matlab. The numerical results of constructing g, g_{x_1} g_{x_2} $g_{x_1x_1}$ $g_{x_1x_2}$ $g_{x_2x_2}$ with N=256, $\overline{\delta}$=0.01 are illustrated in Figs. 1-6., respectively. In Figs. 1-6, the left figures are the original functions and the right figures are the constructed functions. In Figs. 7-9, the constructed errors of g, g_{x_1} g_{x_2} $g_{x_1x_1}$ $g_{x_1x_2}$ $g_{x_2x_2}$ are presented, respectively. In Figs. 1-6, we observe that the reconstructed functions are very similar to those of the corresponding functions. In the following, we also further investigate how the relative errors depend on δ.

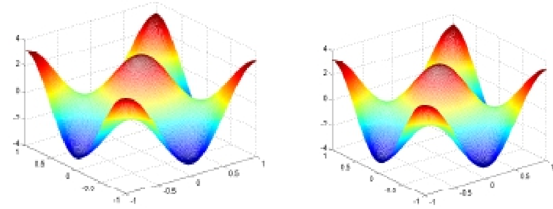


Figure 2: Functions g_{x_1} and its approximation

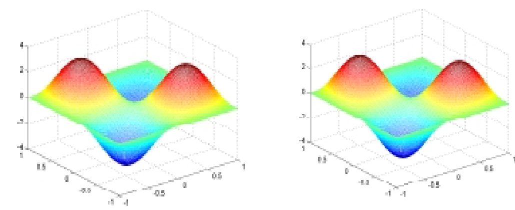


Figure 3: Functions g_{x_2} and its approximation

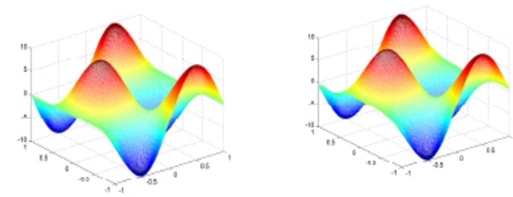


Figure 4: Functions $g_{x_1x_1}$ and its approximation

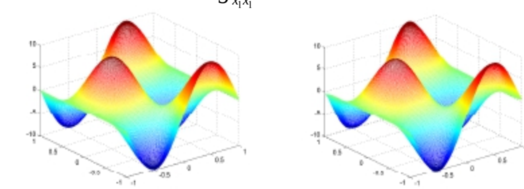


Figure 5: Functions $g_{x_2x_2}$ and its approximation

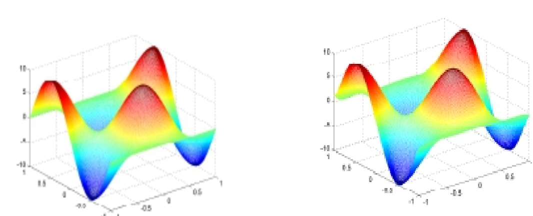


Figure 6: Functions $g_{x_1x_2}$ and its approximation

The relative errors $\varepsilon_g, \varepsilon_{g_{x_1}}, \varepsilon_{g_{x_2}}, \varepsilon_{g_{x_1x_1}}, \varepsilon_{g_{x_2x_2}}, \varepsilon_{g_{x_1x_2}}$ are presented in Table 1 when $\bar{\delta}$ increases form 0.0001 to 0.1 with fixed N=256. Here, the relative error ε_g are defined as

$$\varepsilon_g = \frac{\left\| f_{n_1,\delta} - g \right\|_{N,\omega}}{\left\| g \right\|_{N,\omega}}.$$

We also defined $\varepsilon_{g_{x_1}}, \varepsilon_{g_{x_2}}, \varepsilon_{g_{x_1x_1}}, \varepsilon_{g_{x_2x_2}}, \varepsilon_{g_{x_1x_2}}$, in the same way.

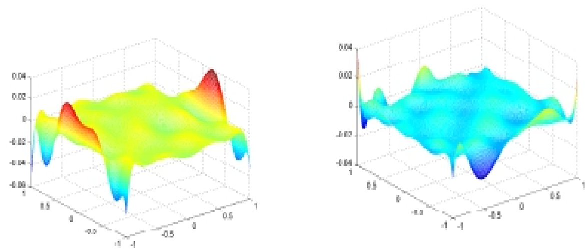


Figure 7: Constructed errors of g_{x_1} and g_{x_2}

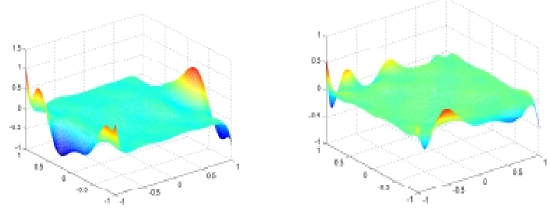


Figure 8: Constructed errors of $g_{x_1x_1}$ and $g_{x_2x_2}$

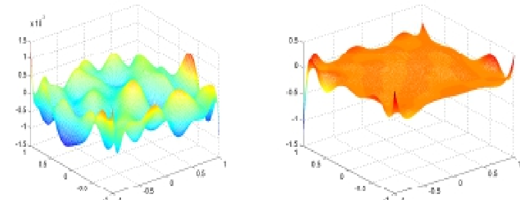


Figure 9: Constructed errors of g and $g_{x_1x_2}$

From Table 1, we can see that when the noise level $\bar{\delta}$ is decreased from 0.1 to 0.0001, the relative errors will decrease too. The above numerical results show that the proposed method is efficient.

6. Conclusion

In this paper, we proposed a new mollification method to reconstruct numerical derivatives from noisy data. The theoretical analyses show that the smoother the genuine solution, the higher the convergence rates of the numerical solution by our methods. Especially if $g \in H^{\infty}(\Omega)$, then the convergence rates of numerical derivatives is $O(\delta)$. Moreover, the solution processes will be uniform for different derivatives, which means that the method is self-adaptive. All the test numerical examples presented in the paper show that the new method works well. The extension of the work to piecewise smooth function is now under investigation.

Table 1: Relative errors with different noise level

$\bar{\delta}$	1e-1	1e-2	1e-3	1e-4
ε_g	3.7147e-3	4.4912e-4	5.7346e-5	5.3339e-6
$\varepsilon_{g_{x_1}}$	1.0487e-2	2.4506e-3	1.7774e-4	2.1542e-5
$\varepsilon_{g_{x_2}}$	1.3426e-2	1.4016e-3	4.2203e-4	3.3407e-5
$\varepsilon_{g_{x_1x_1}}$	4.6777e-2	1.8234e-2	1.3249e-3	1.7913e-4
$\varepsilon_{g_{x_1x_2}}$	7.0255e-2	8.0529e-3	3.9971e-3	3.7298e-4
$\varepsilon_{g_{x_2x_2}}$	2.9172e-2	5.5944e-3	8.7607e-4	1.3671e-4

Acknowledgments

The project is supported by the National Natural Science Foundation of China (No. 11201085).

References

[1] S. Deans, The Radon transform and some of its applications, New York:Wiley-Interscience, 1983.
[2] J. Cheng, Y.C. Hon, Y.B. Wang, "A numerical method for the discontinuous solutions of Abel integral equations", in: Inverse problems and spectral theory: proceedings of Workshop on Spectral Theory of Differential Operators and Inverse Problems, Amer Mathematical Society, 2004,vol.1, pp. 233-240.
[3] R. Gorenflo, S. Vessella, "Analysis and applications Abel Int egral Equations " Berlin: Springer, 1991.
[4] M. Hegland, R. Anderssen, "Resolution enhancement of spec tral using differentiation", Inverse Problems , 2005, vol. 21, pp. 915–93.
[5] M. Hanke, O. Scherzer, Inverse problems light: numerical di fferentiation, American Mathematics Monthly, 2001, vol. 18, no. 6, pp. 512–521.
[6] H. W. Engl, M. Hanke, A. Neubauer, "Regularization of inve rse problems", Berlin: Springer, 1996.
[7] I. Khan, R. Ohba, "New finite difference formulas for numeri cal differentiation", Journal of Computational and Applied Mathematics, 2000, vol. 126, pp. 269–276.
[8] D. A. Murio, M. C. E., Z. S., "Discrete mollification and auto matic numerical differentiation", Computer Mathematics with Application . 1998, vol. 35, no. 3, pp. 1–16.
[9] D.A. Murio, "The mollification method and numerical solution of ill-posed problems", New York: John Wiely & Sons, 1993.
[10] A. Ramm, A. Smirnova, "On stable numerical differentiatio n", Mathematics of Computation, 2001, vol. 235, pp. 1131–1154.

[11] Y. B. Wang, Y. C. Hon, J. Cheng, "Reconstruction of high order derivatives from input data", Journal of Inverse and Ill-Posed Problems, 2006, vol. 16, no. 2, pp. 205-218.
[12] Z. Zhao, G. He, "Reconstruction of high order derivatives b y new mollification methods", Applied Mathematics and Mechanics, 2008, vol. 29, no. 6, pp. 769-788.
[13] M. Bozzini, M. Rossini, "Numerical differentiation of 2D f unctions from noisy data", Computer Mathematics with Application, 2003, vol. 45, no. 1-3, pp. 305-327.
[14] G. Nakamura, S. Wang, Y. Wang, "Numerical differentiati on for the second order derivatives of functions of two varia bles ", Journal of Computational and Applied Mathematics, 2008, vol. 212, no. 2, pp. 341-358.
[15] Y. Wang, T. Wei, "Numerical differentiation for two dimensional scattered data", Journal of Mathematical Analysi s and Applications ,2005, vol. 312, no. 1, pp. 121–137.
[16] T. Wei, Y. Hon, Y. Wang, "Reconstruction of numerical de rivatives from scattered noisy data", Inverse Problems, 2005, vol. 21, pp. 657-669.
[17] D. Hὰo, "A mollification method for ill-posed problems", Numerische Mathematik , 1994, vol. 68, no. 4, pp. 469-506.
[18] D. Hὰo, H. Reinhardt, A. Schneider, "Stable approximation of fractional derivatives of rough functions", BIT Numerica l Mathematics, 1995, vol. 35, no.4, pp. 488–503.
[19] D. Hὰo, H. Reinhardt, F. Seiffarth, "Stable nunerical fractio nal differentiation by mollification", Numerical Functional Analysis and Optimization, 1994, vol.15, no. 5, pp. 635–659.
[20] P. Manselli, K. Miller, "Calculation of the surface temperat ure and heat flux on one side of a wall from measurements on the opposite side", Annali di Matematica Pura ed Applicata, 1980, vol. 123, no. 1, pp. 161–183.
[21] B. Kaltenbacher, "Regularization by projection with a poste riori discretization level choice for linear and nonlinear ill posed problems", Inverse Problems, 2000, vol. 16, pp. 1523–1539.
[22] G. M. Vainikko, "The discrepancy principle for a class of re gularization methods", USSR Computational Mathematics and Mathematical Physics, 1982, vol. 22, no.3, pp.1–19.
[23] R. A. Adams, "Sobolev Spaces,(Pure and Applied Mathema tics) ", New York:Academic Press, 1975.
[24] B. Y. Guo, C. L. Xu, "Hermite pseudospectral method for nonlinear partial differential equations", Mathematical Modelling and Numerical Analysis, 2000, vol. 34, no.4, pp. 859–872.

Ou Xie is an Assistant Professor in the college of Science at the Guangdong Ocean University, China. His research interests are Numerical simulation and numerical algorithms, Computational mathematics.

Zhenyu Zhao holds a PhD from the Shanghai University, Shanghai, He has been an Assistant Professor in the College of Science, Guangdong Ocean University. His research interests include computational method of inverse problems and ill posed problems.

Intelligent communication of WSN based on a multi-agent system

Mostefa Bendjima [1] and Mohamed Feham [2]

[1] University of Tlemcen, STIC Laboratory, Tlemcen, Algeria

[2] University of Tlemcen, STIC Laboratory, Tlemcen, Algeria

Abstract

The abstract A wireless sensor network (WSN) consists of a large number of sensor nodes with limited batteries, which are deployed randomly over an area to collect data. Therefore, it is important to minimize energy consumption of each node, which leads to the extension of the life of the network. As many of data that detected could be redundant or not important, optimization for data collection could be a good technique for saving energy in sensor nodes to extend the network lifetime. Our goal is to design an intelligent WSN that collects a maximum data and processes them intelligently. To achieve this goal we used a system of multi-agent (SMA) to process data, reduce redundancy, eliminate non-useful data and establish cooperation between sensor nodes. We also used mobile agents (MAs) to collect this data and send them to the base station (Sink). Due to the mass of data, and in order to reduce the messages sent between sensor nodes, the network is decomposed into clusters. In addition, nodes cooperate with their neighbors in order to collect the maximum data and eliminate redundant data between sensor nodes. Successive simulations in large-scale WSNs show the ability of the proposed data collection system to extend the life of the network in terms of energy consumption and packet delivery rate.

Keywords: *wireless sensor network, data aggregation, multi-agent system, data processing, communication, clustering.*

1. Introduction

The technological advances made in the field of wireless networks, micro-fabrication and the integration of microprocessors have created a new generation of large-scale sensor networks adapted to a range of varied applications.

A wireless sensor network (WSN) consists of a set of nodes capable of collecting data from a monitored area and transmitting them to a base station (Sink) via a wireless medium. The sensor nodes that we consider are small objects equipped with modules for the communication and acquisition of data, particularly in energy constrained areas. These objects need to save their energy while observing their environment properly. Nodes are constrained in terms of energy resources, and therefore, they cannot emit data directly to the base station. Hence, it is essential to use communications and routing protocols to conserve energy during data transfer to increase the lifetime of the network. It is assumed that WSNs are intelligent, independent, and knowing the context in which they operate. A WSN is often characterized by dense deployment in large-scale environments limited in terms of resources. The limitations are in terms of processing, storage and energy capacity mainly because they are usually powered by batteries. Constraint on the size of a sensor node requires designers to reduce the size of the battery and therefore the amount of energy available. Replacing a battery is rarely possible, for reasons of cost or constraints due to the environment. It is widely recognized that energy limitation is an unavoidable issue in the design of WSNs because it imposes strict constraints on the operation of the network. In fact, energy consumption has become the predominant criterion of performance in this area. If we want the network to function satisfactorily as long as possible, these energy constraints force us to compromise between different activities, both at the node and network levels.

WSNs have given rise to many research issues to improve the performance of the network, including the maximization of their life. WSNs-related literature recognizes that the data is a prominent consumer of energy, the majority of this work has stretched to techniques involving this field. One technique for minimizing energy consumption is the technique of cluster; it is used to partition the network into groups with a Cluster-Head (CH) for each. This latter supports data exchange with the base station, and receives sensored data from all nodes in their cluster to be sent to the Sink.

During the last decade, systems of multi-agent (SMA) have greatly developed, and are applied to various fields such as, simulation and artificial life, robotics, image processing. Ant, spider colonies, etc., are examples of SMA, which are applied in WSN for sensored data processing, routing, detection of shortest paths, etc. Recent works have focused on the use of the paradigm of multi-agent systems to allow a more generic modeling, and to describe more easily new types of sensor nodes. We used this kind of system for managing intelligent communication between the sensor nodes in order to increase the lifetime of the network. Eventually, mobile agents (MAs) have even been integrated in the WSN to collect data and send it to the base station (Sink).

Our goal is to design a network that collects a maximum of data, and minimizes the energy dissipation. This network can be defined as a group of agents able to interact and cooperate to achieve a specific goal.
To achieve this objective, we have integrated, into each sensor node, an agent to process data, reduce redundancy, eliminate non-useful data and enable cooperation between sensor nodes. Each node in the network is seen as an autonomous agent with its own characteristics and attitudes towards the various events they receive. In addition, nodes cooperate with their neighbors in order to collect the maximum data and eliminate redundant ones.

This solution is useful for intelligent processing of data collected by the nodes in terms of redundancy reduction, assessment of the importance of data, and the elimination of non-relevant ones. For more efficiency, we have implemented the clustering technique for better organization, and to send data easily to the Sink. We applied the algorithm Low Energy Adaptive Clustering Hierarchy (LEACH) to decompose the network into clusters, each with a head.

The rest of the paper is organized as follows: Section two provides an overview of the literature in which techniques about effective approaches to save energy in WSN are cited. Section three presents the communication strategy that is based on the grouping of nodes into clusters. In the following section, we present the mobile agent packet structure. The main roles of nodes (agents) during the data collection are described in Section five. Then, in section six we present the parameters used by each node (agent) to calculate data relevance, and to cooperate with other nodes during data collection. Then, Section seven sets forth the purpose of our application, i.e. to establish a system to simulate the communication between a set of sensors and a base station forming a wireless sensor network. Section eight has been devoted to results and their analysis. Finally, Section nine summarizes and concludes this paper.

2. Previous work and problem formulation

Recent advances in wireless communications and electronics have enabled the development of smaller and cheaper sensors. A sensor network is composed of a large number of sensor nodes deployed in a geographical area and that communicate with each other through a wireless medium. The aim behind their use is to gather data from the environment to send them to a base station in order to perform calculations. This gathering must take into account the battery life of each node to maintain network continuity. However, advances in WSN technology enabled the deployment of large amounts of sensors that are smaller and cheaper. The energy constraint is important for data aggregation in WSN, in order to reduce the messages exchanged in the network. The data aggregation techniques have recently been studied, and effective approaches to save energy in the WSN have been developed.

Most efficient proposals are based on the traditional Client/Server (C/S) approach, to manage multi-sensor data fusion in the WSN. Several studies [1]-[8] were performed to optimize the architecture of this model. In this architecture, when a sensor node detects data from the environment, it sends them in their raw state through other Sink nodes to be processed. The transmission of raw data does not eliminate unnecessary or redundant ones, which is energy-costly.

A number of papers have proposed algorithms for data Compression/Decompression (C/D) to reduce the amount of the data transmitted by the sensors. The authors of [2], [3] proposed a correlation algorithm that compresses data in a distributed way. In this proposal, a single node is elected to send raw data to the Sink while the others send coded data only. After the Sink receives data, it decodes them through the correlations between the compressed and uncompressed data. However, it is quite difficult to find a non-complex and energy-efficient coding algorithm suitable for the sensor nodes.

In [4], the authors proposed Data Fusion (DF) of a maximum number of sensors. When a node sends its data to the Sink, the intermediate nodes fuse their data with others coming from the first node. Therefore, this data is merged into a single message instead of many, thereby saving energy. However, the intermediate nodes do not always have to send important data, and do not eliminate the unimportant or redundant ones. In addition, the authors failed to consider the importance of the scalability of such networks.

Other researchers [6] have shown that clustering is a fundamental technique in WSNs. Their objective is to minimize data aggregation processing required at sensor nodes and to move the load to the Sink. Heinzelman and al. [7] proposed a clustering algorithm in which the sensors elect themselves as cluster heads with some probability and diffuse their decisions. Once the data from each node are received, the cluster head forwards them directly to the base station. In this way, it achieves a significant reduction in energy consumption. Unfortunately, the authors did not address the problem of complexity and the amount of energy required to build such cluster-based sensor networks.

In addition to [5] and [6], the authors of [9]-[12] have also proposed a structured strategy. They set up a multicast tree by iteratively adding source nodes to the existing tree until all source nodes are included. Whenever the algorithm detects a remaining source node, which is the closest to the existing tree, it adds the shortest path between this source node and the existing tree. This process continues until all source nodes are included in the tree. However, according to [11] and [12], structured approaches are not practical for dynamic scenarios, due to excessive communication cost and centralized management of the WSN structure.

Inspired by ant colonies, Dorigo *et al.* [13], [14] proposed an Ant Colony Optimization (ACO) algorithm to solve the problem of data aggregation. Every ant explores all possible paths from the source node to the Sink. In the papers [13]-[15], the authors proposed the idea of an agent ant that combines the mobile agent technology with ACO algorithms for network control in the choice of adaptive trace by a data aggregation tree that is built by accumulated pheromone. However, the construction of this appropriate tree depends largely on the deployment of nodes, which is usually random, and consumes a significant amount of energy. As the communication range of a node is limited, the nodes can communicate only with their neighbors in a hop, hence the Euclidean distance between the source node and the receiving node is not reliable.

Researchers have proposed to use system of multi-agent (SMA) as a solution for the adaptation of distributed and complex WSNs. The SMA is a group of agents that interact and cooperate to achieve a specific goal. SMA is well suited to distributed control systems such as WSN. This partly explains the considerable contribution of agent technology when it was introduced in this area.

Researchers [15] have made an intelligent artificial community with intelligent sensors by using this system. Intelligent sensor nodes operate as autonomous agents who develop a network. The authors [16] proposed a group of agents to interact and cooperate in order to reach a specific objective. They proposed a multi-agent approach in which an agent is put in each sensor node to process locally detected data and to cooperate with neighboring agents to eliminate the inter-sensor-nodes redundancy [17]. This problem seems far from being a solution to the density constraint imposed in many WSN applications.

The authors of [1], [7], [8], [17]-[20] proposed the technology of mobile agents (MAs) in WSNs. In these proposals, the MA program refers to a standalone application, able to move between networks nodes by moving the data collected in each node instead of bringing them in their raw state to the Sink by the node itself. It contains an application code, a list of source nodes predefined by the Sink and an empty field to put the data. The calculation model based on MA moves the processing code for data transfer rather than the raw data to the Sink. In addition, the MA is programmed to perform local data processing and filtering on each sensor node according to the data it already carries. The MA then transports the merged data to the next node, and the same procedure is repeated. Due to the bandwidth limitation and density of WSNs, the MA can be used to significantly reduce the cost of communication, in particular the elimination of redundancy, and the reduction of the network overload associated with data transfer [17].

This approach also involves more reasonable use of radio node unit, thus a longer life of the network. When mobile agents are used for data fusion, route selection for agents becomes critical because it greatly affects the overall energy consumption and the cost of data fusion. The drawback of such solutions is the difficulty of creating a list of source nodes and setting the start time of data collection. Another limitation is the definition of areas to be addressed by the MA.

After analyzing the solutions presented above, we can also infer that there is still a lot of work in terms of energy efficiency with attention to the packet delivery ratio and network density. In addition, the solution should be independent of network deployment.

3. Communication strategy

The main purpose of our strategy is to collect the maximum data in a WSN, with improved network performance, including the maximization of its life.

The step of local data processing consumes much less energy than the communication phase; the example presented in [21] illustrates this disparity. Indeed, the energy cost required for transmitting 1KB over a range of 100m is approximately equal to that necessary to perform 03 million instructions at a rate of 100 million instructions per second. Therefore, it is clear and preferable to promote data processing at the node level before their transmission. As many of data that could be detected redundant or not important, optimization for data collection could be a good technique to save sensor nodes energy and extend the network lifetime.

In this work, we propose an intelligent strategy that collects a maximum of data and treats them intelligently. This strategy takes into consideration several parameters to ensure better management of data collection in a WSN.

Among these parameters, we saved energy at each node by reducing the amount of data that is useless and unimportant, and by redundancy elimination.

To achieve this goal we used a multi-agent system to process data, reduce redundancy, eliminate non-useful data and create cooperation between sensor nodes.
Our strategy is described as follows: An agent is introduced in each node, which processes the data locally and judges their importance in order to remove any data that is not useful or redundant. In addition, nodes cooperate with their neighbors in order to collect the maximum data and eliminate redundant data between different sensors. Due to the mass of data, instead of sending data by each node to the Sink separately, we decomposed the network into clusters. Each cluster is composed of Cluster Head, and the data for each cluster are grouped together to be sent. Due to the energy consumption at the highest level of communication, to send an amount of data in a single message is less energy consuming than sending the same amount in several short messages [18]. For this reason, we have proposed mobile agents to concatenate data processed by a node and its neighbors to other nodes in located on the same path to the CH, to send them in a single message.

The main idea is based on grouping nodes in a cluster, and after the Sink elects source nodes, it sends mobile agents to the cluster heads; each MA contains the source nodes of the cluster. Once the CH receives the MA, it will make for each source node in the list a copy of the MA. Then, for each copy of MA is added a list of intermediate sensor nodes between the source node and the CH. The latter will send the mobile agent to the source node so that it flows between the source node and the CH according to the list of intermediate sensor nodes. Each node in the intermediate list invites its neighbors in a single hop to join a session of cooperation for data collection. The MA aggregates the data collected and processed by the list of intermediate nodes and neighboring nodes of each intermediate node in a well-defined strategy.

As shown in Fig. 1, we assume first that during data aggregation an intermediate neighboring node detects data. After processing, it is considered that the neighboring node believes that this data is important and useful. Based on this estimate, it runs a formula defined to make the appropriate decision to cooperate or not. Fig. 1 explains the general scheme of a data collection.

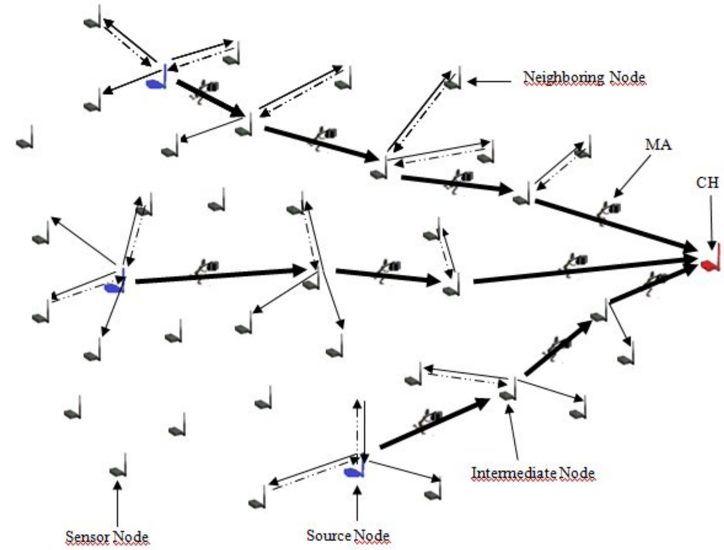


Fig. 1 General scheme of a data collection.

4. Mobile Agent Packet Structure

After the Sink uses direct diffusion to designate the source nodes, it sends mobile agents to the clusters heads; each one containing its source nodes. Our MA is a data packet that circulates in the cluster, and is used to assemble the data collected by the sensor nodes in one cluster. The data contained in an MA packet is shown in Fig. 2.

<table>
<tr><td>Sink_ID</td><td>MA_SeqNum</td><td>CH_ID</td><td>Src_List</td><td>SN_Nbr</td></tr>
<tr><td colspan="2">Processing_Code</td><td colspan="3">Data_Cooperation</td></tr>
</table>

Fig. 2 MA Packet Structure.

Both attributes Sink_ID and MA_SeqNum are used to identify a packet MA. Whenever Sink sends a new packet to MA, it increments the MA_SeqNum. Src_List specifies the source nodes of the cluster head CH_ID, which will be visited by the MA. SN_Nbr attribute is the number of source nodes in the cluster. Processing_Code is used to process the data collected and to manage the MA. Data_Cooperation as well as carry aggregated sensor nodes data.
After the cluster head receives a mobile agent, it will make a copy of MA for every node in the source Src_List, so that the number of source nodes equals the number of MAs.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Online): 1694-0814 | ISSN (Print): 1694-0784
www.IJCSI.org

Copies of MA will change when the CH receives the MA packet from the Sink.

For each copy of MA, CH fills the Src_List attribute by nodes that are in the same path between a source node of the Src_List and CH.

Again for each copy of MA, the SN_Nbr attribute is removed, and we add a Src_Next attribute which determines specifically the sequence of node identifiers that must be visited by the MA. If Src_Next is equal to CH, this means it is the last node visited by MA. Fig. 3 illustrates the case of a mobile agent with its copies.

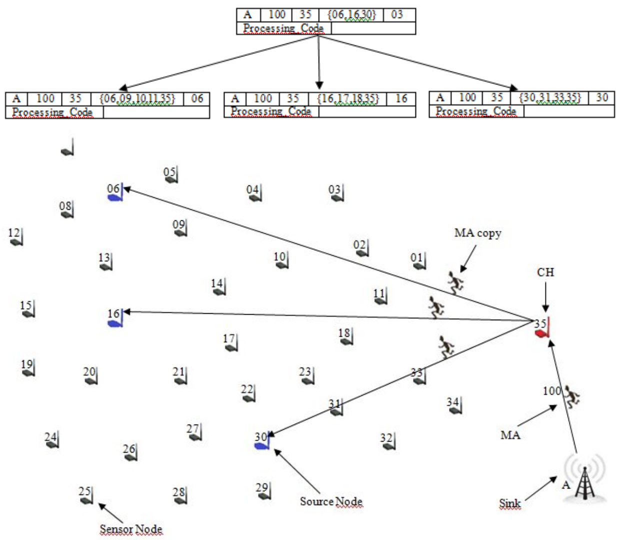


Fig. 3 An example of a MA with copies.

Algorithm 1: Creating copies of a MA at the Cluster-Head level
If CH receives MA **Then**
If MA@ Src_Next <> {} **Then**
For i = 1 **To** MA@ SN_Nbr **Do**
MA[i] = MA; // Make a copy of MA;
Remove MA[i]@ SN_Nbr ; // Delete the attribute SN_Nbr;
Add MA[i]@Src_Next ; // Create the attribute Src_Next;
MA[i]@ Src_Next = MA@Src_List[i];
MA[i]@Src_List = {the intermediate nodes between Src_List[i] and CH};
Sending MA[i] to MA[i]@Src_Next;
EndFor
Else
MA sending to the Sink;
EndIf
EndIf

5. Cooperative data collection

Fig. 4 illustrates the role of sensor nodes (agents) during of data collection sessions, which begins when a source node receives an MA. This source node sends a request for cooperation to its neighbors in a single hop, as shown in step (1). The request for cooperation is a short message, scheduled for a single hop, and the neighbors will be programmed not to replay the message. Thus, a node (agent) neighbor decides to cooperate or not in a very precise formula. After making the appropriate decision, each node sends its cooperating processed data (useful and non-redundant) to the source node. Thus, the source node waits for a fixed time, and then considers that nodes which do not send data are not cooperating. Then, the source node processes its data with the ones transferred by neighboring cooperating nodes to eliminate redundancies. After the elimination of redundancies by the source node, the latter collects its data with data from neighboring cooperating nodes, and sends the MA to the next node in the CH path that is in the next source attribute of MA. These data will be linked in the data attribute of cooperation after the elimination of redundancy, as shown in step (2).

Upon receipt of MA, node (A) sends its request for cooperation to its neighbors in a single hop to collect their processed data, as shown in step (3). After making the appropriate decision, each cooperating node sends its processed data to the node (A). Thus, node (A) waits for a fixed time, then considers that nodes that do not send data are not cooperating. Then, node (A) processes its data with the ones transferred by neighboring cooperating nodes to eliminate redundancies. After that, it sends its data with data from neighboring cooperating nodes in the MA to the next node in the path to CH, as shown in step (4). Node (B), and all intermediate nodes on the path to CH which are found in the list Src_List of MA, repeat the same procedures of preceding nodes until they reach the CH.

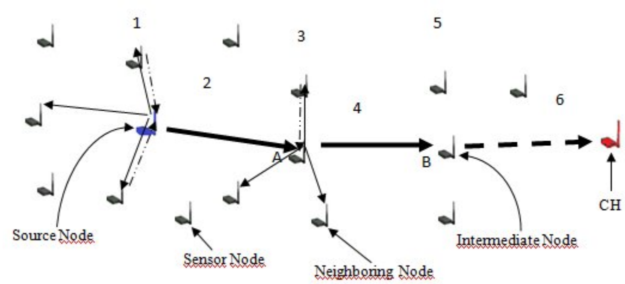


Fig. 4 Cooperation of data collection Proposed beam former.

Algorithm 2: the passage of MA at the intermediate nodes level
While MA arrives at a node not CH **Do**
For i = 1 **to** Number the neighbors of node **Do**
Node broadcasts ReqCoop;
EndFor
If delay = D **Then** // the time out for sent the MA to the next;
For i = 1 to Number of neighbors node cooperating +1 **do** // node + its neighbors cooperating;
If MA is empty **Then**

N=R_1; // Set the data packet MA;
Else
If the node data does not exist in the packet MA **Then**
N=N+ ∑ ρ x Ri ; // add data packet MA;
EndIf
EndIf
EndFor
EndIf
MA@Next_Src = read the new destination for MA@ListSrc ;
Send the MA at MA@NextSrc ; // node same path to CH
End.

Let N be the amount of sensored data accumulated after the MA collects the result of a node, and Ri the size of the data to be processed sensored and accumulated locally by the MA to node i. Hence, we have: $N = R_1$; $N = N + \sum_{k=2}^{i} \rho \times R_k$ (i >=2).

Algorithm 3: Cooperation of a neighboring node
If node receives ReqCoop **Then**
If node contains important data **Then**
Calculate P; // the Pertinence of cooperation;
If P > S **Then** // if the pertinence > to a predefined threshold;
Transfer data;
EndIf
EndIf
EndIf

6. Agent Strategy

We present the parameters used by each node (agent) to calculate the pertinence of cooperation during a data collecting session (P). These parameters are Energy (E), the Distance from the cluster head (D), and the degree of Importance of the data (I).

We express the parameters defined by equation (1) to calculate the pertinence (P) of cooperation. Each parameter is multiplied by a coefficient which is the impact factor according to its importance. Hence, according to a predefined threshold (S), the neighbor node decides to cooperate (P > S) or not (P < S).

$$P = Ce \times E + Cd \times D + Ci \times I \quad (1)$$

Where, Ce, Cd, and Ci are energy coefficients, the distance, and the degree of importance of data respectively.

6.1 Energy (E)

The energy E is only the remaining battery level. The more one node has energy; the more it is requested to cooperate. A low battery level decreases the value of P. In this case, for the agent to cooperate, other parameters have to be at their highest.

6.2 Distance (D)

A head is elected for each group and each node within a group is identified by a particular subnet address. Each elected leader transmits to each node in the group a distance vector separating the group leader. When a node is not close to the head, it is requested to participate in cooperation, but if it is closer, it might lose energy by cooperating with several nodes.

6.3 Degree of Importance of data (I)

The degree of importance of data depends heavily on the desired application. This parameter is calculated by local processing in the node, after the data are gathered. The data processing at the node level can estimate the degree of importance of the data collected. Such data is considered a priority node if it is the first data containing the desired information. In another example, for temperature monitoring, the node saves the last data collected to compare with new ones gathered. If the difference between both is greater than a predetermined threshold, we consider that the data is important. However, if the difference increases, we consider that this data has a higher priority, so the degree increases.

7. Simulation setup

The goal of our application is to simulate the communication between a set of sensors and a base station within a wireless sensor network. Our technique is based on partitioning the network into clusters and on using multi-agent systems and mobile agents as a mechanism to save energy. Thus, we will make an application in terms of efficient energy consumption and packets delivery rate, for a node to call its one hop distant neighbors in order to cooperate during a data collection session.

