An Energy-Efficient and Reliable Data Transmission Scheme in Transmitter-based Energy Harvesting Networks

Kyungrak Lee, Wooyeob Lee, Inwhee Joe

Division of Computer Science and Engineering
Hanyang University
Seoul, 133-791 South Korea
+82-02-2220-1088
iwjoe@hanyang.ac.kr

ABSTRACT

Energy harvesting technology has been studied to overcome a limited power resource problem for a sensor network. This paper proposes a new data transmission period control and reliable data transmission algorithm for energy harvesting based sensor networks. Although previous studies proposed a communication protocol for energy harvesting based sensor networks, it still needs additional discussion. Pr oposed algorithm control a data transmission period and the num ber of data transmission dynamically based on environment information. Through this, energy consumption is reduced and transm ission reliability is improved. The simulation result shows that the proposed algorithm is more efficient when compared with previous energy harvesting based communication standard, Enocean in terms of transmission success rate and residual energy.

Categories and Subject Descriptors

C.2.1 [**COMPUTER-COMMUNICATION NETWORKS**]: Network Architecture and Design – *Wireless communication.*

General Terms

Algorithms, Performance, Design,

Keywords

Energy Harvesting, WSN, Reliability, Energy efficiency

1. INTRODUCTION

A wireless sensor network has a strength compared with wired network in terms of conveni ence and commercialization of deployment and maintenance. For this reason, a wireless sensor network system is considered as a promising technology for environmental monitoring, fire-alarm systems, the military, construction sites, and building management. But limited energy resource of sensor nodes is a critical problem . Recently, to overcome this limitation, energy harvesting technology has been studied for a sensor network. A randomly deployed sensor node is unable to be supplied with additional energy from the outside. Therefore, most sensor nodes is unusable if their own energy is exhausted. But if Energy harvesting technology is engrafted upon sensor network, the utility of wireless sensor network will be improved. For this reason, energy harvesting technology for sensor network has been studied briskly . But it still needs additional discussion.

This paper proposes a new energy efficient and reliable data transmission algorithm in energy harvesting based sensor networks. In Section 2, we present related works of energy harvesting based sensor networks, and in Section 3, we propose a new energy efficient and reliable data transmission algorithm. The simulation results are given in Section 4 and we conclude the paper in Section 5.

2. RELATED WORKS

The Enocean standard is ty pical technology for the Energy Harvesting Sensor Networks. Figure 1 shows the data transmission procedure of the Enocean standard.

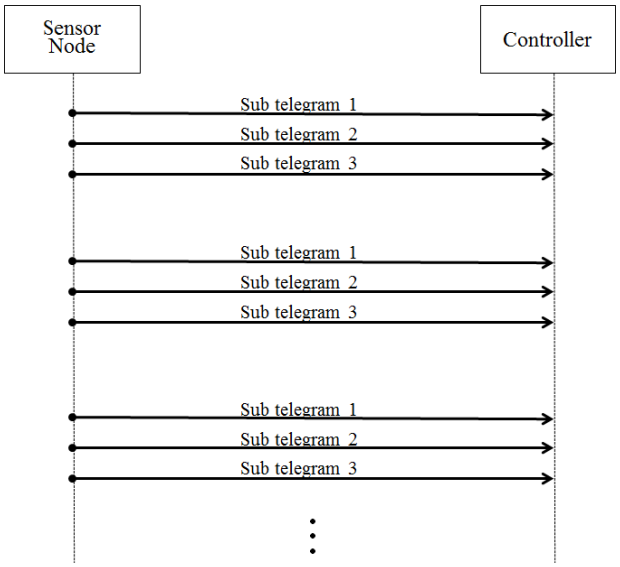


Figure 1. Enocean standard based telegram transmission

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.1

The transport protocol of the Enocean is based on the Slotted Aloha. Therefore, the transmitted message is vulnerable to collision. Therefore, when one telegram is transmitted, because of collision, in order to prevent failure of message transmission further two messages are transmitted. Transmitted message is called as sub-telegram. And three sub-telegrams are the same data. Each sub-telegram is transmitted at designated time slot. And depending on the number value of sub-telegram, distributed range is set differently. So collision is considered to be avoided. However by increasing the number of nodes, the number of transmitted sub-telegrams is increased in entire network. And if the number of distributed slots are exceed, rather perform ance is reduced. In addition, all nodes always transmit three sub-telegrams. So power consumption of the entire network would become very larger than throughput of actual message. It is important issue in terms of power consumption at Energy Harvesting Sensor Network Environment, it is a fatal problem.

3. PROPOSED ALGORITHM

This paper proposes a new energy efficient and reliable data transmission scheme in energy harvesting based sensor networks. It is composed of 2 phase, the number of transmission control and transmission period control.

3.1 Transmission Number Control

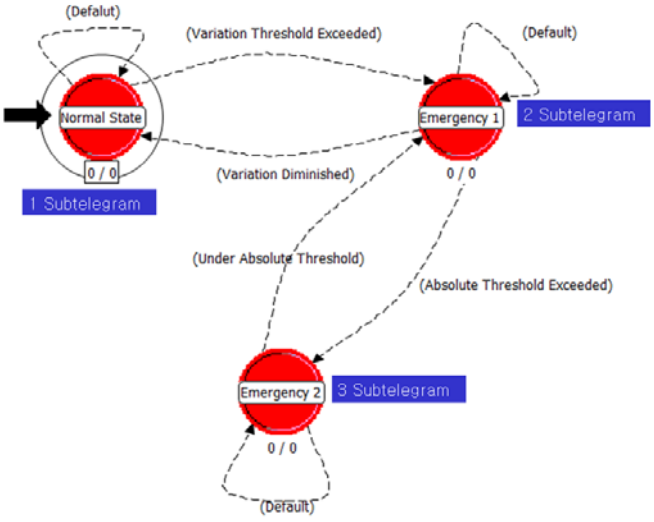


Figure 2. State diagram of transmission number control

A sensor networks which senses an environment periodically requires a time different criteria depending on sensing data. In this paper, we propose an important factor setup algorithm according to sensing data and the number of transmitting Control algorithm according to important factor. The important factor is defined by the absolute value and relative variation of sensing data. A sensor node sends a packet only once in general sensing value. If a sensing value breaks out of general state range, a sensor node sends a packet twice. And, if dangerous state range is sensed, a sensor node sends a packet three times. In addition, although a sensing data belong to same state range compared with previous data, a sensor node sends a packet once more if a variation of data is bigger than variation threshold. The reason of repetitive data transmission according to important factor for reliability is that it is impossible to retransmit a loss-packet when a packet loss occurs in transmitter based energy harvesting sensor networks. In Enocean standard, sensed data is transmitted to controller three times repeatedly for reliability. It is suitable to sparsely deployed sensor networks composed of sensor nodes which have equivalent important factor. But in densely deployed sensor networks, Enocean standard based communication method arouses a low network performance by so many packet transmission processes.

To overcome this disadvantage in densely deployed sensor networks, a new packet transmission algorithm is needed. This paper proposes a reliable and low network overloaded transmission algorithm. In reliable state, we reduce the number of packet transmission gradually for efficient network utilization and low power consumption.

3.2 Transmission Interval Control

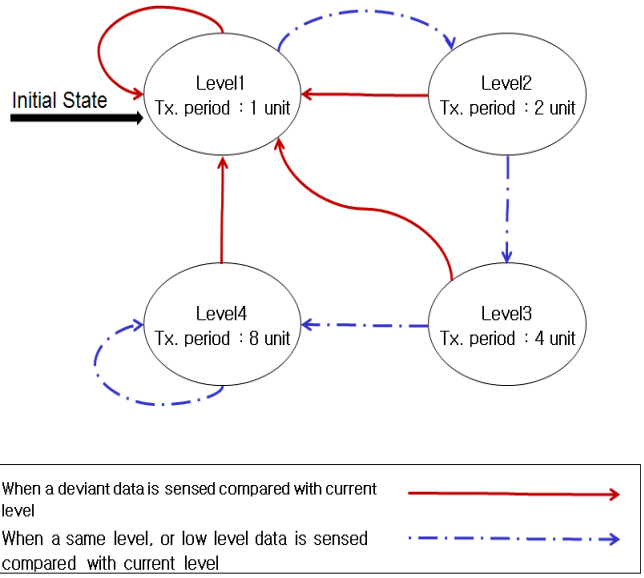


Figure 3. State diagram of transmission interval control

The Sensors consume much power when they transmit or receive data. Therefore it is essential that reducing the number of transmit or receive the data for power consumption. We propose an adaptive control duty cycle algorithm to improve conventional duty cycle algorithm. Figure3 shows a proposed adaptive control duty cycle algorithm. There are 4 kind of Level of Duty cycle and each level has a different duty cycle. Level 1 is the shortest duty cycle which is 1 unit. Duty cycle is increased by two times when they are to be the upper level. Level 4 is the longest duty cycle which is 8 unit. Level is to be level1 regardless the present level when the sensor measures the value which is out of the level's scope. Level is to be the next level and duty cycle is increased by two times when sensor measures the value which is within the level's scope. Through that we reduce the number of the transmission. On the other hand, we get the environment

information quickly by reducing the duty cycle when the sensor measures the value which is out of the threshold.

4. PERFORMANCE EVALUATION

4.1 Simulation Environment

In this paper, we performed a simulation to verify the proposed reliable transmission algorithm. We evaluated the transmission scheme of the original Enocean and our proposed scheme in terms of the PAR (Packet Arrival Rate) and the Rem aining Power. We created the simulation scenario with 20 transmission nodes and 1 receiver node. The parameters used in the simulation are follows.

Table 1. Simulation parameters

Parameter	Value
The number of transmitting telegrams (Not sub telegram)	1000
The number of nodes	20
Sensing threshold	50 Celsius degree
Data rate	125 kbps
Transmission interval (1 unit interval)	1 sec

4.2 Simulation Results and Analysis

As shown in the Fig 4, all nodes using the original Enocean scheme mark the similar values close to 70 %. The s uccess of the transmission means that at leas t one of three s ub-telegrams is arrived successfully.

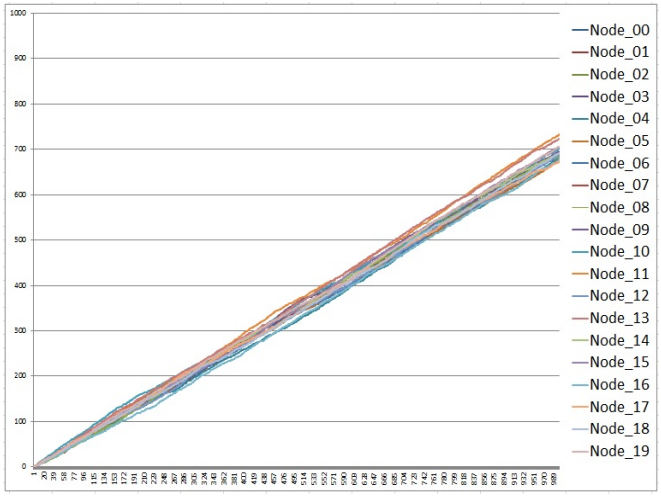


Figure 4. Simulation results for original Enocean

The Fig 5 displays the Performance of the proposed scheme. As the sensing value of the node 18 is 50 Celsius degree, it sends three sub-telegrams because the co llected is treated as critical factor and the rest send only one sub-telegram. The result shows that the PAR of the node 18 marked the value over 95 %, and it means that the perform ance of our s cheme is enhanced approximately 36 % compared to the average PAR of the Original Enocean scheme because of the low probability of the channel collision.

The Fig 6 demonstrates the comparison between the proposed and the original scheme in terms of the power consumption of whole network. As shown in the figure, the power consumption of the proposed scheme is notably decreased because of a sm all number of total sub-telegram transmissions.

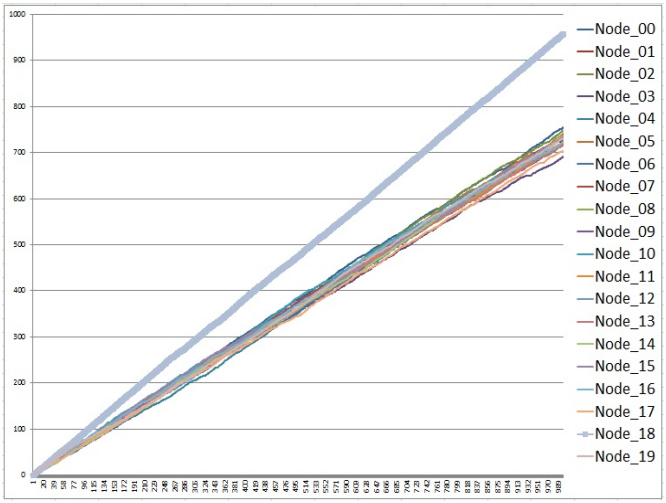


Figure 5. Simulation results for the proposed algorithm

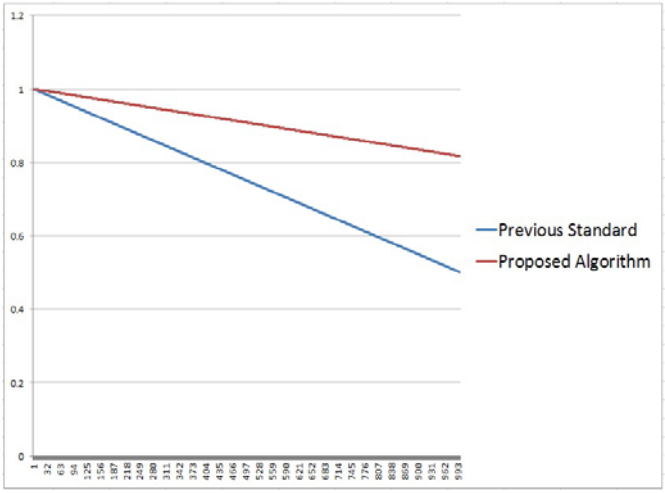


Figure 6. Residual power

4.3 Implementation

We embodied an energy harvesting based wireless sensor network with the proposed algorithm for performance evaluation. Figure 7 shows a whole test-bed composition. It consisted of 1 Controller, which collects end de vice's sensing data, and several end device, which senses envir onmental information with own

sensor. For controller and end device, we use the TCM300C module and the SCM300C module respectively. And, Task manager is connected with the controller to monitoring sensing data of end device. The parameter values for evaluation are equally configured according to Table 1.

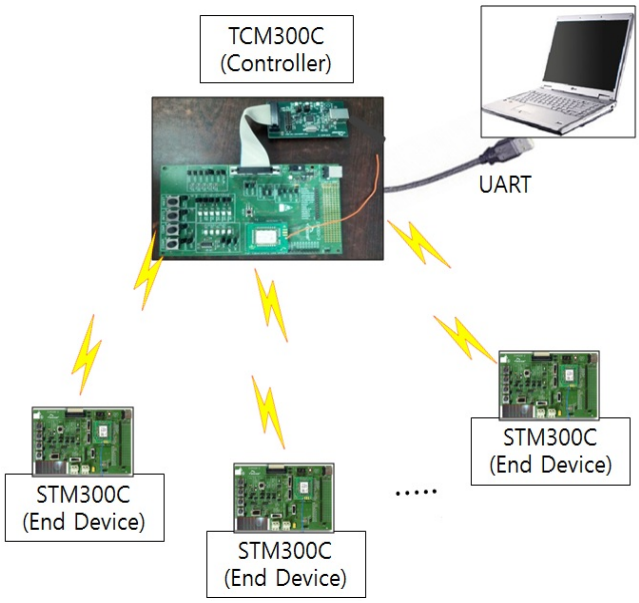


Figure 7. Test-bed setup for the proposed algorithm

Table 2. Test results

Measurement item	Original algorithm	Proposed algorithm
Maximum life time (in Sunlight)	Infinite	Infinite
Maximum life time (in darkness)	1987 sec	315 min

4.4 Experimental Results

We performed a test to verify the energy efficiency of the original and proposed algorithm. We set the default sub-telegram transmission interval (1 unit) up as 1 second, and measured the maximum life time of each end device. In S unlight, both original and proposed algorithm s operated infinitely because the harvesting module generated much more power than consumed power by transmitting sub telegram . But, in darkness, end device's operating with original algorithm is m aintained averagely 1987 seconds. The table 2 shows the test results. On the other hand, the proposed al gorithm based end device shows more long maximum life time, it is approximately 15,750 second.

5. CONCLUSIONS

In this paper, we propose a new energy efficient and reliable data transmission scheme in e nergy harvesting based sensor networks. The proposed schem e improved transmission success rate and an amount of residual energy compared with previous Enocean standard. The sim ulation result shows that our proposed scheme is more effective than the original schem e in terms of the PAR and the energy consumption.

6. ACKNOWLEDGMENTS

This research was supported by the MKE (The Ministry of Knowledge Economy), LG Electronics, under IT/SW Creative research program supervised by the NIPA (National IT Industry Promotion Agency).

7. REFERENCES

[1] J. Ploennigs, U. Ryssel, and K. Kabitzsch. "Performance analysis of the EnOcean wireless sensor network protocol", In ETFA - 15th IEEE Int. Conf. on Emerging Technol. And Factory Autom., pages 1–9, Sept. 2010.

[2] Enocean Alliance, "EnOcean Radio Protocol," Feb.8, 2011.

[3] http://www.enocean.com/case-studies/

An End-User Friendly Architecture for Emergency Networks on Disaster Site

WooYeob Lee, MoonWon Choi, Inwhee Joe
Division of Computer Science and Engineering
Hanyang University
222 Wangsimni-ro, Seongdong-gu, Seoul, Korea
82-2-2220-1088
iwjoe@hanyang.ac.kr

Yeonyi Choi
Department of Fire Safety Management
Shinsung University
Dahak-ro Dangjin-gun, Chungnam, Korea
82-41-3501-423
yychoi@shinsung.ac.kr

ABSTRACT
In this paper, we proposed an emergency network architecture which enables the End-user to use the network resource of the Rescuer Network. We suggested a conceptual approach of the End-user Friendly Architecture and deducted the requirements of the End-user friendly architecture. Finally, we implemented a simple test bed of our suggested architecture.

Categories and Subject Descriptors
D.2.1 [**Computer-Communication Networks**]: Network Architecture and Design – *Network Topology, Wireless Communication*; D.2.3 [**Computer-Communication Networks**]: Network Operations – *Network management, Network monitoring*

General Terms
Management, Design, Experimentation.

Keywords
Emergency Network, End-user, Disaster, Rescuer Network.

1. INTRODUCTION
Recently, there were many disasters in many countries such as the USA, Japan, Philippines, etc. The communication system of those places of disaster became unavailable. This paralysis of the communication system caused not only the long-term problem but also the short-term problem which made it difficult for the rescuers to save people in disaster. Therefore, many countries work for building the Emergency Response Network. Especially, they are researching about constructing the Rescuer Network which is the core part of the Emergency Response Network [1][2][3][4].

The basic purpose of the Rescuer Network is the rescue work in the disaster area which the communication system is totally destructed. The main members of the network are the rescuers. It satisfies both the QoS and the security by building the dedicated network with the dedicated devices. Because of the issue of the QoS and the security, the access of the End-user is basically excluded. However, because of the lack of the communication system, the End-users, who are the targets of the rescue work, cannot provide the helpful information such as their survival and locations. And the possibility of the rescue work is, of course, relatively decreased.

In this paper, we propose a conceptual approach of the End-user friendly architecture. We deducted the requirements for the End-user access and suggested some conceptual solutions with a simple test bed. Finally, we address the future works for this issue.

2. RELATED WORKS
Many countries are researching about the Emergency Response Network. Their works are specified in the states of their country and most of those works are not focus on the communication network but the response process against the disaster situation.

The Homeland Security of USA published the National Incident Management System [1] which contains the detail process of the emergency response in terms of whole country. It provides the procedure to build the disaster response system with the detail descriptions of its components and the resource management manual of it.

Ahmed and Khan et al. [2] proposed the SAFIRE which is the network for the First Responders who perform the rescue works. The SAFIRE used the Publish and Subscribe method for the effective data distribution and they developed the routing and forwarding engine by using the topology information. However, they did not consider and mentioned about the access of the End-user.

Ram G. and Narayanan et al. [4] proposed the joint network system for the recovery and rescue operation in the disaster site. They proposed the network architecture called Portable Recovery Network which enables survivors to report their locations to a Command Center. However, this feature is not for the rescuer network but the Disaster Recovery Network (DRN). The difference between the rescuer network and the DRN is the main user of those networks. Since the rescuer network is established only for the rescuers to perform rescue operations, it is always available in every disaster sites. But the DRN is a special network located in the disaster site to recover the communication system. It means that it is not mandatory to establish the DRN in disaster site.

Although there are many works about the rescuer network, the access of the End-user is mentioned only in case of partial destruction of the infrastructure or in the DRN [4]. As mentioned

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.2

above, as the information of the End-user is helpful for the rescue work, it is necessary to give the access permission within the traffic and security boundary.

3. PROPOSED ARCHITECTURE

3.1 Overview

In this paper, we propose a rescuer network which helps the rescuers to save people by enabling the access of the End-users. We assume that the communication system of the disaster area is totally destructed.

Basically, our proposed architecture is based on the multi-hop mesh network [5] and follows the existing structure: Main/Local Command Center (MCC/LCC), Rescuer Node (RN), Fixed Relay Node (FRN). And we added the End-user (EU) to the rescuer network

3.2 Requirements

Monitoring: It is necessary to monitor the status of the rescuer network and the changes caused by executing the network building process. Monitoring must be available to MCC, LCC and RN. This feature is similar to the existing system, but it need for the function to check the access status of the EU.

Command System: According to the addition of the EU, it is necessary to add essential commands to manage the situation related to the EU. If the EU, who is the target of the rescue work, connects to the RN or the FRN, the LCC should send the rescue command to the rescuers and control the service limitation. Therefore, it is essential to create the command system.

General Access Interface (GAI): Since the device of the EU is not fully compatible with the rescuer's dedicated one, it is necessary to provide a General Interface. The interface should be considered in two terms: Network Interface (NIC) and User Interface (UI). In terms of the NIC, the devices of the EU must include at least one same NIC to the NICs of the RN or the FRN. Different to the case of the NIC, the UI is rarely compatible since there are many S/W platforms. Therefore, it is essential to find the platform-independent UI.

Service Description for the End-user: We must define the services for the EU because of the limited platform and the resource limitations.

Permission Procedure: It is necessary to define the rules for the access of the EU. These rules should be defined as the procedure between the LCC and the RN or the FRN.

Rules for the Use of Network Resources: In order to not causing the traffic problem, it is necessary to make rules for the use of network resources. These rules should consider the network traffic status and the available services of the EU which are selected by considering the status of the EU. And to apply the selected services to the network, it needs the resource scheduling scheme which is essential to minimize the traffic problem.

The Figure 1 demonstrates the conceptual network of our architecture. The detail of each component is described below.

Main Command Center: The role of the MCC is just to monitor the LCCs and their internal members. We assume that the MCC is located far from the disaster area and it is connected the LCC with secure line. As an Exception, if the link between the LCC is disconnected and the members of that LCC try to connect to the MCC via satellite, MCC recognizes that the LCC is not available. Therefore, the MCC manages those members temporally and order to restore the broken LCC.

3.3 End-user Friendly Architecture

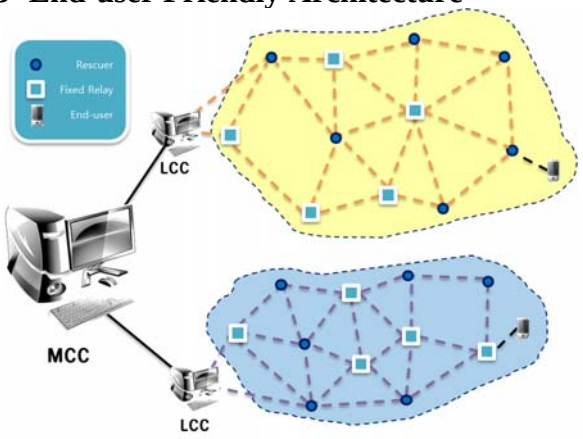


Figure 1. The Architecture of End-user Friendly Network

Local Command Center: The role of the LCC is the manager of a single rescuer network. It is located close to the disaster area and it is connected to the rest members of the network in Mesh manner. Same as the existing architectures, our LCC is fixed node too and it checks whole conditions of the rescuer network. It monitors the rescuer network, manages the services of the rescuers, orders the rescuer work and controls the services of the EU. It has many functions of communication such as the single/group SNS, Voice communication, etc., which are available only for manager or head of the rescuer.

Rescuer Node: The RN is the core member of the rescuer network which is located in the disaster area and connected to other members as Multi-hop Mesh network. We assume that the RN has many types of NICs and the WiFi NIC is used for establishing the mesh network. It can access to the LCC or MCC via satellite depending on the status of the mesh network. Basically, the RN has single Voice/SNS permission. Some of them can take the authority of the Head which is the field manager of RNs and they has have almost same power to the LCC.

To enable the access of the EU, we gave an additional role to the RN. First, It acts as Service Access Point (SAP) of the EU. Since the NIC of the RN is limited, we use WiFi NIC which is the most popular interface.

Because of the security problem, the connect manner of the EU is not the mesh type, but the 1:1 access to the RN or the FRN. And, as mentioned in the previous chapter, we need a general UI. Therefore, in this paper, we removed the platform dependency by using the Web to provide services.

Fixed Relay Node: There are two main functions of the FRN. One is to handle the exceptions and the other is to remove the shadow area. While the RN is in motion, it sets up the FRN in the shadow area or the place where there are no RNs to maintain the connection including the in-door building area. Although the basic role of the FRN is routing and SAP of the EU, in some cases, it acts as the RN or the EU by the login process.

End-user: The EU is the additional member of our proposed architecture which has very limited access permission. Since the main

purpose of the EU is to help the rescue work, it provides the status information of EU itself such as the location. The Smart phones or Lap top computers are used as the devices of the EU and the services are provided via Web pages. If the EU does not have any available devices, it can use the FRN as the device of the EU by logging in with guest account.

3.4 Process Description

3.4.1 Establishment

The establishment of the rescuer network is progressed in three steps.

Disaster Recognition: If the MCC recognizes the disaster situation, it requests the installation of the LCC.

Installation of the LCC and Detachment of the Rescuers: Once the LCC is set up close to the disaster area, from there, the rescuers enter the disaster area with the RN and the FRN

Expansion of the Rescuer Network: The rescuers expand the network area by setting up the FRN while moving the area.

3.4.2 Network Access Process of the End-user

Access Phase: The EU connects to the RN or FRN in 1:1 manner and it belongs to that RN or the FRN

Service Phase: Once the EU is connected to the RN or FRN, it can use limited services via Web pages. This Web page is fixed and other web services are not available because of the security problem. In the initial state, it can only use the SOS service.

3.4.3 Service Management Process of the End-user

SOS Request Phase: If the EU requests the rescue work via the SOS Service, the LCC selects the target EU and orders them to rescue the selected EU

Service Level Management: Once the rescue works is ordered, the service use permission moderates and from that moment, the EU can use the SNS and Voice services. But, they are available only if the traffic condition of the network is good enough. If the condition is bad, the voice service is limited. In case of the Voice Service, the communication reaches to all rescuers who are selected as the rescuer of the current rescue work by the LCC. The responder of the EU is the RN who is the SAP of that EU, but can be transfer to other RN.

4. TESTBED and DISCUSSION

4.1 Testbed

In this Chapter, we introduce our testbed of the rescuer network. Since our testbed is in development, we formed a simple testbed. The basic architecture is same as the previously mentioned network.

Figure 2 is the simple tested with 4 laptops and 1 smart phone. The 4 laptops are MCC, LCC, RN and FRN relatively and the smart phone is the EU. The LCC, RN and FRN are connected as multi-hop mesh network with Wifi Interface. The shown tested is MS Windows-based system. But it is compatible to the Linux platform because the software platform is made by using the Java language. Therefore, our software platform can be operated on many other hardware systems such as the Gateway or Router with OpenWRT, Mac PC with Mac OS, etc.

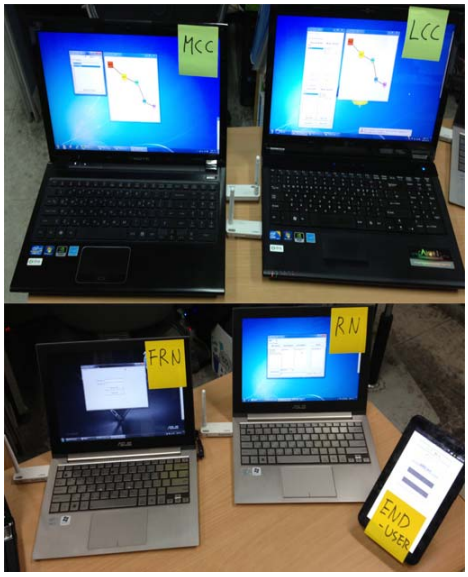


Figure 2. Simple Testbed

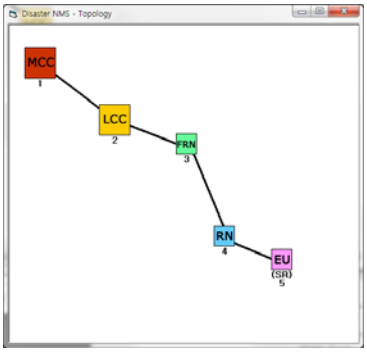


Figure 3. The Network Topology Window

Figure 3 is the whole network topology window which displays on the MCC, LCC and RN. With this window, the MCC can monitor the status of the LCC and can respond the exceptions.

For example, if the FRN becomes disabled, all components belonged to the FRN turn to the dark color and the link between the LCC and FRN disappears.

All members except the EU share the information of the network topology. Since the EU is not a regular member of this network, we cannot give the permission to access the topology.

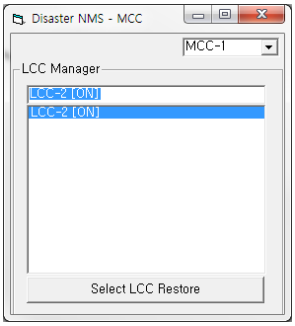


Figure 4. The Control Window of the MCC

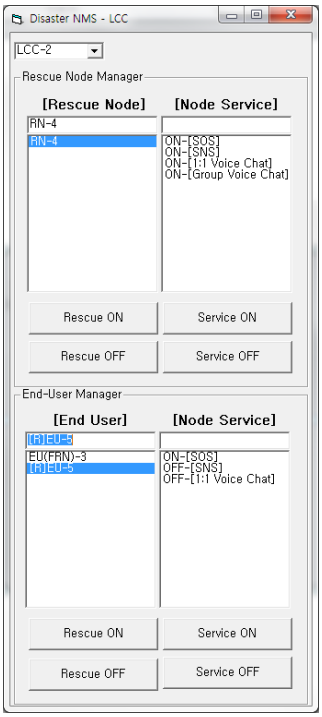


Figure 5. The Control Window of the LCC

Figure 4 is the control window of the MCC. The MCC receives all information of all LCC, but it does not control them. Basically, it monitors the topology window and if the LCC become unavailable, it orders the restore command by using the control window. Then the control window extends to the similar control window to that of the LCC.

Figure 5 is the control window of the LCC. The LCC manages all services of the members and orde rs the rescue work both in manual and automatic manner. For the automatic operation, we need a modeling part but this is not the range of this paper.

And, as mentioned in the previous chapter, the F RN can acts as the EU and it is displayed in the End-user part of the figure 5.

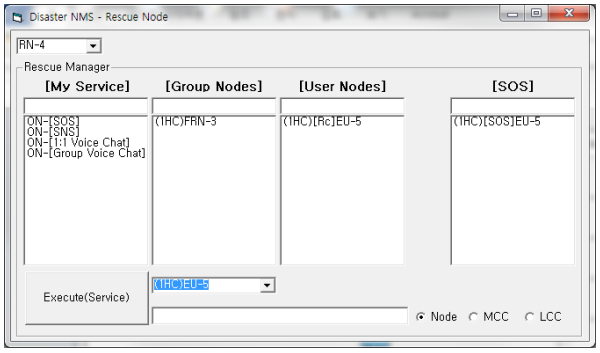


Figure 6. The Control Window of the RN

Figure 6 is the control window of the RN. It displays the available services and the list of the conn ected EUs. Since the RN is the main component of the rescuer network and located in the deep area of the disaster site, this control window display s many factors of the network to help the rescuers for the rescuer operation and for the safety of rescuers themselves.

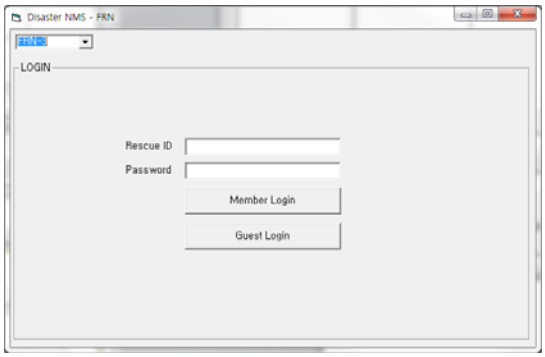


Figure 7. The Login Window of the FRN

Figure 7 is the login window of the FRN. Basically, the FRN does not need extra control window, but if the rescuer or the EU wants to use as the RN or the FRN, as stated above, it must login with rescuer or guest account. If the EU logins to the FRN, the status of FRN changes to EU-FRN and this can be monitored at the control window of the LCC like figure 5.

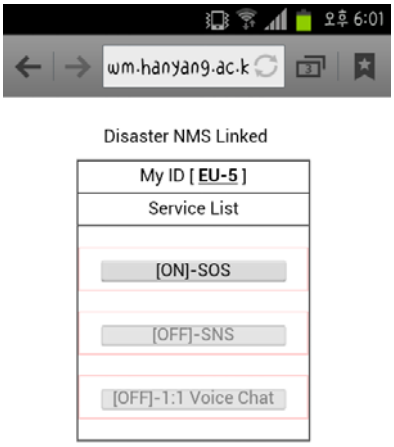


Figure 8. The Service Web Page for the EU

Figure 8 is the Service W eb Page of the EU. With this page, the EU can use the platform-independe nt service. As shown in the figure, according to the m anagement of the LCC, the EU can select the service am ong the perm itted services from web page. But the SOS service is always available.

4.2 Discussion

The future works for our propose architecture and for the testbed are as below.

Communication Algorithm for QoS support: It needs a fitted communication algorithm for QoS support in the Multi-hop Mesh Network. It must include the channel management in MAC Layer, the traffic control by considering the network status, the Dynamic resource reallocation scheme by using the inform ation of targets of the rescue work, etc.

Dedicated Testbed Device: In this paper, we used the general communication devices for the test bed, however, to optimize and to meet the need of QoS, it is essential to produce a dedicated

devices for the rescuer network. And, obviously, the S/W platform for that device must be developed.

The Accuracy of the Information: In this paper, we didn't use the information such as the location values, sensor values, etc. However, to meet the goal of helping rescuer to rescue person easily, it is necessary to use these values. Therefore, we need to algorithm to increase the accuracy of that information since the data is limited because of the traffic problem

5. CONCLUSION

In this paper, we proposed a conceptual architecture of the rescuer network which enables the access of the EU. We deducted the requirements for that and suggested some conceptual solutions for some of them. We also formed a simple testbed for the easy understanding of our proposed architecture. Finally, we addressed the next research topics for our architecture

6. ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation by Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A3012227).

7. REFERENCES

[1] Homeland Security. USA. 2008. National Incident Management System. *Homeland Security*, USA

[2] Jamshaid, K. Khan. O.Z. 2007. SAFIRE: A Self-Organizing Architecture for Information Exchange between First Responders. *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07. 4th Annual IEEE Communications Society Conference on*.

[3] Kanchana K. Apinun T. 2007. A Multimedia Communication System for Collaborative Emergency Response Operation in Disaster-affected Areas. *Technical Report No. TR_2007-1 Internet Education and Research Laboratory (intERLab)*.

[4] Ram G. Lakshmi Narayanan a, Oliver C. Ibe.2012. A joint network for disaster recovery and search and rescue operations. *Computer Networks, Volume 56, Issue 14, Pages 3347–3373.*

[5] I.F. Akyildiz, X. Wangb, W. Wangb.2005. Wireless mesh networks: a survey, *Computer Networks, 47 (2005), pp. 445–487*

Energy Efficient In-Network Data Processing for Mobile Object Tracking System

Jae Sung Choi
DGIST
Daegu, Korea
82-53-785-4783
jschoi@dgist.ac.kr

Byung Rak Son
DGIST
Daegu, Korea
82-53-785-4772
brson@dgist.ac.kr

Dong Ha Lee
DGIST
Daegu, Korea
82-53-785-4720
dhlee@dgist.ac.kr

ABSTRACT
A mobile object tracking system has been widely researched for higher functionality of Location Based Service (LBS). In this paper, we propose robust and fault-tolerance in-network aggregation algorithm using a tree structure for Wireless Sensor Networks (WSNs) based object tracking. The proposed technique improves successful reception rates of tracking information from each sensing node to the base station with minimal delay.