We performed our simulations on a 500m x 500m area with a random distribution of 1000 sensor nodes in 1000 seconds. Thus, we used a single base station that is located to the right of the field, and some source nodes are randomly distributed in the network. We have limited the radio range and the data bit rate of each node to 80m and 1Mbps, respectively, as suggested in [16]. Local processing time is 40ms. The parameters of power transmission and reception, which directly affect the radio range, are selected from the ranges defined in [17].

Table 1: Basic simulation parameters

Simulation parameters	Values
Node distribution	Random
Radio range	80m
Bit rate	1Mbps

Sensored data interval	10 seconds
Simulation time	1000 seconds
Local processing time	40ms
Processing code size	0.4Ko
Raw Data Size	2024 bits
MA Code Size	1024 bits
MA Accessing Delay	10 ms
Data processing Rate	50 Mbps
Raw Data Reduction Ratio	0.8
Aggregation Ratio	0.9
Fusion Factor (ρ)	01
Ce , Cd , Ci	0.5
Network Size	500m x 500m
Number of Sensor Nodes	1000

8. Results and analysis

Mobile agents have been proposed for efficient data dissemination in wireless sensor networks. In traditional, Client/Server architecture based WSN (CSWSN), data from multiple source nodes are transferred to the base station while the mobile agents can be used to significantly reduce the communication costs, which have a significant impact on the effectiveness of WSNs. In addition, to understand the performance of our strategy, we compare it with our previous work [22] Multi-Agent-based Wireless Sensor Network with Clustering (MAWSNC) where we have used the agent–based strategy. However, it is assumed that the nodes are not agents, that is to say, they do not process data as in our approach. This paper proposes to use a strategy based on a multi-agent system for the reduction and aggregation of a maximum of data in sensor network. The proposed strategy is called a Multi-Agent-based System of wireless sensor networks (SMAWSN).

To demonstrate the performance of our approach in wireless sensor networks, we compare it with other approaches, namely CSWSN and MAWSNC according several criteria that we will describe later.

In our simulations, we assume that the sensor nodes have batteries (energy limited), except the Sink, which is assumed to have an infinite supply of energy. We assume again that the Sink and sensor nodes are stationary and that the Sink is located on the right side of the field. To check the scaling property of our algorithms, we select a large-scale network with 1000 nodes.

In this section, we present the main performance criteria and the evaluation of their effectiveness through simulations:

We examine the impact of the number of source nodes on the criteria of energy efficiency. Therefore, energy consumption is the parameter that defines the lifetime of the wireless sensor network. Therefore, we consider this parameter as the most important criterion, according to which we will evaluate the performance of our approach as shown in Fig. 5.

We use E_{CS}, and E_{MA} to describe energy consumption in CSWSN and MAWSNC, respectively. However, in our approach, we add to E_{MA}, energy E_{TR} for local treatment at the source nodes. We fixed the number of source nodes from 05 to 50 per steps of 05, and get a set of results for each case. We represented the results obtained in Fig. 5, which shows the impact of the number of nodes on energy sources, for obtaining sensory data from all source nodes.

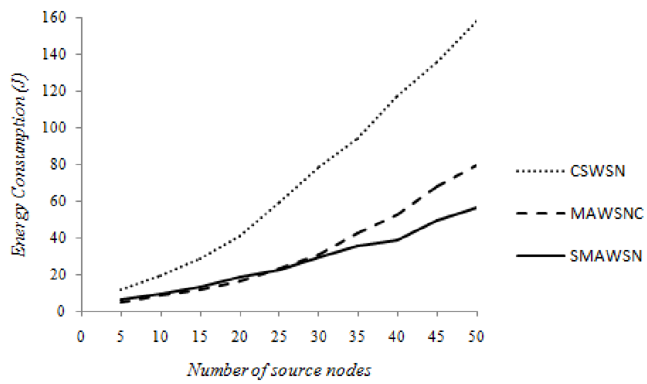


Fig. 5 Comparison of Energy Consumption.

The analysis of the figure above highlights several interesting elements: first, obviously, when the number of source nodes is increased, more energy is needed to perform the duties of each of the three approaches. At the beginning, our approach is always better than CSWSN in terms of energy. However, for an equivalent number of source nodes, the consumption of our approach SMAWSN is always lower than MAWSNC and CSWSN approaches. Moreover, the difference between our approach and the approach CSWSN begins to widen from 05 source nodes, while the gap with other MAWSNC begins at 25 sources nodes. When the number of source nodes is little, energy consumption of our approach is less important compared to the MAWSNC approach, that from 25 source nodes or more, the difference between our approach and the MAWSNC approach becomes increasingly important, and this difference increases continuously with the increase in the number of source nodes. However, from 30 nodes, our approach significantly outperforms the other two approaches; for example with 40 nodes where MAWSNC and CSWSN consume 25% of energy and 67% more than SMAWSN, respectively. However, at 50 source nodes, consumption of our approach is 28.5% lower than in the MAWSNC approach and 64% less compared to the

CSWSN approach. Our SMAWSN approach is therefore more energy efficient.

In addition, and in another experiment, we show the performance comparison of the three approaches in terms of task duration. In a CSWSN, this criterion is the average latency required for carrying messages from source nodes to the base station. In a MAWSNC, the task duration is equivalent to the average end-to-end reporting delay, from the time when a MA is dispatched by the sink to the time when the agent returns to the sink. However, the time includes processing time data by the mobile agent to eliminate redundancy. So our approach for several mobile agents working in parallel, there must be an agent who is the last to return to the Sink, therefore, the duration of the task is the delay of this agent. However, duration includes data processing time at the node level; the longer the cooperation time with neighboring nodes, the more is the mobile agent time. We use T_{CS} and T_{MA} to designate average end-to-end packet delay in CSWSN and MAWSNC, respectively. However, in our approach, we add the T_{TR} time to T_{MA} for local processing at the source nodes level. The results are shown in Fig. 6:

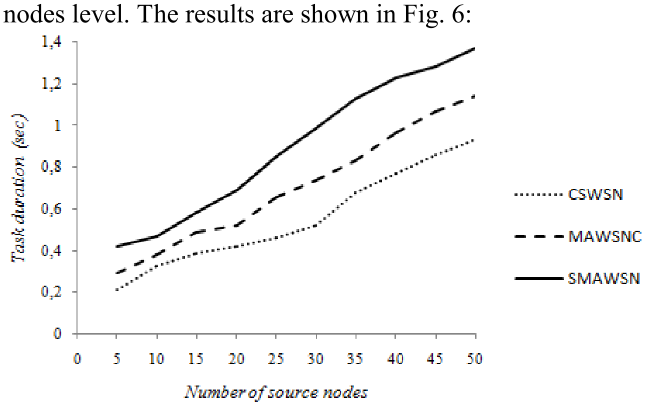


Fig. 6 Comparison of task duration.

In Fig. 6, we show the task duration between the three approaches, and we notice additional latency with SMAWSN in relation to MAWSNC and CSWSN. This gives the false impression that the performance of our approach in terms of duration of a task is poor. However, this is not true, as the duration of a task in the CSWSN approach does not include processing time. If you add the processing time to the latter, we believe that the difference is that in SMAWSN the duration of a task will be relatively small.

This latency is related to the local processing time and the cooperation of neighboring nodes necessary for our approach. We can also notice that the latency difference between SMAWSN and MAWSNC is lower compared to the latency difference between SMAWSN and CSWSN because MAWSNC spends latency time to eliminate redundancies by the mobile agent at each source node, which is not the case with CSWSN which sends the data collected to the base station.

In addition, observing the results, and comparing our approach with CSWSN, which shows the biggest difference, we can notice that values are below 0.45 seconds. This means that the accuracy could be affected for applications that are very sensitive and require less than 0.45 seconds.

In addition, these differences could be explained easily, because with more source nodes to visit, MAs sizes increase and many transmissions will be made with neighboring nodes. However, if we consider the association between these two criteria, the extra time consumption in our approach, shown in Fig. 7, will be easier to understand.

We use E_{CS} and E_{MA} to describe energy consumption in CSWSN and MAWSNC, respectively. We still use T_{MA} and T_{CS} to describe the average end-to-end packet delay in CSWSN and MAWSNC, respectively. However, in our approach, we add energy E_{TR} and time T_{TR} to E_{MA} and T_{MA}, respectively, for local treatment at the source nodes. For this purpose, we adopt the following definitions:

$$E_{CSWSN} = E_{CS} \times T_{CS} \quad (2)$$

$$E_{MAWSNC} = E_{MA} \times T_{MA} \quad (3)$$

$$E_{SMAWSN} = (E_{MA} + E_{TR}) \times (T_{MA} + T_{TR}) \quad (4)$$

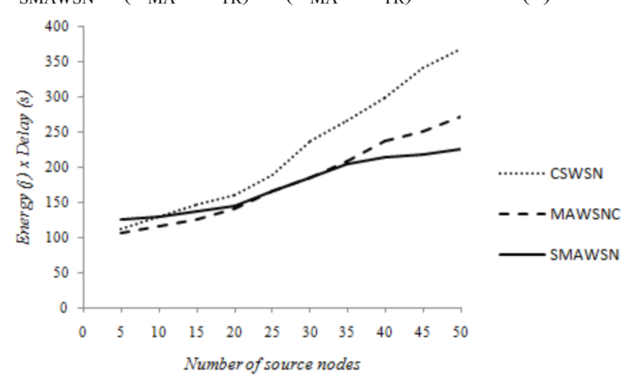


Fig. 7 Comparison of energy x delay.

Fig. 7 compares the performance of the three approaches in terms of energy x delay. Compared to the energy performance, the number of source nodes has a greater impact on this criterion. For time-sensitive applications over energy constrained WSNs, this criterion is defined to facilitate assessment of the overall energy and delay performance of the algorithms. The smaller this value is, the better the performance will be.

First, when the number of source nodes is increased, energy x delay is increased to perform the duties of each of the three approaches. When the number of source nodes is small, energy consumption using our approach is less important compared to other approaches.

It may be noted again that the difference between our SMAWSN proposal and the other two approaches is becoming increasingly important, and this difference increases continuously with the increase of the number of source nodes.

However, for an equivalent number of source nodes, energy consumption using our SMAWSN approach is always lower than that of MAWSNC and CSWSN. Moreover, the difference between our approach and the CSWSN approach begins to widen from 10 source nodes, while the gap with MAWSNC begins at 35 source nodes. Our approach is advantageous with increasing sources nodes; with 50 source nodes; MAWSNC and CSWSN consume 17% and 38.5% more energy than SMAWSN.

In addition, and in another experiment, we changed the size of the data collected at each source node 0.5Ko to 04Ko. The results are expressed in Fig. 8:

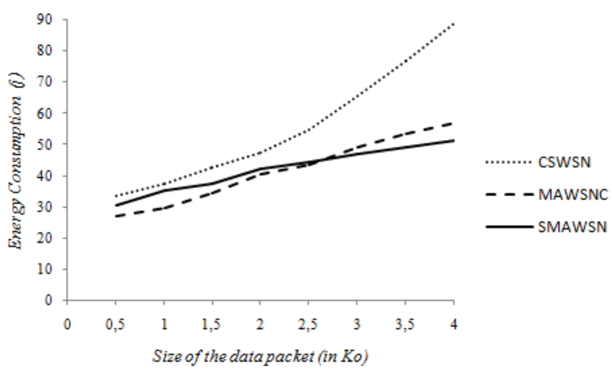


Fig. 8 Energy consumption depending on size.

After analyzing the figure below, we notice that when the packet size is increased, more energy is needed to perform the duties of each of the three approaches.

The first observation is that the difference between our approach and the CSWSN approach begins to widen from 0.5 Ko, while the gap with MAWSNC starts at 2.5 Ko. Moreover, when the packet size is small, energy consumption of our approach is less important compared to the MAWSNC approach, varying between 0.5 Ko and 2.5 Ko, the latter makes additional energy savings of 16% higher than SMAWSN. It may be noted again that from 2.5 Ko the difference between SMAWSN and the two approaches becomes increasingly important, and this difference increases continuously with increasing packet size. However, at 4Ko, energy consumption of SMAWSN is 11% less than in MAWSNC and up to 42% less than CSWSN. This means that the difference in energy consumption gap between our approach and the two others is steadily widening with increasing packet size. By comparison, our SMAWSN proposal solution has better energy efficiency.

9. Conclusion and future works

In an environment where the sensor nodes are close to each other, and where the redundancy is considerable, the sensor nodes generate a large amount of transmission over the wireless channel, which consumes a lot of battery power. Not to send the data in their raw state, as they are captured as in the Client/Server model, and to collect the maximum relevant data, we proposed in this work a solution for data collection using a multi-agent system based on clustering for managing communication between the sensor nodes, and to increase the life of the network. The objective of our work is to compile the maximum possible amount of data from source nodes that are in the same path to the Cluster-Head in a single message.

Our system consists of two types of agents, including stationary and mobile agents. The role of stationary agents consists in integrating an agent in each sensor node, and then each agent processes the locally sensored data by its corresponding sensor node and considers its importance. Then, each source node cooperates with its neighboring nodes in the same cluster to gather their data and eliminate redundancies by a method based on several important parameters for determining the relevance of cooperation. We used as a mobile agent data packet to transmit sensored data to the Cluster-Head to eliminate redundancies between nodes. The result is that a mobile agent gathers data from a source node and its neighboring nodes that are cooperating in the same path to CH. This plan limits communications except for relevant data, and consequently reduces the amount of traffic and energy consumption. Cooperating nodes eliminate redundancies and concatenate all the data processed during a collection session in a single message. This means a gain with regard to amount of data and the load needed to send them. The plan presented is based on a strategy that takes into consideration several parameters deemed important for a longer WSN lifetime. These parameters enable us to determine energy, distance from the CH, and the degree of importance of the data.

Through successive simulations, we have been able to prove that our approach has better performance compared to the Client/Server approach and our previous work [22] Multi-Agent-based Wireless Sensor Network with Clustering (MAWSNC) where we used the strategy of mobile agents. In fact, the only shortcoming of our approach is the latency required to communicate additional data. This latency was required to enable neighboring nodes to cooperate, and can easily be justified by taking into account the gain obtained in terms of packet delivery ratio and energy consumption in dense wireless sensor networks.

As future work, we will attempt to reduce latency to a maximum, and add other parameters to the cooperation formula.

References

[1] M. Chen, T. Kwon, Y. Yuan, and V.C.M. Leung, "Mobile Agent Based Wireless Sensor Networks", University of British Columbia, 2006, pp. 14–21.

[2] J. Chou, D. Petrovic, and K. Ramchandran, "Tracking and exploiting correlations in dense sensor networks", Conference Record of the Thirty-Sixth A silomar Conference on Signals, Systems and Computers, 2002, pp. 39–43.

[3] J. Chou, D. Petrovic, and K. Ramchandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks", Joint Conference of the IEEE Computer and Communications Societies, 2003, pp. 1054–1062.

[4] S.R. Das, "Serial Data Fusion Using Space-filling Curves in Wireless Sensor Networks", In IEEE, ed. First Annual IEEE, Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004, pp. 182–190.

[5] W.H. Liao, Y. Kao, and C.M. Fan, "Data aggregation in wireless sensor networks using ant colony algorithm", Netwerk Computer, 2008, pp. 387–390.

[6] M. Chen, H. Mineno, and T. Mizuno, "Adaptive data aggregation scheme in clustered wireless sensor networks", Comput. Commun, 2008, pp. 3579–3585.

[7] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents", Multidisciplinary University Reseanh Initiative (MURI) under the Omce of Naval Research, And Army Research Laboratory CTA on Communication, and Network under Grant DAADlP-OI-2-CQII, 2003.

[8] Q. Wang, and Y. W. Xu, "Mobile-agent-based collaborative signal and information processing in sensor networks", Proceedings of the IEEE, 2003, pp. 91-98.

[9] J.N. Karaki, R. Mustafa, and A.E. Kamal, "Data aggregation and routing in wireless sensor networks: Optimal and heuristic algorithms". Computer Network, 2009, pp 945–960.

[10]S. Upadhyayula, and S.K.S. Gupta, "Spanning tree based algorithms for low latency and energy efficient data aggregation enhanced converge cast (dac) in wireless sensor networks", AdHoc Network, 2007, pp. 626–648.

[11] S. Jie, and Y. Ning, "A spanning tree algorithm for data aggregation in wireless sensor networks", IEEE explorer, 2008, pp. 5014–5018.

[12]B. Yu, and J. Li, "Distributed data aggregation scheduling in wireless sensor networks", In IEEE INFOCOM, 2008.

[13]S. C. Chu, J. F. Roddick, and J. S. Pan, "Ant colony system with communication strategies", Information Sciences, 2004, pp. 63–76.

[14]M. Dorigo, and L. M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem", IEEE Computational Intelligence Society, 1997, pp. 53–66.

[15]R. Tynan, D. Marsh, and D. Kane, "Intelligent Agents for Wireless Sensor Networks. Adaptive Information Cluster", Smart Media Institute, University College Dublin Belfield, Dublin 4, Ireland, 2004.

[16] A. Sardouk, M. Boulahia, and L. Gaïti, "Information-Importance Based Communication for Large-Scale WSN Data Processing", In: WMNC, Gdansk, Poland, IFIP, 2009.

[17]A. Sardouk, M. Boulahia, and L. Gaïti, "Agent-Cooperation Based Communication Architecture for Wireless Sensor Network", In: IFIP/IEEE Wireless Days/Adhoc and Wireless Sensor Networks, UAE, IFIP, 2008.

[18]H. Qi, Y. Xu, and X. Wang, "Mobile-Agent-Based Collaborative Signal and Information Processing in Sensor Networks", in Proceeding of the IEEE, Vol. 91, NO. 8, 2003, pp. 1172-1183.

[19] Q. Wu, N.S.V Rao, and J. Barhen, "On computing mobile agent routes for data fusion in distributed sensor networks", IEEE Transactions on Knowledge and Data Engineering, Vol.16, NO. 6, 2004, pp. 740-753.

[20]G. Vigna, "Mobile Agents: Ten Reasons For Failure Proceedings", Berkeley, CA, 2004, pp. 298-299.

[21]J. Gregory, and W. J. Kaiser, "Wireless integrated network sensors", Communications of the ACM, 2000, pp. 43-47.

[22]M. Bendjima, M. Feham: "Intelligent Wireless Sensor Network Management Based on a Multi-Agent System", IJCST, 2012, pp. 9-14.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Online): 1694-0814 | ISSN (Print): 1694-0784
www.IJCSI.org

Presentation of a New and Beneficial Method Through Problem Solving Timing of Open Shop by Random Algorithm Gravitational Emulation Local Search

*Ali Asghar Rahmani Hosseinabadi[1], Abbas Bagherian Farahabadi[2], Mohammad Hossein Shokouhi Rostami[3], Ahmad Farzaneh Lateran[4]

[1,2]Department of Computer, Behshahr Branch, Islamic Azad University Behshahr, Iran

[3]Industrial Engineering Department, Iran University of Science & Technology Behshahr, Iran

[4]Industrial Engineering Department, Iran University of Science & Technology Behshahr, Iran

Abstract

One of the most important problems of timing in engineering and industry is timing of open shop. The problem of timing of the open shop induces big and complicated solve space. So, this problem is a kind of NP-Complete. In timing of the open shop, there some works, that each work has several operation. Each operation should do in machine whit should do in the same machine the aim of timing of the open shop is to catch a suitable timing for doing all of the operation, how that enough time to minimize to make-span. In problem solve of timing of the open shop. Until now different algorithm were presented. In this article, a new algorithm that is called TIME_GELS is presented which used of a random. Algorithm Gravitational Emulation Local Search (GELS) for following problem solving. This algorithm is basic of the random local search use of two of the four main parameter of speed and the power of gravity in physics. A suggestive algorithm compared with Genetic Algorithm and result is show that a proposed algorithm has a better efficient and finding the answer very soon.

Keywords: *Timing; Open Shop; Genetic Algorithm; Velocity; Newton law; Gravitational force*

1. Introduction

One of the most important problems of timing in engineering and industry is timing of open shop. This problem was used widely in industry activity. For example imagine the making of an aircraft. To doing of all of the stage this work needs a precise and suitable timing. This problem is a kind of timing of open shop. Problem of timing of open shop includes of a big and complicated solve space, for this reason this is kind of NP-Completed problem. The aim of timing of open shop is catch a suitable timing for make-span operation so that to minimize for enough time to make-span. Unit now different kinds of algorithm one presented that we hot iced some of them:

Dorndorf et al [1] used the bound branch method and innovative algorithms to solve Open System problems. They applied limiting branch methods to avoid non-optimal solutions. Brasel et al [2] presented some innovative algorithms based on adaptation algorithms and interpolation operations. Adaptation and interpolation operations mean which of the operations is used during combined timing with a search method. Other algorithms are also presented by Guerest and Prins [3]. Alcaide et al [4] presented a Tabu search algorithm to minimize make-span in Open Shop timing problem. Liaw [5] also presented a hybrid genetic algorithm, to minimize make-span in open shop timing problem. In addition Prins [6] presented a genetic algorithm cause find a proper solution.

In thin article a new method is presented basic of gravitational force algorithm (GELS) for timing of open shop. A proposed method has more attention on a better and best timing.

In the second part of the problem is expressed the gravitational force algorithm described in Section 3 and 4 in the proposed algorithm In Section 5 simulation results and the conclusions are stated.

2. Problem Description

There some work in timing of open shop that work will has several operations, in the other hand each work contains of some operations. In open shop systems each operation should have done in smellier machine, it means that the first operation of the work of j in should have done in the first machine and the search operation of the work of j should have done in second machine. The aim of timing of

open shop catching a timing of doing of whole of operation, so that as to minimize of the enough time to make-span. The problem of timing of open shop has a following limitation:
1) Each operation should have done in its corresponding machine.
2) Each operation should have processed by one machine at the specific time.
3) every time only one operation of one work can present.
4) There isn't any priority of choosing of performer, it means the presentation of performer happen in each order.
Table 1 is the sample of a timing of open shop that in includes 3 works and each work includes 3 operations that should present through 3 machines.

Table 1: A sample system of open shop

Problem	***Machine1***	***Machine2***	***Machine3***
Job1	7	12	16
Job2	13	10	13
Job3	18	14	2

For example, here the second operation of first work that enough time for performance is 12 should present by second machine. Fig 1 show Gant chart sample system table 1. The same as in Gant chart in fig 1 is identified j 1 o 1 (action 1 work 1) processed in time zero by machine 1, and also j 1 03 (work1, action 3) processed in time 23 by Machine 3 during 16 item. The time of make-span in this case would be sample of 39.

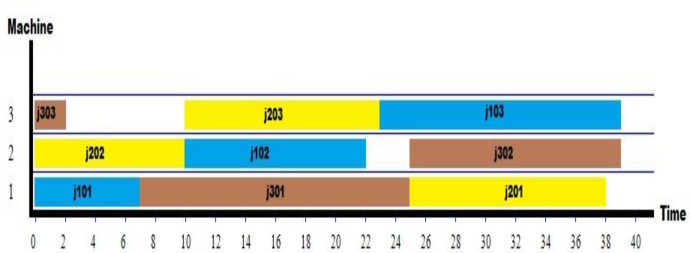


Fig. 1 Gant chart system sample table

3. Gravitational Emulation Local Search (GELS)

In 1995, Voudouris and Tesang [7] offered the algorithm GLS for searching in a search space and solving the example NP-complete for the first time, and in 2004, Barry and Webster [8] offered the algorithm as a powerful algorithm and it was called GELS. This algorithm introduced randomization concept along with two of four primary parameters i.e. velocity and gravity in physics through swapping in terms of groups by using random number in the existing local search algorithm GELS in order to avoid local minima. GELS take as its basis the natural principles of gravitational attraction. Gravity works in nature to cause objects to be pulled towards each other. The more massive the object, the more gravitational "tug" it exerts on other objects. Also, the closer two objects are to each other, the stronger the gravitational forces are between them. This means that a given object will be more strongly attracted to a larger, more massive object than to another object of lesser mass at a given distance, and it will also be more strongly attracted to an object close by than to another, more distant object having the same mass. In GELS the formula of Newton gravitational force theorem between two objects, are involved in:

$$F = \frac{Gm_1 m_2}{r^2} \tag{1}$$

In which m_1 and m_2 are the mass of the first and second object respectively. G is equal to a constant gravitational force which is 6.672; R is the radius of the distance between two objects.
GELS also emulate these processes of nature for searching in a search space. The idea is to imagine the search space as being the universe and object in this universe are the possible solution for the search. Each of these solutions has a "mass" that is represented by its objective function value. The better the solution's objective function value, the higher its mass is. Locations within the search space that do not contain valid solutions are assigned a zero mass [8-10].
In this method, the possible solution in the search space has divided into some sets based on criteria that depend on the kind of problem and each of these sets is called a dimension of the problem solution and for each dimension of the problem solution a value entitled initial velocity has been intended, that it will be explained in continuation. GELS computers the gravitational force between the solution or the objects in the search space by two methods. The first method which is a solution from the local neighborhood space is selected as a current solution and the gravitational force between these two solutions can be computed. The second method applies the formula to all solutions within the neighborhood and tracks the gravitational force between each of them and the current solution individually all solutions. In the movement through the search space, GELS acts in two modes, too. The first mode, allows movement only to solutions within the current local search neighborhood. Each of these

movement modes can be used with each of the computation gravity forces and as a result those four models GELS are made.

GELS maintains a vector, the size of which is determined by the number of dimensions in a solution. This vector's values represent the relative "velocity" in each dimension. The algorithm begins by initializing the current solution, velocity vector and direction of movement. For each dimension, in the velocity vector, random integer between one and the maximum velocity is chosen, and this becomes the value of the element at that dimension. The initial solution can be made as current solution either with user or randomized. For each dimension in the initial velocity vector, concerning the initial velocity vector of the solution dimensions, a direction is selected for the movement that the direction is equal to the solution dimension which has the largest value in the initial velocity vector.

Algorithm includes a pointer object that can move through the search space and a mass which is intended for the pointer object is stable in the whole computations and this object refers to a solution with the largest mass. The algorithm will terminate when one of following two conditions occurs: either all of the elements in the initial velocity vector have gone to zero, or the maximum allowable number of iterations has been completed. In each algorithm iteration as the first method a candidate solution will be selected from the local neighborhood space of the current solution based on the direction of the current movement and the gravity force between the current solution and the candidate solution is computed and then the velocity vector concerning this force will be updated. For the next frequency, the velocity vector is checked and concerning it the movement direction can be chosen. Each iteration algorithm by the second method is completely as similar as the first method, but there is a little difference, as the gravity force and the initial velocity of the update action is computed for each of the candidate solutions instead of the gravity force computation and the update action of the velocity vector for only an obtained candidate solution from the current direction. Newton's formula is used with the alteration that the two masses in the numerator of the equation are replaced by the value of the difference between the objective function value of the candidate solution and that of the current solution. The value of the gravitational force between the two solutions then becomes:

$$F = \frac{G(CU - CA)}{R^2} \tag{2}$$

That, in which, CU and CA are the value of the current solution and the candidate solution. This formula is designed to be a positive value if the objective function value of the current solution is larger than that of the candidate solution and negative if the candidate value is larger. Then the value of this force, positive or negative, can be added to the velocity vector in direction of the current movement. If doing so makes the value exceed the maximum velocity parameter setting, it is set to the maximum. If the update makes the value to go negative, it is set to zero.

Some parameters which can be accessible are as follows:

• Maximum Velocity: the maximum value that any elements within the velocity vector can have been used to prevent velocities that become too large for being usable.

• Radius: it is a radius that can be used in the formula of the gravitational force computation.

• Iteration: the number of iterations of the algorithm that will be allowed to complete before it is automatically terminated [11].

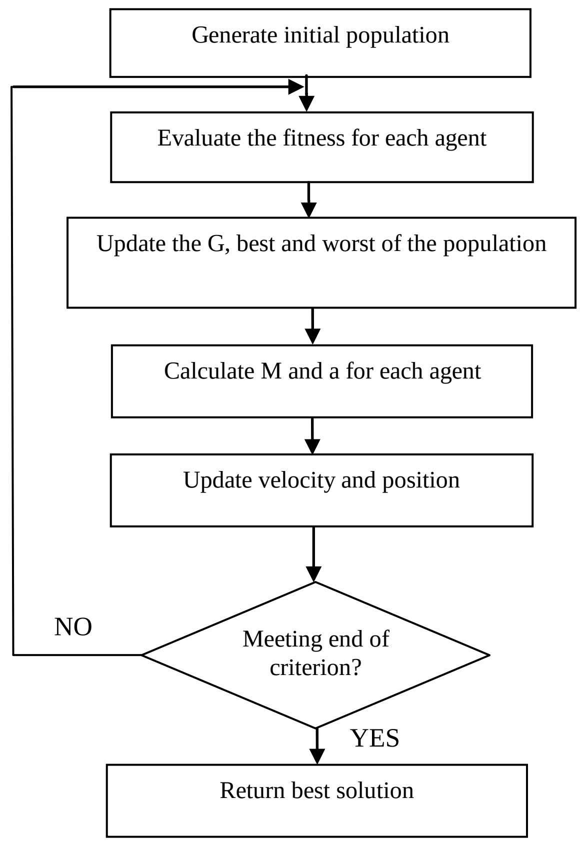


Fig. 2 Gravitational force algorithm steps [10]

4. Proposed Algorithm

In the proposed method of the random algorithm of gravitational emulation local search, (GELS), used as a strategy for problem solving of the open shop. The aim of this algorithm to catching a suitable and benefic timing for

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Online): 1694-0814 | ISSN (Print): 1694-0784
www.IJCSI.org

doing of whole operation of work so that to minimize the enough time for make-span.
Proposed method of algorithm of the gravity power as a strategy used for problem solving of timing of the open shop systems. And it's aim is to minimize final time of performance of finishing work. To understanding proposed algorithm better, methods are explained on the sample table 1. In this case of sample, there are 3 works which each works has 3 operations and works should be presented by 3 machines.

4.1 The show of the produced primary chromosome

In proposed algorithm used for showing chromosomes of one dimension length of the number of whole operation. In this method the showing of chromosome, each gene is a unique integer which is produced following fig:
The operation of each work like a line numbering orderly and gives to each operation a unique number. Table 2 is presented the way of the numbering for sample system table 1.

Table 2: One sample of open shop timing

Problem	***Machine1***	***Machine2***	***Machine3***
Job1	(**1**)7	(**2**)12	(**3**)16
Job2	(**4**)13	(**5**)10	(**6**)13
Job3	(**7**)18	(**8**)14	(**9**)2

Fig 3 is presented a sample chromosome for sample system of the table 1.

9	**5**	**6**	**1**	**2**	**7**	**3**	**8**	**4**

Fig. 3 The structure of sample chromosome

As in the structure of the chromosome you can see in fig 3. Each adornment number will show operation number that should processed by machine. For example the first gene contents is the number of nine that indeed noticed the operator or 3 from work 3. Each chromosome shows timing for make-span.

4.2 The fitness of function

The competence of chromosomes is noticed according to needful time for doing make-span. In proposed algorithm to catch the competence of integral has used formal 3.

$$Fitness = Max_{1} < i <= n \{ Ti \} \tag{3}$$

In this formula n is the number of work. Ti is the time of make-span in timing.

4.3 Definition of dimension solution

In proposed method answer dimensions can imagine equal to the number of presented operations to each machine so that to minimize enough time for make-span. Each operation should have done in a same machine is one dimension of answer, and the answer of neighbor in this dimension is equal to answer that the next operation of the same work should have done in the next machine.

4.4 Definition of neighborhood

In GELS algorithm doesn't do unlike the others algorithm of random searching for the answer of neighbor, so each answer have different neighbors which each one of them is the manner of the special change.
It's called side of movement to word the answer of neighbor and all of the neighbors which used this method, is only the manner of this neighbor.
In proposed method to find the answer of neighbor use this method. Each work contain some operations that in open shop system. Each work contain some operations that in open shop system. Each operation should have done in the same machine, it means first operation of j work should have done in first machine and the second operation of j work should have done in second machine.
In other word the current answer of the neighbor in this dimension equal of the answer which in next operation of same work should have done in next machine.

4.5 Solution problem

In proposed method of random algorithm of the gravitational emulation local search (GELS) used as a strategy for problem solving of open shop systems.
The aim of this algorithm is catching a suitable and benefic timing for make-span so that to minimize the enough time for make-span, so that each operation of the one work should present in the same machine and also each operation will process by one machine in the special time, more over in each time only one operation of one work can present. Solve of this problem, even though it doesn't have much difficulties for less number and machine. So algorithm of gravitational emulation local search is swayable way to solve this problem.
For problem solving of the open shop systems at first we showed notice three the distance, the primary speed and

time of Matris that distance and primary speed Matris create randomly. In speed Matris each one of operation of work is a density and will give a primary speed and then in the next steps speed will change and also time Matris in manner of distance and speed Matris are made by the following this relation:

$$T = \frac{\sqrt{(Y_B - Y_A)^2 + (X_B - X_A)^2}}{V_{in_{A,B}}} \qquad (4)$$

By the use of GELS algorithm in this situation should define a suitable reservation factor. Reservation factor is the number of reserved operations from different work for machines in the future time. The collection of operation transfers only should process by the same machine and there is one solution. This solution can be like two Matris (n*n) that equals to the number of operation of each work and machines to show one answer used. In each line and column of this Matris it has shown one of the operation which is related to it self and the quantity which has kept in each line and column is shown the number of operation which should process by the machine that is related to it. When algorithm completed, the solution for each one of operation, some reservation machines with reservation factor is shown.

5. Simulated Results

To landing algorithm use of language of programming C#.NET and program is presented on the computer with processor GHZ Pentium. Iv 2.4 and Ram with GB 1.
Proposed algorithm (TIME_GELS) has compared with GA algorithm. To compare the following algorithm, there are 8 collection of the designing test that covers small, average and big systems. Giving the name to the this data of the test is Test–J–o. In this giving the name method J shows number of the work, o shows the number of operation in each work in data of the test. The data of the designing test is shown in table 3.