Categories and Subject Descriptors
C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design-Distributed Networks

Keywords
Sensor networks, robustness, tree-based in-network data processing.

1. INTRODUCTION
Recent advances in technologies have made large-scale deployment of low-cost sensor devices which have capabilities of communicating, processing, sensing, and storing. Large-scale networks of these sensor devices are being applied to many applications such as environmental surveillance, traffic monitoring, and target tracking [1-5]. Target tracking becomes a common and important application of wireless sensor networks to various areas (e.g., enemy detection and tracking on a battle field and surveillance of wildlife habitats). Existing research of target tracking is mainly focused on accurate estimation of target location, and classification for multi-targets [4]. On the other hand, there has been less research about efficient, reliable and robust reporting of tracking data from target tracked sensors to the base station [5].

Sensor networks are potentially exposed to communication failures due to limited energy and instability wireless links. Therefore the most of sensor nodes might experience high packet loss rates. According to [6] WSNs have low reliability of communication as over 30% of packet delivery failure rate under a practical condition. In order to cover large monitoring area, the sensor network is designed with multi-hop paradigm in large scare network due to sensor's short communication rage. And it brings critically out decrease of system performance. Target tracking data (i.e., computed fraction of trajectory of sensed target) from each source node has to arrive to the sink or the base station with as possible as higher accuracy and less communication delay. Efficient energy consumption is critical issue of wireless sensor network because sensor nodes have limited energy capacities. Existing in-network aggregation approaches are not suitable to the target tracking system. In multi-path approach, many numbers of packets of tracking information from each sensing node occur packet overflowing in the network. It brings about high energy consumption and long delay from waiting of channel access for communication. On the other hand, the tree structure based approach has one critical problem, such as robustness problem. Communication failure rate seriously affects to accuracy of reported target tracking information from sensor nodes

In this paper, we provide robust in-network data processing for target tracking data via a tree topology under unreliable wireless sensor networks. It provides a technique to maintain a minimal number of redundant data in the network. We address a tree structure based conditional multicasting technique with common ancestor search. It is based on inference of successful packet delivery rate in multi-hop sensor networks. It maximizes packet delivery rate of target tracking to the base station. Also conditional multicasting guarantees minimal extra transmission overhead. The common ancestor search avoids an increase of packet traffic due to redundancy by the conditional multicasting technique.

The rest of this paper is arranged as follows: Section 2 explains important background knowledge and related work. Architecture of WSNs for the mobile object tracking system is illustrated in Section 3. The conditional multicasting with common ancestor search method is proposed in Section 4. Simulation results are presented in Section 5. Finally, the conclusion is given in Section 6.

2. Background Knowledge
In object tracking system on densely wireless sensor distributed area, the system has to detect and sense the target of interest via the sensors, which equip acoustic, seismic, magnetometer, or infra-red (IR) device. Then the system estimates the target location and track in timely manner. However, for those processes, each wireless sensor generates large size of sensory data, which need to be transmitted to the base station. Therefore, the system experiences a high probabilistic of packet overflowing in the

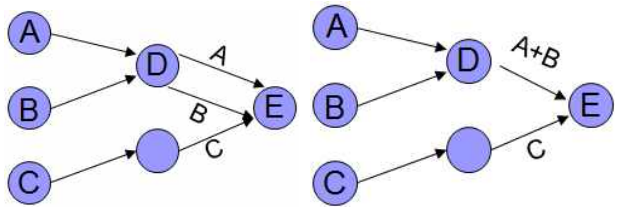


Figure 1. Sample of data processing methods. (a) Centralized data processing. (b) In-Network data processing.

network, and it brings on low energy efficiency due to re-communication to acquire complete packets.

Data processing can be classified to a centralized data processing and an in-network data processing by a role of data relayed node. In case of the central data processing, sensor node only transfers the sensed data to the base station as shown in Figure 1.(a). The base station processes and transforms data for the purpose of the system. This approach guarantees high accuracy and low network delay. But there is higher energy consumption due to more number of packet transmissions. On the other hand as in-network data processing, sensor node gathers data and transforms the data before it sent out the result in the network as shown in Figure 1.(b). This approach has energy efficiency due to transmission of small number of packet for the processed result. But, it have two major problems such as less accuracy and longer delay than the centralized data processing.

The purpose of in-network data processing is that only a minimal amount of data is transferred within the network to minimize traffic and power consumption. Data processing takes place in the network and the results are returned.

2.1 Rings Topology

When sensor nodes are deployed in the specific area, the base station or sink node sets up its own level 0, and it broadcasts a level setup message which includes the unique ID of the message sender and level information to its neighbor nodes, which locates within the base station's communication range. After the neighbor

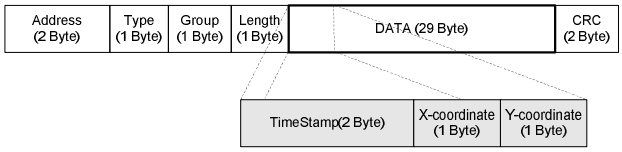


Figure 2. Modification of TinyDB data frame format for the tracking information

node receives the message, the node assigns its level by one greater than received level from the message sender. Then the node generates a new level setup message with its own ID and assigned level, and then broadcasts. Eventually, each node can receive a level setup message, and assign level. In Rings topology, the level denotes a minimal count of hops from the base station to each sensor node. And when each node reports sensory data to the base station, level i's node transmits data to only level i-1's node, and this scheme guarantees each reported data route minimum hop distance among the network to the base station.

2.2 Tree-based Approach and Multipath-based Approach

Tiny Aggregation (TAG) [7] is implemented in the TinyDB system and TinyOS which are optimized DB system and operating system for the sensor device, especially MICA motes. TAG uses in-network aggregation over a tree topology, and each non-leaf node performs in-network aggregation. TAG uses directed diffusion such as aggregating data from children and sending to parents. This method is the simplest and easiest use implementation. Moreover, TAG is a pretty energy-efficient algorithm in wireless sensor networks. But the TAG does not consider the robustness of topology. On the other hand, there are several multipath-based in-network aggregation algorithms such as Naïve ODI algorithm [8]. Naïve ODI algorithm is the most fundamental algorithm for multipath approaches. It is useful for in-network aggregation over any topology. One of its characteristics is duplicate-sensitive aggregation. However, Naïve ODI algorithm is not suitable for large-scale wireless sensor networks.

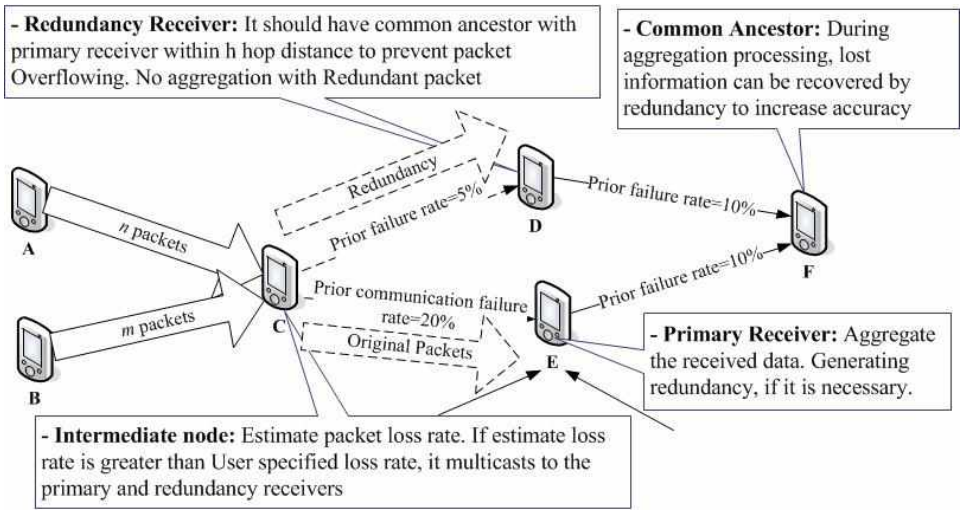


Figure 3. Example of Conditional multicasting with common ancestor search technique.

2.3 Hybrid Approach

Tributary and Delta algorithm [9] is a new approach to in-network aggregation using the advantages of tree- and multipath-based approaches. In the tributary region, the topology assures low or zero approximation error and short message size like tree approach. In the delta region, robustness is guaranteed as in multipath-based approach. Using the dynamical adaptation scheme, the wireless sensor networks' topology can be exchanged depending on the network's current loss rate. In Tributary and Delta, two topologies are dynamically exchanges depending upon the current message loss rate in the network. If the message loss rate is becoming higher, the delta region should be increased (expanding phase) for an increase in accuracy of data. On the other hand, if the loss rate is becoming lower, the tributary region should be increased (shrinking phase) for an increase in energy efficiency. This dynamical adaptation is decided at the Base station. In using the user-specified threshold, only the base station decides whether to shrink or expand.

3. ARCHITECTURE OF WSNs

In dense sensor network in target monitoring area, sensor nodes detect appearance of a target and estimate target's location with uses of intercommunicated neighbor sensors' sensing data within a distributed target tracking application. A subset of nodes which are sources of computed a fraction of trajectory of target in sensing area report tracking data toward the base station. Each intermediate node collects packet from its child nodes, and it accumulates received tracking data and filters out duplicate information in the process of aggregation. Using existed Medium Access Control protocol, each sensor node tries to avoid collision. And before the sender transmits packet to the receiver, the sender notices total packet numbers to the receiver using modified RTS in MAC protocol. There is not packet retransmission due to fast reporting to the base station. For the mobile object tracking information, each packet contains multiple target-tracking data, and each tracking data consists of time stamp and target location (X, Y coordinates) as shown in Figure 2.

4. PROPOSED DATA AGGREGATION METHOD

4.1 Conditional multicasting with common ancestor search

In order to obtain the goal of in-network data processing in the mobile object tracking system using WSNs, intermediate nodes accumulate data and remove redundant data to minimize amount of transfer data to next hop. Because the data processing budget is cheaper than the communication budget, where the budgets refer the cost of power consumption, in-network processing provides more energy efficient data gathering in the ideal condition. However, in practical condition, WSNs potentially inherent communication error, and it can occur serious inaccurate mobile tracking results. In this research, we propose tree structure based conditional multicasting method with common ancestor search process. For robustness and fault tolerant features of networks, the sensor node generates minimal redundant data depended on previous communication failure rate in network using multicasting method.

In the proposed conditional multicasting algorithm, there are several roles of deployed wireless sensors as shown in Figure 3.

```
For All Level l = L to 1
    For all node i
        If i.level = L and i does not have uncle, Then i uses current parent;
        Else If i.level=L and i has more than one uncle
            For All i_u
                If i_u.pin=1 & i_p.pin=1 Then i uses current parent
                Else Then
                    For All i_a & j=i_u
                        If i_a=j_a Then
                            i uses current parent;
                            For All i_a
                                If i_a.level>c_i,j level Then i_a.pin=1
                            For All j_a
                                If j_a level>c_i,j level Then j_a.pin=1
        If there is not common ancestor between i_p and i_u
            If h-2 = 0 Then
                Case 1;
                If OK=false Then
                    Case 2;
                If OK=false, Then i dose not satisfy the condition
            Else If h-2>0 Then
                Case 3;
                If OK=false Then
                    Case 4;
                If OK=false, Then i dose not satisfy the condition
```

//case 1

```
For All i_u & j=i_u
    For All j_u & k=i_p
        If k_p=ju & j.pin=0 Then
            j changes parent to k_p
            j.pin=k.pin=1
            updateAncestorSet(j)
            updateAncestorSet(j_c)
            OK=true
```

//case 2

```
For All i_u & j=i_u & k=i_p
    For all k_u
        If k_u=j_p & k.pin=0 Then
            k changes parent to j_p
            j.pin= k.pin=1
            updateAncestorSet(j)
            updateAncestorSet(k)
            OK=true
```

//case 3

```
For r=h-2 to 0
    For All i_a & m=i_a
        For All m_u & n=m_u & k=m_p
            For All n_u
                If n_u=k_p & n.pin=0 Then
                    n changes parent to k_p
                    n.pin=m.pin=1
                    updateAncestorSet(n)
                    Recursively updateAncestorSet(n_c)
                                        until n_c.level=i.level
                    Recursively pin(n_c) until nc.level = i_p.level
                    Recursively pin(m_c) until mc.level = i_p.level
                    OK=true
```

//case 4

```
For r=h-2 to 0
    For All i_a &m=i_a
        For All m_u & n=m_u & k=i_u
            For All k_a
                If n=k_a & m.pin=0 Then
                    m changes parent to n
                    m.pin=1
                    updateAncestorSet(m)
                    Recursively updateAncestorSet(m_c)
                                        until m_c.level=i.level
                    Recursively pin(n_c) until n_c.level = i_p.level
                    Recursively pin(m_c) until m_c.level = i_p.level
                    OK=true
```

Figure 4. Pseudo code for dynamic common ancestor search

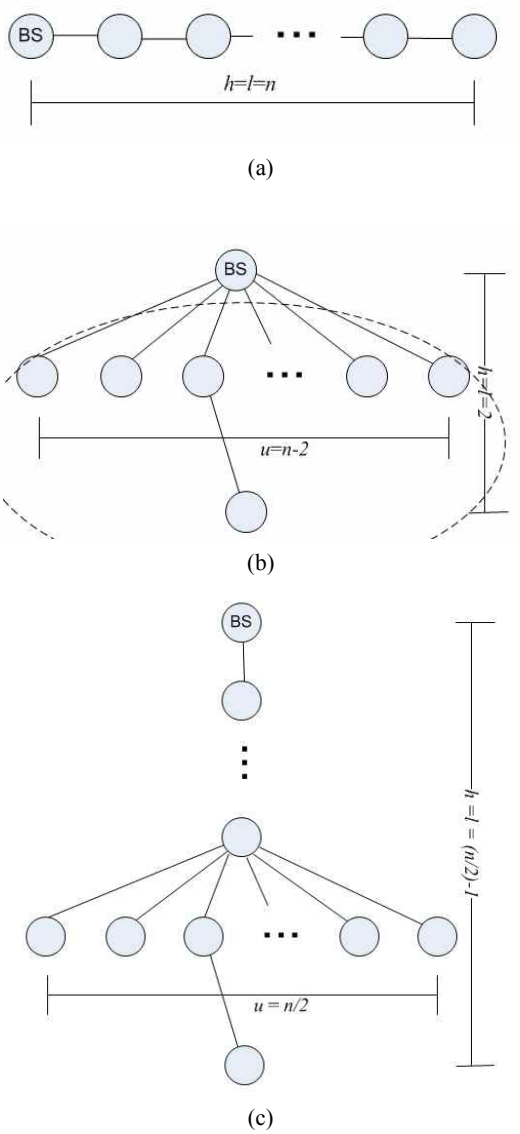


Figure 5. Cases of possible complexity of the proposed algorithm. (a) Liner sensor deployment. (b) Large number of uncles. (c)The number of uncle is the same as n/2 and h and l is equal to (n/2)-1

When the target sensed node transfers the packet to the base station through the tree, an originally assigned receiver (as the primary receiver) in next hop process data and multicast if it is required. Redundancy receiver does not process data and it only relay the redundant packets to its parent due to avoidance of redundant packet overflowing. Common ancestor recovers original data using redundant data, when the original data are experienced by communication failure.

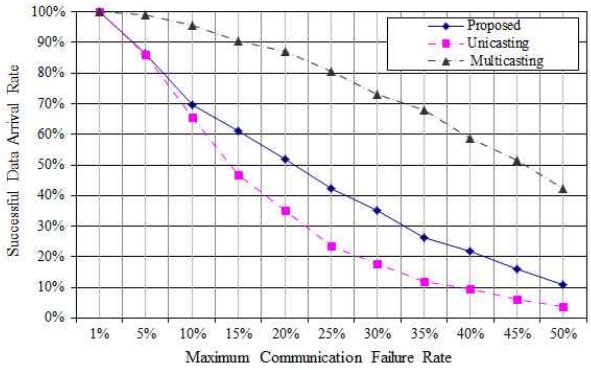
With the use of conditional multicasting, there is no consideration of the actual existence of a common ancestor within the tree. Also if a common ancestor is too far from the pair of nodes, redundant data also experience higher data loss under unreliable wireless sensor networks and the recovery portion should be very small. In the common ancestor search phase, the system builds a special tree for conditional multicasting. It is a type of Shortest Path Tree (SPT), and the SPT is a fundamentally used Rings paradigm. During a level setup process in Rings, every node broadcasts a level setup message one time. And every sensor node can receive the messages from all of its upper level neighbor nodes within communication range. If we assume every node knows its location information, and the level setup message contains the sender's ID, level, and location information, every node can select a parent node which has the shortest distance from the node. And eventually, every node chooses a parent and a shortest path tree is constructed. If a node has more than two upper level neighbors, this pair of upper level nodes has a common ancestor within h-hop distance from the node. Value of h is a given constant and it is the maximum relative hop distance between the common ancestor and the primary receiver. Figure 4 shows a pseudo code for the common ancestor search phase, where i_u denotes a uncle set of node i_s as upper level neighbor set, is an ancestor set within h-hop distance, i_p and i_c denote node i's parent and child node.

4.2 Analysis of the proposed method

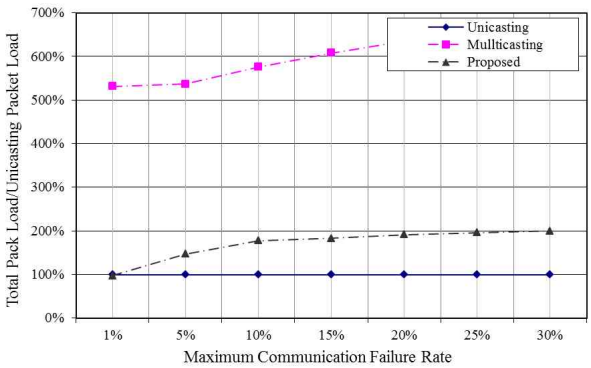
Complexity of the proposed common ancestor searching algorithm is the order $O(nlu^2h^4)$, where l denotes level of the node, n addresses a total number of nodes, u explains the number of uncle nodes, and h denotes the number of hops. Since l, u, and h are functions of n, the order can be of $O(n^8)$ which is seriously high complexity. In this section, we describe the actual possible cases and the upper bounds of running time for each case as shown in Figure 5.

First case is Linear deployment. If the nodes are deployed linearly like Figure 5.(a), h and l values are the same as the number of total nodes. However, in this case, the running time is only O(n) because, in proposed method, if a node has only one upper level neighbor which is a parent node in SPT, the node uses the current parent. Second case is there are large number of uncle nodes in the tree. If there is extreme deployment case such as a node has large number of uncles like Figure 5.(b), the upper bound of the worst running time is roughly $O(n^3)$ because h and l can be constants in this case. Third case is the actual worst case as shown in Figure 5.(c). If a node's u is the same as n/2 and h and l is equal to (n/2)-1, the upper bound of running time is O(n*(n/2)*(n/2)2*(n/2)4). Therefore the complexity is $O(n^8)$.

The proposed common ancestor searching algorithm is not realistic algorithm in wireless sensor network due to high complexity. Therefore the algorithm has to be modified to reduce the upper bound of the worst case running time. I propose two solutions. One is an adaptive common ancestor searching algorithm. The other is zone-based common ancestor searching algorithm. In the current common ancestor searching algorithm, h value is chosen only depending on communication link conditions. But in the adaptive common ancestor searching algorithm, h value is also related on the maximum level value (l) and the number of uncles (u). If (l*u) is greater than the number of total nodes (n), h is assigned the minimum value such as 2. With the use of the adaptive common ancestor searching algorithm, the second case has lower complexity, $O(n^3)$ than original common ancestor searching algorithm. Also I consider about zone-based common ancestor searching algorithm which is similar to a clustering scheme. A sensing area is divided uniform size of zones, and each zone has a head node which is all of pairs of every zone member nodes. Using the zone-based common ancestor searching algorithm, I aim to reduce the complexity less than $O(n^2)$ by using following constraint:



(a)



(b)

Figure 6. Performance comparison under grid sensor displayed network. (a) Successful data arrival rate under various communication failure rate. (b) Expected network traffics.

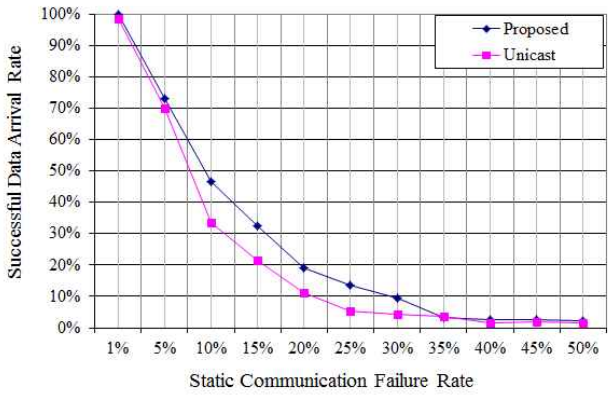
$$\sum_{i \in Set-of-Zones} (l_i^z * u_i^z * h_i^z) \le n$$

,

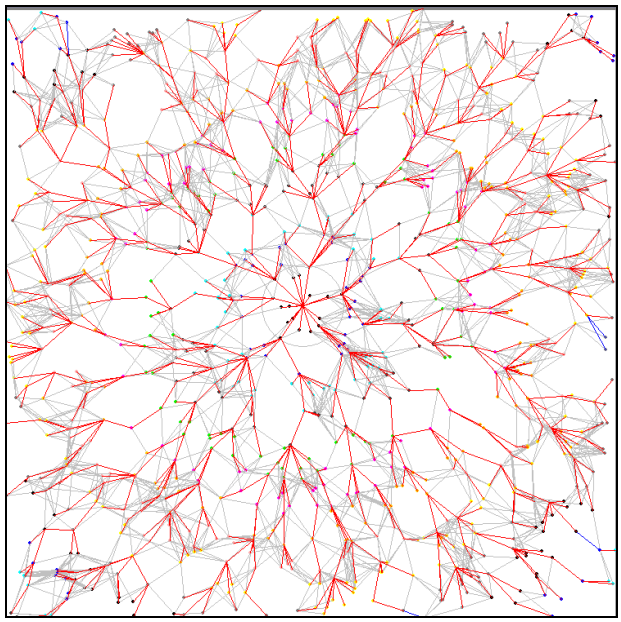
where l_i^z denotes the maximum level in zone i, u_i^z addresses the maximum number of uncles in zone i, and h_i^z explains the maximum possible hop-distance in zone i. In the zone-based common ancestor searching algorithm, a summation of $(l_i^z * u_i^z * h_i^z)$ of all zones is not greater than the number of total nodes, n.

5. ANALYSIS

We simulate and analyze of the proposed algorithm compared with unicasting and multicasting based in-network data processing algorithms. Total 1000 nodes are deployed in the 700m by 750m area and the base station locates at the center of the area. As shown in Figure 6, when the sensors are displayed with GRID manner, multicasting approach shows better rate of successful data arrival than the proposed and unicasting approaches. But multicasting approach brings on high traffic in the network as 600% compared with unicasting. But the proposed algorithm shows higher data arrival rate with less increase of network overload. Figure 7 illustrates comparison of data arrival rates between unicast and the proposed algorithms under random deployed network.



(a)



(b)

Figure 7. Performance comparison under random sensor displayed network.

6. CONCLUSIONS

We propose a Conditional Multicasting technique with dynamic common ancestor search algorithm for robust in-network processing for target tracking data via tree topology. Under varying conditions, the proposed algorithm provides a surely higher successful data arrival rate than the existing tree-based approaches. The algorithm guarantees maximized data recovery with use of redundancies and minimized existence of redundancies within networks. Also it employs minimal extra communication energy cost and minimal increase of traffic overhead.

7. ACKNOWLEDGMENTS

This work was supported by the DGIST R&D Program of the Ministry of Education, Science and Technology of Korea (13-BD-01)

8. REFERENCES

[1] B.R Son, H. Lee, J.E Kim, D.H Lee. A Mutual Position Detection based on Relative Motion Recognition of Multi-Modular Robots, IST 2012, vol 21, April 2012

[2] Ahmed, N.; Rutten, M.; Bessell, T.; Kanhere, S.S.; Gordon, N.; Jha, S.; , "Detection and Tracking Using Particle-Filter-Based Wireless Sensor Networks," Mobile Computing, IEEE Transactions on , vol.9, no.9, pp.1332-1345, Sept. 2010 doi: 10.1109/TMC.2010.83

[3] Samarah, S.; Al-Hajri, M.; Boukerche, A.; , "A Predictive Energy-Efficient Technique to Support Object-Tracking Sensor Networks," Vehicular Technology, IEEE Transactions on , vol.60, no.2, pp.656-663, Feb. 2011 doi: 10.1109/TVT.2010.2102375

[4] Qing Cao, Ting Yan, John Stankovic, and Tarek Abdelzaher. 2005. Analysis of target detection performance for wireless sensor networks. In Proceedings of the First IEEE international conference on Distributed Computing in Sensor Systems (DCOSS'05), Springer-Verlag, Berlin, Heidelberg, 276-292. doi =10.1007/11502593_22.

[5] Chih-Yu Lin; Yu-Chee Tseng; , "Structures for in-network moving object tracking in wireless sensor networks," Broadband Networks, 2004. BroadNets 2004. Proceedings. First International Conference on , vol., no., pp. 718- 727, 25-29 Oct. 2004 doi: 10.1109/BROADNETS.2004.78.

[6] Jerry Zhao and Ramesh Govindan. 2003. Understanding packet delivery performance in dense wireless sensor networks. In Proceedings of the 1st international conference on Embedded networked sensor systems (SenSys '03). ACM, New York, NY, USA, 1-13. doi =10.1145/958491.958493

[7] Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein, and Wei Hong. 2002. TAG: a Tiny AGgregation service for ad-hoc sensor networks. SIGOPS Oper. Syst. Rev. 36, SI (December 2002), 131-146. DOI=10.1145/844128.844142

[8] Philippe Flajolet and G. Nigel Martin. 1985. Probabilistic counting algorithms for data base applications. J. Comput. Syst. Sci. 31, 2 (September 1985), 182-209. doi =10.1016/0022-0000(85)90041-8

[9] Amit Manjhi, Suman Nath, and Phillip B. Gibbons. 2005. Tributaries and deltas: efficient and robust aggregation in sensor network streams. In Proceedings of the 2005 ACM SIGMOD international conference on Management of data (SIGMOD '05). ACM, New York, NY, USA, 287-298. doi =10.1145/1066157.10661

A Heuristic-based Vertical Handover Algorithm using MIH for Individual Users

Seokjoon Hong, Minchul Shin, Inwhee Joe
Division of Computer Science and Engineering
Hanyang University
Seoul, 133-791 South Korea
+82-02-2220-1088
iwjoe@hanyang.ac.kr

ABSTRACT
Current wireless networks include various wireless technologies such as 3G (WCDMA), 4G (LTE), W LAN, WiMAX. The movement of a user within or among different ty pes of networks requires vertical handover. There are many existing algorithms for predicting and executing vertical handover considering various network parameters. However, they usually don't consider the individual mobile user. Generally, an individual mobile user has a unique moving and using network pattern in a day . So, if we consider this personal moving a nd using pattern, we can predict the next handover network. Thus , we propose a heuristic based vertical handover prediction a nd MIH (Media Independent Handover) based handover algorithm for individual user. By using this algorithm, we can reduce handover preparation delay and prevent unnecessary handovers. Also, we provide proof of the performance of the proposed algorithm via OPNET simulation results.

Categories and Subject Descriptors
C.2.3 [**Computer Communication Networks**]: Network Operations – *network management, network monitoring.*

General Terms
Algorithms, Management

Keywords
Vertical handover, individual mobile user, pattern, MIH

1. INTRODUCTION
Mobile communication has become more popular due to the increased availability of portable devices and advanced wireless technology. Moreover, the core network of heterogeneous wireless access networks, e.g. W LAN [1], WiMAX [2,3], and 3GPP (WCDMA, LTE); they are evolving into all-IP based network.

The IEEE 802.21 has proposed the Media Independent Handover (MIH) services [5] to enhance the handovers across heterogeneous access networks, i. e. vertical handover, and to optimize the service (or session) continuity during handovers, i.e. seamless handover. For this reas on, MIH provides generic link layer intelligence and other related network inform ation to upper layers. Particularly, MIH offers a framework of the message flows between handover-related entities to provide inform ation on handover candidate networks and to deliver handover commands.

The information service provides a framework and corresponding mechanisms by which a MIH function entity can discover and obtain network information existing within a geographical area to facilitate the handovers. The inform ation service primarily provides a request/response type of mechanism for information transfer. The inform ation may be stored within the MIH lay er or m aybe presented to som e information server from where the M IH layer can acces s. The inform ation service provides access to static info rmation such as neighboring networks, helping in network discovery. Also, the service may provide access to dy namic information which may optimize link layer connectivity with different networks. This could include link layer parameters such as channel inform ation, MAC addresses of the PoA (Point of A ttachment), security information, network type, operator identifier, service identifier, geographical location, etc.

It is clear that in current and future environments, dynamic context information from network side entities is very important for the vertical handover decision procedures. Context-aware media independent information server is also proposed for optimized seamless handover procedures [6]. The paper addresses a new concept of a context-aware inform ation server that is able to store, manage and deliver real-time dynamic information retrieved from the network and the term inal side entities, such as the user preferences, runni ng services, mobile nodes characteristics and available network resources.

Furthermore, for seamless a nd QoS guaranteed vertical handover, the network selection is critical and there are many proposed algorithms about that. A network selection in an integrated wireless LAN a nd UMTS environment using mathematical modeling and com puting techniques are proposed for integrated cellular/wireless LAN sy stem [7]. The proposed scheme comprises two pa rts, with the firs t applying an analytic hierarchy process (AHP) to decide the relative weights of evaluative criteria s et according to us er preferences and service applications, while the second adopts grey relational analysis (GRA) to rank the network alterna tives with faster and sim pler implementation than AHP. Th e proposed technique can effectively decide the optimum network through making trade-

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.4

offs among network condition, user preference, and service application, while avoiding frequent handoffs. A network selection algorithm considering power consumption in hybrid wireless network is another algorithm which apply AHP and GRA scheme. It considers not only QoS but also lifetim e of m obile node. If user preference is lifetime, the proposed algorithm selects the network that stays longer due to low power consumption [8].

Mobility prediction is also important in vertical handover. There are some previous works in the area of mobility prediction. Tabbane's proposal [9] suggests that the mobile's location may be determined based on its quasi-determ inistic mobility behavior represented as a set of movement patterns stored in a user profile. A pattern matching/ recognition-based mobile motion prediction algorithm (MMP) [10] is suggested which used to estimate the future location of the m obile. The paper treated the problem by developing a hierarchical user mobility model that closely represents the movement behavior of a mobile user, and that, when used with appropriate pattern matching and Kalman filtering techniques, y ields an accurate location prediction algorithm, hierarchical location prediction, which provides necessary information for adva nce resource reservation and advance optimal route establishment in wireless ATM networks.

Generally, an individual mobile user has a unique moving and using network pattern in a day. For example, a man gets up at his home in the morning, goes to his company through public places such as road, station (bus or train), etc. (Figure 1)

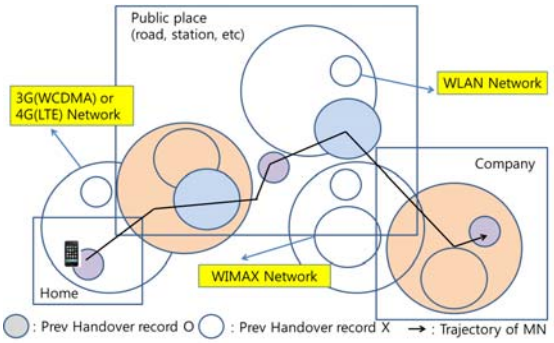


Figure 1. Vertical handover for individual mobile user

Therefore, if MN can rem ember the networks which us ed frequently by individual user and histories about handover then it also can predict the next handover network by user pattern.

In this paper, we propose a heuristic based vertical handover prediction by using MN stack memory and MIH based handover algorithm for individual user. In Section 2, we describe a heuristic based vertical handover predic tion process and MIH based handover algorithm. In Section 3, we evaluate the performance of the proposed algorithm using an OPNET simulation. Finally , we conclude in Section 4.

2. HEURISTIC-BASED VERTICAL HAND-OVER ALGORITHM USING MIH FOR INDIVIDUAL MOBILE USER

2.1 The heuristic-based vertical handover predication

In this section, we describe a heuristic based vertical handover prediction process for individual.

For predicting a handover, first, the mobile node maintains its special stack memory buffer for restoration of handover history. The contents of handover history are information such as time, location, direction as shown in Figure 2.

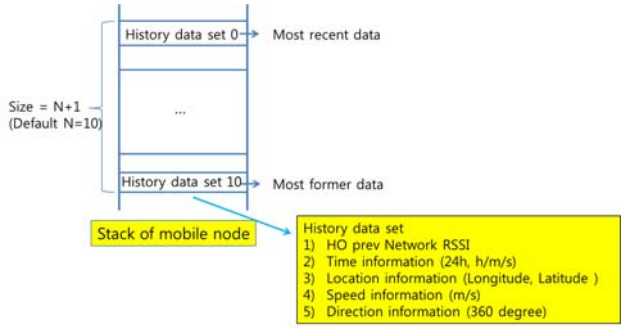


Figure 2. Stack of mobile node for restoring handover history

When current network's RSSI (receive s ignal strength indication)of MN is decreasing, 1 element pushed stack memory . If handover occurs, the stack memory buffer restores its buffer memory to handover history tables which include handover network and previous network. Figure 3 shows handover previous history points and handover execution point.

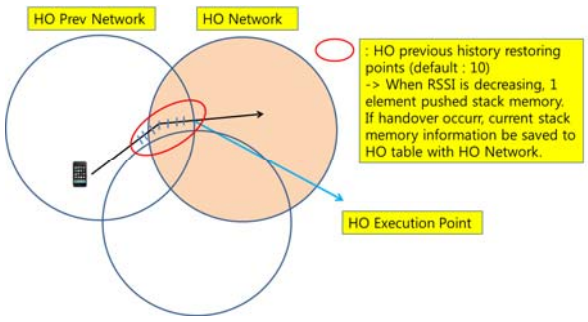


Figure 3. Handover previous history points and handover execution point

Table 1 shows example of handover table after mobile node handover from A network to B network. We also define two parameters which are utilization rate per week, U_w and normalized value of distance between current data of s tack memory and handover table, N_D for handover.

Table 1. Handover table (After MN handover from A to B)

Prev Net	HO Net	U_w	No	RSSI (Prev)	Time (h.m.s)	Location (Lon/Lat)	Deg. (360)	Speed (m/s)	N_D
A	B	20%	0	215	13.1.8	127.041/37.568	120	1.1	0.4
			1	210	13.4.53	127.042/37.567	140	1.2	0.5
			2	178	13.6.7	127.042/37.567	150	0.9	0.3
			3	140	13.10.30	127.042/37.566	140	1.0	0.7
			4	137	13.13.11	127.041/37.565	180	1.1	0.6
			5	109	14.20.13	127.040/37.564	200	1.0	0.7
			6	84	14.26.44	127.039/37.563	190	1.1	0.8
			7	55	14.30.15	127.037/37.562	180	1.2	0.1
			8	36	14.31.30	127.036/37.561	170	1.3	0.2
			9	23	14.31.35	127.035/37.560	160	1.1	0.3

The network utilization rate of network i in a week can be calculated by formula (1).

$$U_w(i) = \frac{\sum_{j=0}^{6} T_u(i,j)}{\sum_{k=0}^{6} T_R(i,k)} \quad (1)$$

$T_U(i,j)$: Average used time per week of i network on j day (0~6 means Sunday to Saturday)

$T_R(i,k)$: Residence time of i network on k day

And handover priority of the network is assigned according to U_w as Table 2. Higher priority of network has large value. If U_w is zero, then priority is also zero. This means that handover to the network should be preventive. On the other hand, if U_w is greater than 0.5(50%), then we can think that the network is frequently used by user and handover is preferred.