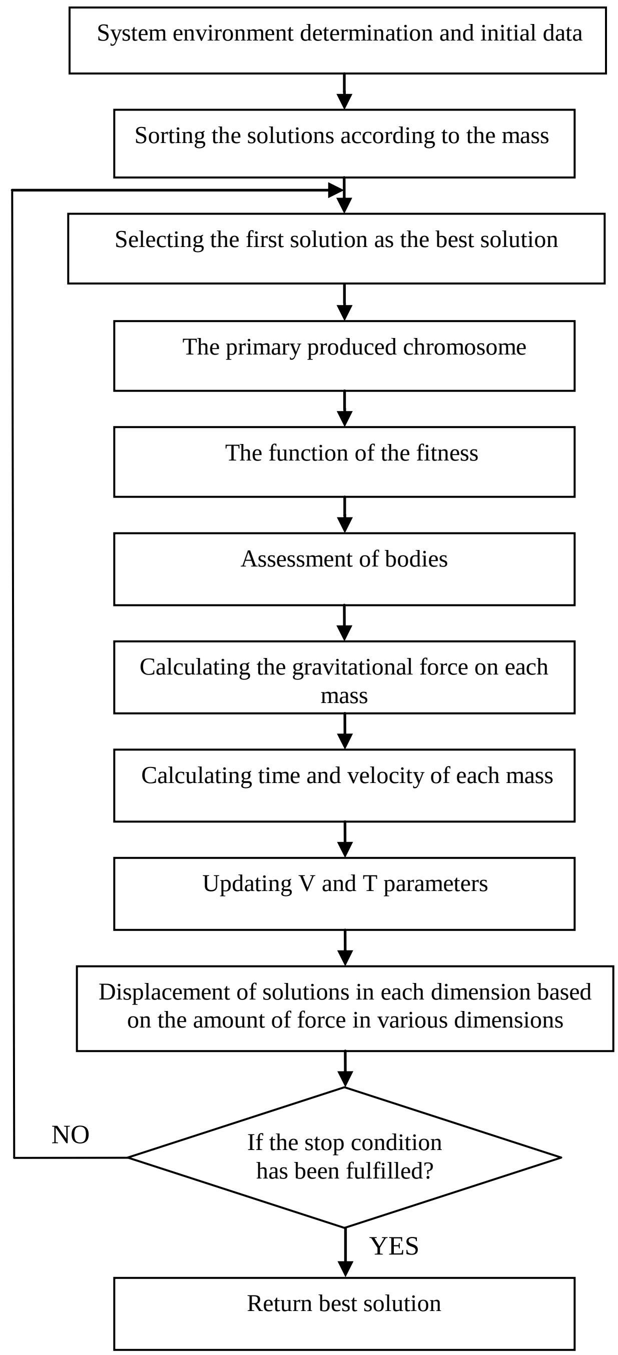


Fig. 4 The proposed algorithm

Table 3: The data of the designing test

Problem	***TIME_GELS***		***GA***	
	Time	***Fitness best***	***Time***	***Fitness best***
Test_4_4	5	175	18	266
Test_5_5	12	192	43	279
Test_8_4	17	341	80	485
Test_8_8	28	410	131	566
Test_10_10	31	598	245	702
Test_15_5	37	767	108	804
Test_20_10	46	842	268	1285
Test_20_20	54	1216	625	1896
Test_30_20	69	906	1832	1042

Table 4 will show the conclusion of presentation of proposed algorithm on the data of the test in table 3. So that you can see, the proposed algorithm finds a good and benefit answer in suitable time.

Table 4: The conclusion of the presentation of proposed algorithm

Test	***The number of the work***	***The number of operation***	***The number of generation***
Test_4_4	4	4	500
Test_5_5	5	5	600
Test_8_4	8	4	1000
Test_8_8	8	8	1200
Test_10_10	10	10	1500
Test_15_5	15	5	1000
Test_20_10	20	10	1000
Test_20_20	20	20	1000
Test_30_20	20	20	2000

Fig 5 Shows the table of competence of comparing of the two algorithm show by collection of the data of the test which is called 8–8. In this data of the test has 8 work that each works has & operation.

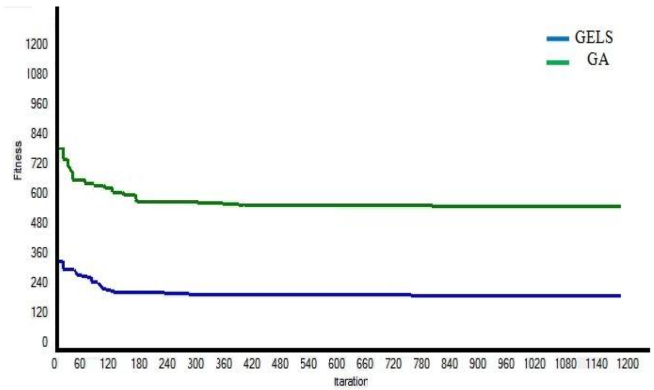


Fig. 5 Table of the competence of comparing of the two algorithm GELS and GA with collection of the data of the test 8–8

Fig 6 Shows the table dispersion of the compared of two algorithms by collection of the data of test which is called 8–8.

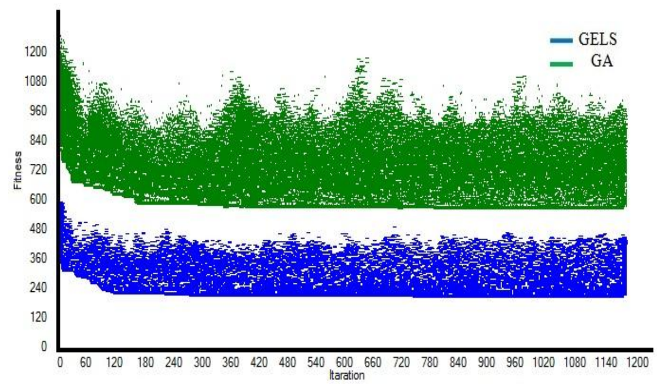


Fig. 6 Table of the dispersion, f the comparing of two algorithms by collection of the data of the test 8–8

Fig 7 Shows the table of competence of compared of the two algorithm show by collection of the data of the test which is called 20–20. In this data of the test has 20 works that each works has & operation.

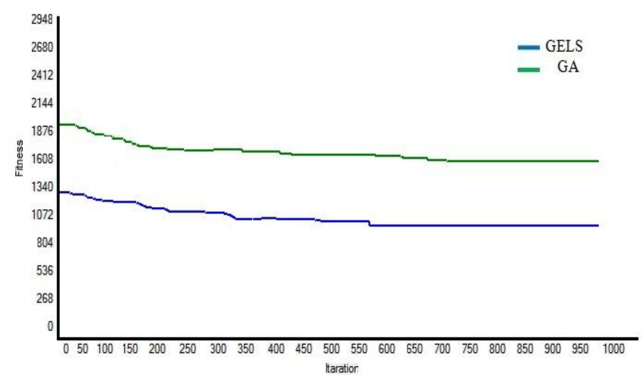


Fig. 7 Table of the competence of comparing of two algorithm GELS and GA with the collection of the data of the test 20–20

Fig 8 Shows the table of dispersion of the comparing of two algorithms by collection of the data of test which is called test 20–20.

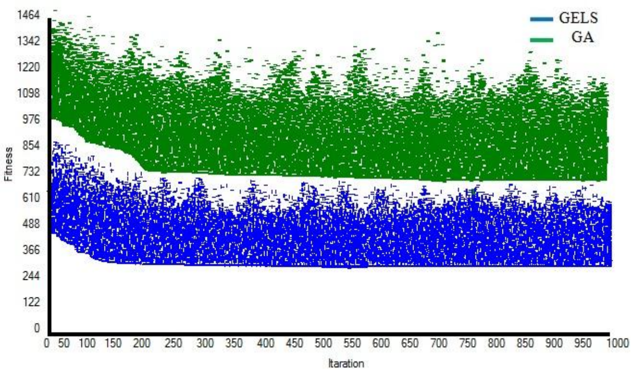


Fig. 8 Table of the dispersion of the comparing of two algorithms by collections of the data of the test 20–20

6. Conclusions

In this article one algorithm of gravitational force which is called TIME_GELS is presented for problem solving of timing of open shop. The lack of speed, time and number are the advantages of this algorithm. The aim of this algorithm to decrease the time of presentation and to catch a benefit and suitable timing for make-span so that to minimize the enough time to make-span. The efficiency of this algorithm has compared whit GA algorithm and the finding has shown that a proposed algorithm has high efficiency and recovery rather than genetic algorithm. And the amount of this recovery is more visible in the big systems. Because of the enough time to find answer and optimum timing is important in problem solving of the timing of the open shop so spending less time in proposed algorithm is one of the disadvantages of this algorithm.

References

[1] U. Dorndorf, E. Pesch and T. Phan-Huy, "Solving the Open Shop Scheduling Problem, Journal of Scheduling, Journal of Scheduling", 2001; Volume 4, Issue 3, 157-174.

[2] H. Brasel, T. Tautenhahn and F. Werner, "Constructive Heuristic Algorithms for the Open-Shop Problem", Computing Springer its journal, 1993; Volume 51, Issue 2, pp. 95-110.

[3] C. Gueret and C. Prins, "Classical and New Heuristics for the Open-Shop Problem: A Computational Evaluation", European Journal of Operational Research, 1998; Volume 107, Issue 2, 306–314.

[4] D. Alcaide, J. Sicilia and D. Vigo, "A Tabu Search Algorithm for the Open Shop problem", Article provided by Springer in its journal, 1997; Volume 5, Issue 2, pp. 283-296.

[5] C. Fang Liaw, "A Hybrid Genetic Algorithm for the Open Shop Scheduling Problem", European Journal of Operational Research, 2000; Volume 124, Issue 1, 28–42.

[6] C. Prins, "Competitive Genetic Algorithms for the Open-Shop Scheduling Problem, Mathematical Methods of Operations Research", 2000; 389-411.

[7] C. Voudouris, E. Tsang, "Guided Local Search", European Journal of Operational Research, Technical Report CSM-247, UK, August 1995; 1-18.

[8] B. Webster, "Solving Combinatorial Optimization Problems Using a New Algorithm Based on Gravitational Attraction", Ph.D. thesis, Melbourne, Florida Institute of Technology, May, 2004; 1-250.

[9] A. S. Rostami, H. M. Bernety, and A. R. Hosseinabadi, "A Novel and Optimized Algorithm to Select Monitoring Sensors by GSA", ICCIA, 2011; 829 –834.

[10] A. R. Hosseinabadi, M. Yazdanpanah, A. S. Rostami, "A New Search Algorithm for Solving Symmetric Traveling Salesman Problem Based on Gravity", World Applied Sciences Journal, 2012; Vol.16 (10), 1387-1392.

[11] B. Barzegar, H. Shirgahi, "Advanced Reservation and Scheduling in Grid Computing Systems by Gravitational Emulation Local Search Algorithm", American Journal of Scientific Research, 2011; Issue 18, pp. 62-70.

Ali Asghar Rahmani Hosseinabadi
Bachelor of Computer - Software Trends, Islamic Azad University, Science & Research Amol 2010.
Elite Member of the National Foundation of Young Researchers Club, Islamic Azad University, Behshahr Center.
His Areas of research: intelligent and innovative Optimization Algorithms are used, Scheduling Flexible Manufacturing Systems, Image Processing, Intelligent Routing, Tsp, Vrp, Time Tabeling.

Abbas Bagherian Farahabadi
Computer Sc - Artificial Intelligence Trends, Islamic Azad University, Tehran Science Research 2009. Faculty Member of Islamic Azad University, Behshahr Branch, Director of The Department of Computer Research Areas: Image processing, Intelligent algorithm

Mohammad Hossein Shokouhi Rostami
Industrial Enginerring, Iran University of Science & Echnology. His Research Interests: Tsp, Vrp, Time Tabeling, Operation Research .

Ahmad Farzaneh Lateran
He is a Software Expert at Electromoj Shomal,. As well as, he has a bachelor's degree in computer software from Iran University of Science and Technology, Iran. His research interests include WEB Technologies, Artificial Intelligence, Sensor Network & Adhoc Network.

E-Government Applications for the Information Society

Athanasios S. Drigas[1] and Lefteris Koukianakis[1]

[1] **Institute of Informatics & Telecommunications, Telecoms Lab - Net Media Lab, N.C.S.R. "Demokritos"**
Agia Paraskevi, Athens, 15310, Greece

Abstract
The incorporation of Information and Communication Technologies (ICTs) in activities and services such as education, culture and medicine constitutes a contemporary reality and governmental activities and services could not be an exception. Apart from expanded citizen services, e-government also offers a complete reshuffle of public organizations and their services through the efficient use of Internet services. This article presents an e-government platform which supports easy public access to governmental information, e-transactions between citizens and public organizations through the acceptance of e-documents submissions (e-applications and e-petitions) and their management through the e-protocol and finally, electronic delivery of a requested document. In addition, a model Greek e-government web portal is presented, which supports the Government to Citizen (G2C) model, which focuses on governmental information dissemination, which includes among others information about ministries, social services, and city news and information.
Keywords: *e-government, e-protocol, e-transactions, e-applications, e-petitions, ICTs.*

1. Introduction

Nowadays electronic government (commonly known as e-government) is a vast area of study for ICT experts and there are many ideas and policies surrounding it. The main concern of e-government applications developers and designers is the user. Therefore, a lot of effort is put into the development of quality user-friendly and user-centric services that accommodate their needs and wishes, which are essential and crucial towards the financial growth of the European Union.

Hence, e-government must fulfill the requirements of its users, namely high-quality services, effectiveness, efficiency, accountability and e-inclusion services for everybody. It is essential that public administrations become more contemporary and more innovative emphasizing on achieving impact and greater user assimilation. In order to make progress in what concerns organizational innovation and it being an ongoing process, the improvement of human resources and skills is an absolute necessity.

The extensive use of ICTs by a government in order to provide and exchange information and services with citizens, entities and businesses is commonly referred to as e-government. In particular, mainly the Internet as well as Web-based applications are being used by increasingly more governments every day in order to alter and facilitate the operation of conventional government services and their transactions with not only the citizens but also with businesses and other entities [1],[2].

The impact of e-government is none other than better government, which is more trustworthy and more reliable as it offers a vast number of advantages such as better policy outcomes, higher quality services, greater engagement with the citizens, greater collaboration between agencies, higher productivity and finally, financial benefits for all.

In much the same way as other e-services (e-learning, e-health and e-commerce), e-government introduces a great wave of technological innovation as well as government reinvention [3],[4],[5],[6].

The use of contemporary ICTs and especially of web-based technologies by e-government applications, provide citizens and businesses with access to governmental information and services and contribute to the improvement of the quality of the services and to the provision of more opportunities to actively participate in democratic processes [7],[8]. In particular, the aforementioned provisions include transactions between government and business (G2B), government and citizen (G2C), government and employee, in various governmental entities such as justice, taxation, welfare, social security, procurement, intra-governmental services etc. [9],[10],[11],[12],[13],[14]. All of the above require technical policies and specifications in order to achieve interoperability, security and IT systems coherence across the public sector [15],[16],[17].

Based on all of the above information and knowledge, this article presents a generic e-government application, which is based on a highly interactive, user-case model (citizen, employee, and administrator) and a flexible-interoperable scheme of assistive ICT tools, which essentially aims towards the design and development of more modular e-transactions. In addition, and within the aforementioned e-government environment, a user-friendly Greek e-government web portal was designed and created, which supports a very important e-government model, namely the Government to Citizen model (G2C), focusing on governmental information dissemination to Greek citizens, which includes among others information about ministries, social services, general information, etc. [18].

2. Structure of the E-Government Environment

The e-government environment consists of three systems: A web portal, the e-protocol system and finally the e-applications/e-petitions system. The last two will be described as one, since the latter may be considered as an extension of the e-protocol system. The governmental organization consists of six departments (planning, havoc compensation, housing, protocol, finances and research). Each department has one director and a large number of employees.

2.1 Web Portal's Environment Tools

The web portal environment includes tools that offer flexibility and adaptability depending on their use. The design of these tools is based on web services, such as discussion forums, chat, message box, e-libraries, which are widespread in the public web community. These tools are distinguished into two groups: ‘‘informative’’ and ‘‘communicative’’. On the one hand, the ‘‘informative’’ tools include services related to the information of governmental functions and their presentation. On the other hand, the ‘‘communicative’’ tools include services that allow the communication of users belonging to the same or different group (session level). The web portal environment enables the management of these tools according to the user groups’ permission. More explicitly, the ‘‘informative’’ tools are the following: announcements, frequently asked questions (F.A.Q.) and e-libraries. Respectively, the ‘‘communicative’’ tools are: discussion forums, message boxes, chat and e-requests. Finally, it must be noted that the environment relates the tools according to the specific user level permissions. These levels are analyzed in the sections to follow.

2.2 User Levels

Seven user levels are distinguished in the web portal environment. Different supporting tools exist in each one of them.

Depending on the corresponding use, these levels have also a different role: Administrator, Manager, Director, Employee, Protocol Administrator (Employee), Registered (Authorized) User and Unauthorized User (Guest). Each of them interacts with the other through the ‘‘informative’’ and ‘‘communicative’’ tools related to each level.

The administrator coordinates and manages the e-government application through the administrative tools. The administrator determines which user level-group has the permission to use the corresponding ‘‘informative’’ and ‘‘communicative’’ tools. Moreover, the administrator can communicate with the other user levels in order to solve issues and has the privilege of updating the system. Finally, the administrator decides about the preparation, design and diffusion of the electronic content to the citizens. Through user friendly and interactive ICT web tools, the administrator authors the governmental content.

The Manager, Director and Employees user levels are described together, as they incorporate many similarities. The manager decides about the preparation, design and diffusion of the electronic content. Moreover, through the communicative tools, the employees cooperate with the directors, the directors with the manager and the manager with the administrator with respect to discussing solutions to problems and to exchange ideas for the better functionality of the system. Finally, these three user levels play an important and diverse role in the e-protocol chain, which will be described later.

The Protocol Administrator (Employee) is responsible for the e-protocol system. Besides the “informative” and “communicative” tools, he/she has the ability to view, change (under conditions) and add applications/petitions to the e-protocol system. The applications/petitions are fully categorized and new categories can be created.

The Registered (Authorized) Users have the ability to see and change specific information regarding their account, can view the progress of their applications/petitions and finally, they can make new applications/petitions that are supported by the e-application/e-petition system.

Finally, Unauthorized Users (Guests) can enter and search the data structure as a means of gathering important information. Finally, they may be informed about the news and events through the news and calendar service.

3. User Tools and Services

3.1 Administrative Tools

The environment provides administrative tools that are divided into two groups as follows: Management of the web portal system and management of the e-protocol and e-applications/e-petitions system. The management of the web portal system incorporates management of the "informative" services and management of the "communicative" services. The management of the informative services is an important issue, as through it the administrator has the flexibility to manage the following ontologies: The users, the main menu description, the e-library, announcements and finally, the frequently asked questions (F.A.Q.). The environment tools enable the administrator to organize the informative content.

Correspondingly, the communicative services group consists of interactive forms through which the administrator manages chat session, the discussion forum and finally, the message box.

Management of the e-protocol and e-applications/e-petitions systems incorporates management of the petitions, their categories and their deadlines. The transactions executed in each group concern retrieval, insertion and update of the corresponding data. All web requests/responses are carried out through interactive and user-friendly forms.

3.2 Manager-Director-Employees Environment Tools

The environment tools for these user levels are divided into three groups: Communicative, Informative and General Services. The group of communicative services is the one that enables these three user levels to communicate with the other user levels. The tools that employ these tasks are: the discussion forum, the message box, and chat. The second group of the informative services consists of tools that enable the fast access and management of the electronic content. This content cannot be accessed by unauthorized users. The general services group includes tools that are different for each user level and play an essential role in the e-protocol chain.

3.3 Protocol Employee Environment Tools

The environment tools of this user level are similar to the ones mentioned in 3.2. In addition, this level has extended tools regarding the e-protocol system. The protocol employee has more privileges in the e-protocol system and can also interact with the e-petitions system. This level is the starting and ending point in the e-protocol chain.

3.4 Registered-Authorized Users Environment Tools

The registered-authorized users have permission to interact with the e-Petitions system. They can submit an application to the agency, as long as it is supported by the system. Moreover, the registered-authorized users have the ability to track the status of the applications they had submitted in the past. Finally, they can view and change some of their account information.

3.5 Guest-Unauthorized Users Environment Tools

The Guests – Unauthorized Users, on the other side, can browse the web portal in order to obtain valuable information regarding the agency and/or the issue(s) they wish to apply for. In order to apply, the guests-unauthorized users have to create an account (register) and interact with the e-Petitions system.

4. Structure Presentation

4.1 General Description

The presented environment is used as the web portal of the Earthquake victims' Compensation Agency. The application serves as a means for the electronic collaboration of the agency's employees as well as for the general informing of citizens regarding the e-services. The basic contribution is the application of the communicative services (discussion forum, chat, message box) as a means of central-based communication of the agency with its employees and with the citizens. The main objective of the developed infrastructure is the diffusion of information from the agency to everyone and the improvement of the e-services to the citizens. The portal's contribution with respect to information and valorization is the diffusion of the agency's information and services to the simple Internet user.

4.2 The Core of the E-Protocol System

The e-Protocol system accepts petitions from various sources such as deposits, faxes, standard mail, e-mail and from the Internet. In the case where the petition's source is the Internet, the applicant receives a confirmation number and directions in order for his/her application to be fully registered. This mechanism is intended to avoid fake applications entering the e-protocol system.

Once a new petition has entered the system, the application is regarded as a new task that must be assigned to someone in order to process it. In the beginning it is assigned by the Protocol Administrator to the Department Manager, who in his/her turn assigns it to one or more Department Directors and the latter to one or more employees. Finally, it reaches the Protocol Administrator who completes it and sends it to the Correspondence Office. All steps are automated and the system has been designed so as to minimize the need of human intervention. For instance, the users are notified by the system when a new task is assigned to them.

4.3 Security

The transactions are made under secure communications (SSL) and there is an idle timeout of 20 minutes. If there is no activity during this period, the system automatically logs out the user.

5. A G2C Web Portal

As it was mentioned before, within the presented e-government environment, a user-friendly Greek e-government web portal was designed and created, which supports a very important e-government model, the Government to Citizen model (G2C), focusing on governmental information dissemination to Greek citizens, which includes among others information about ministries, social services, and city news and information.

5.1 Structure of Content

The government to Citizen Web portal that will be described contains three links (government info, Greek news and info-city). The structure content of the system consists of the following steps:

In the first step, there was the task of collecting all the necessary information regarding the objectives of the portal. That is to say information regarding all of the Ministries of Greece, the hospitals, pharmacies, cinemas, news agencies and other social services in the city was collected, in order to be assessed.

In the second step, there was the task of categorizing this information after the assessment, in order to have a more clear view of how to organize it and present it in the best possible way. The presentation of the information within a web portal is materialized bearing in mind the needs of the user-visitor at all times. The main aim is to render the web portal accessible and user friendly to the user-visitor, who will be able to easily and quickly navigate through the web pages of his/her choice, without having to spend time unnecessarily on searching. In this case, the information regarding the Greek Ministries, the Greek News Agencies and the info-city was presented in separate web pages.

The third step involved the studying of the appropriate e-content. This task was undertaken aiming at presenting only the most interesting and necessary information for the users-visitors rather than feeding them with redundant information. Following this task, the e-content was written and embedded into the system.

The final task had to do with the in depth study of the standards and technologies used for the creation of this web portal. Such technologies include HTML for the design of the web pages and the FLASH technology, which was used in order to provide zoom in and zoom out services to individuals with visual impairments. The "design for all" approach was adopted constituting the web environment user friendly equally to disabled individuals as it is for normal citizens.

5.2 G2C Web Portal Presentation

The G2C Web Portal main page namely, "E-government Services", contains the three links, regarding information about life in Greece. The Main Page of the system has a very simple structure with a single column of the three links, in order to help the user-visitor navigate more easily. All three of these links, will be analyzed in detail, namely, "Government Info", "Cities Info" and finally "Gr News".

By clicking on the link "Government Info" on the main page, the user enters the correspondent page. There, the user can find a list of links of all the Greek Ministries. When clicking on one of the Ministries, the user will have access to information such as the address, the telephone and fax numbers and finally the web address of each Ministry, which appears on the right hand side of the page.

Hence, the user-visitor has the possibility of initially obtaining basic information about a Ministry of his/her choice such as its phone number and address. More specifically, the personal phone numbers of the Minister, Deputy Minister and General Secretary of the Ministry can also be found in this page. For more information regarding the specific Ministry, the user has the possibility of visiting the Ministries web site, by simply clicking on the provided link.

Correspondingly, by clicking on the link "Info City", the user enters the corresponding page. This page presents the implementation of an info-city for the two largest cities in Greece namely, Athens and Salonika, information that is essential for both a citizen and a visitor.

After choosing his/her preferred language (Greek or English), the user enters another page, whereby the page instigates the user to choose one of the two aforementioned cities (Athens or Salonika). After the selection of the city, a page is loaded where the user can choose one of the following social services, to get informed about vital and useful information regarding the city's activities and services of his/her interest, such as theatres, cinemas, pharmacies, hospital, useful phone numbers, inter-city motor buses, airlines, radio-taxis, railways, radio stations, TV channels, museums and art galleries.

By clicking on one of the above links, the user can get all the necessary information regarding these institutions. For example by clicking on the link "Cinemas", the user accesses information about all the cinemas in the city, their addresses and their telephone numbers.

As it was mentioned before, the FLASH technology, which was used in order to provide zoom in and zoom out services to individuals with visual impairments. By right clicking the mouse and selecting the "zoom in" function, the user views a larger version of the web page. Reversely, by choosing the zoom out function, the user can view the original version of the web page.

Finally, by clicking on the link "Gr News", the corresponding page loads. This page provides Greek news to the web users-visitors and to the overseas Greek citizens. It contains news from well-known Greek radio stations and news agencies and acts as a Greek news library. This is the first ever Greek media web portal. By clicking on one of the image links, the user can access information regarding the correspondent radio station or news agency such as their program as well as news feedbacks.

6. Conclusions

The aim of e-government and of EU governments in general is the simplification, acceleration and improvement of governmental e-services and e-provisions to the citizen, to businesses and to other interested entities. The E-Europe policy emphasizes on online government with the use of modern ICTs in order to ensure the easy, efficient and user-friendly online access to governmental information, services and decision-making procedures. In other words, the basic aim is the design and development of web-based services towards the improvement of access to public information and services, the improvement of e-transactions transparency and security and finally, to ensure the access to specific data in order that the citizens can actively participate in decision making processes.

It must be mentioned that when discussing improvement of governmental services to the citizen, it is from both a quantitative and a qualitative point of view. In addition, it means the restructuring as well as the reengineering of government entities and their respective provided services since now they incorporate state-of-the-art ICT technologies and particularly web-based tools and services.

The life cycle of an e-government transaction, begins with the easy access to governmental information and services, then moves on to the actual e-transaction between the citizen and the governmental body before the e-delivery of the requested document to the citizen.

In order to achieve a successful life cycle it is essential to design and develop an e-government platform that can support e-protocol and e-applications/e-petitions functions. This paper presented an e-government platform that was developed for a Greek public entity that supports the aforementioned functions and which presents the practical use and importance of the integration of ICTs into the different aspects of traditional government. More particularly, this e-government platform provides public information diffusion, it accepts e-document submissions and it manages them through e-protocol and finally, it supports easy communication among the entity's departments as well as robust and user-friendly management, storage, search, retrieval, handling and delivery of the e-documents.

References

[1] K. Metaxiotis, and J. Psarras, "E-Government: New Concept, Big Challenge, Success Stories", Electronic Government: An International Journal, Vol. 1, No. 2, 2004, pp.141-151.
[2] Z. Fang, "E-Government in Digital Era: Concept, Practice, and Development", International Journal of the Computer, the Internet and Management, Vol. 10, No. 2, 2002, pp.1-22.
[3] A. Drigas, J. Vrettaros, D. Kouremenos, and P. Stavrou, "E-learning Environment for Deaf People in the E-Commerce and New Technologies Sector", in WSEAS Transactions on Information Science and Applications, 2004, Vol. 1, No. 5, pp. 1189.
[4] A. Drigas, J. Vrettaros, and D. Kouremenos, "An E-Learning Management System for the Deaf People", in WSEAS Transactions on Advances in Engineering Education, 2005, Vol. 1, No. 2, pp. 20-24.
[5] A. Drigas, "E-Course Support and Delivery for E-Psychology", in WSEAS Transactions on Advances in Engineering Education, 2005, Vol. 1, No. 2, pp. 25-28.
[6] A. Drigas, and L. Koukianakis, "A Modular Environment for E-Learning and E-Psychology Applications", in WSEAS Transactions on Information Science and Applications, 2004, Vol. 6, No. 3, pp. 2062-2067.

[7] A. Xenakis, and A. Macintosh, "G2G Collaboration to Support the Deployment of E-Voting in the UK: A Discussion Paper", Electronic Government, Lecture Notes in Computer Science, Vol. 3183, 2004, pp. 240-245.
[8] A. Macintosh, E. Robson, E. Smith, and A. Whyte, "Electronic Democracy and Young People", Social Science Computer Review, Vol. 21, No. 1, 2003, pp. 43-54.
[9] E.P. Kelly, and W.J. Tastel, "E-Government and the Judicial System: Online Access to Case Information", Electronic Government: An International Journal, Vol. 1, No. 2, 2004, pp. 166-178.
[10]S.J. Barnes, R. Vidgen, "Interactive E-Government Services: Modelling User Perceptions with eQual", Electronic Government: An International Journal, Vol. 1, No. 2, 2004, pp. 213-228.
[11]P. Henman, "E-Government and the Electronic Transformation of Modes of Rule: The Case of Partnerships", Journal of Systemics, Cybernetics and Informatics, Vol. 2, No. 2, 2004, pp. 19-24.
[12] S. Verma, "Electronic Government Procurement: A Legal Perspective on the Indian Situation", Electronic Government: An International Journal, Vol. 1, No. 3, 2004, pp. 328-334.
[13] R.H. Wild, and K.A. Griggs, "A Web Portal/Decision Support System Architecture for Collaborative Intra-Governmental Planning", Electronic Government: An International Journal, Vol. 1, No. 1, 2004, pp. 61-76.
[14] P. Kanthawongs, "An Analysis of the Information Needs For E-Parliament Systems", in WSEAS Transactions on Information Science and Applications, 2004, Vol. 1, No. 5, pp. 1237-1242.
[15] J. Borras, "International Technical Standards for E-Government", Electronic Journal of E-Government, Vol. 2, No. 2, 2004, pp. 75-80.
[16] Y.S. Chen, P.P. Chong, and B. Zhang, "Cyber Security Management and E-Government", Electronic Government: An International Journal, Vol. 1, No. 3, 2004, pp. 316-327.
[17] H. Abie, B. Foyn, J. Bing, B. Blobel, P. Pharow, J. Delgado, S. Karnouskos, O. Pitkanen, D. Tzovaras, "The Need for a Digital Rights Management Framework for the Next Generation of E-Government Services", Electronic Government: An International Journal, Vol. 1, No. 1, 2004, pp. 8-28.
[18] A. Drigas, and L. Koukianakis, "Government Online: An E-Government Platform to Improve Public Administration Operations and Services Delivery to the Citizen", Visioning and Engineering the Knowledge Society. A Web Science Perspective, Lecture Notes in Computer Science, Vol. 5736, 2009, pp. 523-532.

Athanasios Drigas is a Senior Researcher at N.C.S.R. Demokritos. He is the Coordinator of Telecoms and founder of Net Media Lab since 1996. From 1985 to 1999 he was Operational manager of the Greek Academic network. He has been the Coordinator of Several International Projects, in the fields of ICTs, and e-services (e-learning, e-psychology, e-government, e-inclusion, e-culture etc). He has published more than 200 articles, 7 books, 25 educational CD-Roms and several patents. He has been a member of several International committees for the design and coordination of Network and ICT activities and of international conferences and journals.

Lefteris Koukianakis is the Head of Network and Multimedia Applications of Net Media of N.C.S.R. “Demokritos”, since 2000. He is the author of 20 International scientific articles, 3 books and 25 CD-ROMs. His research interests include e-services (e-learning, e-culture, e-government, e-psychology, e-health, e-business, e-commerce etc), as well as Java and Mobile applications.

Business goal oriented approach for Adaptive Learning System

Houda Zouari Ounaies[1], Yassine Jamoussi[2], and Henda Ben Ghezala[3]

**[1] ENSI, National School of Computer Science
Manouba, 2010, Tunisia**

**[2] ENSI, National School of Computer Science
Manouba, 2010, Tunisia**

**[3] ENSI, National School of Computer Science
Manouba, 2010, Tunisia**

Abstract

Several adaptive learning systems are currently available. Nevertheless, most existing e-learning platforms lack efficient alignment to decision makers. Our approach is two-fold. Firstly, we aim to integrate business goals in the selection process of learning concepts. Secondly, we make use of Case Based Reasoning (CBR) approach to learn from past experiences and construct effective adaptive system.

Keywords: *adaptive learning systems, business goal, CBR.*

1. Introduction

Web development in recent years has resulted in a number of increasingly important users of web applications. The characteristics of these users, their preferences and objectives are often distinct and variable over time. It is in this context that adaptive Web based systems are born. E-learning is the most concerned field by such systems [1] [2] [3] [4]. It revolves around a central objective is that learning process should be adapted to learner profiles in order to promote knowledge construction. The field of Adaptive Learning Systems (ALS) is the meeting point of these two areas: the adaptive hypermedia and the e-learning.

Several stakeholders with different perspectives on ALS services are decisive in making such project succeed. In our study, we looked more closely at the problem of developing adaptive learning systems that deal with business goals.

The principle of "one size fits all" is generally the basis of traditional learning materials [5]. The same course is addressed to a number of heterogeneous learners. Now adaptivity aims to provide personalized learning that is intrinsically linked to learner profiles. These systems can be used in the context of digital campus university, distance learning courses for enterprises employees or individual initiative simply. The goals of this training are quite different, ranging from the simple acquisition of knowledge, to the obtention of qualification or certification. Satisfy both decision makers expectations who are responsible for initiating the training and learners who actually follow this training is essential to ensure the success of these systems. Learning materials provided to learners often are not aligned with either individual skills or organizational objectives [6]. One factor that prevents the integration of adaptive learning system in the business world is the non-alignment of these systems with the customers business objectives.

During the nineties adaptive hypermedia systems for e-learning have attracted growing interest at the individual, organizational and social scale's. A decade later, a lot of frustration and dissatisfaction have been reported. Multiple failures in the field have been identified in recent years [7]. In 2006, for example, the initiative of prestigious universities as Stanford University, the University of Oxford, Yale University and Princeton University failed, and the project Alliance for Lifelong Learning, established in 2000, was officially closed in March 2006. This high failure rate, dropout before the end of the course and the low return on investment, are all symptoms reflecting the immaturity of these pedagogical devices.

The paper introduces a new approach that consider business goals in adaptive selection process of learning concept. In the following section we will introduce adaptive hypermedia reference model.

2. Adaptive hypermedia reference model

The field of adaptive hypermedia systems can be seen as the most representative type of adaptive systems. Several reference models have been developed in this area. As it is described in the Munich Reference Model [8] (figure 1), adaptive hypermedia system is based on a three explicit models: the domain model, the user model and the adaptation model. The user model separates the attributes into two categories depending on whether they depend on the domain (knowledge, learning outcomes, etc.) or not (preferences, career, etc.). The domain model represents concrete elements of the domain. The adaptation model is based on the rules of type condition / action. The conditions that trigger rules are events perceived by the system such as an action performed by the user. Actions include updating the information stored in the user model and selecting appropriate documents that will be presented to the user.