Table 2. Handover priority table

Uw (value or range)	Priority	HO Category
0	0	Preventive
0< Uw < 0.5	1	Normal
0.5 < Uw <= 1	2	Preferred

$$N_D(i,j,k) = \frac{D_{\max}(j) - X(i,j,k)}{D_{\max}(j) - D_{\min}(j)} \quad (2)$$

$$Avg\_N_D(i) = \sum_{j=0}^{4}\sum_{k=0}^{9} N_D(i,j,k) \quad (3)$$

We can calculate norm alized difference value and average normalized value of network i between current s tack data and previous handover table data by using formula (2), (3). And the difference range of each entity of MN is as Table 3. If the average normalized difference value of network i has lowest value among the candidate networks, then in all probability the network i,will be the next handover network.

Table 3. Difference between previous data of each entity

MN Entity (i)	Difference with history (D) range
RSSI(0)	0 ≤ D ≤ 255
Time(1)	0 ≤ D ≤ 3600 (second)
Location(2)	0 ≤ D ≤ 2R (R : radius of cell)
Degree(3)	0 ≤ D ≤ 360
Speed(4)	0 ≤ D ≤ 40 (m/s)

2.2 The mobile-initiated handover procedure using MIH

The mobile-initiated handover preparation procedure operates as follows and shown in Figure 4.

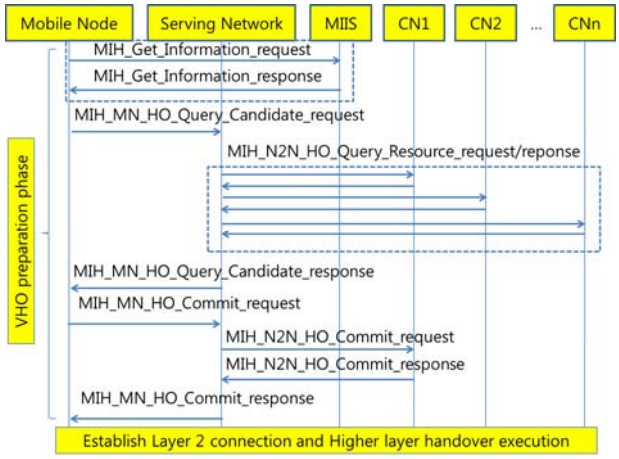


Figure 4. Existing MIH based vertical handover preparation signaling flow

2.2.1 Mobile-initiated HO procedure using MIH

1. The mobile node queries information about neighboring networks by sending an MI H_Get_Infomation request message to the Information Server.
2. The mobile node triggers a m obile-initiated handover by sending an MIH_MN_HO_Ca ndidate_Query request message to the Serving Networ k. This request contains the information of potential candidate networks.
3. The serving network queries the availability of resources at the candidate networks by sending an MIH_N2N_HO_Query_Resource_request message to one or multiple candidate networks.
4. The candidate networks respond with an MIH_N2N_HO_Query_Resources response message and the serving network notifies the mobile node of the resulting resource availability at the candidate networks through an MIH_MN_HO_Candidate_Query response message.
5. The mobile node decides on the target of the handover and notifies the serving network of the decided target network information by sending the MIH_MN_HO_Commit request message.
6. The serving network sends the MIH_N2N_HO_Commit request message to the target network to request resource preparation at the target network.

On the other hand, with our proposed vertical handover algorithm, a heuristic based ver tical handover prediction can take the place of MIIS functionality for searching neighbor network as shown in Figure 5.

So, vertical handover preparation delay of mobile node and network traffic load for communicating with MIIS can be reduced with the proposed algorithm.

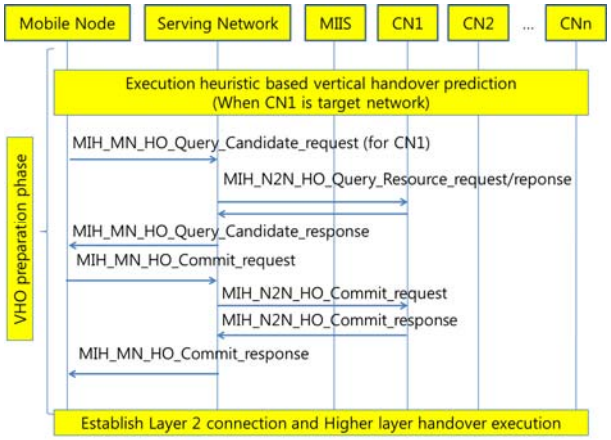


Figure 5. Proposed heuristic based vertical handover preparation signaling flow.

Figure 6 shows proposed vertical handover algorithms. It is composed of the heuristic base d vertical handover prediction and the mobile-initiated handover procedure using MIH.

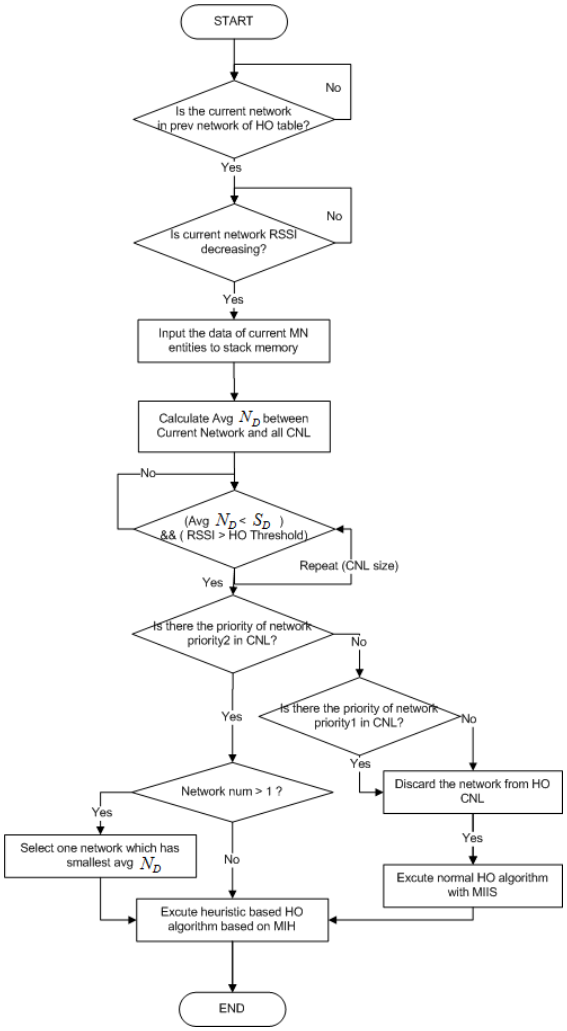


Figure 6. Total handover flowchart

If current network is in previous network of handover table and current network RSSI is decreasing, the process input the data of current MN entities to stack memory.

After that, it calculates average norm al difference between current network and all CNL (Candidate Network List). If the average normal difference is smaller than predefined s imilarity difference (default setting: 0.5), S_D and RSSI is greater than handover threshold, it checks the priority according to utilization of network.

If there is priority 2 network, the process executes handover to the network by using heuristic based handover algorithm based on MIH. And if there is priority 0 network (blacklist), the process discard the network among the candidate network list.

In other case, it execute normal handover algorithm based on MIH and MIIS server.

2.3 The energy efficiency scheme based on proposed handover algorithm

Because our proposed algorithm can predict the next handover network, it also can save battery consumption by turning on the modem for the handover network before handover and turning off the other network modem except the modem for current network after handover .

Figure 7 shows proposed handover based energy efficiency scheme flowchart.

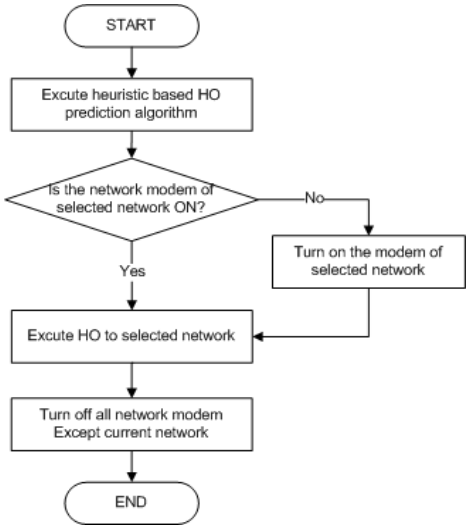


Figure 7. Energy efficiency algorithm with proposed HO algorithm

3. PERFORMANCE EVALUATION

In this section, we describe the performance evaluation of the proposed algorithm through simulation using the OPNET simulator. Figure 8 shows the network model of hy brid network that consists of various networks such as 3G(WCDMA), 4G(LTE), WLAN, WiMAX. And for simulating vertical handover, we suggested five trajectories of MN.

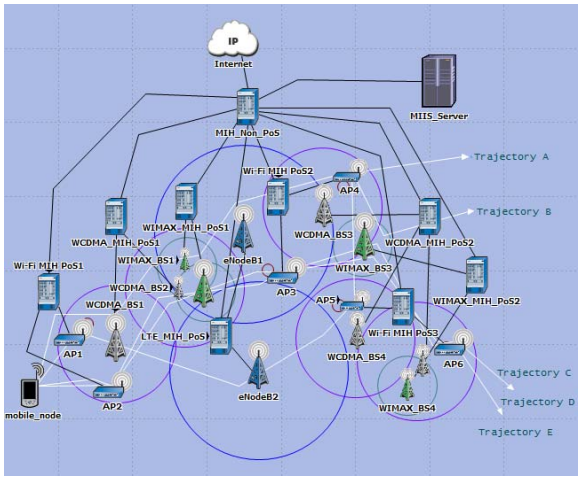


Figure 8. OPNET network model

Table 4 shows assumed values of the simulation parameters for OPNET simulation. And we als o assumed average power consumption using real modem specification reference [11-14].

Table 4. OPNET simulation parameters

Simulation parameter		Value
Average mobile node speed		10km/h
Cell radius (LTE,WCDMA, WiMAX,WLAN)		3, 2, 1, 0.1 km
Trajectory of mobile node (Only one trajectory at a time)		Uniform Distribution (A~E)
Link delay in network node		1ms for wired link (per link) No delay for wireless link
Handover perventive network (Blacklist, Utilization of network : 0%)		AP4, AP5, WIMAX1
Handover preferred network		All networks except Blacklist
MN battery capacity		2000 mAh
Average power consumption (LTE,WCDMA, WiMAX, WLAN)	Active	500mA, 150mA, 300mA 450mA
	Idle (Stanby)	65mA, 45mA, 160mA, 300mA

Figure 9 shows total handover numbers of the mobile node with number of replications in simulation. Because the existing MIH based vertical handover algor ithm considers receive signal strength (RSSI) mainly, the mobile node executes handover frequently in this network model.

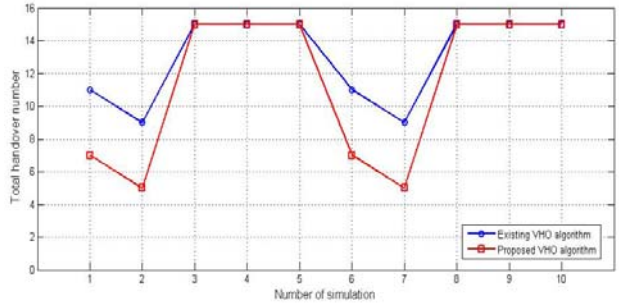


Figure 9. Total handover number of the mobile node with number of replications in simulation

However, the proposed vertical handover algorithm predicts and prepares the handover network by previous restored handover history table. It can also prevent unnecessary handover by using blacklist concept. Hence, the total handover number is smaller than the existing vertical handover algorithm.

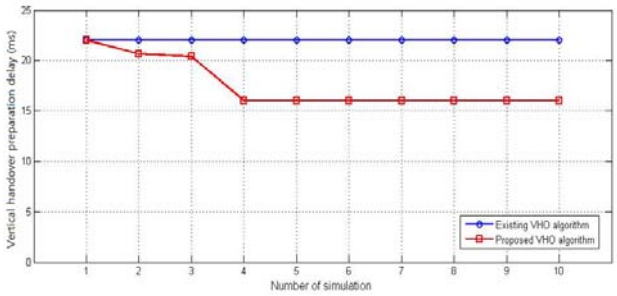


Figure 10. Vertical handover preparation delay with number of replications in simulation

Figure 10 shows vertical handover preparation delay with number of replications in simula tion. If MN uses the MIH based conventional vertical handover algorithm, MN needs to communicate with MIIS (Media Independent Information Server) for preparing handover.

But because MN can predict the next handover network for itself by using proposed algorithm, MN doesn't need to communicate with MIIS. Theref ore handover preparation delay can be reduced.

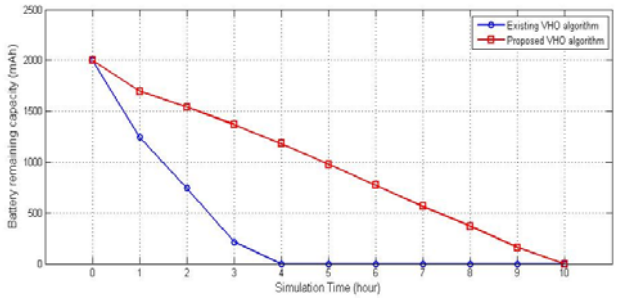


Figure 11. Battery remaining capacity (mAh)

And the last result, figure 11 shows power consumption of MN as increasing simulation time (hour).

If MN uses existing handover al gorithm, it consumes active mode power of current network and idle mode power of the other network. But if energy efficient applied handover algorithm is used, MN can save idle mode power by turning off the other network modem.

4. CONCLUSIONS

In this paper, we proposed a heuristic based vertical handover prediction process and MIH based handover algorithm for individual user. We also demons trated the perform ance of the proposed algorithm through simulation using OPNET simulator. The OPNET simulation results showed lower handover preparation delay as well as total handover number with the proposed algorithm than with the existing vertical handover algorithm. And it also can save battery power consumption by applying energy efficiency scheme to proposed algorithm. The results show that the proposed al gorithm is efficient than existing vertical handover algorithm.

5. ACKNOWLEDGMENTS

This work was supported by Basic Science Research Program through the National Research Foundation by Korea (NRF) funded by the Ministry of Edu cation, Science and Technology (2012-0005507).

6. REFERENCES

[1] IEEE Standard for Local and Metropolitan Area Networks, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Std, 2007, p. 802.

[2] IEEE Standard for Local and Metropolitan Area Networks; Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Std. 802.16-2004, 2004.

[3] IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, IEEE Std. 802.16e-2005, 2005.

[4] D. Johnson, C. Perkins, J. Arkkok, “Mobility Support in IPv6”, IETF RFC 3775, June 2004

[5] IEEE Standard for Local and Metropolitan Area Networks, Media Independent Handover Services, IEEE 802.21-2008,2009

[6] Pedro Neves, Joao Soares, Susana Sargento, Hugo Pires, Francisco Fontes, “Context-aware media independent information server for optimized seamless handover procedures”, Computer Networks 55, pp. 1498–1519 , 2011.

[7] Q. Song, A. Jamalipour,”Network selection in an integrated Wireless LAN and UMTS environment using mathematical modelling and computing techniques,” IEEE wireless communications, pp.42-48, June 2005.

[8] I. Joe, W.-T. Kim, and S. Hong. A network selection algorithm considering power consumption in hybrid wireless networks. IEICE Transactions on Communications, e91-b(1):314--317, 2008.

[9] S. Tabbane, “An alternative strategy for location tracking,” IEEE J. Select. Areas Commun., vol. 13, June 1995

[10] T. Liu, P. Bahl, and I. Chlamtac, “Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks,” IEEE J. Selected Areas in Comm., vol. 16, no. 6, pp. 922-936, August 1998

[11] Sierra wireless, AirPrime EM7700

[12] Xiamen Four-Faith Communication Technology, F8414 Zigbee+WCDMA/HSDPA/HSUPA IP MODEM TS

[13] C-motech, CBU-410s

[14] User manual. D-Link AirPlusWireless G PCI Adapter DWL-G510

Study of Keylogger Information Sniffing by Using Hooking Technology

In-woo Park
Dept. of Excellence Engineering
Hoseo Graduate School of Venture
+82-10-5274-3511
cowboyiw@hanmail.net

*Dea-woo Park
Dept. of Excellence Engineering
Hoseo Graduate School of Venture
+82-10-8299-4455
prof_pdw@naver.com

ABSTRACT
Hackers use the keylogger technology to hack Internet banking user's PC. When the user enters personal information with a computer keyboard, they steals personal information through sniffing by means of hooking technology. This study aims to analyze hacking attacks by using the existing Netbus, and to code an attack program for designing a prior process and hooking. This study also aims to prove personal information hacking by sending social engineering e-mails to install a hooking program in user's PC and to sniff personal information by the keylogger. Another aim of this study is to compare and analyze the aforementioned hacking with the existing hacking technology. This study will contribute to developing hacking attack technology, and technology to protect national cyber security against hacking attacks and defense.

Categories and Subject Descriptors
K.6.5 [**Management of Computing and Information Systems**]: Security and Protection (D.4.6, K.4.2) – *Authentication, Invasive software (e.g., viruses, worms, Trojan horses), Physical security, Unauthorized access (e.g., hacking, phreaking)*.

General Terms
Security.

Keywords
Netbus, Keylogger, Keylogging, Hacking, Hooking, Sniffing.

1. INTRODUCTION
The Internet Infringement Trend and Monthly published in August, 2012, by KISA reports account information leak by the keylogger technology in some online gaming sites (World of warcraft, Dungeon & Fighter, Mabinogi, Maplestory, etc.) to hack personal accounts and incur financial damages [1].

On April 25, 2005, a hacker attacked Internet banking user's PC for the Foreign Exchange Bank by means of the keylogger technology, called NetDevil, to withdraw 50 million won [2].

Now, the keylogger technology and hooking is modified into a malicious code, 'Citadel', for stealing money to threaten cyber security [3].

* Corresponding Author

Now, the keylogger technology and hooking is modified into a malicious code, 'Citadel', for stealing money to threaten cyber security [3].

The analysis of the event reveals that the keylogger technology is to hack by sniffing when a user enters personal information through a computer keyboard. Hackers can sniff specific program ID, password, contents of e-mails, victim's confidential documents, and the details of dialogue windows by using the keylogger technology. The keylogger technology followed by financial damages through the attack can be fatal to victims.

Therefore, it is necessary to study the hooking technology by the keylogger technology and to take secure measures in order to ensure safe electronic financial transactions involving Internet transactions, and ensure national cyber security.

This study aims to analyze Netbus which is a hacking program generally used, and to design and code a program by means of the LIFO(Last in First Out) architecture against the keylogger by using the hooking technology. It is intended to transmit e-mails through social engineering, to install program in the background for sniffing personal information. This process is tested and analyzed.

This study is composed of: I. Introduction describes the necessity and configuration of this study; II. Related studies describes the hooking technology and the keylogger technology; III. Designing hooking analysis and coding hacking program; IV. Analysis of sniffing keylogger information; V. Conclusion draws a conclusion and suggests future studies.

2. Related studied
2.1 Hooking technology
The hooking technology is to intercept the information entered by means of a computer keyboard, and is a hacking technology to intercept the information between the keyboard and a computer body, unlike existing computer viruses or hacking which access computer hardware bodies to steal information.

Various types of hooking is used depending on used methods and targets, and hooking is divided into message hooking and code hooking. The code hooking includes API hooking, native API hooking, and interrupt hooking. The filter driver of the device driver is associated with hooking [4].

● User-Mode hooking

• IAT (Import Address Table) Hooking: this is to change the API address on IAT into hacker's own function address, and to return the original API address at the end of its own function as shown in Fig. 1. This is a method the most generally used by computer viruses [5].

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.5

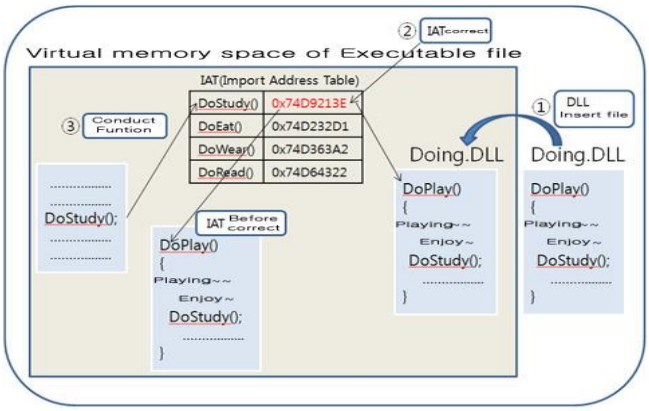


Figure 1. IAT Hooking process

• Inline Function Hooking (Detour Hooking) : hackers change the first 5 bytes of the API to be used into the code for Imp to their own function address, modify the changed code of the original API in their code, and return it to the API start location. It is not easy to detect it because it is more intelligent than IAT hooking.

● Kernel-Mode Hooking (root kit)

• SSDT (System Service Descriptor Table Modification): hackers change the address indicated by SSDT to the address of the hooking function, and return the address of the original kernel API after invoking the function as shown in Fig. 2 [6]. This is a method more than 50% of root kits use, and is used the most for concealing processes and files.

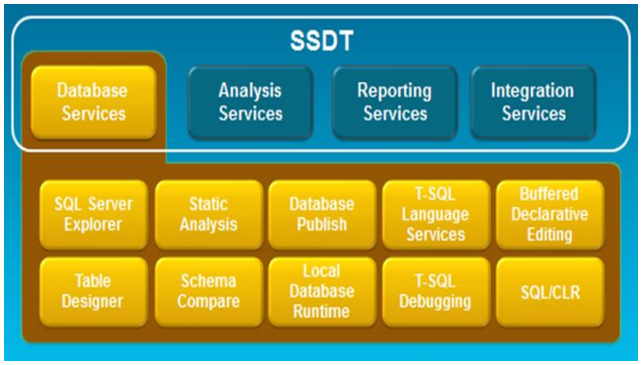


Figure 2. SSDT method

• DKOM (Direct Kernel Object Modification) :

This is a method of concealing the process executed by operating the kernel object, thread, service, port, drivers and handle entries in the execution list (PsActiveProcessHead, PsActiveModuleHead).

• SYSENTER: INT 2E (for Windows 2000)/ SYSENTER is used to go from the user mode to the system call. The subsystem service handler is stored in the registry IA32_SYSENTER_EIP. Hackers install the kernel driver to modify relevant values to invoke a root kit and then to return the original value.

• Filter Device Driver: this is a method of registering a filter device driver on the lower part of a security product. This is executed before any anti-virus vaccines because it is loaded in the boot time.

• Runtime Detour Patching : this is a method of operating the kernel memory to make the memory point indicate a root kit and then to hook kernel functions. For example, hackers write the address representing them for the IDT register for causing exception and controlling the Exception Handle to implement hooking.

• IRP table Modification : the dispatch routine for controlling the I/O Request packets used when a device driver processes network packets or writes files is stored in DEVICE_OBJECT. The root kit used by viruses can select the location of DRIVER_OBJECT in DEVICE_OBJECT by using the API IoGetDeviceObjectPointer. That is, they first execute their own root kit before other Original Driver Call is operated to manipulate the call result.

2.2 Keylogger technology

The Keylogging technology is to intercept and record the details entered by users in their PC by means of a keyboard by stealth. Keylogging can be carried out by hardware or software, electronically, or by using sound technology. Fig. 3 shows how hackers steal user's information through keylogging.

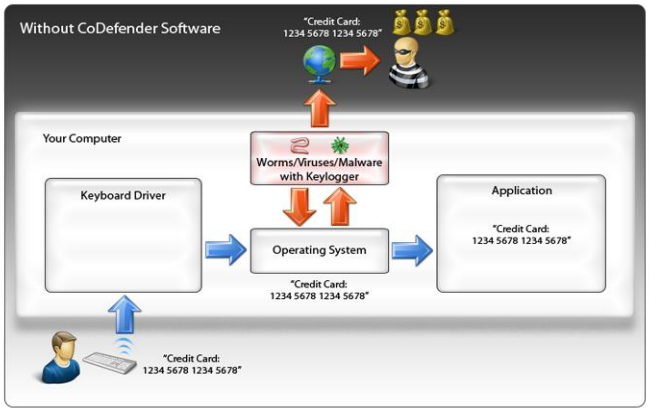


Figure 3. Keylogging process

Software type keylogging is that hackers reside in programs to save information, and hardware type keylogging is to save information through separate hardware device.

2.3 Sniffing technology

The sniffing technology is to eavesdrop on other users' packet exchange on a computer network. Sniffing is a threatening attack in an environment in which the network is shared in web hosting or the IDC (Internet Data Center). When a system is attacked, it is possible to use the system to eavesdrop on and identify other users' ID and password.

Sniffers can receive packets broadcast across the entire network of the Ethernet [7].

Packet sniffers use network monitoring tools, for example, tcpdump, snoop, or sniffer to analyze packets transmitted and received across the network to identify information. They can analyze all external and internal hosts connected to the network.

Sniffers attack systems vulnerable to security to get the ROOT right, and then install a backdoor, install and execute network monitoring tools which can intercept the first 128 characters of the sessions ftp, telnet and rlogin.

The best method of avoiding sniffing is to encrypt data [8].

3. Analyzing and designing hooking and coding hacking program

3.1 Analyzing and designing hooking and coding hacking program

Netbus is a type of hacking programs, and an example of computer viruses which can install a backdoor to manipulate other people's computer in a remote place.

This is a hacking program that hackers can carry out remote control through the backdoor when Netbus is executed in user's infected computer.

Netbus can control activities of a targeted computer to be hacked, and can force the targeted computer to stop its keylogging and its operation.

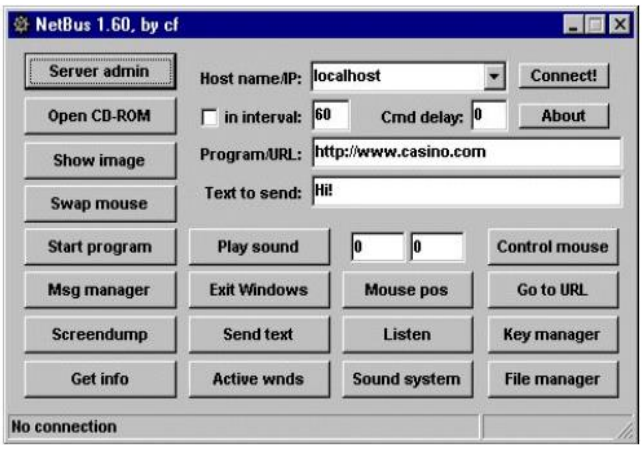


Figure 4. ExecutingNetbus.

The principle of Netbus is hacking and Troy which is a malicious program to drain infected computer information to the outside and sniff user's information. Once executed, it is automatically executed each time Windows is rebooted. The reason is that this file is recorded in the registry HKEY_LOCAL_MACHINE/SOFTWARE/Microsoft/Windows/ CurrentVersion/Run which is a part of the start program loaded while a computer starts concurrently with execution thereof.

3.2 Designing prior process to Netbus

Netbus is divided into a server program and a client program. Where the server patch.exe (Troy) is embedded and executed in a hacked computer, it is automatically executed each time Windows is rebooted. Design the client program Netbus.exe so that it can manipulate the hacked computer.

However, the hooking procedure features an LIFO (Last In First Out) architecture like a stack. If so, the hooking program on top executed last receives input first to operate before the existing Netbus operates to sniff information..

3.3 Coding program for hooking

The designed attack program executes the procedure before the existing hooking program is executed to hook keylogger information. If the hooking procedure is executed, it operates before other hooking functions as a hooking chain operate.

As shown in Fig. 5, the attack program coded in this study has the right for information sniffing and control before the existing Netbus program is executed to own the right of attack.

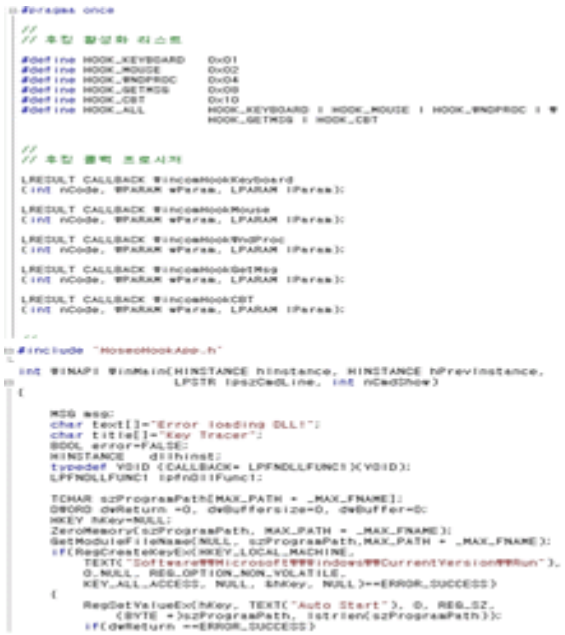


Figure 5. Coding source for attack program.

4.1 Installing hooking program

Send a social engineering e-mail to a user to be attacked. Attach files of alumni name and phone number list and a list of staffs for promotion to the e-mail. Once the user click the list of attached files, the attack program coded in this study is concurrently and automatically installed.

When the attack program is executed, the values is concurrently registered in the registry as shown in Fig. 6. The program is also executed in the background so that the user cannot recognize the installation process in his/her computer.

Figure 6. Registering attack program in registry

4.2 Sniffing personal information by keylogger

Operation of the hooking program installed in the e-mail user's computer enables the attacker to sniff user's personal information on attacker's computer screen through keylogging.

The hacker sniffs attacked victim's personal information on the Notepad through keylogging of the hooking program, as shown in Fig. 7.

Figure 7. Sniffing personal information through keylogging

4.3 Analyzing sniffing information

The information of the ID and password entered for the site where the attacked user logged in is sniffed by means of Notepad and transferred to the attacker. As shown in Fig. 8, attacked user's personal information is intercepted by the attacker

Figure 8. Collecting personal information

4.4 Comparing with existing Netbus attack

A comparison of performance was made between Netbus which is an existing hacking program and the hooking program.

As shown in Fig. 9, keylogging was made from attacker's computer by using the Netbus program which is an existing hacking program. The information in the domain entered on the left side of NAVER is hooked with the existing Netbus program.

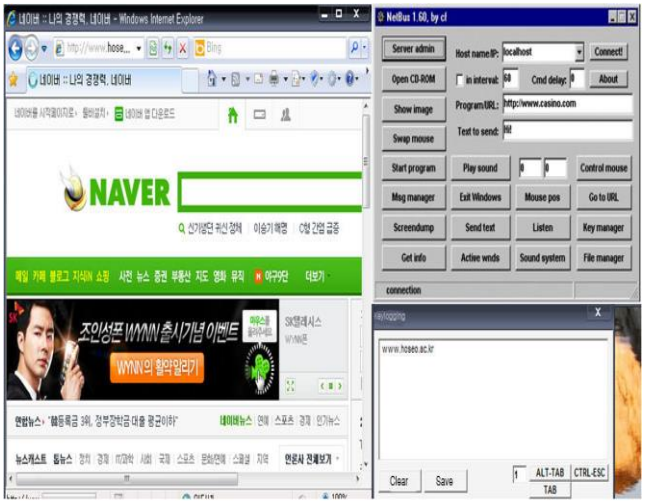


Figure 9. Keylogging by existing Netbus.

Fig. 10, when the attack hooking program coded in this study is operated, it sniffs user's keylogging information before Netbus to prevent the personal information from being transferred to the attacker.

That is, while the hooking program is operated with the message "Filtering Keyboard." in the center of the current NAVER screen, the keylogging information is not hooked to the existing Netbus window, and goes to the same Notepad.

As a result, the advantage is proved that the right of keylogging control is first acquired before the existing Netbus hacking program takes it.

Figure 10. Suggested attack program operates before Netbus keylogging operates.

5. Conclusion

Personal information including ID, password and account numbers are stolen by means of keylogger technology in online gaming or online financial transactions. However, since the hooking attack program is executed in the background, users cannot recognize the process.

This study analyzes the existing Netbus to design and code an hacking attack program to first acquire the right of control. A process was executed to infect the attacker through social engineering e-mails and to sniff user's personal information. As a result, the analysis revealed that the keylogging and hooking attack technology is improved, and the right of controlling attacked users is acquired.

It is necessary to further study technology for stopping operation of the sniffing process for personal information and backtracking attackers if a keylogger is detected.

6. REFERENCES

[1] Korea Internet & Security Agency. 2012. Internet Hacking Trend and Analysis Monthly in August 2012. *Monthly report, Korea Internet & Security Agency*.

[2] Huge amount of money stolen by first Internet banking hacking. http://news.naver.com/main/read.nhn?mode=LS2D&mid=sec&sid1=101&sid2=259&oid=001&aid=0001019113.

[3] Warning to users against banking information stealing malicious code 'Citadel', http://www.kcsnews.co.kr/news/articleView.html?idxno=6314.

[4] C.H.Lee. 2011. API programming. C.H.Lee, Korea.

[5] S.J.Hwang, G.H.Park. 2011. A Keyboard Security Method Based on a Subclassing. *Article of Korea Multimedia Society*. 15-23.

[6] S.W.Kim. 2007. Destructive Power of Hacking. S.W.Kim, Korea.

[7] Yin H, Poosankam P, Hanna S, Song D. HookScout. 2012. proactive binary-centric hook detection. *Proceedings of the 7th conference on detection of intrusions and malware & vulnerability assessment*. Yin et al, Jult.

[8] Jinpeng Weia, Calton Pub. 2012. Toward a general defense against kernel queue hooking attacks. *Computers & security*. 176-191. Atlanta.

Word Insect (butterfly) Robot in Link Structure for Nuri Course

Young-Suk Park
Dept. of Excellence Engineering
Hoseo Graduate School of Venture
Korea
+82-10-2510-1004
melisa02@hanmail.net

Jae-Han Shin
Ministry of Education
Science and Technology
Korea
+82-10-9372-2619
han3645@mest.go.kr

*Dea-Woo Park
Dept. of Excellence Engineering
Hoseo Graduate School of Venture
Korea
+82-10-8299-4455
prof_pdw@naver.com

ABSTRACT
The Nuri course is a common course for 5-year-old children's education and childcare developed by the Ministry of Education, Science and Technology in 2012. It is to use educational robots for early learning and creativity education of children who will be national human resources and the leader in the future and the age of high-tech technology. It is thus necessary to develop educational robots in duplex link structure to be used in the Nuri course and to study how to produce them. This study aims to develop and produce a butterfly robot which is an educational insect robot for the Nuri course in link structure. That is, the aim is to design the link structure of the butterfly robot, and an operating program, and to produce a body controller, the body and wings. This study will contribute to national development in the field of education through early learning and creativity education for national human resources, and to improving national competitiveness through innovated educational robots of Korea.

Categories and Subject Descriptors
K.6.5 [**Management of Computing and Information Systems**]: Security and Protection (D.4.6, K.4.2) – *Authentication, Invasive software (e.g., viruses, worms, Trojan horses), Physical security, Unauthorized access (e.g., hacking, phreaking)*.

General Terms
Security.

Keywords
Nuri Education Course, Link Structure of the Robot, Duplex Link Structure, Butterfly Robot, Robot Education, creativity education.

1. INTRODUCTION
The strength of a nation depends on global competition among nations by super-high speed communication networks and free trade competition in the 21st century. The Korean government recognizes that education for improving scientific thinking power and creativity while children have experiences and operation educational objects in the field of children's education for early learning of national human resources and in order to raise creative human resources.

Creativity education is recognized as a method of solving various educational issues resulting from knowledge education, and of promoting the independent role in the changing world.

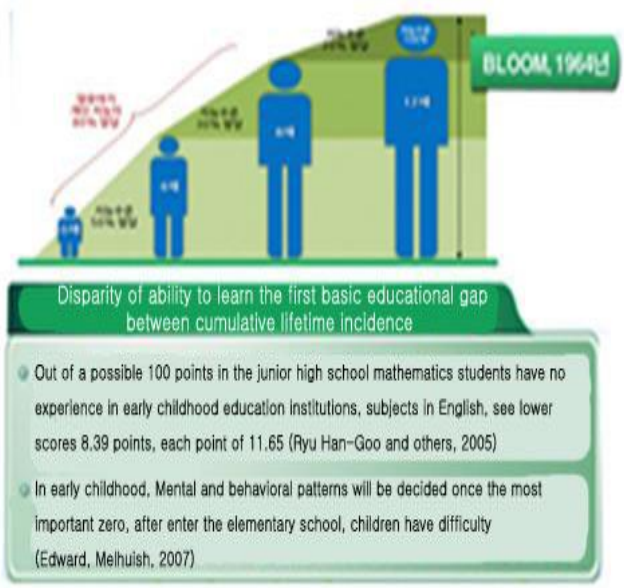


Figure 1. Importance of 5th Education· Childcare.