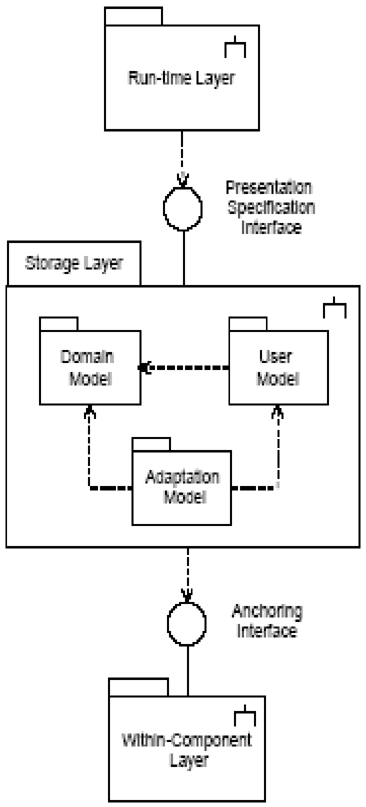


Fig. 1 Adaptive hypermedia model [8]

"Run-Time" layer describes the mechanisms supporting the user interaction with the system. A set of functions are included to ensure the presentation of pages, constructed by a set of adaptation rules and using the concepts in the domain model. These pages are adapted according to the individual user model. "Within-Component" layer describes the concrete implementation of the content and structure in hypertext nodes. This layer was introduced to isolate the other layers details of specification data and media.

3. Business goal oriented approach

Our solution is two-fold. First, we aim to deal with business goals during the adaptation process. Secondly, we adopt Case Based Reasoning (CBR) approach to capitalize from past experiences in order to construct effective adaptive system.

3.1 Business goals

The business side in adaptive learning process is essentially companies which undertake distance learning for their employees in order to achieve strategic goals. The Business Motivation Model (BMM) is an OMG standard [9] which models means, ends, action plans and factors that influence a business process. We are interested in our work in the part called "End" of the meta-model (see figure 2). The end may be either a vision or desired result. Vision is an overall image of what the organization wants to be or become. Desired Results are Goals and Objectives target that the enterprise intends to maintain or sustain. Business actors in a system of e-learning are, on the one hand, policy makers which initiate the formation, on the other hand, employees which are the end users of the system. Compared to an objective, a goal tends to be a longer term, more qualitative than quantitative and general rather than specific. The goal is usually a business strategy, expressed by policy makers. Compared to a goal, an objective tends to be short-term, quantitative rather than qualitative and specific rather than general. It is the way to achieve the goal so it is done by the learners. An objective must be measurable and completed within a predetermined time space.

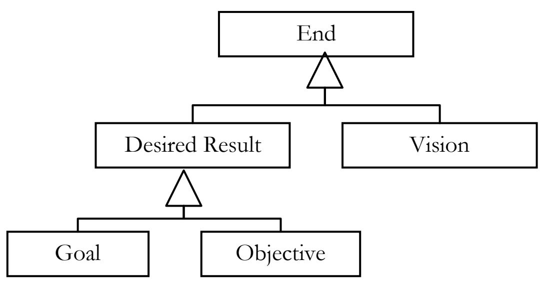


Fig. 2 Business meta-model [9]

Meeting the client's goals is an essential criterion for the sustainability of developed systems. In many cases, e-learning market has failed to meet the expectations of customers. Therefore, measuring the extent to which adaptive learning systems are suitable is crucial. We propose to construct "Goal Alignment Degree" (GAD) measure. For that reason we consider the meta-models associated with the two concepts "Achieved Business Goal" and "Desired Business Goal" in order to identify quantifiable elements.

GAD measure is the ratio of achieved performance for one targeted knowledge level related to the organization learners (Ach_Goal), compared to the rate of desired success for this targeted knowledge level (Desired_Goal_Rate). This is expressed by the following formula:

$$\text{GAD} = \frac{Ach\_Goal}{Desired\_Goal\_Rate} \qquad (1)$$

Meta-model based approach is adopted in order to identify quantifiable elements used for "Achieved Business Goal" and "Desired Business Goal" description.

3.1.1 Desired Business Goal

In an e-learning system, policy makers' goals are mainly related to learner outcomes. As shown in figure 3, the desired business goal is described by a desired achievement which is characterized by quantifiable elements that are success rate and targeted knowledge level throughout a learning session. Learning objective can be certification or a qualification. As it can be an informational learning that basically aims to improve learner's knowledge. In both cases it is a knowledge acquisition related to one or more learning concepts.

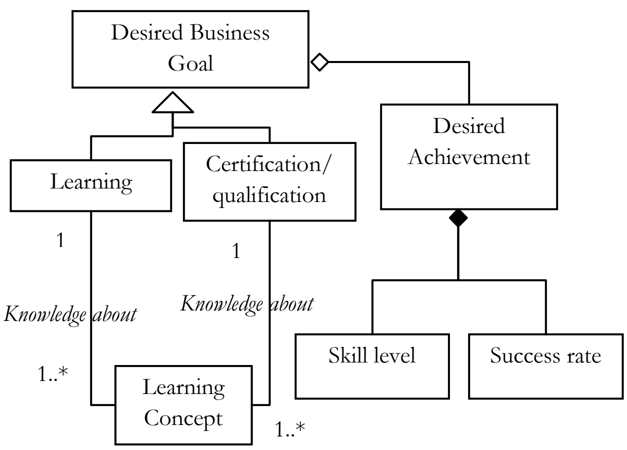


Fig. 3 Desired business goal concept

3.1.2 Achieved business goal

A learning goal is actually realized through a set of objectives that must be achieved during a learning scenario. The latter describes the learning unit that can be of variable granularity [10]:

- A basic activity is the finest granularity of a learning situation. An elementary activity can pursue a specific learning objective or contribute to it when integrated within a structured sequence; Example: Call web page, post a contribution to a forum or a problem solving through a simulation.

- Activities sequence corresponds to an average size of a learning situation in which several basic activities or sequences are organized to achieve a specific learning in terms of knowledge or skills. This structure must be able to express the relationship of sequence and parallelism, to describe the sequence and conditions or to specify the flow data process associated, Example: "classical" Sequence <Following a presentation, do application exercises, self evaluate>.

- Pedagogical unit is the highest granularity of learning situation in which a set of learning sequences are assembled to form a logical unit on a given learning theme. Example: A course, a module, a license, a master, etc.

Achieved goal is described in terms of accomplished success state regarding to decision makers targeted level. These elements are described by the meta-model in Figure 4:

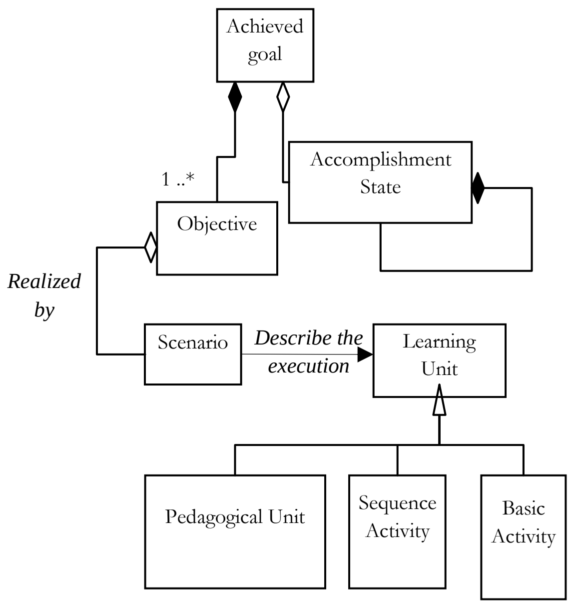


Fig. 4 Achieved goal meta-model

To assess the accomplishment state for a given goal we propose to apply propagation rules that describe the state reached by the goal in terms of its objectives accomplishment. Rules are described in first-order logic and the following predicates are used:

State_Obj(Obj, E) : shows the status E of an objective Obj for the whole learners' session. Two states are possible: "realized" or "unrealized".

Participate(Goal, Obj) : illustrates that an objective Obj contribute to the achievement of Goal.

State_Goal(Goal, Ui): returns the accomplishment value of a goal for a user Ui.

State_Acc_Obj(Obj, Ui) : returns objective achievement state related to a user Ui.

The propagation rule is triggered when the objective reaches the state "realized". The term "Cont_Obj" indicates the objective contribution proportion to achieve goal. This rule is used to assess a goal achievement state.

∀ Ui ∀ Obj
State_Obj(Obj, realized) =>
(∀ Goal Participate(Goal, Obj) =>
(State_Goal (Goal, Ui) = (State_Goal(Goal, Ui) + Cont_Obj * State_Acc_Obj(Obj, Ui)))

The goal achievement state (Ach_Goal) is the success rate realized by all learners involved in the training session and this regarding to knowledge level required by the decision maker. The goal of the decision maker is not bound to a single learner, but to all learners.

Assessment of business goal achievement provides feedback about ALS alignment with business needs and then adaptive behavior should be triggered accordingly. Case Based Reasoning (CBR) can be adopted to enhance the adaptation performance and learn from past experiences.

3.2 Adaptive behavior based on CBR approach

CBR is based on knowledge reuse and past experience capitalization to solve problems. As described in [11], a new problem is solved by finding a similar past case, and reusing it in the new problem situation. Past experiences are stored in a Case Base. One case is composed by both problem situation and past experienced solution. A new solution is obtained by comparison between new situations and past cases. In the following section, we describe a learning situation in an adaptive learning system context.

Measures characterize the learning situation. One situation is related to one session launched by a learner. It is described first by user model state which reflects the learner knowledge about concepts of the application domain and the Goal Alignment Degree (GAD).

Adaptive learning system presents functionalities fitting user needs. Each function corresponds to a set of orchestrated Web Services. Figure 5 illustrates UML class diagram describing adaptive learning situation.

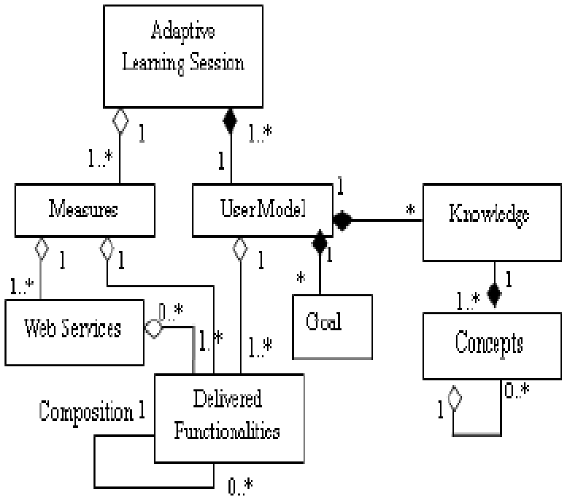


Fig. 5 Adaptive Learning Situation

Different learning scenarios are possible. Measurement criteria characterize each specific situation and reflect the validity of delivered functionalities adaptation. Solutions to resolve possible failures are proposed. Adjustments results are stored as a new case in the CBR system.

In the context of our research, we have implemented an automated case creation tool. Information from the current state of the user model, the delivered functionalities and related Web services are stored automatically in the Case Base.

This CBR project aims to develop a system that is able to create cases using a mostly automated process, and to develop a Java-based case-based reasoning engine that can be integrated to any adaptive Web-based educational system. The cases capture the adaptation performance expertise of learning system and best adjustment practices that can be reused in similar situation.

4. Conclusion

This paper has introduced a business goal centered-approach for adaptive learning system. The objective is to align the system to the decision makers needs. We propose to reinforce the adaptation process by adopting Case-Based Reasoning. Adaptive learning situations are stored in the Case Base. Measures related to delivered adaptive functionalities and to business goal alignment characterize each situation. CBR allows learning from past experiences and enhances adaptation decision making. Currently, we are studying the integration of automated and real-time case creation in adaptive e-learning environment.

References

[1] F. Patterno and C. Mancini. Designing Web Interfaces Adaptable to Different Types of Use. In Proceedings of the Workshop Museums and the Web, 1999.

[2] A. Kavcic. Fuzzy user modeling for adaptation in educational hypermedia. IEEE transactions on systems, man and cybernetics. Part C, Applications and reviews, 34(4), 2004.

[3] N. Stash, Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System. Phd thesis. Eindhoven University of Technology, 2007.

[4] N. Stash, D. Smits, C. Romero, and S. Ventura. Authoring and Management Tools for Adaptive Educational Hypermedia Systems: The AHA! Case Study. Book Series Studies in Computational Intelligence, 62, 285-308, Springer, 2007.

[5] A. Gunasekaran, R.D. McNeil and D. Shaul. E-learning: research and applications. Industrial and Commercial Training Journal, 34(2): 44 – 53, 2002.

[6] K. Leyking, P. Chikova, and P. Loos. Competency - and Process-Driven e-Learning – a Model-Based Approach. The Electronic Journal of e-Learning, 5(3):183-194, 2007.

[7] D. Keegan, J. Lossenko, I. Mázár,, P.F. Michels, M.F. Paulsen, T., Rekkedal, J. Toska, and D. Zarka. E-learning initiatives that did not reach targeted goals. Megatrends Project, NKI Publishing House, Norway, 2007.

[8] N. Koch. *Software Engineering for Adaptive Hypermedia Systems: Reference Model. Modeling Techniques and Development Process*. Phd Thesis, 2001.

[9] Business Motivation Model (BMM) Specification, OMG, technical Specification V1.1, 2010.

[10] J.P. Pernin and A. Lejeune. Modèles pour la réutilisation de scénarios d'apprentissage. Actes du colloque TICE Méditerranée, Nice, 2004.

[11] Aamodt, A., Plaza, E.: AICom - Artificial Intelligence Communications, IOS Press, Vol. 7: 1, (1994) 39-59.

Dynamic Tracking Protocol for Maneuvering Targets in Sensor Network

Fathy Aamer, and Seham Ebrahim

Department of Information Technology, Faculty of Computer & Information Cairo University
Cairo, Egypt
Department of Computer Engineering, Modern Academy of computer engineering
Cairo, Egypt

Abstract

Target tracking is one of the most important applications in wireless sensor networks (WSNs). Energy efficiency is one of the most important design goals for target tracking. We propose and evaluate a distributed, energy-efficient, light-weight framework dynamic clustering algorithm for target localization and tracking in wireless sensor networks. Since radio communication is the most energy-consuming operation, this framework aims to reduce the number of messages and the number of message collisions, while providing refined accuracy. The tracking protocol is also adaptive to the target velocity to ensure high accuracy with minimum power consumption. We envision a hierarchical sensor network that is composed of a static backbone of sparsely placed high-capability sensors which will assume the role of a gateway upon triggered by certain signal events and moderately to densely populate low-end sensors whose function is to provide sensor information to their gateway upon request. A cluster is formed when the activation message is received from the gateway. All the active nodes send their data to the next predicted node. The last active node (source node SN) at certain hop count sends its data to the gateway. The election of reference node is based on the data received no overhead due election. Acoustic signal strength detected by the active nodes exceeds a pre-determined threshold introduce a back off time proportional to its value. This back off time reduces collision rates. Using NS2 and Matlab to introduce the accuracy of target location, total power consumption for different proposed protocols, the effect of changing frequency rate at different target speeds and the drop ratio for different protocols.

Keywords: *Wireless sensor networks; target tracking; energy consumption; Reporting frequency*

1. Introduction

Wireless sensor networks are systems of small, low-powered networked sensing devices deployed over an interested area to monitor interested events and perform application specific tasks in response to the detected events. One of the most significant and elementary application is localization and tracking moving targets. The type of interested signals includes temperature, sound, light, magnetism and seismic vibration. A sensing modality is determined based on the types of targets to be tracked.

We study the issue of tracking mobile targets using wireless sensor networks. The whole tracking process is divided into the positioning stage and the tracking stage. In the process of target tracking, a lot of factors [13] affect energy consumption, which includes the number of moving targets, the speed of moving targets, data reporting frequency, tracking data accuracy, data collection frequency, and so on. Obviously, as design goals, the sensor nodes surrounding the moving target should be able to promptly provide reliable status information about the moving target and the area around it in an energy efficient way, and the sensor nodes should report this information to the sink in a fast and energy efficient way. In this paper, we propose a Dynamic target Tracking Protocol (DTP) for single-target tracking in WSNs. In summary, the key contributions of this paper are as follows:

- The active node group which consists of sensor nodes in the tracking state can track the target dynamically by using the node state transition mechanism and weight calculation mechanism.
- The sensor nodes should participate in tracking target or not depends on the received activation message.
- The data reporting frequency can be adjusted according to the velocity of the target to reduce unnecessary data transmission and improve the energy efficiency.
- Dynamic election of the source node (that collect the data from detected node and send it to the gateway) based on the information included in the received message.

The rest of the paper is organized as follows: Section 2 summarizes some existing target tracking protocols in WSNs. Section 3 discusses the basic ideas behind the proposed DTP protocol, including node state transition, node weight calculation, the initialization and reconstruction of the communication. The system architecture is discussed in Section 4. Section 5 includes the calculation of the main system parameters. We simulate and evaluate the proposed DTP protocol in Section 6. Section 7 concludes the paper.

2. Related Works

There are many dynamic clustering algorithms in Object Tracking Sensor Network (OTSN). However, there may be overlap area between two clusters which would cause redundant data and unnecessary data transmissions from the cluster head to base station. Thus, an algorithm is needed to reduce these unnecessary data transmissions. In-Sook Lee[8] proposed an algorithm to avoid redundant data and reduce energy consumption based on prediction result of moving objects. It also introduces an algorithm for cluster formation to minimize the overlap area of clusters. H. Yang [9] presents a Distributed Predictive Tracking protocol. The protocol uses a static clustering based approach for tracking mechanism. The speed of the target is a variable parameter which they use to evaluate its impact to the tracking quality. They assume high dense network and algorithm to avoid target lose in case of wrong prediction. The energy consumption of the tracking algorithm and communication cost is evaluated. Aysegul [10] proposed a Target tracking system that predicts future locations of the target and awakens the corresponding leader nodes so that the nodes along the trajectory self organize to form clusters to collect data related to the target in advance and thus reduce the target misses. The algorithm first provides detection of the target and forms a cluster with the neighboring nodes around it. After the selection of the cluster header, the coordinates of the targets is estimated using localization methods and cooperation between the cluster nodes under the control of the leader node. Two phase timer based are used in leader election algorithm which elects the node closest to the target as the leader node. Xiaofei[11] proposed a herd-based target tracking protocol. A sensor node has three states, namely, sleeping state, sensing state, and tracking state. Each sensor node is associated with a weight to be used to make a state transition among the three states. When a target moves into a monitoring area, a cluster node is selected as the herd head that is responsible for reporting the target information to the sink in the network. The sensor node can adjust the frequency of data reporting according to the velocity of the target. Khin [12] proposed a model to identify the lifetime of target tracking wireless sensor network. The model is static cluster based architecture and aims to provide two factors. First, it is to increase the lifetime of target tracking wireless sensor network. Secondly, it is to enable good localization result with low energy consumption for each sensor in the network. Xingbo Yu [2] proposes a quality aware information collection protocol in sensor networks for tracking applications. A sensor senses the environment and communicates its readings to the server periodically. A grid based network topology was simulated to evaluate the performance of their adaptive tracking protocols. They used four separate protocols to compare the total energy consumption in the sensor network over the period of tracking. The protocol explores trade-off between energy and application quality to significantly reduce energy consumption in the sensor network thereby enhancing the lifetimes of sensor nodes.

Most tracking protocols addresses target localization. In many cases, however, targets are not cooperative with sensors, e.g., enemy vehicles and unregistered victims in disaster areas. In the non-cooperative cases, however, the absence of the original signal strength information prevents the use of absolute distance estimates. Instead, one can estimate the original signal strength by collecting and analyzing a number of sensing data, which often requires non-linear optimization techniques. Qianyu Ye [3] addresses the measurements and accuracy of location calculation. In many circumstances, it is difficult to obtain the distance from the observation nodes to the surveillant target. As a result, using distance as a factor to determine the target location is not always feasible. They proposed a target tracking algorithm based on the sensing information which is in proportion to or in inverse proportion to the exponent of distance. The proportions of sensing information on different sensors are used to draw the loci of the target. The target location can be approximated by these loci. Zhijun Yu [7] proposed tracking protocol for acoustic target tracking in wireless sensor networks. The location of the target is calculated depending on time difference of arrival (TDOA) calculation. The distributed processing algorithm is based on mobile agent computing paradigm and sequential Bayesian estimation. To decrease the wireless communications, they proposed to represent the belief by parameterized methods such as Gaussian approximation or Gaussian mixture model approximation. They presented an attraction force function to handle the mobile agent migration planning problem, which is a combination of the node residual energy, useful information, and communication cost. Jeongkeun Lee [1] proposed light-weight localization algorithm, dubbed Ratiometric Vector Iteration (RVI) that is based on relative distance ratio estimates rather than absolute distance estimates. Vinodkrishnan [4] presents a wireless sensor network protocol, Trail that supports distance-sensitive tracking of mobile objects for in-network subscribers upon demand. Information about closer objects is required more often and more quickly than that of farther objects. Livio

[5] presents localization platform that exploits a single-hop WSN, based on a Microchip MCU and a Cypress RF device, to tracking its moving nodes. Specifically, the nodes of the network are divided in three sets. The algorithms adopted to manage the network and to localize moving nodes are discussed. A working prototype based upon the hardware platform, software and protocol described in this paper has been deployed and tested, and some results are shown. Guanghui [6] considered the issue of how to track mobile targets with certain level of *quality of monitoring* (QoM), while conserving power. They addressed the target tracking problem by taking into account of both the *coverage* and the *QoM*. Also studied analytically whether or not the detection/observation made by a single sensor suffices to tracking the target in a reasonably populated sensor network. In particular, for the sake of determining the relay area, the scheme requires that each sensor node knows its own position, and a localization algorithm has to be included. This may increase both the computational and communication overhead. In [18] Adaptive Target Tracking Protocol (ATTP) has been proposed which at each time step, employs two helpful tools. First, an extended Kalman filter (EKF)-based estimation technique to predict the tracking error and second, an energy consumption model to estimate energy consumption based on different number of nodes and sampling time intervals. After that by using these estimations selects the best number of nodes and sampling time interval according to an objective function which is defined based on tracking accuracy and energy consumption. In [19] proposed a new noise model for distance measurement to account for both additive noises and multiplicative noises. Also they discussed how to numerically obtain the necessary noise statistics using the least square (LS) method. Second, proposed a new measurement conversion method using maximum likelihood estimation. This method is based on the triangulation idea, commonly used in global positioning systems (GPS), and recently extended to sensor localization in wireless sensor networks. But here they applied maximum likelihood estimation to obtain a good linear estimate of the target position in the Cartesian coordinates and an approximate covariance matrix of the converted measurement noise. Finally, the converted measurement and noise covariance matrix are then used in a standard Kalman filter to update the target state estimate recursively. In our previous work [20] the different tracking parameters were introduced and how it affects on tracking accuracy. In this paper the power reduction using network structure and MAC protocol are introduced and also two cluster head election algorithms to reduce power consumption.

3. The Dynamic Tracking Protocol

Tracking protocols in general address certain application civilians or military. According to required application the network topology, sensor types and required accuracy is designed. In our tracking protocol the application is tracking small vacuolar its speed changing between zero to 20m/sec. The network topology is similar to that proposed in [14,15,12] and presented in section 4. The architecture aims to design and implement a wireless sensor network that enables energy efficient detection and tracking of events.

The network includes four gateways at the network corners, and acoustic nodes. Each gateway includes dual photo sensor. These gateways guard the network sides. The sensor nodes have two states sleep or active. They are sleep but during the detection operation gateways activate the nodes near the target trajectory. These activated nodes start to activate other nodes. After certain hop count the calculated data is sent to the gateway which this node belong to. The node which sends the data is Source Node (SN).

The proposed tracking protocol achieves dynamic election of SN based on the information in the received message. This method reduced overhead election in communication and time, but introduces redundancy, as more than one SN can be elected. The tracking protocol activates only the near nodes to the targets trajectory. Nodes are normally in sleeping mode. The four gateways at the corners of the monitoring area are in sensing state all the time. At the detection of a target the two gateways surround the target start to activate *n* nodes. These nodes start to calculate the RSS between them and the target. Each active node starts to activate at least 2 of its neighbor depending on their ID. We propose two cases of activation and also two cases of SN election. These cases are introduced in section 3.2.

The Tracking protocol is summarized in the following steps:

1- The target passes the monitoring area from one side.
2- The two gateways at this side start the activation of the acoustic sensor nodes as in algorithm 1
3- The activation message includes target calculated distance from the acoustic node, the active nodes ID (includes its ID and other nodes), the initial time step (tracking resolution) and the initial reporting frequency
4- At the reception of this activation message, acoustic nodes turn into active mode.
5- Each active node knows its neighbor and active nodes.
6- According to the RSS which is translated to estimate distance, it sends activation message according to two cases:

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Online): 1694-0814 | ISSN (Print): 1694-0784
www.IJCSI.org

a. Activation message is sent to its entire neighbor near at this distance including pervious active nodes, RSS, time step and reporting frequency decremented by one.
b. Activation message is sent to only acoustic nodes in sleeping mode including pervious active nodes, RSS, time step and reporting frequency decremented by one.

7- When the hop count of frequency reporting reach zero. The decision of the SN which sends all target data to its gateway has two cases:
   a. The last activated nodes send their data. Where any previous activated nodes are rejected from the SN as there are newer nodes than them.
   b. The last activated nodes and higher weighting factor send their data.

8- The gateway updates the time step for higher quality if necessary and reporting frequency according to the target speed. The estimate positions of the target are calculated.

9- The updated message is sent to the next predicted active nodes.

10- The active nodes return to sleeping mode as the RSS is decreasing and reaches its minimum threshold.

3.1 The Gateway

There are four gateways at the corner of the network topology. Each gateway has a photo sensor with dual beam. Gateways are powerful node and active all the detection time. The network is divided into four clusters each cluster join one of the gateways.
The functions of the gateway:

- In sensing mode all time.
- Start the activation of sensing the nearest nodes to the targets
- Define the period SN state to send data
- Define the recovery process
- Includes localization algorithm to do the processing
- Define the election period to the SN (in case of target stop moving)

The network distribution is uniform has N nodes. The probability of finding n_s nodes near the target for RSS s , gateway1and2 *gd1,gd2* is .

(1)

1

For acceptable accuracy as introduced in [20] n_s nodes is considered at n_s =5. Number of nodes near the target with received signal s and beam length r. if the network dentistry is ¢

$$N_{active} = \int_{r}^{r} \int_{s-r}^{s+r} ¢\, dxdy \qquad (2)$$

The two gateways which the target inter the monitoring area from their side start to calculate at least 5 nodes. These nodes satisfy the iteration of algorithm1. This algorithm may run at one gateway if the target is passing in its cluster or in both gateways in case passing between them.

Algorithm 1

d_{thresh}=0.25r.
Switch cluster
Case one cluster
While n_{active}< 5
Calculate the nodes satisfy this relation

$$\left|d_{gd1,n} - d_{tar}\right| \le d_{thesh} \oplus \left|d_{gd2,n} - d_{tar}\right| \le d_{thesh}$$

If n_{active}<5
d_{thresh}= d_{thresh}+5;
End;
End;
Case between clusters
While n_{active}< 5
Calculate the nodes satisfy this relation

$$\left|d_{gd1,n} - d_{tar}\right| \le d_{thesh} and \left|d_{gd2,n} - d_{tar}\right| \le d_{thesh}$$

If n_{active}<5
d_{thresh}= d_{thresh}+5;
End;
End;
End;

The format of the activation message sent to the active nodes by gateway

Nodes ID in sensing mode	Time the sensing mode starts To each node	Resolution: count decremented by each received node and time step decrements occur

The first field includes the ID of all active nodes by the gateway or SNs. The second field is the time it starts sensing mode. The third field stats the resolution of the tracking. The time step is time between successive detection. The resolution is the reporting frequency. The count here means after how many hop count it needs the data.

3.2 Sensor Nodes

Initially, we assume that each sensor has its own localization and stationary awareness. The sensing area of a sensor is shaped as a circle with a radius r centered at the location of the sensor. We also assume that sensors can directly communicate with the neighboring sensors within a radius at least *2r* because the communication range is twice more than the sensing range. The system uses uniform distribution for sensor placement. All sensor nodes have the same initial power, receiving, transmitting, sleeping and sensing power. The nodes have two states active and sleep state as described in table 1. They are initially set at the sleep modes and turned to active modes only when there is a target. Therefore, they can save their energy and make their life time long.

Table 1: Modes of Sensor Nodes

Processing Mode	Active Mode	Sleeping Mode
Tasks perform in the corresponding mode	All parts of the sensor are fully powered. Sensor performs sensing the environment, processing, receiving packet from cluster head, and transmitting RSS value to source node and then to gateway.	It is energy saving mode. Sensor only listens an interrupt to go active mode.

Addition of nodes in the detection process:

The two gateways around the target send the target position to the activated nodes. The nodes change its state from sleep to active mode. The RSS of the target is measured. According to its value and pervious calculated target position it can estimate the target direction. Each active node knows:

1- The active nodes at their time step and previous time steps
2- The first target position calculated by gateways
3- The RSS of each activated nodes
4- Estimated target direction

There are two cases in node addition. Each active node estimates the new nodes which going to be activated. In the first case of nodes addition includes new and previous activated nodes. In the second case neglects the previous activated nodes. So the activation messages are reduced and they reach zero in case of slow speed or target stop.

The number of included active nodes has a limit. At reaching the limited size the nodes with minimum RSS are removed to add the newer nodes with higher RSS. The limit of number of nodes included is $l_{act,node}$

$$l_{act,node} = \overline{N}_{active} \times 0.25 \times f_{rep} \qquad (3)$$

Where $\overline{N}_{active}$ the mean of estimated active node, and f_{rep} is the reporting frequency which is calculated in section 5.3

Format of the message of active sensing nodes

Nodes ID sensing mode	Time the sensing mode starts to each node	The distance of the target	Resolution count decremented by each received node and time the decrements occur	SN_{ch} ID and the Time it is elected

It has two more fields, the estimated distance of the target and ID of the SN_{ch} with its election Time. There are two methods proposed to elect SN as cluster head (SN_{ch}). The energy consumption of each of them is calculated and also the redundancy of each method is introduced.

The election of SN_{ch}

The election of SN_{ch} is a self decision based on the data received from the incoming message. The SN_{ch} election introduces redundancy in their number. The number of SN_{ch} depends on the estimated target velocity, distance, its kinematic mode, number of neighbors (network connectivity) and time step.

$$\text{Network connectivity} = ¢ \times \int_{y_i}^{y_i+r} \int_{x_i}^{x_i+r} dxdy \qquad (4)$$

Where ¢ is the network density, (x_i, y_i) is the node position. In the proposed topology in section 4.1 which is a square connectivity are 4 for hexagon topology are 6 and increased for dense network. Increasing connectivity increase accuracy but in our proposed protocol increase redundancy of SN numbers. To overcome redundancy in case of dense network a weighting factor is introduced and discussed in details. Connectivity also affects the calculation of $\overline{N}_{active}$. It has maximum value 4 if we neglecting the estimated direction of the target and 2 including estimated direction.

The first case

The SN_{ch} is elected as the last activated node. In case the sensor nodes send activation to its entire neighbor near the target, which are 4 in our topology. To activate the entire four nodes that means the target is on the active node. But for actual case only two of them is activated

The second case without weighting factor

The probability of choosing more than SN_{ch} depend on target velocity. The slow target stays in the same detection region for time. As result know new active node are

introduced and all old active node becomes SN_{ch} as the hop count reach zero.
The detection region is a circle quadrate of radius *r*. for linear motion the target moves distance $v\Delta T$ at each time step. The distance between the first detection (d_{RSS}) and departure from the detection region *r* is $r\text{-}d_{RSS}$. The redundancy number of SN_{ch} N_{rd} for network connectivity Net_{conn} is calculated as follow.

$$N_{rd}\, v\Delta T \approx (r - d_{RSS})Net_{conn}$$

$$N_{rd} \approx \frac{(r - d_{RSS})Net_{conn}}{v\Delta T} \qquad (5)$$

In case of nonlinear motion the traveled distance inside the detection region is the line integral over the pass. The maximum redundancy in the first case for both kinematic models is equal to the connectivity. In the second case for both kinematics differs according to target parameters.

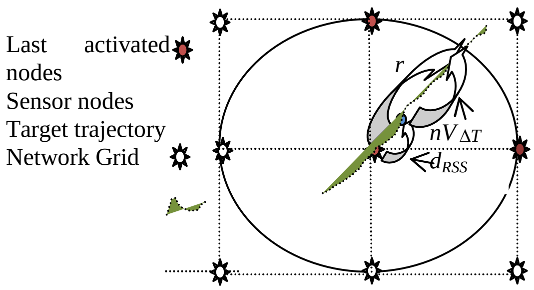


Fig. 1 The Election f SN_{ch}

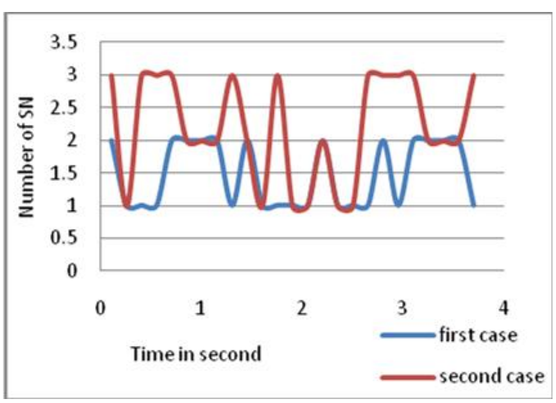


Fig. 2 the redundancy of SN_{ch} in the two cases of reporting algorithms

Second case with weighting factor

At the activation of new sensor node, its activated node sends to each pair different weighting factor. This factor is random value from 1 to 10 as the target direction is unknown to the activated node. The sent values must be different. Each active node start to compare its weighting factor with the other nodes since they are included in the same received message.

The weighting factor is neglected by the node in case if target received signal is double the threshold value. The node considers itself a SN_{ch}.
The introduced SN for both first and second case of DTP simulation runs for ΔT =0.5 sec, r=15m and v=20m/sec shown in figure 2. The maximum number of redundancy of SN_{ch} in first case is 2 while the second case introduces 3.

4. System Architecture

In this section, the proposed wireless sensor network topology is presented in figure 3. Models introduced in tracking process as Localization model and target kinematics is included in this section.

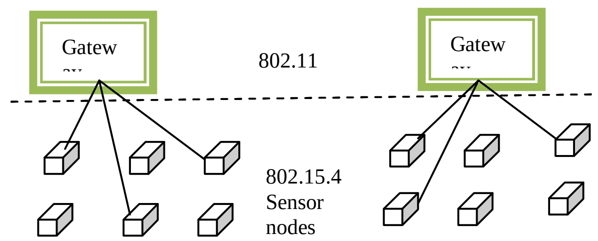


Fig. 3 the system architecture of the network

4.1 Network Topology

The network topology is grid with 100 nodes as introduced in figure 4. The monitoring area is *100mX100m*. There are four gateways at the corner. The beam length is 15*m*. Each node has two states sensing modes (active) and sleeping. The nodes can communicate with their neighbors and know the distance between them. The gateway at the corners is active all the time while all other nodes are sleeping until they receive activation message. The gateway has photo detectors, dual beam detection. It's rang out is 30~60 m. Their response time starts from 50 msec. Each active node calculates the distance between it and the target using Received Signal Strength (RSS).

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Online): 1694-0814 | ISSN (Print): 1694-0784
www.IJCSI.org

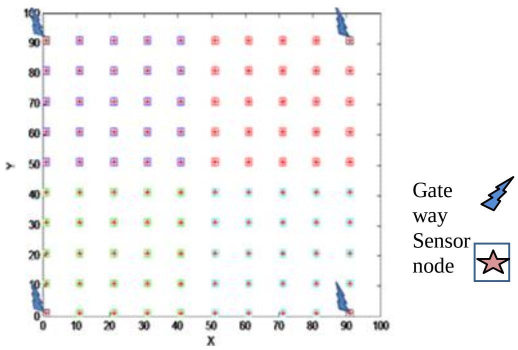


Fig. 4 Network topology has four clusters for each gateway

4.2 Localization Model

The fundamental principle applied in the energy-based approach is that the signal strength (i.e., energy) of a received signal decreases exponentially with the propagation distance:

$$s_i = a \,.\, \|x - x_i\|^{-\sim} + Q_i \qquad where \quad 0 < i < N$$

(6)

S_i is the received signal strength by the i^{th} sensor node and $a \in \Re$ is the (unknown) strength of an acoustic signal from the target, $x \in \Re^2$ is the target position yet to be determined, $x_i \in \Re^2$ is the known position of the i^{th} sensor, i ~ s the (known) attenuation coefficient, and Q_i is the white Gaussian noise with zero-mean and $\pm^2$ variance . We consider eXtreme Scale Mote measured values to evaluate our algorithm. In [16] for acoustic signal less than 1.38 $«w/m^2$, range more than 20m and event duration more than 2.3sec.

The estimated distance d_{RSS} has an error d_{err}.

$$d_{RSS} = d_{act} + d_{error} \tag{7}$$

The introduced error must be less than the change of RSS which proportional to the change of target position. To avoid activation of wrong nodes we assume that

$$\Delta RSS \propto v\Delta T \tag{8}$$

$$2d_{error} \prec \frac{1}{2} v\Delta T \tag{9}$$

$$d_{error} \prec 0.25 v\Delta T$$

Where v is target velocity, ΔT time step.

4.3The Target Models

Two kinematics models are introduced for quick maneuver detection. Constant Velocity Model (CV). This model is the most commonly used. The target is assumed to move with nearly constant velocity. Coordinated Turn Rate Model (CT) where set of accelerations modeled are those in the direction normal to the velocity, which model constant turns μ . The nonlinear model is the case of considering a fifth state ω instead of constant value.

For notational simplicity, $x_t \cong \{x_t, \dot{x}_t, y_t, \dot{y}_t\}$ refers to the state (coordinates and the velocities).

All models are in the generic state-space form

$$x_{t+1} = F\, x_t + G\, u_t \tag{10}$$

$$F_{CV} = diag\begin{bmatrix} 1 & \Delta T \\ 0 & 1 \end{bmatrix}, \quad G = diag\begin{bmatrix} \frac{\Delta T^2}{2} \\ \Delta T \end{bmatrix}$$

$$F_{CT} = \begin{pmatrix} 1 & \frac{\sin \mu\Delta T}{\mu} & 0 & -\frac{1-\cos\mu\Delta T}{\mu} \\ 0 & \cos\mu\Delta T & 0 & -\sin\mu\Delta T \\ 0 & \frac{1-\cos\mu\Delta T}{\mu} & 1 & \frac{\sin\mu\Delta T}{\mu} \\ 0 & \sin\mu\Delta T & 0 & \cos\mu\Delta T \end{pmatrix} \tag{11}$$

Where $u_t \sim N(0, diag(\pm_x^2, \pm_y^2))$.

5 System Parameters

Performance improving of the tracking protocol depends on calibration of some parameters. These parameters control accuracy and reduction of power consumption and so increase the network life time.

5.1 Time step calculation ΔT

The minimum value of time step is the T_{cycle} as calculated by Jeongkeun Lee [1]. The maximum value is calculated depending on the estimated target velocity v where the target can be seen by the node when passing by at least once. The inter node distance d_{in} for sensing rang r is $\frac{r}{\sqrt{2}}$

$$v = \frac{current\ position - previous\ position}{\Delta T}$$

$$r\sqrt{2} \cong 2\Delta T v \tag{12}$$

$$\Delta T \cong \frac{r}{v\sqrt{2}}$$

5.2 The window size and backoff time

The window size w_{size} here includes the mean time delay in processing, queuing, carrier sense time and transmission. The back of time T_{bf} is a random variable normal distribution proportional to the RSS at each node.

$$T_{bf} = f(«) - f(RSS) \quad where \quad f(x) = N(« = Rss_{\max}, \pm = 0.2) \tag{13}$$

To reduce energy consumption, the time step must be varied according to target velocity.

$$T_{cycle} \le \Delta T \le \frac{r}{v\sqrt{2}} \quad (14)$$

5.3 Reporting frequency

After localizing the target, the SN is obliged to report the estimated location to the sink node. Usually, the sink is a gateway connecting the sensor network to subscribers. Therefore, reporting to the sink normally incurs multi hop message relaying and multiple transmissions which depend on the hop distance between the SN and the gateway. If the target is stationary or moves around within a small area, the estimated location does not change notably as time goes by and the same location estimate (or possibly with small deviation) is repeatedly reported to the gateway. In these cases, it is not necessary and even wasteful to report (almost) same location estimates repeatedly. Rather than reporting every time leader performs localization, we propose to dynamically schedule a reporting frequency considering the target's kinematics. The reporting frequency is a hop count decremented each ΔT that avoids clock synchronization. The maximum communication rang is R_{max} .The network cluster diagonal is for regular shape topology has maximum side length l number of regular clusters *ch* is

$$d_{\max} = \frac{2l\sqrt{2}}{ch} \quad (15)$$

The reporting frequency f_{rep} is proportional to the target velocity and maximum travelled distance without reporting.

$$f_{rep} = \frac{v}{traveled\ \text{distance}}, \quad \text{traveled distance} = n_{hop} v \Delta T;$$

$$where\ n_{hop}\ \text{maximumhop count between gateway and SN}$$

$$n_{hop} \approx \frac{d_{\max}}{R_{\max}}$$

$$f_{rep} = \frac{1}{n_{hop}\Delta T} = \frac{R_{\max}}{d_{\max}\Delta T} \quad (16)$$

Considering these parameters time step weighting factor backoff time and reporting frequency increase the QOS and minimize power consumption.

5.4 The Energy Model

A sensor consumes energy in three ways: the energy used to run circuitry, the energy used to send radio transmission and energy used in computational. The energy used to run circuitry is proportional to the number of bits in the message. Say that the message length is k bit, define E_{elec} be the energy per bit. Then the energy used to run circuitry is $E_{elec}k$. This energy is the same for receiver and transmitter [17]. The energy for transmitter to send k bits over distance d is $£_{amp}\, k\, d^{\sim}$, where $£_{amp}$ is the energy constant for the radio transmission and ~ is the path loss exponent. The total energy for a receiver to handle a k-bit message is

$E_{Rx}(k) = E_{elec}k$ =24 «w for 2.5msec (18)

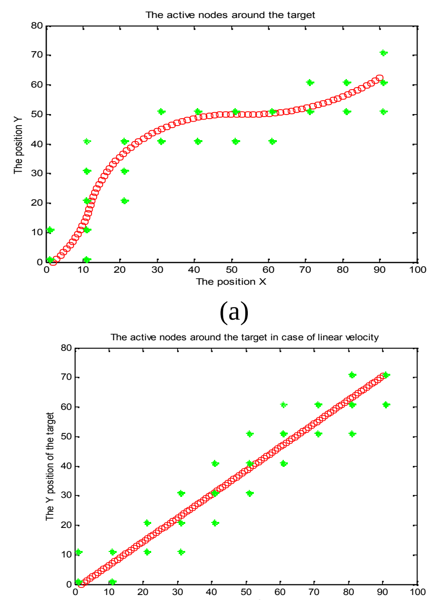
The total energy for a transmitter to send a k-bit message over distance d is,

$E_{Tx}(k; d) = E_{elec}k + £_{amp}\, k\, d^{\sim}$ =48 «w for 2.5msec (19)

The computational energy for acoustic sensor [16] with parameters number of samples256 and sampling frequency 8,192 *hz* and expected covered path length 2,300*ms* results an average power of 370 «w.

6 Performance Evaluations

The network topology is described in section 4.1. The sensor nodes communicate using MAC protocol (IEEE 802-15.4). Nodes start sensing mode by receiving a signal from the cluster head or its neighbor and end sensing mode as there is no detection. All nodes have the same type. Starting energy is 1000 joule, Idle power 0.01 watt, Sleep power 0.001 watt and Sending power 0.2 watt.In figure 5 the target trajectory for two kinematics model with velocity 20m/sec. The green points are the activated node near the target during the monitoring time.



(a)

(b)

Figure 5 (a) the nonlinear motion trajectory surrounded by the activated nodes, (b) the linear motion trajectory surrounded by active nodes both has velocity of 20m/sec

The total power consumption in the two cases without including the weighting factor are introduced in figure 6. The accumulation of power consumption in first case increase with time. Second case introduces less power consumption.

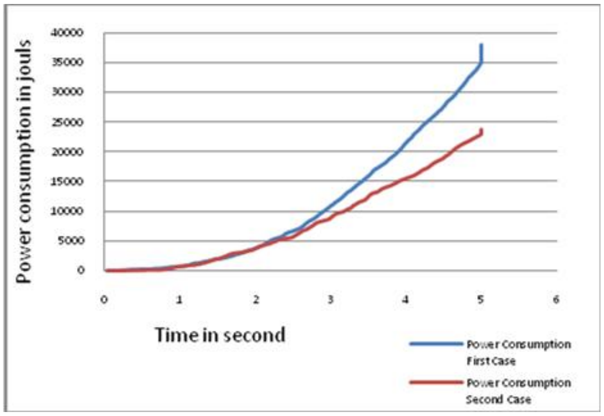


Fig. 6 total power consumption including gateways communication

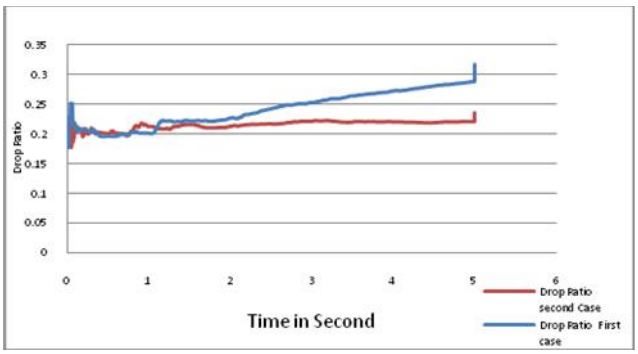


Fig. 7 Total dropping rations without considering the backoff time

The dropping ratio of the second case also is less which also introduces reduction in power consumption.

7 CONCLUSION

Target tracking is an application dependent protocol. As it depends on the problem formulation as environment, network topology, types of sensor nodes included in network (same type or different types) and targets type. The proposed DTP minimize power consumption using cross layer as it benefits from MAC protocol. Also the relation between different parameters as target velocity, sensing beam, network connectivity, sensing time step, reporting frequency and target hit are joined in a simple equation to give a relative design value. Two election algorithms are introduced. In first election algorithm SN_{ch} is elected as the last activated node. In second case the sensor nodes send activation to its entire neighbor near the target, which are 4 in our topology. At the activation of new sensor node, its activated node sends to each pair different weighting factor. This factor is random value from 1 to 10 as the target direction is unknown to the activated node. For both algorithms the impact on redundancy in SN_{ch} is simulated.The second algorithm has minimum power consumption and also introduces less packet drop ratio. In this paper the main objective is high accuracy with minimum power consumption. The tracking protocols benefits from the other sensor network protocols to reduce power consumption and also reduce packet drop ratio. In the future work changing reporting frequency and introducing multi target from the same type to evaluate the power consumption and redundancy rate.

References

[1] Jeongkeun Lee , Kideok Cho, Seungjae Lee, Taekyoung Kwon, Yanghee Choi. Distributed and energy-efficient target localization and tracking in wireless sensor networks. Computer Communications 29 (2006) 2494–2505

[2] Xingbo Yu, Koushik Niyogi, Sharad Mehrotra, Nalini Venkatasubramanian**.** Adaptive Target Tracking in sensor Networks. *University of California Irvine.* [Online] http://www.ics.uci.edu/~xyu/pub/cndsXYUUCI.pdf.

[3] Qianyu Ye, Yu Liu, Lin Zhang, Chan-hyun Youn**.** Target tracking by proportion of distance in wireless sensor network. *: Mobility networks, systems and applications.* POSTER SESSION, 2007, pp. 195-201.

[4] Vinodkrishnan Kulathumani, Anish Arora, Mukundan Sridharan**.** Trail: A distance-sensitive Sensor Network service for distributed object tracking. *ACM Transactions on Sensor Networks (TOSN).* 2009, Vol. 5, 2.

[5] Livio Denegri, Sandro Zappatore, Franco Daoli**.** Sensor Network based localization for continuous tracking applications. 2007.

[6] C.Hou, Guanghui He and Jennifer**.** tracking Targets with Quality in wireless sensor network. [Online] http://www.ieee-icnp.org/2005/papers/06_heg_tracking.pdf.

[7] Zhijun Yu, Jianming Wei, and Haitao Liu. An Energy-Efficient Target Tracking Framework in wireless sensor networks. *Hindawi Publishing Corporation EURASID Journal on Advances in Signal Processing.* 524145, 2009.

[8] In-Sook Lee, Zhen Fu, WenCheng Yang and Myong-soon Park**.** An Efficient Dynamic Clustering Algorithm for Object Tracking in Wireless Sensor Networks. *Complex systems and applications Modeling control and simulations.* 2007, Vol. 14, pp. 1484-1488.

[9] B.Sikdar, H. Yang and. A Protocol for Tracking Mobile Targets using Sensor Networks. *Proceedings of IEEE Workshop on Sensor Network Protocols and Applications.* 2003.

[10] Aysegul Alaybeyodu, Kayhan Erciyes,Aylinkantarci and Orhan Dagdeviren. Tracking fast Moving Targets in wireless Sensor Networks. [Online] http://www.tr.ietejournals.org.

[11] Xiaofei Xing, Guojun Wang and Jie Wu**.** Herd-Based Target Tracking Protocol in Wireless Sensor Networks. *Proceedings of the 4th International Conference on wireless Algorithms.System and Applications.* 2009.

[12] Khin Thanada Soe**.** Increasing Lifetime of Target Tracking Wireless Sensor Networks. *World Academy of Science, Engineering and Technology.* 42 , 2008.

[13] Xu, Y., Winter, J., Lee, W.: Prediction-Based Strategies for Energy Saving in Object Tracking Sensor Networks, In: Proceedings of the 2004 IEEE International Conference on Mobile Data Management, (2004) 346-357

[14] Songhwai Oh, Phoebus Chen, Michael Manzo, and Shankar Sastry, HYPERLINK "http://www.eecs.berkeley.edu/%7Esho/papers/icra06_iwsnrts.html " Instrumenting Wireless Sensor Networks for Real-time Surveillance **,** in *Proc. of the International Conference on Robotics and Automation* (ICRA), Orlando, FL, May 2006.

[15] Chen *et al"* Experiments in Instrumenting Wireless Sensor Networks for Real-Time Surveillance (video)" **,** in *Proc. of the International Conference on Robotics and Automation* (ICRA), Orlando, FL, May 2006.

[16] Prabal Dutta, Mike Grimmer, Anish Arora, Steven Bibyk, and David Culler, Design of a wireless sensor network platform for detecting rare, random, and ephemeral events Source ,Proceedings of the 4th international symposium on Information processing in sensor networks, Sponsors: IEEE Signal Processing Society,ACM, 2005

[17] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, **"**Energy-Efficient Communication Protocol for Wireless Microsensor Networks'' Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000

[18] Elham Ahmadi, Masoud Sabaei, "A New Adaptive Target Tracking Protocol inWireless Sensor Networks", International Journal of Computer Applications (0975 – 8887) Volume 28– No.10, August 2011

[19] Xingbo Wang, Minyue Fu, Huanshui Zhang, "Target Tracking in Wireless Sensor Networks Based on the Combination of KF and MLE Using Distance Measurements", IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 11, NO. 4, APRIL 2012.

[20] Seham Meawed Aly , Raafat El Fouly, Hoda Baraka,"Real Time Localization Algorithm and Tracking Protocol For Maneuvering Targets In Non-Uniform Sensor Networks", Intelligent solutions in embedded systems,2009 seventh workshop on, 2009 , page: 21-27

Hierarchal Variable Switching Sets of Interacting Multiple Model for Tracking Maneuvering Targets in Sensor Network

Fathy Aamer, and Seham Ebrahim

Department of Information Technology, Faculty of Computer & Information Cairo University
Cairo, Egypt

Department of Computer Engineering, Modern Academy of computer engineering
Cairo, Egypt

Abstract

Tracking maneuvering targets introduce two major directions to improve the Multiple Model (MM) approach: Develop a better MM algorithm and design a better model set. The Interacting Multiple Model (IMM) estimator is a suboptimal hybrid filter that has been shown to be one of the most cost-effective hybrid state estimation schemes. The main feature of this algorithm is the ability to estimate the state of a dynamic system with several behavior modes which can “switch” from one to another. In particular, the use of too many models is performance-wise as bad as that of too few models. In this paper we show that the use of too many models is performance-wise as bad as that of too few models. To overcome this we divide the models into a small number of sets, tuning these sets during operation at the right operating set. We proposed Hierarchal Switching sets of IMM (HSIMM). The state space of the nonlinear variable is divided into sets each set has its own IMM. The connection between them is the switching algorithm which manages the activation and termination of sets. Also the re-initialization process overcomes the error accumulation due to the targets changes from one model to another. This switching can introduce a number of different models while no restriction on their number. The activation of sets depends on the threshold value of set likely hood. As the likely hood of the set is higher than threshold it is active otherwise it is minimized. The result is the weighted sum of the output of active sets. The computational time is minimum than introduced by IMM and VIMM. HSIMM introduces less error as the noise increase and there is no need for re adjustment to the Covariance as the noise increase so it is more robust against noise and introduces minimum computational time.

Keywords: *Interacting Multiple Model (IMM), Probabilistic Data Association, Sensor Network.*

1. Introduction

Multiple-model approach provides the state-of the-art solutions to many problems involving estimation, filtering, control, and/or modeling. One of the most important problems in the application of the multiple-model approach is the design of the model set used in a multiple-model algorithm. There are two types of model-set design: online and offline. Offline design is for the total model set or the initial model set in a variable-structure approach, as well as for the fixed structure approach. In a fixed-structure algorithm, the model set used cannot vary and is determined *a priori* by model-set design. In a variable-structure algorithm, the model set in effect at any time is determined by an adaptation process, known as *model-set adaptation*, which may be viewed as an online (real-time) design process and will depend on the total model set determined *a priori* if such a set exists. In this paper we study the IMM with a large number of operating modes and introduce its performance for different modes changes. Then we replace this IMM with another structure of IMM. The structure includes sets of IMM. Each set includes part of these operating modes. At initialization all the sets are active but the right set which introduce lower innovation will be active while the others will be switched off When the system change to a different active set the diverging of the active set will cause system initialization and activation to all sets and start to tune to the right set. This algorithm overcome the problem of large number of modes in IMM, activate only the right set not all sets which introduce less computation complexity, and allow variation of time step to large values to reduce energy consumption in Sensor Network while tracking.

In section 2 the related work is presented. We introduce IMM in section 3. Variable structure IMM is in section 4 while the HSIMM in section 5. The Results of IMM of The First Tracking Problem is in section 6.The design of HSIMM is introduced in section 7 while its results of the first tracking problem in section 8. The second tracking dynamics and its results are introduced in section 9, finally section 10 presents conclusion and future work.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Online): 1694-0814 | ISSN (Print): 1694-0784
www.IJCSI.org

2. The Related Work

The IMM was introduced in [1] which summarized the state-of-the-art of the IMM and its variants. They discuss and compare the base-line IMM algorithm with variable-structure variants, multiple sensor variants, correlated noise variants, glint-noise influenced variants, and others to know more about IMM. But it is shown theoretically that the use of too many models is performance-wise as bad as that of too few models. In [7] they introduced difficulties of excess of measure while [8] introduced the difficulties of using IMM in Radar system. In [2] they introduced limitations of the fixed structure algorithm. Then presented theoretical results pertaining to the two ways of overcoming these limitations. select/construct a better set of models and/or use a variable set of models. The new approach was illustrated in non stationary noise identification problem. In our previous work [12] considering Extended Kalman filter and IMM in the same tracking problem to introduce the accuracy and time delay of the two tracking algorithm. In this paper structure of IMM sets are introduced to overcome limited number of sets and nonlinearity of target motion model. The structure includes number of IMM (set of modes) working separately. But the right set will be considered while the others will be switched off.

3. The Interacting Multiple Model

The system is described by the model:

$$x(k+1) = F(k, m(k+1))x(k) + G(k, m(k+1)) * u(k, m(k+1)), \ldots \quad (1)$$

$$z(k) = H(k)x(k) + w(k), \ldots \quad (2)$$

Where $x \in \Re^{n_x}$ is the system state vector, $z \in \Re^{n_z}$ is the measurement vector, $u \in \Re^{n_u}$ and $w \in \Re^{n_z}$ are mutually uncorrelated, white zero mean Gaussian noises with coveriances Q_u and *Rw* respectively. The parameter m_k presents the current system mode. *F* is the system dynamic matrix, and *H* is the measuring matrix.
Because the accurate system model is unknown, the system is described by a number of models. The event that the i^{th} model m_i is actual at time *k* is denoted as
$M_i(k)=\{m(k)=m_i\}$.
It is assumed that the system model sequence is a Markov chain with transition probabilities

$$P\{m_j(k+1) \mid m_i(k)\} = P_{rij}(k)$$

where $\sum_{j=1}^{r} P_{rij}(k) = 1, \qquad i = 1,2.....,r$

Designing model set is the major element affects the performance of the IMM. These models represent different mode of operation of the system. If the system operating mode is fare from these models it doesn't converge. Or if we add models near to each other in parameters the estimated state may converge to the wrong model and also models probabilities after converging to the wrong model cause system divergence.
We introduce two models one linear motion and one coordinated turn model for quick maneuver detection.
1) Constant Velocity Model: This model is the most commonly used. The target is assumed to move with a constant velocity. For notational simplicity, $x_t \cong \{x_t, \dot{x}_t, y_t, \dot{y}_t\}$ refers to the state (coordinates and the velocities) of a single target following this motion model and u_t is the corresponding motion noise.

$$x_{t+1} = \begin{Bmatrix} 1 & \Delta T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta T \\ 0 & 0 & 0 & 1 \end{Bmatrix} x_t + \begin{Bmatrix} \frac{\Delta T^2}{2} & 0 \\ \Delta T & 0 \\ 0 & \frac{\Delta T^2}{2} \\ 0 & \Delta T \end{Bmatrix} u_t \quad (3)$$

where $u_t \sim N(0, diag(\pm_x^2, \pm_y^2))$.
2) Coordinated Turn Rate Model: This model assumes that the target moves with a constant speed (norm of the velocity vector) and a constant known turn rate μ. Again, we denote x_t as the state of a single target from this class and u_t as the corresponding motion noise. To introduce the case of a target moving with varying μ change from 0.2 to 1.8 we divide the IMM into 10 modes, one linear and other 9 modes at different μ

$$x_{t+1} = \begin{pmatrix} 1 & \frac{\sin\mu\Delta T}{\mu} & 0 & -\frac{1-\cos\mu\Delta T}{\mu} \\ 0 & \cos\mu\Delta T & 0 & -\sin\mu\Delta T \\ 0 & \frac{1-\cos\mu\Delta T}{\mu} & 1 & \frac{\sin\mu\Delta T}{\mu} \\ 0 & \sin\mu\Delta T & 0 & \cos\mu\Delta T \end{pmatrix} x_t + \begin{pmatrix} \frac{\Delta T^2}{2} & 0 \\ \Delta T & 0 \\ 0 & \frac{\Delta T^2}{2} \\ 0 & \Delta T \end{pmatrix} u_t \quad (4)$$

Where u_t has the same Gaussian distribution as in (3).

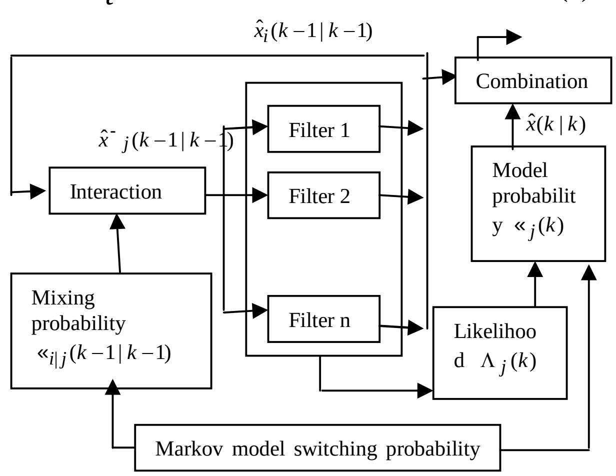


Fig. 1 Scheme of IMM algorithm

A Markov transition matrix is used to specify the probability that the target is in one of the modes of operation. The model probabilities are updated at each new measurement, and the resulting weighting factors are used in calculating the state. One cycle of a practical IMM algorithm consists of the following steps which in [3],[4],[9]

One Cycle of the IMM Estimator

1. Model-conditioned re initialization (for $i = 1, 2,\ldots,;M$):

Predicted model probability: $\hat{\mu}_{k|k-1}^{(i)} \overset{\Delta}{=} P\{m_k^{(i)} \mid z^{k-1}\} = \sum_j \pi_{ji} \mu_{k-1}^{(j)}$

Mixing weight: $\mu_{k-1}^{j|i} \overset{\Delta}{=} P\{m_{k-1}^{(j)} \mid m_k^{(i)}, z^{k-1}\} = \pi_{ji} \mu_{k-1}^{(j)} / \mu_{k|k-1}^{(i)}$

Mixing estimate: $\bar{x}_{k-1|k-1}^{(i)} \overset{\Delta}{=} E[x_{k-1} \mid m_k^{(i)}, z^{k-1}] = \sum_j \hat{x}_{k-1|k-1}^{(j)} \mu_{k-1}^{j|i}$

Mixing covariance: $\bar{P}_{k-1|k-1}^{(i)} = \sum_j \begin{matrix}[P_{k-1|k-1}^{(j)} + (\bar{x}_{k-1|k-1}^{(i)} - \hat{x}_{k-1|k-1}^{(j)}) \\ (\bar{x}_{k-1|k-1}^{(i)} - \hat{x}_{k-1|k-1}^{(j)})']\mu_{k-1}^{j|i}\end{matrix}$

2. Model-conditioned filtering (for $i = 1, 2,\ldots M$):

Predicted state: $\hat{x}_{k|k-1}^{(i)} = F_{k-1}^{(i)} \bar{x}_{k-1|k-1}^{(i)} + G_{k-1}^{(i)} \bar{w}_{k-1}^{(i)}$

Predicted covariance: $P_{k|k-1}^{(i)} = F_{k-1}^{(i)} \bar{P}_{k-1|k-1}^{(i)} (F_{k-1}^{(i)})' + G_{k-1}^{(i)} Q_{k-1}^{(i)} (G_{k-1}^{(i)})'$

Measurement residual: $\tilde{z}_k^{(i)} = z_k - H_k^{(i)} \hat{x}_{k|k-1}^{(i)} - \bar{v}_k^{(i)}$

Residual covariance: $S^{(i)} = H_k^{(i)} P_{k|k-1}^{(i)} (H_k^{(i)})' + R_k^{(i)}$

Filter gain: $K_k^{(i)} = P_{k|k-1}^{(i)} (H_k^{(i)})' (S_k^{(i)})^{-1}$

Updated state: $\hat{x}_{k|k}^{(i)} = x_{k|k-1}^{(i)} + k_k^{(i)} \tilde{z}_k^{(i)}$

Updated covariance: $P_{k|k}^{i} = P_{k|k-1}^{(i)} - k_k^{(i)} S_k^{(i)} (k_k^{(i)})'$

3. Model probability update (for $i = 1,2,\ldots M$):

Model likelihood: $L_k^{(i)} \overset{\Delta}{=} p[\tilde{z}_k^{(i)} \mid m_k^{(i)}, z^{k-1}] \overset{assume}{=} \dfrac{\exp[-(1/2)(z_k^{(i)})'(S_k^{(i)})^{-1}\tilde{z}_k^{(i)}]}{\left|2\pi S_k^{(i)}\right|^{1/2}}$

Model probability: $\mu_k^{(i)} \overset{\Delta}{=} P[m_k^{(i)} / z^k] = \dfrac{\mu_{k|k-1}^{(i)} L_k^{(i)}}{\sum_j \hat{\mu}_{k|k-1}^{(j)} L_k^{(j)}}$

4. Estimate fusion: $\hat{x}_{k|k} \overset{\Delta}{=} E[x_k \mid z^k] = \sum_i \hat{x}_{k|k}^{(i)} \mu_k^{(i)}$

Overall estimate: $\hat{x}_{k|k} \overset{\Delta}{=} E[x_k \mid z^k] = \sum_i \hat{x}_{k|k}^{(i)} \mu_k^{(i)}$

Overall covariance: $P_{k|k} = \sum_i \left[P_{k|k}^{(i)} + (\hat{x}_{k|k} - \hat{x}_{k|k}^{(i)})(\hat{x}_{k|k} - \hat{x}_{k|k}^{(i)})'\right] \mu_k^{(i)}$

4. Variable Structure Multiple Model Estimation

General speaking, a fixed structure MM (FSMM) algorithm is one with a fixed set of models while a variable structure MM (VSMM) algorithm is one with a variable set of models. The set of models used by MM algorithm at time k is denote by M_k and the total set of models is denoted **M.** As M is the union of all M_k's. The MM algorithm is said to have a fixed structure if the model set M_k used is fixed over time (i.e. $M_k = \mathbf{M}$; $\forall k$). Otherwise it is said to have a variable structure.

VSIMM Recursion

1. Model-set conditioned (re)initialization [$\forall m_i \in M_k$]:

Predicted model probability:
$\hat{\mu}_{k|k-1}^{(i)} \overset{\Delta}{=} P\{m_k^{(i)} \mid M_k, M_{k-1}, z^{k-1}\} = \sum_{m_j \in M_k} \pi_{ji} \mu_{k-1}^{(j)}$

Mixing weight: $\mu_{k-1}^{j|i} \overset{\Delta}{=} P\{m_{k-1}^{(j)} \mid m_k^{(i)}, M_{k-1}, z^{k-1}\} = \pi_{ji} \mu_{k-1}^{(j)} / \mu_{k|k-1}^{(i)}$

Mixing estimate: $\bar{x}_{k-1|k-1}^{(i)} \overset{\Delta}{=} E[x_{k-1} \mid m_k^{(i)}, z^{k-1}] = \sum_{m_j \in M_k} \hat{x}_{k-1|k-1}^{(j)} \mu_{k-1}^{j|i}$

Mixing covariance:
$\bar{P}_{k-1|k-1}^{(i)} = \sum_{m_j \in M_k} [P_{k-1|k-1}^{(j)} + (\bar{x}_{k-1|k-1}^{(i)} - \hat{x}_{k-1|k-1}^{(j)})(\bar{x}_{k-1|k-1}^{(i)} - \hat{x}_{k-1|k-1}^{(j)})']\mu_{k-1}^{j|i}$

2. Model-conditioned filtering [$\forall m_i \in M_k$]:

Predicted state: $\hat{x}_{k|k-1}^{(i)} \overset{\Delta}{=} E[x_k \mid m_k^{(i)}, M_{k-1}, z^{k-1}] = F_{k-1}^{(i)} \bar{x}_{k-1|k-1}^{(i)} + G_{k-1}^{(i)} \bar{w}_{k-1}^{(i)}$

Predicted covariance: $P_{k|k-1}^{(i)} = F_{k-1}^{(i)} \bar{P}_{k-1|k-1}^{(i)} (F_{k-1}^{(i)})' + G_{k-1}^{(i)} Q_{k-1}^{(i)} (G_{k-1}^{(i)})'$

Measurement residual:
$\tilde{z}_k^{(i)} = z_k - E[z_k \mid m_k^{(i)}, M_{k-1}, z_{k-1}] = z_k - H_k^{(i)} \hat{x}_{k|k-1}^{(i)} - \bar{v}_k^{(i)}$

Residual covariance: $S^{(i)} = H_k^{(i)} P_{k|k-1}^{(i)} (H_k^{(i)})' + R_k^{(i)}$

Filter gain: $K_k^{(i)} = P_{k|k-1}^{(i)} (H_k^{(i)})' (S_k^{(i)})^{-1}$

Updated state: $\hat{x}_{k|k}^{(i)} \overset{\Delta}{=} E[x_k \mid m_k^{i}, M_{k-1}, z^k] = x_{k|k-1}^{(i)} + k_k^{(i)} \tilde{z}_k^{(i)}$

Updated covariance: $P_{k|k}^{i} = P_{k|k-1}^{(i)} - k_k^{(i)} S_k^{(i)} (k_k^{(i)})'$

3. Model probability update (for i = 1; 2; : : : ;M):

Model likelihood:
$L_k^{(i)} \overset{\Delta}{=} p[\tilde{z}_k^{(i)} \mid m_k^{(i)}, z^{k-1}] \overset{assume}{=} \dfrac{\exp[-(1/2)(z_k^{(i)})'(S_k^{(i)})^{-1}\tilde{z}_k^{(i)}]}{\left|2\pi S_k^{(i)}\right|^{1/2}}$

Model probability: $\mu_k^{(i)} \overset{\Delta}{=} P[m_k^{(i)} / z^k] = \dfrac{\mu_{k|k-1}^{(i)} L_k^{(i)}}{\sum_j \hat{\mu}_{k|k-1}^{(j)} L_k^{(j)}}$

4. Estimate fusion: $\hat{x}_{k|k} \overset{\Delta}{=} E[x_k \mid z^k] = \sum_i \hat{x}_{k|k}^{(i)} \mu_k^{(i)}$

Overall estimate: $\hat{x}_{k|k} \overset{\Delta}{=} E[x_k \mid z^k] = \sum_i \hat{x}_{k|k}^{(i)} \mu_k^{(i)}$

Overall covariance: $P_{k|k} = \sum_i \left[P_{k|k}^{(i)} + (\hat{x}_{k|k} - \hat{x}_{k|k}^{(i)})(\hat{x}_{k|k} - \hat{x}_{k|k}^{(i)})'\right] \mu_k^{(i)}$

As the outcome of the advances during the past three decades, the state-of-the art FSMM estimators usually perform quite well for problems that can be handled by the use of a small set of models. Consequently, they have found a great success in solving many state estimation problems compounded with structural or parametric uncertainty, particularly in target tracking. However, when they are applied to solve many real-world problems (e.g., many practical target-tracking problems), it is often the case that the use of only a few models is not good enough. Although further development is certainly possible, the FSMM estimation techniques have arrived at such a stage

that great improvement can no longer be expected. This perception is based on an understanding of the fundamental limitations of the FSMM approach.
These limitations stem from the following facts:

- It assumes fundamentally that the system mode at any time can be represented (with a sufficient accuracy) by one of a fixed set of models that can be determined in advance.
- The set of possible system modes is not fixed. It depends on the hybrid state of the system at the previous time.
- As shown in [34],[38], use of more models in an FSMM estimator does not necessarily improve performance; in fact, the performance will deteriorate if too many models are used.
- It cannot incorporate certain types of a priori information.
- Clearly, the amount of computational resource required by an FSMM estimator increases dramatically with the number of models used.