As shown in Figure 1, the Nuri course promoted by the Korean government is a governmental policy for creativity education, and the educational robot is promoted as one of creativity education policies.

The teaching method and the learning method in the Nuri course include the child-directed principle, the play-centered principle, the interest & absorption-centered principle, and the mutual action principle.

The insect robot in link structure is for children's activity with the robot produced by them in this respect. Creativity education by means of the insect robot facilitates understanding the motion of machines around us, and the insect robot is an integrated teaching tool so that children can investigate the interesting world of insects, play together with other children for interaction.

* Corresponding Author

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.6

Therefore, it is needed to study how to develop and produce educational insect robots in link structure for the Nuri course.

This study is composed of : 1. Introduction describes the necessity of this study; 2. Robot education for children; 3. Designing link structure and operating program for the butterfly robot for children to be interested in the insect robot; 4. Producing body controller for the butterfly robot body and wings, comparing and analyzing how the insect robot contents are applied to children's creativity; and 5. Conclusion and future studies.

2. RELATED STUDY

Therefore, it is needed to study how to develop and produce educational insect robots in link structure for the Nuri course.

2.1 Childcare Act

The Childcare project aims to raise infants and children whose parents cannot look after them due to their work, diseases or other reasons as a healthy member of the society by protecting them and desirable education. It also aims to contribute to improving domestic welfare so that the parents can carry out smooth economic and social activities as well [1].

The Childcare project is based on the Childcare Act, and individual and specific matters are specified in the Enforcement Ordinance (presidential decree) and the Enforcement Rule (decree of the Ministry of Health and Welfare) thereof [2].

2.2 Nuri Course

The 'Nuri course for 5-year old children' is a course that enables 5-year-old children enrolled in one of kindergartens and preschools to learn the same contents wherever they are enrolled. The early childhood education and childcare for 5 year-old children is currently divided into kindergartens and preschools, but the Nuri course is a new common course by integrating the two courses. Parents can select any one of kindergartens and preschools for their children's education [3].

After introducing the Nuri course for 5-year-old children, one-day operation time in kindergartens and preschools which varies in each institution is organized with 3 to 5 hours for the standard course and the time for self-controlling courses. Directors of preschools and kindergartens can flexibly organize the time for self-controlling course to be ideal for their institution. It is expected that kindergartens and preschools will enhance parents' satisfaction with respect to education and childcare service through flexible operation focusing on parents [4].

2.3 Educational Robot

The educational robot is divided into learner's robots and teaching robots. The learner's robot is used for education, and the teaching robot provides educational contents to play the role of a teacher [5].

The educational robot is an intelligent robot which is divided into 'teaching assistant robots' which assist teachers or communicate with users for teaching, and 'teaching aid robots' which are produced by users for creativity education [6].

Examples of the teaching aid robot include 'Mindstorms' available from LEGO of Denmark, 'HUNAROBO' from SRC, 'OLLO' from ROBOTIS, and products from ROBOROBO of Korea. Yoojin Robot recently released another exemplary robot for English education [7].

3. DESIGNING EDUCATIONAL INSECT ROBOT IN LINK STRUCTURE FOR NURI COURSE

The educational insect robot for children to which IT fusion technology was applied was designed while fully reflecting the contents of education in the Nuri course so that children can lead their play, have activity and learn from their play.

3.1 Analyzing educational insect robot for Nuri course

We selected the subject, the sub subject and key contents of education, to be connected to an ideal robot, and included them well in the educational activities. The activities were configured and applied so that children could analyze insects seen every season in order to include the educational contents of body movement, health, communication, social relationship, art experience and investigation of nature, which are 5 areas of the Nuri course together with the contents of creativity and humanity education.

3.2 Designing educational insect robot in link structure for Nuri course

The complex motion of machines is analyzed to configure and design it as a simple combination of motions. The link structure which is a motion converter, one of useful and general devices, is used so that children can learn the mechanism of leg motion of insects, and assemble robots of insects easily observed around us for children's respect to lives, interest and the spirit of inquiries.

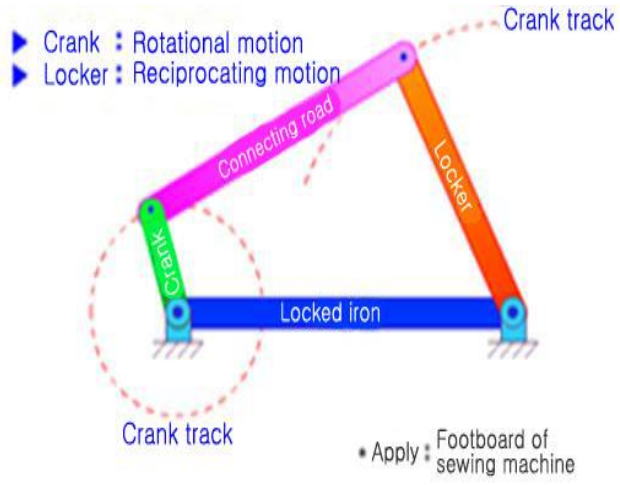


Figure 2. Link Structure by Crank-Rocker.

The above Fig.2 shows a link structure in which the rocker implements reciprocation led by the rotation of the crank. This principle of link structure is used to convert rotation of a motor to reciprocation of legs to assemble the legs of insects for operation. Children can be interested in and easily understand the complex structure of machines.

Fig.3 shows an assistant game developed to support educational activities in the Nuri course, which looks like a game board of a butterfly robot to be used for teaching and learning. This game board can improve teacher's teaching efficiency and convenience by learning focusing on self-led activities.

Figure 3. Game Board of Robot (Butterfly).

3.3 Designing program for operating insect robot

Operation of an insect robot is triggered by pressing the button Start to select a desired program, and the change of the program selection mode is identified with sound from Do to La which is a 6-tone scale.

* The selection mode 'Do' is to select the program edited by a user: this is not to use a program already saved, but the mode in which a child programs a desired one in the GUI environment. An R-logic program chip is produced as a card for programming exercise in the card-play manner in order to execute the program in a computer as the child has learned in the card-play.

* The selection mode 'Re' is a free mode: a mode for movement in the forward, backward, right and left direction without any operation.

* The selection mode 'Mi' is a remote control mode: enables a child to control the robot with a radio remote controller.

* The selection mode 'Fa' is a stoker mode: a mode to enable the robot to follow the direction sensed by a IR sensor equipped in the robot.

* The selection mode 'Sol' is an avoid mode: a mode to avoid obstacles.

* The selection mode 'La' is a LineTracer mode: a mode to enable the robot to follow the black line as shown in Fig.4.

Figure 4. LineTracer of Robot (Butterfly).

4. PRODUCING EDUCATIONAL INSECT ROBOT (BUTTERFLY) IN LINK STRUCTURE FOR NURI COURSE

4.1 Objective of producing educational insect robot for Nuri course

The objective is to suggest a method of integrating the link structure which is a basic structure of motion conversion and gaming with children's learning with insect robots for children's interest and fun in order to enhance educational effects.

1) In the step of recognition, suggest various activities to lead children's thinking into the activity subject.

2) This is intended so that children can examine and creatively solve problems for themselves to enhance creative problem solving capability.

3) Introduce activity through expanded activities and the fun of gaming into learning by robots, and guide children to produce and program a robot for themselves. This aims for robot control and problem solving, creative robot programming design and execution to enhance the spirit of challenge and investigation.

4.2 Producing educational insect robot of link structure for Nuri course

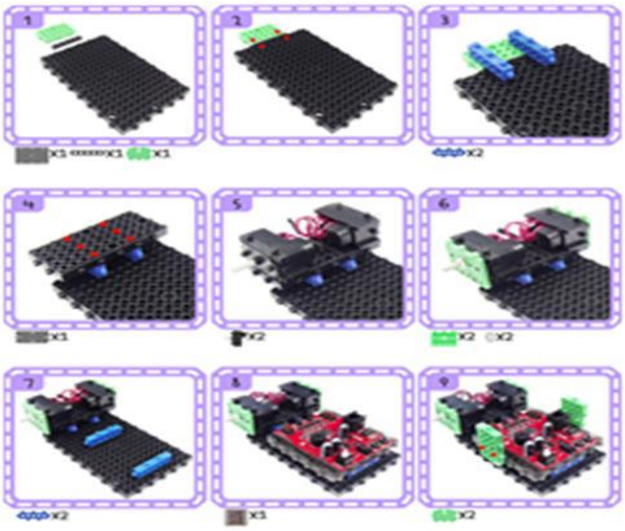


Figure 5. Producing Body Controller of Robot (Butterfly).

Provide a printed plan for the educational butterfly robot body. Design, produce and provide blocks to provide a motor and a robot controller.

Train children in advance to carefully handle the butterfly as if it is a live insect and dies if it is carelessly handled.

Teachers and children assemble the butterfly robot blocks to connect the motor and the controller while checking the plan on the book as shown in Fig.5.

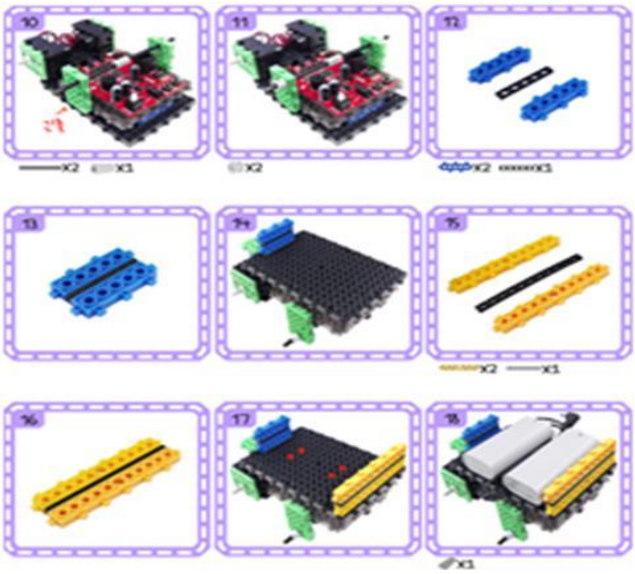


Figure 6. Producing Body of Insect Robot (Butterfly).

Connect the shaft for connecting the legs of the link structure on the abdomen of the butterfly, and equip the battery case on its back.

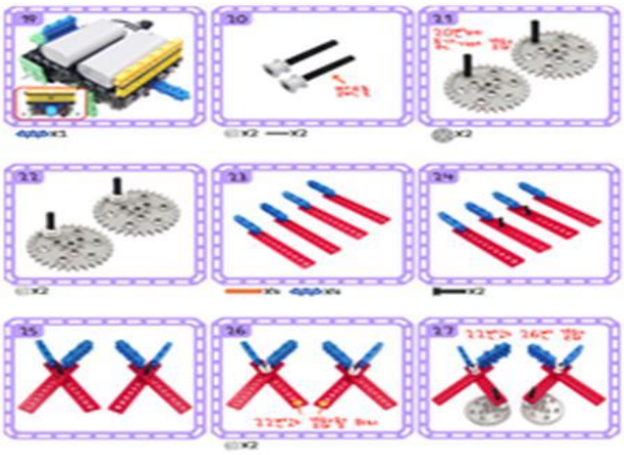


Figure 7. Producing Body of Insect Robot (Butterfly).

Use the gear, the shaft and the bush to assemble 2 legs of the link structure as shown in Fig.7.

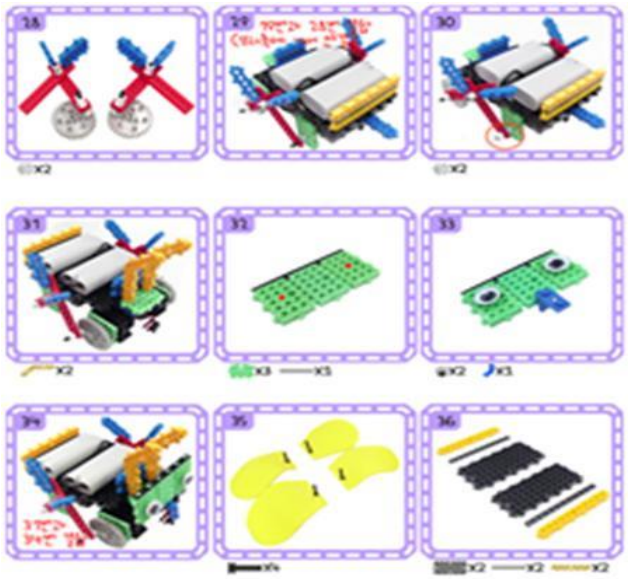


Figure 8. Producing Wing Body of Robot (Butterfly).

Connect the assembled 2 legs of the link structure to both sides of the motor, and assemble the antenna and wing blocks to complete a butterfly robot as shown in Fig.8.

Connect the electric wire and the power supply to the completed robot for operation as shown in Fig.9.

Figure 9. Completed Robot (Butterfly) Assembly.

4.3 Analyzing activity with and opinion about educational insect robot (butterfly)

The analysis was commissioned to researchers in the Ministry of Education, Science and Technology to examine activities with and operation of the educational insect robot (butterfly) of this study.

Table 1. Activity Subject: Activity of Educational Robot (Butterfly)

Step	Activity (subject)	Description
1	Identify task (subject of learning)	Show photos of butterflies, and have a talk with children about male and female butterflies, how they get honey from flowers, get and propagate pollen shown in books or CDs.
2	Investigate (creative thinking)	Imagine one cycle of butterfly's life so that it can be presented with the robot by means of various gestures and facial expressions.
3	Solve (producing a robot)	Show the actual completed robot in the 3D screen with the AR system to attract interest and motivate children for learning. Assemble the VR robot and produce an actual robot.
4	Apply (operating the robot)	Input a program edited through programming exercise, or select and operate a saved program.
5	Conclusion (play, game and opinion)	Perform expansion activity while gaming on the activity board with the robot produced by children for themselves. Talk about their opinion about their own and friends' robot after the activity.

An analysis and a comparison was made of opinions of the experts in the Children's Education Development Center and teachers involved in the Nuri course to know how the contents of the educational insect robot (butterfly) for the Nuri course are applied to children's creativity.

* Sensitivity - the ability to widen new investigation areas

The robot was applied and produced for the process of being interested in the moving robot.

* Fluency - the ability of creating ideas as many as possible in a specific circumstance concerned.

The robot was applied and produced for the process of robot design for problem solving.

* Flexibility - the robot was designed and produced to enable children to create as many ideas as possible in a specific circumstance concerned.

* Creativity-the robot was produced in a creative design process for improving functions with the ability of creating novel and unique ideas, not typical ideas.

* Sophistication - the robot was produced by applying design and logical error discovery with the ability of developing typical unrefined ideas into more sophisticated ideas.

5. CONCLUSION

The Nuri course is a national common education course for children introduced to enhance national responsibility for early childhood which is the first phase of person's education.

We analyzed the educational insect robot to ensure systematic connection of the contents of children's creativity and humanity education who will be the leader of the future society with advanced technology, and studied a process of designing and producing an interesting butterfly robot. An analysis and a comparison was made of operation of the educational insect robot, and opinions of experts and teachers involved in the Nuri course.

This study will contribute to disseminating educational robots in children's education and thus to developing national human resource education through advanced children's education and the effect of teaching and learning.

It is necessary to further study a method of quantitative comparison and analysis of field application of teaching aids to the children's education process, and to convert them into animal robots.

6. REFERENCES

[1] Eun Soo Shin, Eun Hye Park. 2012. Designing and Implementing Early Childhood Education Curriculum on the basis of 2011. *The Korean Society for Early Childhood Teacher Education*. vol. 16, no. 3, 71-91.

[2] Eun Hye Park, Eun Soo Shin. 2012. Analysis of the NURI System on the basis of 2011 ISCED (International Standard Classification of Education). *The Korean Society for Early Childhood Teacher Education*. vol. 16, no. 2, 341-356.

[3] Jeong Wuk Lee, Ji Ae Yang. 2012. A Study on Early Childhood Teachers` Perceptions of the Nuri Policy and Curriculum for Five Year-Olds. *The Korean Society for Early Childhood Teacher Education*. vol. 16, no. 4, 167-192.

[4] Jee Hyun Lee, Hong Ju Jun, Eun Hye Park. 2012. Analysis on the Continuity and Sequence of the Educational Contents in the National Language Curricula for Young Children. *The Korean Society for Early Childhood Teacher Education*. vol. 16, no. 4, 253-279.

[5] Young-Ok Kim. 2012. R-Learning and Childhood Education. *The Korean Society for Early Childhood Education Newsletter*, no. 50.

[6] Jeong Wuk Lee, Min Jung Lee, Kyung Sook Ahn, Soo Jin Lim. 2011. Influence of R-Learning Based Education on Kindergarten and Kindergarten Teachers. *The Korean Society for Early Childhood Teacher Education*. vol. 15, no. 5, 423-444.

[7] Jeong Wuk Lee, Min Jung Lee, Kyung Sook Ahn, Soo Jin Lim. 2011. A qualitative study of the exploration of characteristics on the application of R-Learning based education in kindergarten classroom. *The Korean Society for Early Childhood Teacher Education*. vol. 31, no. 6, 353-378

DID Image System of (Clock) Contents with 3D Hologram

Sung-Yong Yang
Dept. of Excellence Engineering
Hoseo Graduate School of Venture
+82-10-4411-8577
ysyktg@naver.com

*Dea-Woo Park
Dept. of Excellence Engineering
Hoseo Graduate School of Venture
+82-10-8299-4455
prof_pdw@naver.com

ABSTRACT
A screen of 2D contents by means of the DID image system is composed of planar images, still images and text. A product with 2D contents delivers planar uni-directional information to customers. Therefore, 2D contents by means of the DID image system are hard to reflect the value of products, which means it is limited in terms of information delivery capability. This study relates to a DID image system for product refinement, purchasing power and recognition which uses the 3D hologram DID. This study aims to design a 3D hologram, a DID image system, a touch screen, and a CMS, and to implement a DID image system with 3D hologram contents. An analysis is made of refinement, purchasing power, recognition, and satisfaction with respect to 2D contents, 3D contents and 3D hologram DID. This study will contribute to enhance global competitiveness by developing cyber e-transactions and cyber-images.

Categories and Subject Descriptors
K.6.5 [**Management of Computing and Information Systems**]: Security and Protection (D.4.6, K.4.2) – *Authentication, Invasive software (e.g., viruses, worms, Trojan horses), Physical security, Unauthorized access (e.g., hacking, phreaking)*.

General Terms
Security.

Keywords
DID(Digital Information Display), Content, 3D Hologram, Touch-Sensitive, Display.