5. Hierarchal Switching sets of IMM

The variable IMM increase the accuracy of the estimated position but still doesn't solve the problem of increasing number of models to estimate a wide range of variation. It works as IMM but takes only the results of models with relativity higher model probability.
In our proposed algorithm we mix the advantage of small number of models of IMM and variation of its activation over a wide range. The space plane of the nonlinear variable (e.g. acceleration or turning angle) is divided into sets. Each set has its own IMM with its transition matrix and model probability as if it stands alone. The likelihood of each set is calculated according to its introduced innovation. The output of sets is calculated as in VIMM. A threshold value to the introduced innovation cause the switching between the sets. The right set will be on while the other will be off which reduced the computation time. As the right set deviates from being the right one, all other sets will start to work while this set will be off. The overall computation time is less than that of including all sets as in IMM and VIMM. The accuracy isn't as IMM or VIMM due to the activation of off sets. Overall the Hierarchal Variable Switching of IMM overcomes the limitation of increasing number of models of IMM with high stability against measuring noise and also reduces the computation time.

5.1 Model-Set Probability and Likelihood

As in VIMM the *marginal* likelihood of a model-set M_j at time k is the sum of the predicted probabilities times the marginal likelihoods, both of all the *models* in M_j

$$L_k^{M_j} \stackrel{\Delta}{=} p[\tilde{z}_k \mid s \in M_j, z^{k-1}] = \sum_{m_i \in M_j} \underbrace{P[\tilde{z}_k \mid s = m_i, z^{k-1}]}_{\text{Model marginal likelihood}} \; \underbrace{P\{s = m_i \mid s \in M_j, z^{k-1}\}}_{\text{Predicted model Probability}} \quad (5)$$

Where $\tilde{z}$ the measurement residual and s is is the mode in effect during the time period over which the test is performed. Note that s has to be assumed constant because a hypothesis cannot be time variant. The *joint* likelihood of the model-set

M_j is defined as $L_{M_j}^k \stackrel{\Delta}{=} p\{\tilde{z} \mid s \in M_j]$. The (posterior) probability that the true mode is in a model-set M_j is defined as

$$M_k^{M_j} \stackrel{\Delta}{=} P\{s \in M_j \mid s \in \mathrm{M}_k, \tilde{z}\} = \sum_{m_j \in M_j} P\{s = m_i \mid s \in \mathrm{M}_k, \tilde{z}\} \quad (6)$$

Which is the sum of the probabilities of all models in M_j, where M_k is the total model-set in effect at time k, which includes M_j as a subset and is problem dependent. The model probability $P\{s = m_i \mid s \in \mathrm{M}_k, \tilde{z}\}$ for each model m_i is typically available from an IMM estimator.

5.2 Initialization of New Models and Filters

The key to the optimal initialization of new models and filters is the concept of state dependency of the system mode set. It simply states that given the current mode (and base state), the set of possible modes at the next time is a subset of the mode space, which is determined by the (Markovian) mode transition law. The optimal assignment of the initial probability to a new model accounts only for the probabilities of those models that may jump (switch) to this new model, and the optimal initial state estimate for a filter based on a (new or old) model is determined only from the estimates (and the probabilities) of the filters based on those models that may jump to the model.

Specifically, the optimal initialization of a filter based on a new (or old) model m_n can be done as follows. When calculating $E[x_k \mid m_k^{(n)}; M^{k-1}, \tilde{z}]$, only the previous estimates $\hat{x}_{k-1|k-1}^{(l)}$ based on models in the set E_n should be used, where E_n is the set of models in M_{k-1} that are allowed to switch to m_n:

$$E_n = \{m_l : m_l \in M_{k-1}, ®_{l,n} \neq 0\} \quad (7)$$

Specifically, the initial estimate for time k cycle of the filter based on model m_n can be obtained as

$$\begin{aligned}\bar{x}_{k-1|k-1} &= E\left[x_{k-1} \mid m_k^{(n)}, M^{k-1}, z^{k-1}\right]\\ &= \sum_{m_l\in E_n} E\left[x_{k-1} \mid m_{k-1}^{(l)}, M^{k-2}, z^{k-1}\right] P\left\{m_{k-1}^{(l)} \mid m_k^{(n)}, M^{k-1}\right\}\\ &= \sum_{m_l\in E_n} \hat{x}^{(l)}_{k-1|k-1} \text{«}^{l|n}_{k-1}\end{aligned} \tag{8}$$

Where

$$\text{«}^{l|n}_{k-1} = P\left\{m_{k-1}^{(l)} \mid m_k^{(n)}, M^{k-1}\right\} = \frac{\text{®}_{l,n}\text{«}^{(l)}_{k-1}}{\sum_{m_i\in E_n}\text{®}_{in}\text{«}^{(l)}_{k-1}} \tag{9}$$

5.3 Adding and Removing Sets

Perform N model-set sequential likelihood ratio test. $\left(H_0 : s\in M_0\ vs\ H_1 : s\in M_1\right)$,....
$\left(H_{N-1} : s\in M_{N-1}\ vs\ H_N\ \ s\in M_N\right)$
These tests are implemented by using thresholds $B = \frac{\sim}{1-}$ and A=-∞ . This step ends when only one of the hypotheses H_1, H_2;……,H_N remains. Specifically:

- Reject set that includes M_i for which $\Lambda_i^k \overset{\Delta}{=} L_M^k / L_{M_i}^k \geq B$
- Continue to the next time cycle to test for the remaining pairs with one more measurement until only one of the hypotheses H_1,H_2,…..,H_N, say H_j , is not rejected.

5.4 Re-initialization of the off sets

The likelihood functions for filter j is as follows:

$$\Lambda_j(k) = N\left[\tilde{z}_j(k);0;S_j(k)\right] = \left|2\text{®}S_j(k)\right|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\tilde{z}_j^T(k)S_j^{-1}(k)\tilde{z}_j(k)\right] \tag{10}$$

Where $\tilde{z}_j(k) = z(k) - \hat{z}(k\mid k-1)$ is the innovation for filter j and $S_j(k)$ is the covariance matrix.
Switching between active sets as

$$\Lambda_j(k) - \Lambda_j(k-1) > \sim \tag{11}$$

The model set of M_j is change to off state and initiate the other two sets.
One Cycle of Switching sets of IMM

Start one cycle for each set of IMM
$\Lambda(k)_i = \max\left(\Lambda(k)_{i,j}\right)$ *where* $j = 1,2,....$ *and* $i = 1,2,3$
For i=1:3
If $\Lambda(k)_i >$ *Set I change to off state*
Else active
End.
If $\Lambda(k)_i - \Lambda(k-1)_i > \sim$
Change the state to off and activate the other sets
Else set is active.
End.
End.
Compare all $\Lambda_i(k)$ *the smallest* $\Lambda_i(k)$ *takes its highest mode I probability to be the set probability*
Set probability of the other two sets=1- probability of minimum $\Lambda_i(k)$
If the remaining sets are active
Distribute this value between them according to their maximum model probability.
Else
Set the probability of the off sets to zero
End.
For active set which has the maximum probability we take its output
$\hat{x}_{k|k} \overset{\Delta}{=} E[x_k \mid z^k] = \sum_i \hat{x}^{(i)}_{k|k}\text{«}^{(i)}_k$

6 The Results of IMM of The First Tracking Problem

We tested our model using Matlab 2010Ra. intel core 2 duo., under windows vista environment. The following results is for 10 modes one for linear motion and the others are at different ω [0.2,0.4,0.6,0.8,0.9,1.1,1.4,1.6,1.8], Q=diagonal(0.5^2) R=diagonal(100). The results are the average of 200 run. The target trajectory described in

Table1: target trajectory

scenario	First trajectory	Second trajectory
0<k<90	ω=1.4	ω=1.4
90<k<150	ω =0.2	ω=1.6
150<k<200	linear	linear
200<k<300	ω=1.9	ω=1

The transition matrix has a diagonal of 0.82 while all models start with equal model probability

Table 2: The results of the first target trajectory

IMM						
RMSEx	RMSEy	RMSEvx	RMXEvy	Value	ΔT	Exe.time
0.6174 2.9391	0.6039 2.711	0.3242 1.416	0.2970 1.528	Mean Max	2.3	0.01443
0.6275 2.041	0.691 2.9366	0.4422 1.504	0.3774 1.6223	Mean Max	2.5	0.01443
0.7423 3.3473	0.6081 2.4812	0.2973 0.9727	0.3163 1.4287	Mean Max	2.9	0.014402
1.0326 4.9509	1.0102 4.509	0.3486 1.2098	0.4908 1.2983	Mean Max	3.3	0.014498

Table3: the results of the second target trajectory

IMM						
RMSE x	RMSEy	RMSEv x	RMXEvy	Value	ΔT	Exe time
0.6174 2.9391	0.6039 2.711	0.3242 1.416	0.2970 1.528	Mean Max	2.3	0.01443
0.6275 2.041	0.691 2.9366	0.4422 1.504	0.3774 1.6223	Mean Max	2.5	0.01443
0.7423 3.3473	0.6081 2.4812	0.2973 0.9727	0.3163 1.4287	Mean Max	2.9	0.01440
1.0326 4.9509	1.0102 4.509	0.3486 1.2098	0.4908 1.2983	Mean Max	3.3	0.01449 8

7. Hierarchal Switching of IMM

In the above examples we change models with ω near to each other while the change of linear model can take place at any k . the sudden jump of ω to far values cause system to diverge. Also in this model we choose ω that can introduce good results with each other not all values of them cause system converge or good tracking. Also Not wide range of the time step variation is available.

Two overcome these limitation we introduce a variable set of models. We take the advantage of good tracking of small sets of IMM and divide our structure into three sets

The first set includes $\omega_{1=}$[0.2 0.4 0.6] .The second set includes $\omega_{2=}$[0.8 0.9 1.1]

The third set includes linear model with $\omega_{3=}$[1.4 1.6 1.8] .The variable structure of IMM algorithm is shown in figure 2

The transition matrix of the first two set and model probability as follow

$$\Pr = \begin{bmatrix} 0.98 & 0.01 & 0.01 \\ 0.01 & 0.98 & 0.01 \\ 0.01 & 0.01 & 0.98 \end{bmatrix} \qquad « = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.3 \end{bmatrix}$$

The transition matrix of the third set has 0.97 diagonal and equal model probability

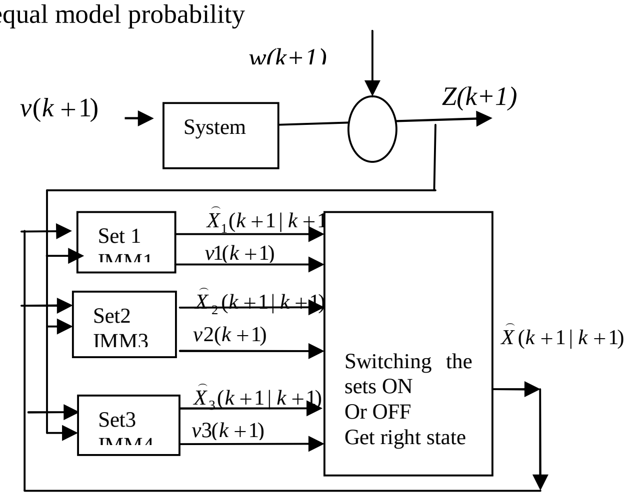


Fig.2 The Hierarchal structure of IMM for the three sets

8. The results of Hierarchal Structure of IMM for The First Tracking problem

We tested our model using Matlab Ra2010 on intel core 2 duo., under windows vista environment. The following results is for 10 modes one for linear motion and the others are at different ω [0.2,0.4,0.6,0.8,0.9,1.1,1.4,1.6,1.8], Q=diagonal(0.5^2) R=diagonal(100). The results are listed in table 4. The system characterized by unknown changeable structure as in the previous first example.

Table:4 The results of the first tracking problem

IMM						
RMSEx	RMSEy	RMSEvx	RMXEvy	value	ΔT	Exe. time
2.2637 18.505 3	1.1498 5.039	0.6093 3.6603	0.4307 2.5497	Mean Max	2.3	0.0066
2.6597 7.2371	2.4558 7.3775	1.0951 4.0049	0.9815 4.078	Mean Max	2.5	0.0066
0.6341 5.1443	0.6489 5.034	0.5999 1.8247	0.6096 1.5653	Mean Max	2.9	0.0067
2.5489 7.929	2.7573 7.603	0.8608 3.2303	0.8229 3.0638	Mean Max	3.3	0.0065
3.4279 14.176 2	3.2624 13.3039	1.1856 3.9484	1.2166 3.9018	Mean Max	4	0.0067
2.4515 8.2294	2.6506 9.7918	0.8125 3.1517	0.9002 5.5053	Mean Max	5	0.0065

If we change ω to ω_1=[7/dt 8/dt 9/dt] ω_2=[4/dt 5/dt 6/dt] and ω_3 =[1/dt 2/dt 3/dt] with linear model. These values include most of the turning angles that have good separation between modes of operation. The other values included in these modes. The RMSE for the same parameters of the IMM are listed in table 5 while the results of RMSE if we choose R=I 25, are listed in table 6

Table 5: The result of the first tracking problem

HSIMM						
RMSEx	RMSEy	RMSEvx	RMSEvy	value	ΔT	Exe time
17.6978 1.8674	18.2421 2.1699	2.2899 0.5395	2.6145 0.5922	Max Mean	2.3	0.0071
7.6344 0.872	12.4231 1.1034	3.0548 0.6205	1.9364 0.4502	Max Mean	2.5	0.0068
10.5143 0.9774	15.9724 1.2805	2.2557 0.4727	1.8583 0.4726	Max Mean	2.9	0.007
15.8625 1.1525	11.5765 1.0396	2.3474 0.4897	2.4372 0.4882	Max Mean	3.3	0.007
13.3789 1.0023	7.4274 0.9793	2.8286 0.9544	2.2394 0.8300	Max Mean	4	0.0067
11.5338 1.6121	14.1307 1.3044	1.814 0.6895	3.4595 0.8429	Max Mean	5	0.0067

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Online): 1694-0814 | ISSN (Print): 1694-0784
www.IJCSI.org

Table 6: The result of the Other sets for the same target trajectory

HSIMM						
RMSE x	RMSEy	RMSEv x	RMSEvy	Value	ΔT	Exe.time
4.0811 0 .4532	7.2529 0.7541	1.4272 0.4739	1.4737 0 .4404	Max Mean	2.3	0.0067
4.893 0 .7274	6.7546 0.5493	1.8966 0.4879	1.5485 0.4558	Max Mean	2.5	0.0068
5.1443 0.6341	5.034 0.6489	1.8247 0.5999	1.5653 0.6096	Max Mean	2.9	0.0067
7.3538 0.6939	4.0377 0.7537	2.7559 0.7178	2.2064 0.6819	Max Mean	3.3	0.0068
4.773 0.7348	6.9591 1.0398	3.2547 0.6877	3.3066 0.6414	Max Mean	4	0.0061
9.954 1.5052	5.7863 1.0657	2.9827 1.0743	3.3516 1.0657	Max Mean	5	0.0068

9. Results of The second Tracking Problem

The target measurement model

$$x(k+1) = Fx(k) + G[a(k) + w(k)] \qquad (12)$$

$$z(k+1) = Hx(k+1) + v(K+1); k = 0,1,2,.... \qquad (13)$$

Where $x = (x, v_x, y, v_y)'$ denotes the targets state $a = (a_x, a_y)'$ is the acceleration, $w \sim N[0,Q]$ is the acceleration process noise, $z = (z_x, z_y)$ is the measurement, $v \sim N[0,R]$ is the random measurement error and

$$F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad G = \begin{bmatrix} T^2/2 & 0 \\ T & 0 \\ 0 & T^2/2 \\ 0 & T \end{bmatrix}$$

The unknown true acceleration is assumed piecewise constant, varying over a given continuous planar region A^c. In the MM framework, we consider a generic finite set (grid) of acceleration values: $A^{(r)} \stackrel{\Delta}{=} \{a^{(i)} \in A^c : i = 1,2,...r\}$ which defines the total model set. We approximate the evolution of the true acceleration over the quantized set $A^{(r)}$ via a Markov chain model, that is, $a_k \in A^{(r)}$ with given $P\{a_0 = a^{(i)}\} = P_i$ and $P\{a_k = a^{(i)} \mid a_{k-1} = a^{(i)}\} = ®_{ij}$ for $i, j = 1,2,....r$.

Consider the following target-tracking example, adopted from [8],[18],[17],[19]. A target moves in the horizontal plane that may have piecewise-constant acceleration with a maximum value of 4g (40m/s2) in any direction. Assume that the following set of 13 time-invariant models, characterized by the expected acceleration vector *a*, is used:

$$\begin{Bmatrix} a_1 = 20[0,0] & a_2 = 20[1,0] & a_3 = 20[0,1] & a_4 = 20[-1,1] \\ a_5 = 20[0,-1] & a_6 = 20[1,1] & a_7 = 20[-1,1] & a_8 = 20[-1,-1] \\ a_9 = 20[1,-1] & a_{10} = 20[2,0] & a_{11} = 20[0,2] & a_{12} = 20[-2,0] \\ a_{13} = 20[0,-2] & & & \end{Bmatrix}$$

The transition relations among models are easily understood in terms of the directed graph (i.e., digraph) representation of an MM, introduced in [13], [15], [14], [16]. The topology of model set $A^{(13)}$ is depicted in Figure 3. Each model is viewed as a point in the mode (acceleration) space. An arrow from one model to another indicates a legitimate model switch (self-loops are omitted) with non-zero probability.

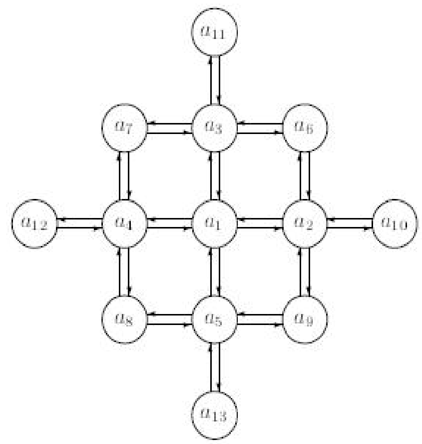


Fig. 3 Digraph representation of 13-model set

8.1 Design of VIMM and HSIMM

For both VIMM and HSIMM we divide the sets into three sets. $M_1 = \{a_2, a_3, a_{11}, a_6, a_{10}\}$, $M_2 = \{a_9, a_5, a_{13}, a_8\}$ and $M_3 = \{a_1, a_4, a_7, a_{12}\}$

The probability transition matrix of IMM13 and VIMM has diagonal of 0,8

The model probability is

$$« = [0.08 \quad 0.08 \quad 0.08 \quad 0.076 \quad 0.076 \quad 0.076 \quad 0.076 \quad 0.076 \quad 0.076 \quad 0.076 \quad 0.076 \quad 0.076 \quad 0.076]$$

For HVSIMM we have three sets each set has its own transition matrix and model probability as follow

Set1

$$® = \begin{bmatrix} 0.8 & 0.04 & 0.04 & 0.04 & 0.04 \\ 0.04 & 0.8 & 0.04 & 0.04 & 0.04 \\ 0.04 & 0.04 & 0.8 & 0.04 & 0.04 \\ 0.04 & 0.04 & 0.04 & 0.8 & 0.04 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.8 \end{bmatrix} \qquad « = [0.2 \quad 0.2 \quad 0.2 \quad 0.2 \quad 0.2]$$

Set2 and Set3

$$® = \begin{bmatrix} 0.97 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.97 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.97 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.97 \end{bmatrix} \qquad « = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}$$

The sets probability is initialized by $«_{Set} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \end{bmatrix}$. It is changed according to the maximum model probability of the model of sets. The active set takes higher value while the other is distributed according to their higher model probability.

8.2 Performance Evaluation

Test scenarios

The performances of the IMM, VIMM and HSIMM tracking algorithms were investigated first over a large number of deterministic maneuver scenarios with fixed acceleration sequences. Deterministic scenarios serve to evaluate algorithms' peak errors, steady-state errors and response times. We present two of them, referred to as DS1 and DS2, in the sequel. Their acceleration values are given in Table 7

Deterministic Scenarios' Parameters

The other parameters for both scenarios are *T = 1s; Q = O; R* =1250*I; x0* = [8000; 25; 8000; 200]. Note that while the acceleration values in DS1 are relatively close to the fixed grid points of IMM13, in DS2 they are deliberately chosen far apart from the grid points. As such, for the fixed structure estimator IMM13 the scenario DS2 is more difficult than DS1.

Table 7: The Targets Dynamics

Scenario	DS1		DS2	
K	$a_x(k)$	$a_y(k)$	$a_x(k)$	$a_y(k)$
1-29	0	0	0	0
30-45	8	22	8	22
46-55	2	37	12	27
56-80	0	0	0	0
81-98	25	2	15	2
99-119	-2	19	-2	9
120-139	0	-1	0	-1
140-149	38	-1	28	-1
150-160	0	0	0	0

Performance measure:

The *accuracy* of the algorithms was measured in terms of position and velocity root-mean-square errors (RMSE):

$$RMS\_error(k) = \sqrt{\frac{1}{M}\sum_{j=1}^{M}(x(k)-x^{j}(k))(x(k)-x^{j}(k)^{T}}$$

$$\mathrm{RMSE}_k^{x} = \left(\frac{1}{M}\sum_{i=1}^{M}\left(\left[x_k^i - \hat{x}_{k|k}^i\right]*\left[x_k^i - \hat{x}_{k|k}^i\right]'\right)\right)^{\frac{1}{2}}$$

$$\mathrm{RMSE}_k^{y} = \left(\frac{1}{M}\sum_{i=1}^{M}\left(\left[y_k^i - \hat{y}_{k|k}^i\right]*\left[y_k^i - \hat{y}_{k|k}^i\right]'\right)\right)^{\frac{1}{2}}$$

$$\mathrm{RMSE}_k^{vx} = \left(\frac{1}{M}\sum_{i=1}^{M}\left(\left[\dot{x}_k^i - \hat{\dot{x}}_{k|k}^i\right]^2 + \left[\dot{x}_k^i - \hat{\dot{x}}_{k|k}^i\right]^2\right)\right)^{\frac{1}{2}}$$

$$\mathrm{RMSE}_k^{vy} = \left(\frac{1}{M}\sum_{i=1}^{M}\left(\left[\dot{y}_k^i - \hat{\dot{y}}_{k|k}^i\right]^2 + \left[\dot{y}_k^i - \hat{\dot{y}}_{k|k}^i\right]^2\right)\right)^{\frac{1}{2}}$$

,

Where

(x_k^i, y_k^i) true position, $(\dot{x}_k^i, \dot{y}_k^i)$ true velocity, $(\hat{x}_k^i, \hat{y}_k^i)$ estimated position $(\hat{\dot{x}}_k^i, \hat{\dot{y}}_k^i)$ and the estimated velocity.

The performances of the three MM tracking algorithms are investigated first over a large number of deterministic maneuver scenarios with fixed acceleration sequences. Deterministic scenarios serve to evaluate algorithms' peak errors, steady-state errors and response times. We present two of them, referred to as DS1 and DS2, in the sequel. Their acceleration values are given in Table 7. The other parameters for both scenarios are *T=1 sec, Q=0,R=1250I*

Note that while the acceleration values in DS1 are relatively close to the fixed grid points of IMM13, in DS2 they are deliberately chosen far apart from the grid points. As such, for the fixed structure estimator IMM13 the scenario DS2 is more difficult than DS1.

Table 8: The results of target dynamics in table 7

	IMM13		VIMM13		HSIMM13	
	DS1	DS2	DS1	DS2	DS1	DS2
RMSx	0.0182	0.0118	0.0.094	0.0892	5.6850	4.2039
RMSy	0.0291	0.0276	0.312	0.3131	3.6809	5.6190
RMSvx	103.3092	74.8224	105.6351	75.1330	23.8764	16.0580
RMXvy	97.7809	93.3336	103.9023	99.9368	15.2076	10.5261

Table 9: The execution Time of target dynamics in table 7

	IMM13		VIMM13		HSIMM	
	DS1	DS2	DS1	DS2	DS1	DS2
Time	0.0.0075	0.0072	0.0085	0.0073	0.0045	0.0019

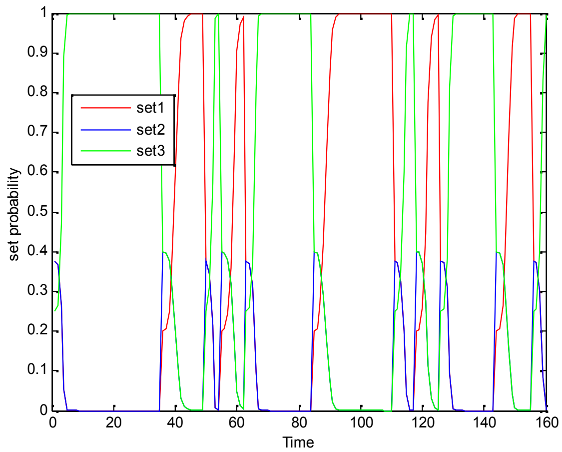


Figure 4 Activation between the sets as their set probability change for DS1 of HSIMM13

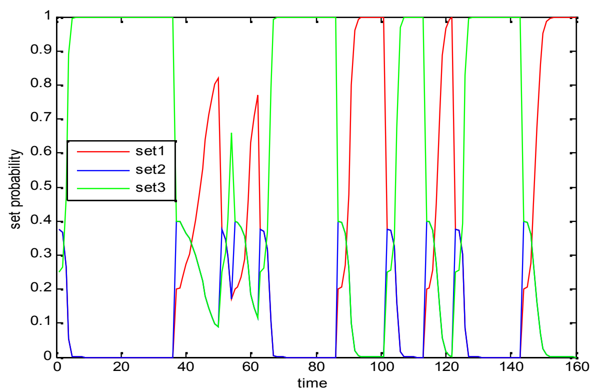


Figure 5 Activation between the sets as their set probability change for DS2 of HSIMM13

The proposed HSIMM introduced less computational time and also minimum RMS error as shown in table 5, 6,8 and 9. But it needs to be re-initialized to overcome error accumulation.

The activation between included sets is achieved by the introduced threshold value of innovation. The switching algorithm as shown in figure 4 and 5 for DS1 and DS2 is effective.

10. Conclusion

As the number of the IMM increase the algorithm stability decrease. Or in other word As the parameters change the system doesn't converge to different values. As we show in IMM with 10 models. The change of ω and ΔT . Also the choosing of their values may cause system to converge to the wrong model of ω. Also as the time step change the (increase) the system doesn't converge as ω and time step change together. Not all the models change during operating from model to another allowed.

In structure set of IMM we first choose the near values of ω in the same model to avoid converging to the wrong set. The small numbers of models increase the system stability. It doesn't diverge at the changing of time step or different values of ω. Also changing from any model to another is allowed. The advantages of the structure set of IMM are introducing varieties of motion models and also varieties of time step values. Introduce variety of Model change during operation. Introduce large number of modes of operation so we can avoid using the nonlinear models with their calibration hardness. It also introduce less computation time than introduced by the large number of IMM since we only activate the right set.

The error introduced by the structure HSIMM is due to the initialization at the beginning before converging to the right set. This error can be reduced by the refinement process if we take the saved values of the right set but it doesn't suite the real time process. The HSIMM also introduce relatively similar errors at velocity components compared to other algorithms. The computational time is minimum than introduced by IMM and VIMM. HSIMM introduces less error as the noise increase and there is no need for re adjustment to the Covariance as the noise increase so it is more robust against noise and introduces minimum computational time.

References

[1] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan. Interacting multiple model methods in target tracking: A survey. *IEEE Transactions on Aerospace and Electronic systems*, 34(1):103{123, January 1998.

[2] X.R. Li and Y. Bar-Shalom. Multiple-model estimation with variable structure. *IEEE Transactions on Automatic Control*, 41(4):478{493, April 1996.

[3] Hongshe Dang, Chongzhao Han Xi'an, Shaan and Dominique GRUYER "Combining of IMM filtering and DS data association for multi-target tracking" www.fusion2004.foi.se/papers/IF04-0876.pdf

[4] Ludmila Mihaylova and Emil Semerdjiev, "An Interacting Multiple Model Algorithm for Stochastic Systems Control" www.citeulike.org/user/sourada/article/347493 - 16k

[5] Leigh A. Johnston and Vikram Krishnamurthy, "An Improvement to the Interacting Multiple Model (IMM) Algorithm, IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 49, NO. 12, DECEMBER 2001 PP 2909-2923.

[6] Joseph J. LaViola Jr. "A Comparison of Unscented and Extended Kalman Filtering for Estimating Quaternion Motion". www.cs.brown.edu/people/jjl/pubs/laviola_acc2003.pdf -

[7] Songhwai Oh and Shankar Sastry, " An Efficient Algorithm for Tracking Multiple Maneuvering Targets". *Proc. of the IEEE International Conference on Decision and Control*, Seville, Spain, Dec. 2005.

[8] Ludmila Mihaylova 1 Emil Semerdjiev1 X. Rong Li ," Detection and Localization of Faults in System Dynamics

by IMM Estimator.
http://citeseer.ist.psu.edu/cache/papers/cs/22073/

[9] Youmin Zhang and X. Rong Li., “Detection and Diagnosis of sensor and actuator failures using IMM estimator”, IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS .vol.34, No. 4 October 1998. PP 1293-1313.

[10] Donka Angelova 1 , Emil Semerdjiev 1 , Ludmila Mihaylova 1 X. Rong Li , “An IMMPDAF Solution to Benchmark Problem for Tracking in Clutter and Standoff Jammer”, http://citeseer.ifi.unizh.ch/cache/papers/cs/22073/

[11] Simon J. Julier Jeffrey K. Uhlmann,” A New Extension of the Kalman Filter to Nonlinear Systems”. media/pdf/Julier1997_SPIE_KF.pdf/**kalman**www.cs.unc.edu/~welch/

[12] Seham Meawed Aly , Raafat El Fouly, Hoda Barak, “Extended Kalman filtering and interactive multiple model for tracking maneuvering targets in sensor networks,” Intelligent solutions in embedded systems,2009 seventh workshop on, 2009 , page: 149-156

[13] X. R. Li, V. P. Jilkov, and J.-F. Ru, “Multiple-model estimation with variable structure—part VI: expected-mode augmentation,” IEEE Transactions on Aerospace and Electronic Systems, 41(3): 853–867, July 2005

[14] X. R. Li, Z.-L. Zhao, and X. B. Li, “General model-set design methods for multiple-model approach,” IEEE Transactions on Automatic Control, 50(9):1260–1276, September 2005

[15] X. R. Li, "Engineer's guide to variable-structure multiple-model estimation for tracking," Y. Bar-Shalom and W. D. Blair, eds., Multitarget-Multisensor Tracking: Applications and Advances, Vol. 3, Chapter 10, Artech House, Boston, 2000, pp. 499-567

[16] Xiao-Rong-Li, Yaakov Bar-Shalom, “Multiple- Model Estimation with Variable Structure “ ,IEEE Transactions on Automatic Control, Vol. 41, No, 4 April 1996. pp.478-493.

[17] X. R. Li, "Engineer's guide to variable-structure multiple-model estimation for tracking," Y. Bar-Shalom and W. D. Blair, eds., Multitarget-Multisensor Tracking: Applications and Advances, Vol. 3, Chapter 10, Artech House, Boston, 2000, pp. 499-567

[18] X. R. Li, and Y. M. Zhang, “Multiple-Model Estimation With Variable Structure —Part V: Likely-Model Set Algorithm,” *IEEE Trans. Aerospace and Electronic Systems*, Vol. AES-36, Apr. 2000.

[19] X. R. Li and V. P. Jilkov, "Expected-mode augmentation for multiple-model estimation," Proc. 2001 International Conf. Information Fusion, Montreal, QC, Canada, August 2001, pp. WeB1-3-WeB1-10

Framework of Software Quality Management Using Object Oriented Software Agent

Anand Kr Pandey [1], Dr. A.S. Saxena [2] and Rashmi Pandey [3]

[1] Research Scholar, Mewar University, Chittorgarh, Rajasthan, India

[2] Prof. Dept of CSE, M.P.C.T., Gwalior, Madhya Pradesh, India

[3] School of Computer Application, ITM University, Gwalior, Madhya Pradesh, India

Abstract

Development of software is a scientific and economic problem, particularly the design of complex systems whichrequire evolving methods and approaches. Agent technology is currently one of the most active and vibrant areas of IT research and development. Object-oriented Software Engineering (OOSE) has become an active area of research in recent years. In this paper, we review the framework of software quality management using object-oriented methodology concepts for software agents.The software specification acts as a bridge between customers, architects, software developers and testers. Using object-oriented concept of software agent and its standard it may offer benefits even if the system is implemented without an object-based language or framework . We propose and discuss a software agent framework, specifically to support software quality management. Although still in its initial phases, research indicates some promise in enabling software developers to meet market expectations and produce projects timeously, within budget and to users' satisfaction. However, the software quality management environment has also changed and is continuously evolving. Currently software projects are developed and deployed in distributed, pervasive and collaborative environments and its quality should be managed by applying its best standard. From the point of view of software engineering this framework and its standards are applying for developing the software projects.We discuss the standard and benefits that can be gained by using object-oriented concepts, and where the concepts require further development.

Keywords: Agent, Framework, Design, Software Quality Management, Autonomy

1. Introduction

The software quality management and its standard acts as a bridge between customers, architects, software developers and testers. The terms "objects" and "oriented" in something like the modern sense of object-oriented programming seem to make their first appearance at MIT in the late 1950s and early 1960s. This process is the part of software analysis and software analysis often occurs after requirements gathering, and before architectural design. Software analysis is the activity of understanding the requirements, so that decisions can be made about how to architect it. Software quality management and assurance is gaining increasing attention throughout the software lifecycle. While software testing has become mainstream in professional software development, sophisticated quality management in the early phases of software development is still not evolved that far. Existing methods for software analysis and quality management include data flow diagrams, state transition diagrams, and object-oriented modeling. These methods have all been applied successfully, and all have their own restrictions and limitations.

Now a days software agents are probably the fastest growing area of Information Technology (IT). Software agents, by definition, are active, independent components. Most agents are designed to act as or for the user to help execute some task or operation [1]. In this paper, we discuss the review of designing framework for software quality management and its standards. The development of Software Project Management (SPM) as an independent application area and field of study. SPM includes, amongst other things, the management of all issues involved in the development of a software project, namely scope and objective identification, planning, evaluation, project development approaches, software effort and cost estimation, activity planning, monitoring and control, risk management, resource allocation, as well as managing contracts, teams of people and quality [2]. In this work, we focus on our *software specification framework* and demonstrate its strong integration of constructive and analytic software engineering methodology. For better and efficient software quality management and programming service, software can be defined by differentiating basic environment for launching, basic service and extended service. Software agent technology offers a promising solution to addressing software quality management

Problems in a distributed environment. According to this technology, software agents are used to support the development of SPM systems in which data, control, expertise, or resources are distributed. Software agent technology provides a natural metaphor for support in a distributed team environment, where software agents can help the project manager and team members to monitor and coordinate tasks, to apply quality control measures, to validate and verify, as well as to ensure proper change control [1]. We specifically concern ourselves with the question of how software agents can be used to improve quality management in a distributed environment. Here we discussed about software quality and constructing quality using specification method. The next section contains a background study and a discussion on software quality management in the context of the software project management framework.

2. What is Software Quality Management

The first question arise that what is software quality? The first is that quality means "meeting requirements." of customer, because customer defines quality as to whether the product or service does what the customer needs. The professional views about software quality reflects conformance to requirement and fitness for use. Software quality is something that everyone wants. Manager know that they want high quality, software developers know they want to produce a quality product, and user's insist that software work consistently and be reliable [2]. We can say that a quality product is one which meets its requirements and satisfies the user. The aim of Software Quality Management (SQM) is to manage the quality of software and of its development process. To mantain the quality we have to ensure that the required level of quality is achieved in a software product and for manage all the activities and processes we have to encourage a company-wide "Quality Culture" where quality is viewed as everyone's responsibility.

The Software Management Body of Knowledge defines software quality management as the processes required to ensure that the software will satisfy the needs for which it was undertaken. It includes all activities of the overall management function that determine the qualitative policy, objectives and responsibility, and implements these by means of quality planning, quality assurance, quality control and quality improvement, within the quality system [3]. Quality management not only includes the concepts, tools and methods of quality assurance, but also validation and verification. The SQM should have following features.

(i) **Quality planning:** determining which quality standards are relevant to this specific project and deciding how these standards will be met.

(ii) **Quality assurance:** involves evaluating overall performance regularly to ensure conformance to the set standards. Quality audits or reviews can support this function.

(iii) **Quality control:** monitoring the activities and end results of the project to ensure compliance with the standards utilizing various available tools and techniques.

However, software quality management should not be considered as a separate developmental phase but should be an inextricable part of all phases and all processes during software project management.

3. Role of Object Oriented Software Agent in SQM

Software agents, by definition, are active, independent components. Most agents are designed to act as or for the user to help execute some task or operation. The object-oriented software agent paradigm has some flaws related to design and implementation of multi-agent systems, we also believe that it is still the most practical programming language to implement the agent technology. Each developer and researcher in the agents field adopts their own definition of an Agent.Object oriented software agent technology offers a promising solution to addressing software quality management problems in a distributed environment. According to this technology, software agents are used to support the development of SPM systems in which data, control, expertise, or resources are distributed [4]. Software agent technology provides a natural metaphor for support in a distributed team environment, where software agents can help the project manager and team members to monitor and coordinate tasks, to apply quality control measures, to validate and verify, as well as to ensure proper change control. Software agents can be grouped, according to specific characteristics, into different software agent classes. Object-oriented concept developed as the dominant programming methodology in the early and mid 1990s when programming languages supporting the techniques became widely available.

In this paper the use of object oriented software agents is investigated as a potential tool for improving the quality management of SPM processes. Whether or not an agent has a user interface, depends on whether it collaborates with humans, other agents or hosts. Software agent technology is being explored as a promising way to support and implement complex distributed systems. In this section, we briefly consider how agent technology is currently being deployed in SQM by considering some application examples and standards [3]. This research approach is based on the fact that object-oriented software engineering has proved to be extremely powerful for building complex systems, which promote modularity, maintainability, and reusability.

3.1 Features Of Software Agents

There is a minimum set of common features that typify a software agent.Various features have proposed different definitions of agents, these commonly include concepts such as [5]:

- *Autonomy:* Agents should be able to perform the majority of their problem solving tasks without the direct intervention of humans or other agents, and they should have a degree of control over their own actions and their own internal state.
- *Social ability:* Agents should be able to interact with other software agents and humans in order to complete their own problem solving and to help others with their activities where appropriate.
- *Perceptive:* A software agent is perceptive; it is able to perceive and respond to changes in its environment.
- *Reactivity*: Agents perceive the context in which they operate and react to it appropriately.
- *Monitoring:* A software agent can be used to monitor a computer system and software system to make sure that it is performing and functioning correctly.
- *Interaction:* An object oriented agent can communicates with the environment and other agents by means of sensors and effectors.

The autonomy characteristic of a software agent distinguishes it from general software programs. Autonomy in agents implies that the software agent has the ability to perform its tasks without direct control, or at least with minimum supervision, in which case it will be a semi-autonomous software agent.

4 SQM Framework

The software agents are used to control and monitor several activities execution at various sites in an open source platform supporting distributed software engineering processes.

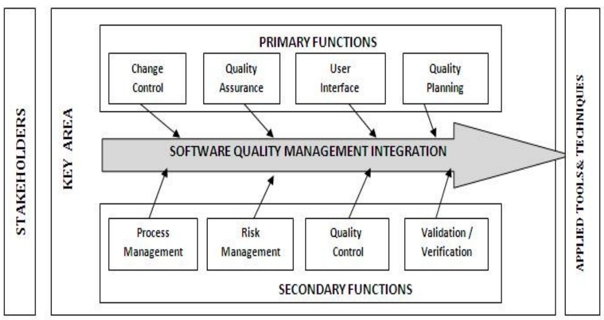


Fig 1: Framework of Software Quality Management

The design of the framework of SQM using object oriented model for overall system, based on components (specialised agents), which simplifies the design and programming of agents. The proposed approach for designing framework encourages the separate handling of object oriented modeling concerns, and provides a disciplined scheme for their composition [6]. An object oriented software agent can collaborate with other agents in order to use the counterpart services via some communication language, it provide the abstractions that explicitly capture the structure. This framework technology support the ability to build applications by selecting and assembling objects from libraries.

Above figure 1 shows the framework for SQM. The following specialised factors of working agents are used in our discussion making for the software quality management framework that we present in this section. Here multiple methods are used for variable parametres. The following responsible factors are as:

(i) *Change Control:* Change control is a systematic approach to managing all changes made to a product or system for designing framework. It is a formal process used to ensure that changes to a product or system are introduced in a controlled and coordinated manner.

(ii)

(iii) *Quality Assurance:* It is a process-centered approach to ensuring that a company or organization is providing the best possible products or services. consists of a means of monitoring the software engineering processes and methods used to ensure quality.

(iv) *User Interface:* The user interface is one of the most important parts of any program because it determines how easily you can make the program do what you want. A powerful program with a poorly designed user interface has little value. The user interface, in the SQM is the space where interaction between user and model occurs.

(v) *Quality planning* consists of determining which quality standards are relevant to this specific project and deciding how these standards will be met.

(vi) *Process management:* This is the ensemble of activities of planning and monitoring the performance of a process during the activities of applied tools and techniques on SQM.

(vii) *Risk Management:* Software Quality Management (SQM) is concerned with all the processes, methods, and practices that affect quality during designing framework when producing, supporting, and operating software. Essentially, SQM is risk management, identifying and addressing all the factors that can negatively impact all those processes, methods, and practices. While often the most visible form of managing risk, testing developed code is only one of many aspects that SQM considers. The strategies to manage risk typically include transferring the risk to another party or avoiding the risk or reducing the negative effect or probability of the risk, or even accepting some or all of the potential or actual consequences of a particular risk.

(viii) *Quality Control:* Quality control (QC) is a procedure or set of rules and functions intended to ensure that a designed framework or performed service adheres to a defined set of quality criteria or meets the requirements of the client or customer.

(ix) *Validation & verifucation:* In software quality management, software testing, and software engineering, verification and validation (V&V) is the process of checking that a software system meets specifications and that it fulfills its intended purpose. It may also be referred to as software quality control. Verification and Validation are independent procedures that are used together for checking that a product, service, or system meets requirements and specifications and that it fulfills its intended purpose.

5 Conclusion

In this paper we investigated an approach for designing framework of software quality management using object oriented software agent technology to address the challenges proposed in the software project management arena. The concept of a agent is most useful as a tool to analyze systems, not as a prescription. We focused on one of the key elements of Software quality Management, namely software quality management, and designed a generic agent framework to address all the tasks of this key element. This framework forms a basis for other key elements, and could be adapted into individual frameworks and then coordinated by an overall multi-agent system to achieve the objectives of Software Project Management. Our framework follows an approach of agent teams being composed of specialised software agents.

References

[1] A. Graesser and S. Franklin, "Is it an agent, or just a Program?" A Taxonomy for Autonomous Agents". In Proceedings of the Third International Workshop On Agents Theories, Architecture, and Languages, Springer-Verlag, 1996.

[2] Chen, F, Nunamaker, J. F., Romano, N. C. & Briggs, R. O. (2003). "A collaborative project management architecture". Proceedings of the 36th Hawaii International Conference on System Sciences. Big Island, Hawaii.

[3] D'Inverno, M. & Luck, M. (2001). "Understanding Agent Systems". Berlin: Springer-Verlag.

[4] Leonard Foner, "What's an Agent Anyway?", Agents Memo 93-01, Agents Group MIT Media Lab, 1993.

[5] Michael Wooldridge and Nick Jennings, Dept. of Computer science, Manchester Metropolitan University, IEEE Review, January 1996.

[6] P. Maes, Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back. MIT Press, 1991, ISBN 0-262-63135-0.

First Author Anand K. Pandey, MCA, M.Tech., is a research scholar for Ph.D. program from Mewad University, Rajasthan. He has been published 05 papers in reffered journals. He has 10 years of teaching experiance and also written a book titled "Simulation and Modelling" for BE, MCA and M.Tech. students.

Second Author Dr. A. S. Saxena, prof in Dept of C.S.E. at M.P.C.T. Gwalior has 40 years of teaching experiance and also published 10 research papers in reffered journals. He has been guided 06 Ph.D. students.

Third Author Rashmi Pandey, MCA, is currently working as a Asst. Prof. in computer application dept at Institute of Applied Science and Computer Application, Gwalior. have been published 6 papers in reffered journals. She has 7 years of teaching experiance.

The Sentiment Trend Analysis of Twitter Based on Set Pair Contact Degree

Chunying Zhang[1], Jing Wang [2]

**1 College of Science, Hebei United University
Hebei, Tangshan, China**

**[2] College of Science, Hebei United University
Hebei, Tangshan, China**

Abstract

Sentiment trend of twitter users have a great influence on their friends and the crowd listened. This paper directs at the user sentiment state of twitter, the unique medium, and applies set pair analysis method for trend analysis. First, we begin with set pair contact degree, then based on set pair affective computing model to make comparison with the size relationship of same degree, difference degree, opposition degree of the emotion, to build the user sentiment trend analysis model; Secondly, we analyze the influence for the user's own sentiment trend when the value changed of difference coefficient i ; thirdly, after analyze to obtain one user's sentiment orientation threshold as prerequisite for user behavior prediction. Finally, setting an example to calculate the sentiment trend of one twitter, then to get the conclusion is that the analysis of user emotion from a three-dimensional angle is more realistic than the single angle.

Keywords: *Twitter information, sentiment trend analysis, twitter form, contact degree, sentiment orientation threshold.*

1. Introduction

Accompanied by the Web2.0 era, the network became an important carrier for the users to reflect the social public opinion; more and more people choose twitter and Tencent QQ to express their views and opinions for some certain events. The twitter is of wide dissemination range has become one timely communication tool and be concerned with the characteristic of its brief and pithy. Twitter has opened multiple API enables a large number of users to instantly update their personal information by phone, network. Nowadays, the form of twitter has become more diverse, for example, the ordinary text, the new network word, emoticon, picture information, link and so on. Based on the twitter platform to study has become the first choice for most scholars, because of twitter information is of the convenience, timeliness, and multi-style. Therefore, as the research field rise in recent years [1-3], affective computing[4] also chose this platform to study. For example, the work [5] is based on affective computing to study the detection method for twitter incident, to find a method to detect unexpected events; work [6] makes sentiment analysis on hot events for twitter to achieve sentiment trend analysis for hot events. The Chinese twitter topic extraction algorithm of the user's interest topic for the massive twitter information has achieved the prototype system-BTopicMiner of Chinese twitter topic mining [7]. Thereinto, the foreign scholar Kim [8] divided views expression into four semantic components: theme, expresser, representation and emotion. The process of sentiment analysis automatically determines the relationship among the natural language text which reflected the specific view and evaluation for the things in the text. Affective computing is the basis of text orientation analysis [9-10], and makes sentiment analysis and mining for text in essence; the goal of sentiment analysis is to identify the subjective sentences in the text, making judgment for its sentiment trend, utilizing the existing sentiment analysis [11-12] theory to make judgment for twitter users sentiment orientation, providing the data support for user behavior prediction is particularly important. Therefore, in consideration of the above analysis, this paper based on twitter platform for user to conduct affective computing and make trend forecast, by the data form to achieve the better purpose for predictive analysis. First of all, the first part from different levels to introduce the form and content of the twitter, And from the different effect of expression to classify the twitter form, positive emotions, uncertainties emotion, negative emotion. in the third part, according to the set pair sentiment level model that the author has proposed to calculate the sentiment state polarity of the users, Finally, based on the data, beginning with the size of the same degree, difference degree, opposition degree a,b,c in contact degree theory to obtain the trend analysis table of

Ⅰ level(support), Ⅱ level(neutral), Ⅲ level(opposition), Under the influence of the size relation between a, b, c to give which attitude is the main trend in the trend analysis table, Part IV calculates an instance twitter to forecast the results of trend analysis, and make a detail analysis on uncertain sentiment coefficient take the different value, the sentiment changes of the users showed. Then make the detailed description for each part, as follows.

1. The introduction of twitter form

Along with the development of the Internet era, the twitter form has also changed. Nowadays, the twitter form has five parts: Ordinary text, the new network word, emoticon, picture, and link. The following figure is one twitter intercepted from QQ client which contains the above five parts.

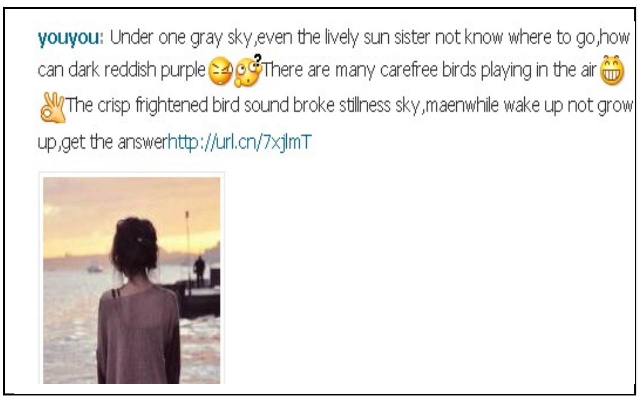


Fig. 1 An entry microbolg.

1.1 The introduction of twitter text

With the constitute form of twitter constantly enrich, the type of text information has become diverse, from the initial form of ordinary text to the emergence of new network word text. work [13]construct the polarity dictionary to have sentiment analysis for the existing problems of the prior sentiment dictionary, including: a basic dictionary, the domain dictionary ,network word dictionary, and the polarity dictionary of modifying word dictionary, and gives the polar strength formula to get the emotion polarity of the text, also mentioned the relative degree size of some adverbs, but in this process do not discuss the polarity of a single word; work [14] gives the definition of twitter text and summed the content of twitter is extremely unique, not only has the characteristics of "short", "colloquial" , "network", "icon" the network text has, also have significant dialogue characteristics in the first post and follow-up, for these characteristics to analyze in the level of linguistic analysis and semantic analysis level, and make a research on the twitter text processing. In this article, starting with each sentiment vocabulary, then dividing into three categories: positive emotion vocabulary, negative sentiment vocabulary, uncertainty sentiment vocabulary.

As in Fig. 1, the text contains the new network words: dark reddish purple, porridge, drops, and me. The dark reddish purple means like this, the porridge means like. And they all belonging to the positive sentiment vocabulary; drop is yes; pair is me. In the calculation of this text, we will convert the new network word into ordinary word to analyze calculation.

1.2 The introduction of twitter text

On September 19, 1982, the Scott - Fireman Professor of American Carnegie - Mellon University inputted such a string of ASCII characters: ":-)." in the electronic bulletin board for the first time. The first computer smiley was born in the history of mankind. Since then, the network emoticons are popular on the Internet world, and widely accepted in society. The new vocabulary to describe these emoticons in the English language in the late 20th century, that is skillfully merge emotion and icon to become the new word "Emoticon". The appearance of emoticons can express more for twitter user's sentiment state at the time in graphical, from the initial simple sentiment symbol to the animation style, during the crawl process, such emoticons will shows the different meanings of this expression in the form of the text, Through the integration of emoticons can get the expression of a complete language, for example:
< (- ^ -)>means angry, I'm angry!! Knock it off me, you know?

Foreign scholar Jonathon Read utilized the Twitter API to obtain a large number of emoticons, and then pointed out smile as the active text and frown as the negative text, so as to achieve the emotion classification in work [15], this paper will divide emoticons into three categories: positive sentiment symbol, negative sentiment symbol and uncertain sentiment symbol. We take into account the sentiment polarity expressed by some sentiment symbol is not all certain, for example, carried with the text message as a "pinched face" ,for this type of emoticons polarity is not obvious, setting them as uncertain sentiment emoticon in the actual analysis. That is, the attitude of author is to remain neutral, in other words, these emoticon can be converted into support or oppose which due to some certain factors. In Fig. 1, for example the text of it is OK; it can be classified into the positive emoticon that represents an willingness to agree; emoticons from the appearance point of view is a sweating state, combined with the text meaning represents a negative sentiment emoticon, integrating with text interpretation of the symbol itself is judged as one negative sentiment emoticon.

1.3 The introduction of twitter picture information

The constitute of twitter in addition to ordinary text, new network words, emoticons also have some pictures, and the picture attached after the 140 word bear the role of transmitting the information to express emotions and the beauty[16], sometimes even these pictures convey richer information than the text, and also can be used as an annotation of text content. In “7·23”Yongwen line major railway accident, the "grassroots" photographer push the twitter picture onto the stage again. This incident highlights the widely used of twitter picture in news dissemination. The work states that the coexistence formats of the picture is professional and grassroots in future, and the trend of the hot network picture is faster and faster, more and more, it has entered the fast food era of picture [17]. This shows that the twitter picture has occupied a very important position in the information expression. As a result, it is very important for the picture analysis. According to the sentiment orientation of the picture content, dividing the pictures into three categories: positive emotion picture information, negative sentiment picture information, uncertainty sentiment picture information.

1.4 The introduction of twitter link

The link also known as the hyperlink, the hyperlink is the connection relation of one page point to an object, and used as object hyperlink in a Web page, it maybe a piece of text, an image or an emoticon. When viewers click the link, the target of the link will be displayed in the browser, and depending on the type of target to open or run. There are some articles has mentioned the link as a new form of twitter in our vision among the existing twitter content analysis, This paper regards the link as one twitter form to detailed analyze, According to polarity of the object of twitter link, dividing into positive sentiment link, negative sentiment link, uncertainties sentiment link.

With the brief introduction above four parts of twitter content, next, this paper utilizes the contact number for users’ sentiment attitude to expand trend analysis.

2. The contact number and trend

2.1 The contact number theory

The contact number [18] was first proposed by Professor Zhao Keqin in 1989, it is the mathematical analysis tool of set pair analysis, which is based on the set pair and the contact degree as the basic concept, depicting and studying one analysis techniques of the certainty, uncertainty and its conversion rule of widely exiting in systems [19-20]. The key thought is: the certain relations and uncertain relations of any system is a unity of opposite body, known as the “certain- uncertain systems”. Contact number puts same degree, difference degree, and opposition degree of the research object and given reference set into one mathematical expression, its general form is:

$$U = A + Bi + Cj \tag{1}$$

Thereinto A 、B 、C are non-negative real number, $j = -1, i \in [-1,1]$ depending on the situation to have value in the range, commanding $N = A + B + C$, N is the contact norm, let N divided the two sides of formulas (1), then taking

$$u = \frac{U}{N}, a = \frac{A}{N}, b = \frac{B}{N}, c = \frac{C}{N} \tag{2}$$

So

$$u = a + bi + cj \tag{3}$$

Therefore, the formulas (3) is Called the contact number expression, a is the same degree, b is the difference degree, c is the confrontation degree, i is the difference coefficient, j is the confrontation coefficient. Therefore, putting positive, negative, uncertain together to estimate the state and trend of the object being studied.

2.2 The trend posture

The trend [21] is the concept that reflecting the size relations order of the same degree (a), the difference degree (b), the confrontation degree (c), and can be divided into the same potential, the balance potential and the anti-potential. In addition, each potential can also be subdivided by the degree of size. In contact degree $u = a + bi + cj$, if $a/c > 1$, that is $a > c$ called the same potential, if $a/c = 1$, that is $a = c$ called the balance potential, if $a/c < 1$, that is $a < c$ called the anti-potential.

In view of this, for twitter form to conduct trend analysis based on contact number, thereby obtaining the orientation threshold of user sentiment state.

3. The sentiment trend and analysis of twitter

3.1 The set pair affective computing model

First, this part gives the set pair sentiment computing model that proposed by the authors in the study of sentiment orientation threshold.

$$\begin{aligned} A &= \alpha A_1 + \beta A_2 + \gamma A_3 \\ &= \alpha(a_1 + b_1 i + c_1 j) + \beta(a_2 + b_2 i + c_2 j) + \lambda(a_3 + b_3 i + c_3 j) \\ &= (\alpha a_1 + \beta a_2 + \lambda a_3) + (\alpha b_1 + \beta b_2 + \lambda b_3) i + (\alpha c_1 + \beta c_2 + \lambda c_3) j \end{aligned} \quad (4)$$

On this basis, utilizing the contact degree to conduct sentiment trend analysis for the single twitter, in this process can get one detailed explanation of the user' attitude.

3.2 The sentiment trend analysis

The main purpose of twitter text analysis [22] is to dig out attitude and their sentiment polarity of the user. The significance of mining not only can identify the potential users of online shopping to help businesses make decisions; but also make predictions on some of the major decisions of the election for the relevant departments to offer data support. Therefore, after the judgment of the polarity of the twitter constitutes form, at the end, we give the sentiment orientation of the users utilizing the set pair trend table.

In actual emotion analysis，Each vocabulary is not the same degree of sentiment orientation，For example, positive affect vocabulary，In the theory of basic emotions，Human emotion is divided into many basic types，Such as joy, sadness, anger, etc. To describe sad can use disappointment, regret, sadness, grief even despair with a variety of different strength level, Similar, joy can also be described by pleasant, happy, joy as well as carnival. Thus need more than strength of sentiment words to describe each class of emotions. The degree of a person's sentiment is indistinguishable from the point of view of knowledge representation. Even more refined classification of each emotion will be similar results. Then leading sentiment orientation that something expressed will be different, So this is due to the system inherent uncertainty leads to uncertainty of the sentiment elements , In this article, we divided sentiment polarity into positive emotions, uncertain emotions, negative emotions, Inductive analysis of the polarity of the emotions expressed by the plain text, the new word, emoticons, pictures. Thus, single twitter user sentiment orientation Contact is defined as follows:

$$\begin{aligned} U &= A + BI + CJ \\ &= (\alpha a_1 + \beta a_2 + \gamma a_3) + (\alpha b_1 + \beta b_2 + \gamma b_3) I + (\alpha c_1 + \beta c_2 + \gamma c_3) J \end{aligned} \quad (5)$$

In the above formula, indicates all the positive emotions of user twitter supportive, including positive sentiment vocabulary, positive emoticons, and positive sentiment pictures. Indicating that all uncertainty sentiment part of twitter neutral attitude, including uncertainty sentiment vocabulary, uncertain emoticons signal and uncertain sentiment pictures. It indicates negative sentiment part of user twitter opposed. Including the negative sentiment vocabulary, negative emotions, and negative affective pictures. Integration by twitter constitutes different parts work together to analyze the attitudes of the orientation of users in twitter.

On Contact degrees analysis of the Relationship of A, B, C and I, J different values to analysis, Can be grading consider based on the size of the relationship between user attitude, According to the value of I, at the time of $I = -1$, Uncertainties B will be completely converted to C, thus we can accord the change of I to carried grasp on the sentiment tendencies relationship trend, we carry on trend analysis use the definition of the potential degree of personal sentiment orientation. The expression of potential as follows:

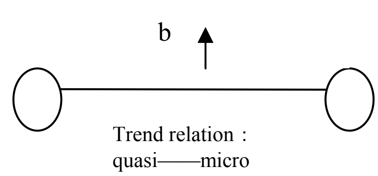
$$trend = A / C \quad (6)$$

We stipulated, if $A/C > 1$, then personal feelings tend to belong to a supportive attitude; if $A/C = 1$, then personal feelings tend to belong to a neutral attitude; if $A/C < 1$, then personal feelings tend to belong to the opposed attitude.

Therefore, according to the three cases of personal sentiment attitude, then we define three different relations level, such as: Level I (support), Ⅱ level (neutral), Ⅲ level (against), which shows the relationship trends change of the personal sentiment orientation, the incremental of relationship level which make the degree of user sentiment orientation changed, the trend of the sentiment orientation grading analysis, as shown in following table1:

Based on the above analysis, among the three factors of affecting user sentiment orientation, the strength relation of a, b, c will change due to the value of i changes in uncertain sentiment part, finally, leading to the trend of certain part of support, neutral, opposition that the users hold. Therefore, under the effect of uncertainty, the attitude of opposition and support that the users hold will be affected in some certain extent. Thus it can be seen; the uncertain affecting factors occupy a very important position in the entire relation metrics.

In sentiment trend analysis table, for the relation class Ⅰ level, Ⅱ level, Ⅲlevel respectively, with the growing of the uncertain factor b ,the relation strength from quasi-relation to strong relation, weak relation and has been degraded to the micro-relation.



Property 1: The value of i takes two boundary values:

Table 1: The trend of sentiment orientation

Relation level	**The size of A.B.C**	**Intension of relation**	**Meaning Description**	**Sentiment orientation threshold**
level Ⅰ (support)	A>C,B=0	standard	determined trend of supporting	positive
	A>C>B,B≠0	strong	Mainly to the trend of supporting	
	A>B>C	weak	the trend of supporting is weak	
	A>C,B>A	tiny	the trend of supporting is very weak	
level II (neutral)	A=C,B=0	standard	the trend of supporting and opposing head-to-head	neutral
	A=C,A>B>0	strong	the trend of supporting and opposing are equal	
	A=B=C	weak	the trend of supporting and opposing are equal, but uncertain	
	A=C,B>A	tiny	the trend of supporting and opposing are equal, due to the uncertainty is very weak	
levelIII (against)	A<C,B=0	standard	determined trend of opposing	Negative
	A<C,A>B>0	strong	Mainly to the trend of opposing	
	A<C,B>A	weak	the trend of opposing is weak	
	A<C<B	tiny	due to the uncertainty the against trend is very weak	

When $i=-1$, uncertainties b will be fully converted to c, at this time, there is only exists the positive sentiment part and negative sentiment part (b and c) expressed as the quasi relation of the relation strength;

When $i=1$, uncertainties b will be fully converted to a, at this time, there is only exists the positive sentiment part and negative sentiment part (b and a), expressed as the quasi relation of the relation strength;

Property 2: The value of i takes among the threshold range:

When $i\in[-1,0)$, at this time, the uncertain sentiment expression tends to the negative emotion;

When $i\in[0,1)$, at this time, the uncertain sentiment expression tends to the positive emotion;

When $i=0$, at this time, there is only exists the positive sentiment part and negative sentiment part.

Thus it can be seen, the value i plays a decisive role in the analysis process of sentiment polarity.

The above table has given relation of a,b,c, we can get a certain kind of attitude the user hold and can be get the relation strength of this attitude from the table, and thus well divided into different sentiment orientation threshold, to make the results more intuitive and clear. Next, by an instance to verify the effectiveness of the method.

4. Instance of verification

We set the Fig. 1 as an example to analyze and discuss the trend of user sentiment state, which is the twitter that contains the text, emoticons, picture and link. The vocabulary is divided into the general vocabulary and sentiment vocabulary, to summarize the sentiment vocabulary according to the number and the weight belongs to the class of such vocabulary, for the emoticons and pictures to summary are also based on the number and weight, The link is based on its specific expression and polarity of an object after clicks the link, making summarization on the basis of the number and weight. The specific data are indicated in following table2:

For the data and weight in table II to calculate, from the perspective of the effect of the expression to give the size relation of α,β,γ is $\beta=0.4>\gamma=0.3>\alpha=0.2$, that is

Table 2: The summarization of vocabulary, emoticon and picture information

Polarity / **Weight Range**	**Positive**			**Uncertain**			**Negative**		
	v	e	p	v	e	p	v	e	p
$(0-0.2]$	3	1		1					
$(0.2-0.4]$	1		2	3	1	1	3		
$(0.4-0.6]$	1	1						1	

(In the above table v= vocabulary, e= emoticons, p= picture information)

the effect of twitter is different, from the intuitive effect to analyze, the effect of emoticons expression is stronger than the picture, the picture is stronger than ordinary text. According to the formula (5) to get the size of a,b,c , $a=\frac{14}{75}, b=\frac{31}{300}, c=\frac{2}{25}$, that is $a>b>c$, the sentiment contact degree is $u=\frac{14}{75}+\frac{31}{300}i+\frac{2}{25}j$.

By the existing set pair affective computing model we can see the sentiment orientation of the user is positive, from the number of the form data in table II, we can obviously see that, among the form of three polarity orientation, the number of positive occupy majority, Which shows the user sentiment state is positive, that is the user for their own pleasure state; Combination with table one according to the trend analysis table, we can see the result of twitter analysis belongs to level I (support), that is the attitude of the twitter hold is support, it belongs to the positive sentiment orientation threshold. Making comparison with the size relation of a,b,c , the relation among them is weak relation, the trend of supporting is weak. By the further analyzing, that is the uncertain sentiment expression plays a regulatory role in this twitter. The specific analysis is as follows:

(1)When $i=-1$, at this moment, the uncertain sentiment portion is converted to a negative emotion sentiment portion, the expression of sentiment contact degree has converted to $u=\frac{14}{75}+(\frac{31}{300}+\frac{2}{25})j$. the relation is $a=\frac{56}{300}>(b+c)=\frac{55}{300}$, As can be seen from the data, at this time the positive and negative feelings is flat, that is, due to the change of the uncertain sentiment part, finally, making the positive emotion converts to neutral emotion.

(2) When $i=0$, at this time, there is no uncertain sentiment in sentiment performance part, the expression of sentiment contact degree has converted to $u=\frac{14}{75}+\frac{2}{25}j$. the relation is $a=\frac{56}{300}>c=\frac{24}{300}$, That is, at this time, there is no uncertain emotion and the sentiment orientation of the user shows positive emotion.

(3) When $i=1$, at this moment, the uncertain sentiment portion is converted to a positive emotion sentiment portion, the expression of sentiment contact degree has converted to $u=(\frac{14}{75}+\frac{31}{300})+\frac{2}{25}j$. the relation is $a+b=\frac{87}{300}>c=\frac{24}{300}$, Comparing the calculated data with the original data, we can see that the degree of sentiment expression of positive emotions is more significant.

From the above analysis, for our subsequent judgment for sentiment polarity classification, about the uncertain emotion also should be taken into account, that is, when i takes different value, it will affect the uncertain emotion, and the relation strength of a,b,c will affect the trend of all the sentiment state.

5. Conclusion and the next step

First, this article introduces four form of twitter; Secondly, on the basis of the set pair affective computing model for trend analysis based on the contact degree. According to the size relation of same degree, difference degree and opposition degree to give the trend of user sentiment change in certain sentiment threshold. Finally, discussing the different value of i in the case, the user sentiment state will show the different sentiment orientation, it demonstrates that the uncertain sentiment occupy an important position in user's sentiment state analysis. In addition, this paper is based on one single twitter to discuss, to make analysis on the specific form, to give the sentiment polarity. The next step is based on the multiple twitter to analyze, considering the report, comments, and replies of users on the twitter will be affected by the users of social association, and analyzes the branch problem among the users' sentiment state of multiple twitter, including the discussion of the transition probability.

Acknowledgments

This paper is supported by Hebei province natural science fund. (F2012209019)

References

[1]Johan Bollen, Huina Mao, Xiaojun, Zeng. Twitter mood predicts the stock market. Journal of Computational Science, Vol.2, 2011, pp.1-8.

[2]L.Barbosa, J.Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data". COLING 2010, pp.36-44.

[3] Qiang YanLianren Wu, Lan Zheng. Social network based twitter user behavior analysis. Physica A: Statistical Mechanics and its Applications. 2012.

[4]Pizard R W. Perceptual Computing Technical Report. Affective Computing [R]. Cambridge: MIT Media Lab, 1995.

[5]Zhang Lumin, Jia Yan, Zhou Bin. Bursty event detection in twitter based on sentiment computi ng. Netinfo Security. Vol.8, 2012, pp.143-145.

[6]Song Shuangyong, Li Qiudan, Lu Dongyuan. Hot event sentiment analysis method in microbolg. Computer science. Vol.39, No. (6A), 2012, pp.226-228,260.

[7]Li Jin, Zhang Hua, Wu Haoxiong, Xiang Jun. BTopicMiner: domain-specific topic mining system for Chinese twitter. Journal of Computer Applications. Vol.32, No.8, 2012, pp.2346 －2349.

[8] Kim S M, Hovy E. Determining the Sentiment of Opinions [A]//Proceedings of COLING-04[C].Geneva, Switzerland, 2004,pp.1367-1373.

[9] Tian Shengli, Xiong Delan. Research on Emotion Attitude Orientation classification of Chinese Web Page Author. Journal of Xinyang Normal University Natural Science Edition. Vol.22, No.2, 2009, pp, 307-309.

[10] Zhu Jian. Text sentiment analysis based on ensemble learning method. Computer Engineering and Applications. 2012.

[11]Akshi Kumar, Teeja Mary Sebastian. Sentiment Analysis on Twitter. IJCSI International Journal of Computer Science Issues. Vol.9, No.3, 2012, pp, 372-378.

[12] A.Agarwal, B.Xie, L.Vovsha, O.Rambow, R.Passonneau, "Sentiment Analysis of Twitter Data". In Proceedings of the ACL2011 Workshop on Languages in Social Media, 2011, pp.30-38.

[13] Zhang Chenggong, Liu Peiyu, Zhu Zhenfang, Fang Ming. A sentiment analysis method based on a polarity lexicon. Journal of Shandong University (Natural Science). Vol.47, No.3, 2010, pp. 47-50.

[14] Zhang Jianfeng, Xia Yunqing, Yao Jianmin. A review towards microtext processing. Journal of Chinese information processing. Vol.26, No.4, 2012, pp.21-27.

[15] Jonathon Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification[C]// Proceedings of the ACL Student Research Workshop. Association for Computational Linguis-tics, Morristown, NJ, 2005.

[16] He Xi. The comparison analysis of spreading of microbolg and newspaper picture. Business and culture. Vol.5, 2012, pp.236-240.

[17] Zhai Zhengxuan. The trend of picture format and the strategies of paper media in the twitter era. Chinese journalist. No.8, 2011, pp.119-120.

[18] Zhao Keqin. Set pair analysis and its application. Hang Zhou: Zhejiang Science and Technology Press, Vol.13, No.47, 1994, pp.67-72.

[19]Zhang, Chunying, Guo, Jingfeng, Liang, Ruitao, 2011, Set pair community situation analysis and dynamic mining algorithms of web social network, ICIC Express Letters, Vol.5, No.12, 2011,pp.4519-4524.

[20]Zhang, Chunying, Liang, Ruitao, Liu, Lu, Wang, Jing, 2011,Set pair community mining and situation analysis based on web social network, 2011 International Conference on Advanced in Control Engineering and Information Science, CEIS 2011, 2011/8/18-2011/8/19, pp. 3456-3460, Dali, Yunnam, China(EI)

[21] Zhao Surui. Evaluation of Situation Analysis Model and Application Based on the Connection Number. Journal of Gansu Lianhe University (natural Sciences). Vol.21, No.6, 2007, pp.57-59.

[22]Xing Xinyan. The motional analysis research based on sequence model. Dalian University of Technology. 2010.

Chunying Zhang is a PhD in computer application technology of Yanshan University, Master and Bachelor of computer science and technology. She is currently working as a university professor in department of computer science at Hebei United University, and also as the master tutor. She is the members of the experts working group of specialty instruction guidance Committee Sub-Committee of the ministry education college and computer science and technology, a member of the steering committee of computer teaching in Hebei Province; an executive director of the Machine Learning institute in Hebei Province, a director of computer educational research association in Hebei Province, an member of China Computer Federation and other society duties. Her major research directions are: social network analysis, intelligent information processing, and Web mining and other fields. In recent years, she has published more than 50 papers on important domestic, international journals and international conferences, and more than 30 articles are retrieved by the three retrieves.

Jing Wang is doing Master of Science in Applied mathematics from Hebei United University, she has done her bachelor in applied mathematics, and she is currently working in the field of social network analysis.

Testing the Usability of the HSPA Wireless Broadband in the Middle East: Jordan and Saudi Arabia as Case Studies

Suleiman Almasri [1] and Saleh Alshomrani [2]

[1] **Information Technology Department, Faculty of Computing and Information Technology, King abdulaziz University, North Jeddah, Saudi Arabia**

[2] **Faculty of Computing and Information Technology, King abdulaziz University, North Jeddah, Saudi Arabia**

Abstract

The number of Wireless Internet Broadband users in the Middle East was emerged due to the advances of Mobile technologies. For example, 3.5G (beyond third generation) provides very advanced services such as high wireless internet speed and good quality video/voice telephone calls. In developing countries, this service is still at the initial phase. There are still some obstacles facing people of these countries to use this new service. This paper provides an overview about 3.5G in these countries: Jordan and Saudi Arabia as case studies. It also discusses people's expectations about this service via conducting a questionnaire targeted Internet users. The paper also presents a network assessment experimental test carried out to evaluate the network usability and QoS at selected places of the area where this service is provided. The test was performed on three main network performance metrics: throughput, delay and packet loss. The results showed that the QoS of the network is irregular i.e. it is acceptable at some locations, whereas it is poor at other locations.

Keywords: *Wireless Broadband; 3.5G; HSPA; Network QoS; Network Usability; Jordan; Saudi Arabia.*

1. Introduction

The past decade has brought about many changes in the way people access information. New technologies provided high speed of wireless internet connection. These technologies provided new range of multimedia services such as Video Calls. In this paper we study one of these important technologies which is called HSPA (High Speed Packet Access). At earlier stage, the widely used broadband technology was the DSL (Digital Subscriber Line) which is provided via cables and was playing an important role in connecting people. Recently, HSPA took the lead in broadband connection, not only for the high speed which can be achieved but also for the reason that the whole service is provided wirelessly using the local cellular network [1]. HSPA protocol is one of the Universal Mobile Telecommunications Systems (UMTS) 3G technologies leading the way by providing users with mobile broadband services [2].

This Usability study was performed to test 3.5G UMTS technology using Orange-Jordan and Mobily- Saudi data cards, which utilises HSPA technology in these two developing countries in the Middle East. The test was conducted in different areas of Amman the capital of Jordan and Jeddah the second largest city of Saudi Arabia.

In order to realise the aim of this study, this paper is structured as follows: Section 2 illustrates the Research background. Section 3 discusses the internet in Saudi Arabia. Section 4 discusses the internet in Jordan. Section 5 presents the data collection methodology as well as the outcome of a questionnaire conducted in Jordan. Section 6 illustrates the results of the experimental tests which have been conducted during this project. Finally, section 7 concludes and summarizes the aim, objectives and results of the whole study.

2. Research Background

HSPA is a 3.5G packet air interface protocol which enriches the user experience to a new level of applications, such as video streaming, position location, mobile gaming, and web access. HSPA is designed to provide lower latency with Round Trip Delays of 70 ms, enabling great interactive applications. Theoretically, it allows up to 3.6 Mbps peak data rate for a Category 6 Mobile per user with a classical receiver and up to 14.4 Mbps peak data rate for a Category 10 mobile per user with a special receiver [2].

HSPA offers a new High Speed Downlink Shared Channel (HS-DSCH), commonly used by a number of users. In addition, it introduces the use of adaptive modulation and coding and the use of fast retransmission based on hybrid technique to handle error packets. HSPA also introduces a high speed transmission at the physical layer (2 ms) which is sometimes called Transmission Time Interval (TTI) [2].

Due to the time scheduling aspect of HSPA, Fast Cell Selection (FCS) is used instead of soft handover whilst moving from one cell to another. The mobile device chooses the best cell which provides the best service on the downlink [2,3].

In HSPA the base station determines which device to send data to in the next 2 ms time frame, hence making the most efficient use of the available bandwidth. The base station also determines how much data to send to user devices based on their link account. The HSPA system uses 16 codes, so the base station then determines how many codes to assign to each device within the cell at any given 2 ms time slot, in order to determine the total speed data that should be sent. The base station can assign all the 2 ms time slots and all other 15 codes to a single user device in the cell, and if that device is available within good signal conditions, the maximum data bit rate could be achieved [1,2].

3. Internet in Saudi Arabia

Saudi Arabia is the largest country in the Middle East occupying four-fifths of the Arabian area (see Fig. 1). In south-western Asia, the country is also at the crossroads of three continents: Europe, Asia and Africa. Population growth in Saudi Arabia is recording high rates, grown by an average of 4.3% per annum over the period from 1977 to 2000, rising from 8.060 million in 1977 to 21.5 million in 2000. The factors that contributed to the rise in the population of Saudi Arabia included a marked improvement in the standard of living, health and social conditions. During the last three decades, a large number of non-Saudi labors who participated in the implementation of development programs contributed also to raise these rates [10].

Saudi Arabia and Syria were the slowest states in the region to allow public access to the Internet [11]. The number of internet users in 2000 was around 200,000, while in 2005 it was around 1,500,000 users. The percentage growth between 2000 and 2005 is 650% [11].

Mobily is one of the ISPs in Saudi Arabia [12]. They started providing HSPA service recently. In this study we utilized their network card to test the usability of the 3.5G new technology in the region.

4. Internet in Jordan

Jordan is a developing country located at the Middle East (see the map at Fig. 1). The population of this country is around 5.5 million. Approximately, half of them are based in Amman (the capital) and surrounding area [4]. Jordan was linked to the Internet in 1994. The number of Internet users in 2005 was 457,000 users [11]. Jordan even made the Guinness Book of World Records for the highest concentration of Internet cafés anywhere in the world. There are more than 200 Internet cafés on a single street in Irbid, Jordan [11].

Jordan is one of the countries which have entered recently the world of third generation cellular wireless networking. This service has been provided during the first half of 2010 by Orange Telecom [13]. But the development of this service in Jordan is still at the first stage. Hence, the number of users is still very small relatively [5].

This research aims to compare the people's expectations regarding the 3.5G in the developing countries to the real measurements of this technology which is being provided to them.

Fig. 1 The map of the Middle East where the experiments conducted

5. Data Collection

In order to measure the importance of wireless Internet in developing countries (e.g. Jordan and Saudi Arabia), a group of 100 participants have been selected and asked to fill a questionnaire. This questionnaire was divided into two main parts. The first of which contains questions about the use of Internet in general. The second part contains questions in relation to the people's expectations about the Wireless broadband (e.g. HSPA). The participants were firstly asked a group of questions about the time they spend connecting to the Internet. Then another group of questions have been asked regarding their preferable internet connectivity (Wireless in opposition to Wired). Finally, they have been asked a group of questions concerning the new service itself (HSPA).

Fig. 2 shows the answers of the first question asking: how often do you use the internet in general? The participant's answers showed that most of them (58%) use the internet frequently and 34% use it sometimes and the rest, which is about 8% only, use the Internet rarely. These answers proof that the sample has been selected correctly.
The second question of the first group was asking whether the participants know about this new service or not. As can be seen

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013
ISSN (Online): 1694-0814 | ISSN (Print): 1694-0784
www.IJCSI.org

in Fig. 3, 85% of them said that they know it whereas only 15% do not. This answer clearly indicates that the sample has been selected correctly and their opinions are having an important effect.

The next question was whether they prefer the Asynchronous Digital Subscriber Line (ADSL), which is the most advanced wired broadband service or 3.5G which is the most advanced wireless broadband service. The answers which are illustrated in Fig. 4 showed that 82% of the sample would like to use the wireless technology whereas 18% only would prefer the wired one. This result gives a clear idea about people's willing to enter the world of 3.5G.

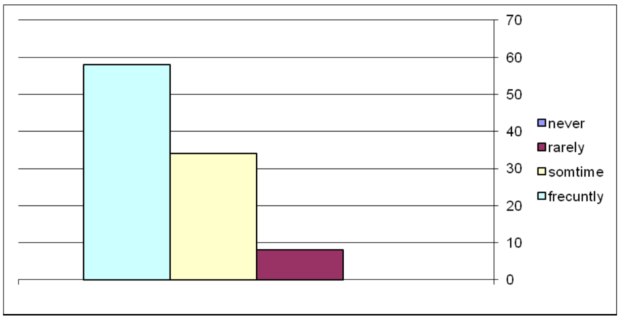


Fig. 2 The Internet Usage

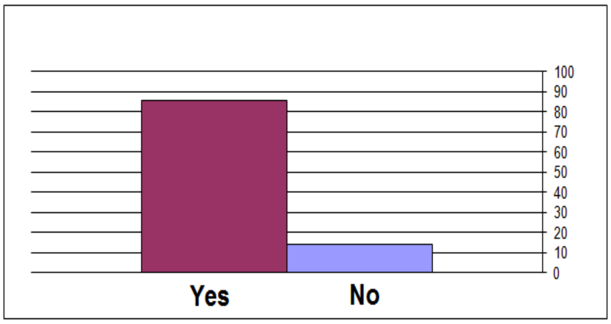


Fig. 3 Number of participants who know 3G+ service

The next part of the questionnaire contains questions about the services provided within the HSPA technology. The participants were asked firstly if they would like to have a wireless connection speed up to 7.2 Mbps. Secondly, they were asked if they would like to make video calls using this new technology. Finally, they were asked if they would like to have better voice call quality. Their answers regarding these three questions are listed in Table 1. As can be seen in the table, 88% of the participants have selected "Yes" for the first question, whereas 12% said "No" for the same question. This result shows that people need to have a broadband Internet on the move while travelling from one place to another.

Besides, 76% of them have selected "Yes" for the second question, where 24% said "No". This result shows that this service will be successful as people would like to use video calls rather than current regular voice calls.

Moreover, 90% of the sample said "Yes" for the third question and 10% of them selected "No". Thus, the majority of people is having problems with the current 2G voice calls and would like to have improved voice calls which can be provided by HSPA [6].

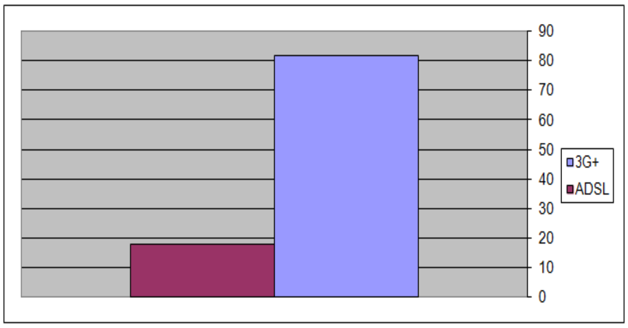


Fig. 4 Preference of 3G+ against ADSL

Table 1: Part two of the questionnaire

Question	Yes	No
Would you like to have a wireless connection speed up to 7.2 Mbps?	88	12
Would you like to make video telephone calls?	76	24
Would you like to have better voice call quality?	90	10

6. Experimental Test.

The second step of this research work is evaluating the performance of the wireless HSPA service in different locations at different times. Therefore, five different places in Amman (Jordan) and Jeddah (Saudi) have been chosen according to how urban the area is. These locations are shown in the map at Fig. 6 and Fig 11.

According to [7], three main issues of wireless network must be tested to evaluate its performance: throughput, delay and packet loss. Consequently, these issues have been tested within this research work many times at each selected location and the average measurement was taken.

The test bed which was utilised during the experimental phase is shown in Fig. 5.A Compaq laptop with Core 2 Due ® processor and a 2GB RAM was connected to a HSPA connector. This connector is the modem which facilitates the Internet access. Data can be sent and received through a cellular (3G) Network. Also, the server which was used to send/receive data to/from is located at Petra University (Jordan) and King AbdulAziz University (Saudi) was contacted via the Internet.

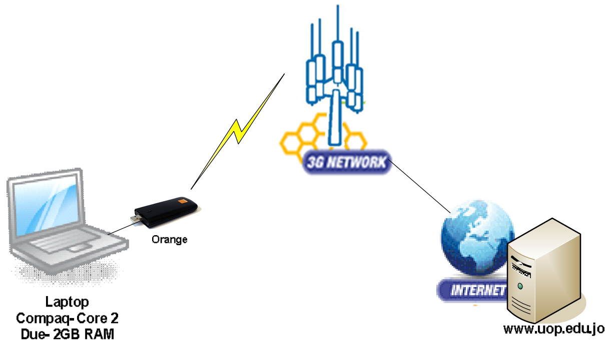


Fig. 5 Setup test of the system

6.1. Experiments in Amman (Jordan)

As can be noticed in Fig 6, the selected positions are located at the western area of Amman. This is because the 3G network is currently available at this area only. Orange™ announced that the whole city will be covered in the near future.

6.1.1. Throughput

Throughput can be calculated by dividing Size of downloaded data over the time taken to download it. Therefore, a 4.47 MB file was used to measure the time taken to download it from a server over the internet, and then this time was recorded. The test was also performed at different times and in different locations. HSPA supports 5, 10, and 15 different codes, with the objective of providing users with adoptive bandwidth allocation depending on their needs. This would give a theoretical throughput of up to 14 Mbps (960 Kbps per code) to users, assuming a user has been allocated with all 15 codes [8]. However, this would not be achieved unless there is only one user located just near the base station.

Fig. 7 shows the average measured time taken to download the file at each location. In this situation, the best throughput was experienced at location 4 whereas the poorest was measured at location 3. In general, the test results at locations 1, 2, 4, and 5 were close to the expected values.

6.1.2. Delay

Delay is the time taken for sending information from sender to destination over a wireless network. In order to test this delay, a ping command was used to send ICMP to measure the Round Trip Time (RTT) [9].

In the first scenario, 100 ICMP's were sent while both Uplink and Downlink are idle. This test was conducted by sending packets to a web site (www.uop.edu.jo) using HSPA link at each selected location. Fig. 8 shows the results of this test. The best case was measured at location 3 and location 4, whereas the poorest was experienced at location 1.

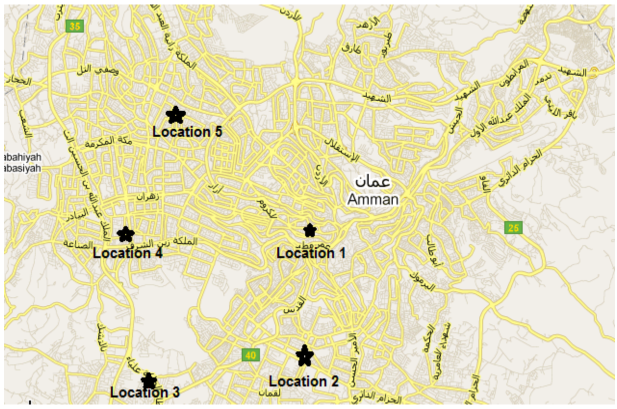


Fig. 6 The five locations of Amman where the experiments took place

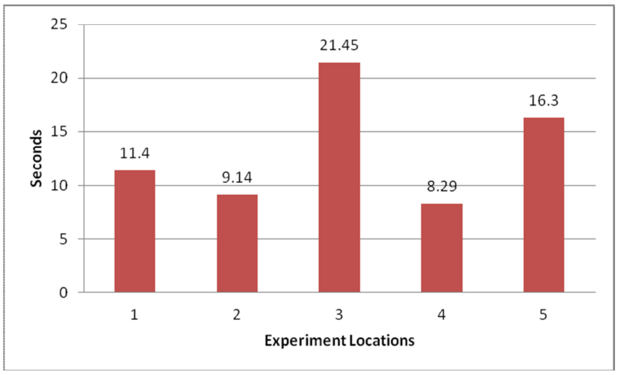


Fig. 7 Average time to download the file measured at each location

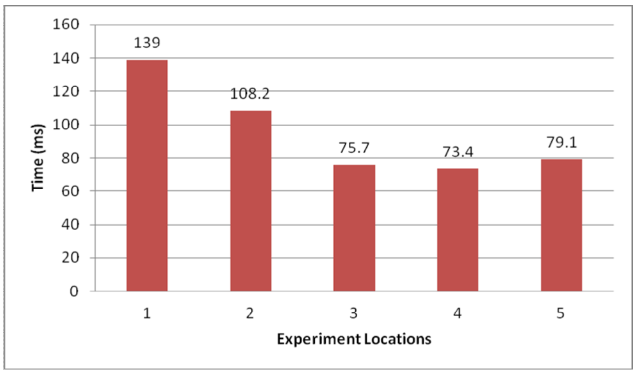


Fig. 8 The average RTT while both Uplink and Downlink are idle at each location

The second scenario of this experimental test is calculating the RTT while the Downlink is busy. This test was performed by pinging the website (www.uop.edu.jo) while downloading data from the same server. The test was repeated many times in different times at each selected location. As can be seen in Fig. 9, the best average RTT was experienced at location 3 whereas the poorest average value was found at location 5.

The third experimental scenario is measuring the RTT while the Uplink is busy. This test was performed by pinging the same website while data is being uploaded to that web server. The test was repeated many times at different times at the selected locations. As can be seen in Fig. 10, the best average RTT was measured at location 3, whereas the poorest was found at location 1 and location 2.

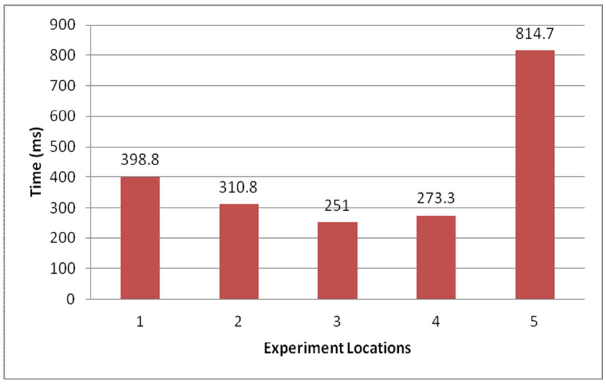


Fig. 9 The Average RTT while downloading data at each location

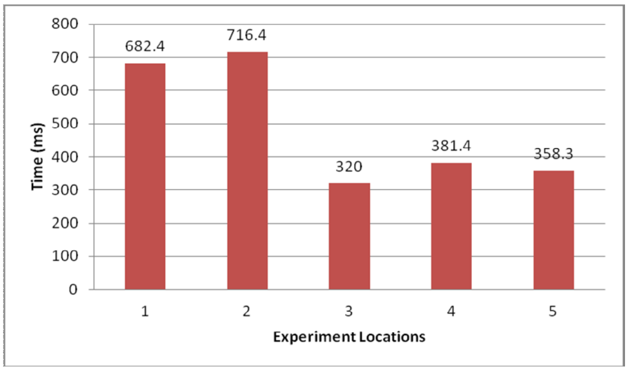


Fig. 10 The average RTT while uploading data at each location

6.1.3. Packet Loss

It is very important to test the packet loss because the more packets the system loses the more re-transmission processes it performs. Likewise, more re-transmission actions carried out will cause more delay in the network. Thus, the test was conducted to monitor the packet loss during different times of day, on different locations and different scenarios. More than 100 packets, in groups of 4, were sent to the website (www.uop.edu.jo), and then all of the packets that did not return were recorded. The test was conducted through two different scenarios: one sending the packets while downloading or uploading a file (busy network), and another without any file being downloaded or uploaded (free network).

The good news is that the average number of the lost packets during these scenarios was very small (only 4.2). These packets where lost while the Downlink was busy, but the average number in all other scenarios was less than 1 packet.

6.2. Experiments in Jeddah (Saudi Arabia)

As can be noticed in Fig 11, the selected positions are located at different areas of Jeddah. This city is located at the western area of Saudi Arabia. It is the second largest city in the country after the capital of Saudi Arabia (Riyadh).

Location 1 is considered to represent a semi urban area. Location 2 is very close to sea with semi urban area. Location 3 has been chosen in an urban district and Location 4 is representing a very crowded urban area. Location 5 is representing very low density areas. It is located outside Jeddah on a highway to Almadinah Almunawwarah City.

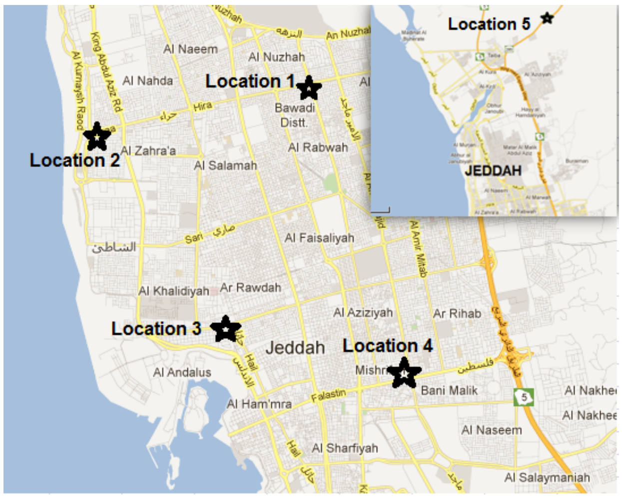


Fig. 11 The five locations of Jeddah where the experiments took place

6.2.1. Throughput

A 2 MB file was used to measure the time taken to download from a server over the internet, and then this time was recorded. The test was performed at different times and in different locations.

Fig. 12 shows the average measured time taken to download the file at each location. In this situation, the best throughput was experienced at location 4 whereas the poorest was measured at location 3. In general, the test results at locations 1, 2 and 4 were close to the expected values. At Location 5 no 3G signal could be received. It seems that the service doesn't cover suburban environments.

6.2.2. Delay

In the first scenario, 100 ICMP's were sent while both Uplink and Downlink are idle. This test was conducted by sending packets to a web site (www.kau.edu.sa) using HSPA link at each selected location. Fig. 13 shows the results of this test. The best case was measured at location 1, whereas the poorest was experienced at location 4 and no service was received at location 5.

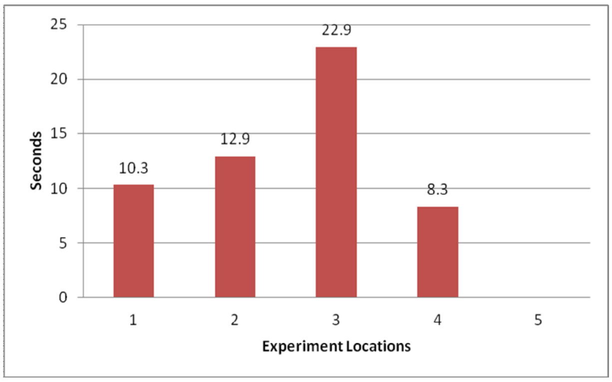


Fig. 12 Average time to download the file measured at each location (Jeddah)

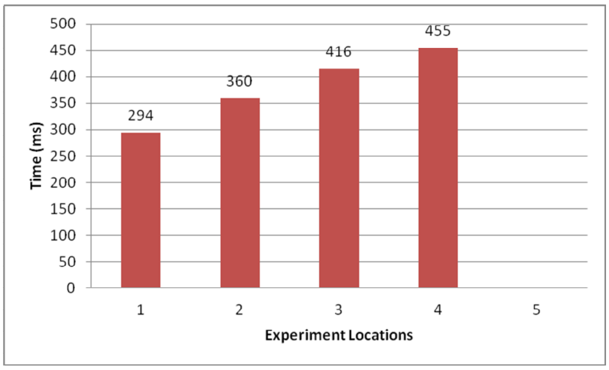


Fig. 13 The average RTT while both Uplink and Downlink are idle at each location (Jeddah)

The second scenario of this experimental test is calculating the RTT while the Downlink is busy. This test was performed by pinging the website (www.kau.edu.sa) while downloading data from the same server. The test was repeated many times in different times at each selected location. As can be seen in Fig. 14, the best average RTT was experienced at location 2 whereas the poorest average value was found at location 4.

The third experimental scenario is measuring the RTT while the Uplink is busy. This test was performed by pinging the same website while data is being uploaded to that web server. The test was repeated many times at different times at the selected locations. As can be seen in Fig. 15, the best average RTT was measured at location 3, whereas the poorest was found at location 4.

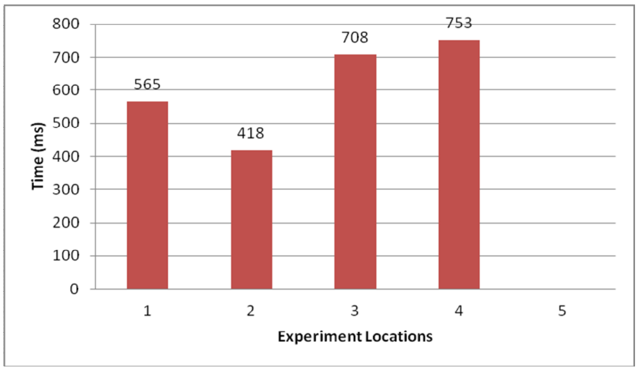


Fig. 14 The Average RTT while downloading data at each location (Jeddah)

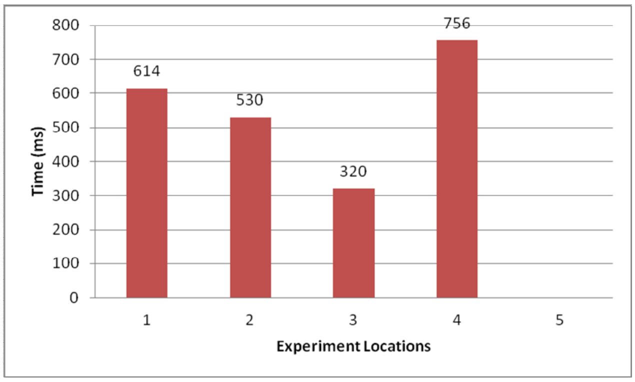


Fig. 15 The average RTT while uploading data at each location (Jeddah)

6.2.3. Packet Loss

During this experimental test, no packet losses have been experienced. This good news is related to the good strength of signal received in locations 1,2,3, and 4.
The outcome of this experimental test shows that the 3.5G network performance provides fluctuated measurements. The QoS is sometimes accepted, whereas at other location or other time of the day it is not. This is because of many reasons: terrain, number of users, weather, etc. As a result, HSPA like any other wireless communication medium is still suffer from such obstacles.

7. Conclusion

In this research work, two main tasks were performed: Questionnaire and real time experiments. The questionnaire methodology and part of the experimental tests have been conducted in Jordan whereas the other part was performed in Saudi Arabia. The outcomes of the questionnaire showed that people of developing countries are willing to use the 3.5G service and are well prepared to enter this world. It showed also that their expectations are very high according to their knowledge about this service. Afterwards, a group of tests was carried out to evaluate the performance of this service and to check if the Network QoS will meet people's expectations or not.

The experimental test results which was conducted in the two countries showed that the QoS of the HSPA network is irregular. Therefore, it would be accepted by users at some locations only. Users might prefer the wired service (DSL) at the locations where the wireless service is poor.

References

[1] M. Jurvansuu, J. Prokkola, M. Hanski and P. Perala. 'HSDPA Performance in Live Networks'. ICC '07. IEEE International Conference on Communications, 2007. Glasgow, UK. Pp. 467 – 471.

[2] M. Necker, A. Weber. 'Impact of Iub Flow Control on HSDPA System Performance'. IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. Vol. 3 pp. 1703 – 1707.

[3] C. Andersson, 'GRPS and 3G Wireless Applications', Willey, 2001, pg 158.

[4] Jordanian Ministry of Tourism. Web site: http://www.tourism.jo/Home/index.htm. Accessed on 1 Feb. 2011.

[5] F. Qawasmeh and M. Bataineh, 'The Impact of Technological Choices on Competitive Strategy: Orange Jordan as a Case Study'. American Journal of Scientific Research. EuroJournals Publishing, Inc. 2010. ISSN 1450-223X Issue 9 (2010), pp.30-46.

[6] B. Beheshti. 'Study of the technology migration path of the cellular wireless industry from 3G to 3.5G and beyond. Long Island Systems', Applications and Technology, 2005. Print ISBN: 0-7803-8930-1. Pp. 15 – 28.

[7] M. Claypool, R. Kinicki , W. Lee , M. Li , G. Ratner, 'Characterization by measurement of a CDMA 1x EVDO network', Proceedings of the 2nd annual international workshop on Wireless internet, p.2-es, August 02-05, 2006, Boston, Massachusetts.

[8] I. Forkel, H. Klenner, 'High Speed Downlink Packet Access (HSDPA)--a means of increasing downlink capacity in WCDMA cellular networks', Proceedings of the European Wireless Conference, Barcelona, Spain, 2004.

[9] S. Benford, 'Coping with Uncertainty in a Location-Based Game', IEEE Pervasive Computing, September 2003, pp. 34-41.

[10] I. Dincer, B. Al-Rashed, 'Energy analysis of Saudi Arabia', International Journal of Energy Research 26 (3), 263–278.

[11] L. Deborah Wheeler, 'Empowering publics: Information Technology and democratization in the Arab World—Lessons from Internet cafés and beyond', Oxford Internet Institute , Research Report No. 11, July 2006.

[12] Mobily Telecom. Internet web site: http://www.mobily.com.sa. Accessed on: 27/12/2012.

[13] Orange Telecom. http://www.orange.jo. Accessed on: 27/12/2012.

Dr. Suleiman Almasri is a faculty of Information Technology Department at King Abdulaziz University. He is also serving now as the head of department of Information Technology– North Jeddah Branch at King Abdulaziz University. He earned his Bachelor degree in Computer Science (BSc) from Philadelphia University, Jordan 1998. He received his Master degree in Computer Science from Amman University, Jordan 2005. He also earned his Ph.D. in Computer Science from Anglia Ruskin University 2009, UK, in the field of Mobile Computing and Location Based Services, and he is actively working in this area.

Dr. Saleh Alshomrani is a faculty of Information Systems Department at King Abdulaziz University. He is also serving now as the Vice-Dean of Faculty of Computing and Information Technology– North Jeddah Branch at King Abdulaziz University. He earned his Bachelor degree in Computer Science (BSc) from King Abdulaziz University, Saudi Arabia 1997. He received his Master degree in Computer Science from Ohio University, USA 2001. He Also earned his Ph.D. in Computer Science from Kent State University 2008, Ohio, USA, in the field of Internet and Web-based Distributed Systems, and he is actively working in this area.

IJCSI CALL FOR PAPERS SEPTEMBER 2013 ISSUE

Volume 10, Issue 5

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 indexed by Google Scholar, Cornell's University Library, DBLP, ScientificCommons, CiteSeerX, Bielefeld Academic Search Engine (BASE), SCIRUS, EBSCO, ProQuest and more.

Deadline: 10th September 2013
Online Publication: 30th September 2013

- 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

For more topics, please see http://www.ijcsi.org/call-for-papers.php

All submitted papers will be judged based on their quality by the technical committee and reviewers. Papers that describe on-going research and experimentation are encouraged.
All paper submissions will be handled electronically and detailed instructions on submission procedure are available on IJCSI website (www.IJCSI.org).

For more information, please visit the journal website (www.IJCSI.org)

IJCSI

The International Journal of Computer Science Issues (IJCSI) is a well-established and notable venue for publishing high quality research papers as recognized by various universities and international professional bodies. IJCSI is a refereed open access international journal for publishing scientific papers in all areas of computer science research. The purpose of establishing IJCSI is to provide assistance in the 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 ongoing 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
10. DBLP
11. EBSCO
12. ProQuest