1. INTRODUCTION
The current DID (Digital Information Display) market is small in comparison with the TV market, and is developing the cyber space thanks to activated Internet e-transactions.

The 2D DID image systems are used for advertisement and show information about products in the field of transportation including subways, buses, airports and terminals, schools, classrooms, shops and shopping malls.

Most of the contents of the DID image system are provided as 2D images.

However, the DID image system market is currently growing fast, and the global market is growing on the basis of fast technology innovation.

* Corresponding Author

In Table 1, the sales volume in the public display market is predicted to grow from 147 million dollars in 2011 to 486 million dollars in 2015.

In the report of "Evolution of Stereoscopic 3D image into Hologram" publicized by the Korea Creative Contents Agency in December, 2011, it is described that advanced countries including the U.S., Japan and European countries have invested a significant amount of money and time to develop hologram technology.

Table 1. Prospect of public display sale (in million dollars)

year	2010	2011	2012	2013	2014	2015
Public displays	44.2	147.9	214.1	295.8	379.2	486.2

* Source: www.displaysearch.com

It is necessary to study the DID image system which uses 3D holograms to maximize the effect of information delivery by maximizing product refinement, purchasing power and recognition for potential consumers in the currently growing online transaction market and the mobile market.

This study aims to produce contents by means of 3D holograms, not the type of planar displays, for the DID image system technology.

1. Introduction describes the necessity and configuration of this study; 2. Related studies describes 3D holograms, the DID image system, and implementation technology; 3. Designing 3D hologram DID image system describes the DID image system and designing CMS; 4. Contents of 3D hologram DID image system describes implementation, test and analysis; and 5. Conclusion draws a conclusion and suggests future studies.

2. Related Study
2.1 3D hologram
The hologram technology is for saving the size and phase distributions of light emitted from an object by means of the phenomenon of light interference and for reproducing it [1].

Holograms are classified into transmission holograms and reflection holograms. In the transmission holograms, the light from a reference light source and the light reflected from an object is in the same direction. In the reflection holograms, the direction of the aforementioned two lights is in the opposition direction [2].

2.2 DID image system
DID is an abbreviation of digital information display and is a general name of displays for displaying public information [3].

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.7

The DID has been limited in terms of information delivery capability because of simple and planar information display and planar images due to divided planes of layer configuration, still images, screen presentation mainly by text, and the uni-directional information delivery system to result in no elements to guide information consumers' positive reaction and eyes.

The DID is currently evolving into a medium for both advertisement and public relations for guiding user's active participation and creating new experiences.

2.3 Technology for implementing DID image system

The technology for implementing a DID image system is divided into a server group, a set-top box group and a display group, and has the following components shown in following Table 2.

Table 2. Components of DID image system

	Details
Server group	* Integrated manager server: monitors software for managing set-top boxes, schedule and contents, environment setup and set-top boxes. Video wall control server: image signal server input into the image system. * Contents server: stores and manages multimedia contents. DB server: manages logs and data.
Set-top box	* Gateway: multicast traffic management. * Set-top box: receives the schedule established in a server to display it on the screen.
Display	* LCD: general LCD display * DID: display of high-class appearance

3. Designing 3D hologram DID image system

3.1 Designing 3D hologram DID image system

3D hologram DID image systems are required for indoor displays used in digital guidance in hotel lobbies, video conference, presentation, museums, and theaters.

Fig. 1 shows an image system designed for a DID image system with a management server, contents and a display in order to provide high-class product services to potential consumers. The image system displays contents through various displays to a plurality of clients by saving and managing contents in a management server after producing product contents and then transmitting them through a network.

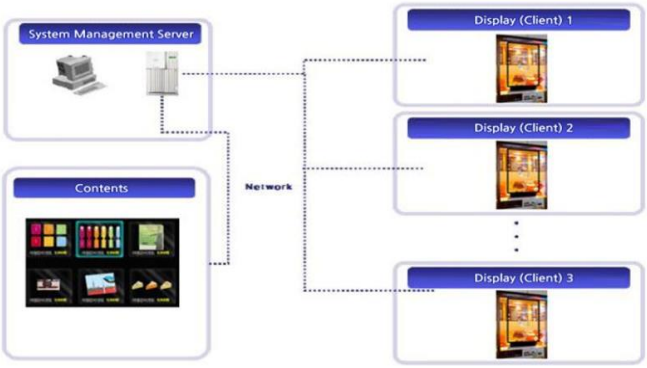


Figure 1. 3D hologram DID configuration.

3.2 Designing CMS(Content Management System)

The CMS target system for producing and managing contents to provide the 3D hologram DID image service is configured and designed as described below.

Designing an implementation system requires a DID server for operating the DID image system and a local CMS for updating created contents on the web. The updated contents are provided through the 3D hologram DID image system.

The contents of data located in servers after producing the DID contents are managed through the CMS. The CMS can manage technology for providing contents displayed by a manager on the screen of each hologram device, registration, modification, deletion, inquiry and schedule. This technology enables 3D hologram DID images to be displayed.

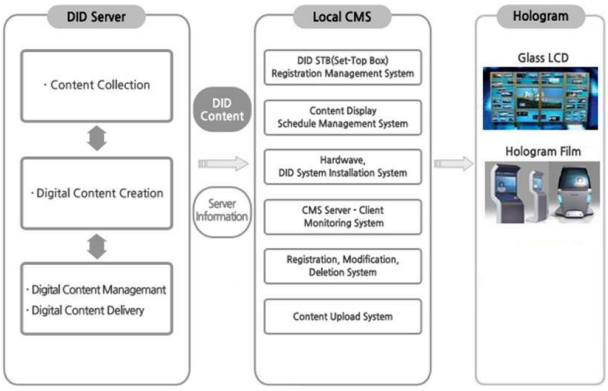


Figure 2. 3D hologram CMS configuration

4. Contents of 3D Hologram DID Image System

4.1 Implementing 3D hologram DID products

The product contents of the 3D hologram DID image system for products is divided into the hologram part and the touch screen part and implemented as shown in Fig. 3.

The hologram part display 3D hologram images and contents of a products, and the hologram images gives spatial and stereoscopic animation effects. The contents selected in the touch screen part is displayed in the hologram part and product contents move.

Potential buyers can touch the touch screen part to select product contents, and the selected product contents provide detailed information and product information desired by the potential buyers through the DID images.

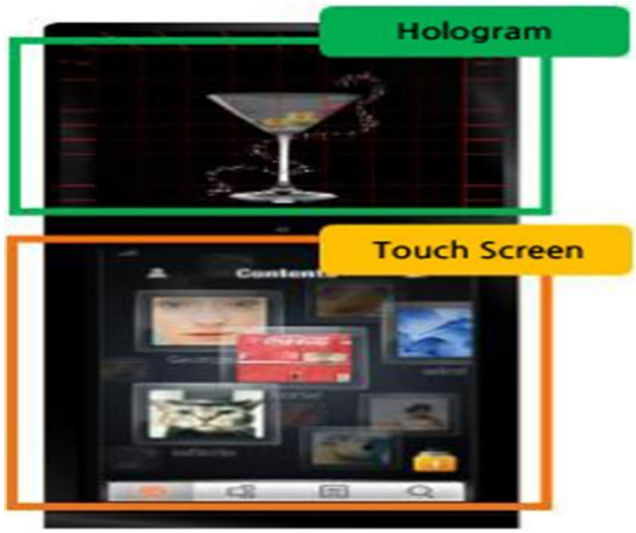


Figure 3. Product contents of 3D hologram DID.

Contents presentation is illustrated in the following Fig. 4, and the contents are produced in 3D with a 3DMAX tool.

It is possible to expand, shrink the clock image interactively, and the clock image looks like a real clock. The 3D hologram DID image experiences less image damages if it is expanded in comparison with 2D product images. It is possible to change time of the clock and observe the second hand movement.

4.2 Implementing 3D hologram DID product image system

Fig. 5 shows the resulting DID image system with 3D holograms. In the upper half part, the DID product contents system is introduced to display 3D holograms, and, in the lower half part, the contents of related products are displayed, through the DID image system. The manager can frequently update the product contents displayed in the lower part by means of the CMS.

Figure 5. Window of 3D hologram DID image system

4.3 Testing and analyzing 3D hologram DID

The "clock" contents made in advance were used by means of the existing 2D contents, 3D contents, and 3D hologram DID image system service for an analysis and investigation about 4 items of refinement, purchasing power and recognition about the product. The method used was to show and compare the above system with the 3D hologram DID image system to potential 207 buyers who visited the shop from January 1, 2012 to August 31, 2012 for a questionnaire.

In Table 3, the 3D hologram DID image system is the best in comparison with 2D contents and 3D contents in terms of refinement and satisfaction. However, for recognition in association with light, both 2D and 3D contents revealed similar result. The connection with purchasing power was not identified.

Therefore, it is considered that the 3D hologram DID image system was effective for expensive goods and those products released early in the market.

However, demonstrative verification is required in the Internet and mobile e-transaction market related to general product sale in order to lead the 3D hologram DID image system to substantial purchasing power.

Table 3. Comparison and analysis of product (clock) contents

	2D contents	3D contents	3D hologram DID	Analysis & comparison
Refinement	low	medium	high	Very high refinement in hologram images in comparison with 2D,3D images
Purchasing power	low	medium	medium	Higher refinement than 2D, but hard to identify

				connection with purchasing power
Recognition	medium	high	medium	Light tends to show slightly low recognition in the hologram environment
Satisfaction	low	medium	high	The highest satisfaction in hologram images in terms of product images

5. Conclusion

Software providers are entering the market of developing DID contents and management systems. In particular, it is highly likely that DID image systems will be developed with 3D holograms led by the expanding DID market.

The 3D hologram DID image system suggested in this study was proved to be the best in terms of refinement and satisfaction about the 3D hologram contents in comparison with 2D and 3D contents.

However, disadvantages are the method of producing 3D contents, high expenses, limitations in installation, and low recognition due to light.

It is necessary to further study improved sharpness of light, and how to improve contents production tools for low cost and easy production in order to build a 3D hologram DID image system.

6. REFERENCES

[1] Jae-Hyoung Park. 2011. Hologram using Integral Imaging Technology Generated. *Korea Society Broadcast Engineers Magazine*. vol. 16, no. 2.

[2] Duk-Gyu Lee, Yoon-Seok Lee, Dong-Seok Yang. 2012. Comparison of Trnsmission Hologram Images between Reconstruction Beams with Different Colors. *Bulletin of Science Education*. vol. 27, no. 2.

[3] Chung, Yoo Kyung. 2012. UI Design Style Guideline Suggestion for the Manager Interface on Digital Information Display. *Journal of Digital Design*. vol. 12, no. 1.

[4] Ki-Sung Hong. 2012. Digital Hologram Contents Manipulation and Synthesis. *The Journal of the Korean Institute of Information and Communication Engineering*. vol. 16, no. 1.

[5] Kyung Sook Lee. 2007. Next-generation display and device industry's 2020 vision and strategy. *Korea Institute for Industrial Economics & Trad.* Policy Resources.

[6] Fusion SW commercialization project. 2012. Video interaction using three-dimensional techniques DID implementation of SW development. *National Institute of Pension Administrators*.

[7] Ki-Sung Hong, Young-Ho Seo, Dong-Wook Kim. 2012. Digital Hologram Contents Manipulation and Synthesis. *The Journal of the Korean Institute of Information and Communication Engineering*. vol. 16, no. 1.

[8] Kim, Tae-Geun. 2010. Views of three-dimensional holographic imaging system. *Optical Science and Technology*. vol. 14, no. 4.

[9] Chung, Yoo Kyung. 2011. The study on goodness of it through preferences of MI(management interface) icon of DID(Digital Information Display). *Journal of Digital Design*. vol. 11, no. 4.

Two-Tier Failure Detector for Large-Scale Disaster Wireless Sensor Networks in CPS

Sungmoon Chung, Inwhee Joe
Department of Electronics and Computer Engineering,
Hanyang University
17 Haengdang-dong, Sungdong-gu, Seoul, Korea
+82-2-2200-1088
iwjoe@hanyang.ac.kr

Yeonyi Choi
Department of Fire Safety Management
Shinsung University
Dahak-ro Dangjin-gun, Chungnam, Korea
+82-41-3501-423
yychoi@shinsung.ac.kr

ABSTRACT
In this paper, we proposed structure of failure detector for Large Scale Disaster Wireless Sensor Networks in CPS and cluster head/sink selection algorithm. Therefore our proposed adapt the network situation. As a result this particular proposed improve network lifetime. Also it provides situational awareness in real-time, which allows intelligent decisions to be made quickly and reliability in time critical environment such as the disaster areas.

Categories and Subject Descriptors
C.2.1 [**Network Architecture and Design**]: Network communications, Network topology

General Terms
Algorithms, Performance, Reliability

Keywords
Failure Detector, Cyber Physical System

1. INTRODUCTION
The disaster events take place frequently for the past few years, disaster-tolerance of information system becomes more and more important [1]. In the early period, local disaster system is enough for the need, with people's improving requirements for disaster tolerance, the remote disaster system is appeared which can monitor disaster area in remote center. Every year we can see the news another natural disaster completely devastate a city or town. Also we hear about the problems rescue teams face with having very little knowledge of how the area was affected, where possible survivors may be, and how dangerous a particular area is for rescue or clean-up workers to be deployed in. The efficient solution is that with the CPS (cyber-physical system) proposed.

A CPS is a network of sensors or actuators capable of computation, communication, and control that relies highly on the integration of these capabilities for its operations and interactions with the physical environment where it is deployed [2]. Sensors or actuators of CPS can perform actuation by themselves and by integrated network control according to the physical environment.

Although each individual sensor and actuator is unsuitable to actuate the environment on its own, with the cooperation among the individual sensors and actuators, the network has the ability to carry out tasks that otherwise would be impossible by a single node. Especially there are many restricts to apply CPS to Wireless Sensor Networks (WSN). It is infeasible to actuate itself because of its limited hardware. Also collision occurs when the sensors broadcast their sensing data and buffer overflow occurs due to the limited buffer in large scale WSN. Therefore there are a number of significant problems without considering these problems. Having the ability to get important information with guaranteed Quality of Service (QoS) is a valuable asset in the disaster area because it is related to life and death closely.

Therefore we proposed structure of failure detector for Large Scale Disaster Wireless Sensor Networks in CPS and cluster head/sink selection algorithm. This particular proposed provide reduced collision and improve network lifetime. Also it provides situational awareness in real-time, which allows intelligent decisions to be made quickly and reliability.

2. RELATED WORKS
The previous researches of failure detector for disaster are mainly concentrated in the model and algorithm of failure detection. Chandra and Toueg [3] studied the failure detection firstly. They present the system model and qualitative analysis with the theoretical guide for later research. Chen etc. [4] studied the failure detection metrics which is basis of failure detection algorithm quantitative analysis. Chen, Bertier, Hayashibara, Xiong studied the failure detection algorithms and present representative algorithms of Chen-FD [4], Bertier-FD [5], φ –FD [6], EDFD [7]. Also He Jun and Wang Tao [8] studied to set up the failure detection model in the asynchronous distributed environment and describe the conditions when the states of failure detection transform. Mao etc. [9] studied data monitor collection, failure detection processing and system migration. However these previous researches studied the failure detection without considering CPS. Furthermore they didn't design the failure detector considering specific network. As a result there are many problems when the failure detector is applied in practical environment.

3. DESIGN OF FAILURE DETECTOR
A few main reasons the failure detector is necessary to addressing this Large Scale Disaster Wireless Sensor Networks in CPS are the following. (1) Sensing data and information of

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.8

indicate sensor situation should collected by entity which can perform modeling and control the network for actuation. (2) The real-time is guaranteed between sensors and entity which can perform modeling and control the network for real-time. (3) Failure control should be done promptly by actuation when the failure detector failure. (4) Collision, buffer overflow should be considered for Large Scale Wireless Sensor Networks. (5) Power consumption should be considered for power-constrained sensor nodes.

3.1 Structure of Failure Detector for Large Scale Disaster Networks in CPS

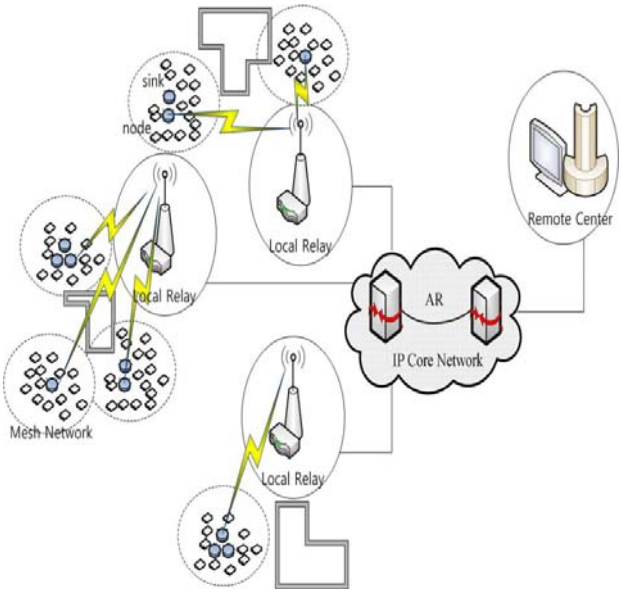


Figure 1. Structure of Failure Detector for Large Scale Disaster Networks in CPS

- The network topology of senor field is Wireless Mesh Network. Therefore rescuers or savers can collect the environment situation from the closet node to them in the event of a disaster.
- Cluster head collect sensing data and information of indicate sensor situation. After that the cluster head send these collected data to sink. Cluster head/sink is updated per updated period and when the failures are detected.
- There are multi available sink in the senor field. They provide selective link option between sink and local relay.
- Sink has a failure detector of cluster-level. Therefore network can actuate environment situation promptly at cluster-level when the failures are detected. .
- Local relay collect data from the sink and send data to remote center by IP core Network.

Remote center has a failure detector of system-level. Therefore it can do high level modeling and actuation

3.2 Two-tier deployment of Failure detector

(1) After Sensors form the Wireless Mesh Network, cluster head is selected by cluster selection algorithm.

(2) Sink is selected by sink selection algorithm.

(3) Cluster head collects data from the sensors which are deployed in the cluster.

(4) Cluster head send collected data to sink.

(5) Failure detector of cluster-level determines that it is fail or not.

(6) If it is not fail, sink send data to local relay and local relay send data to remote center.

(7) If it is fail, failure detector of cluster-level controls fail promptly and send data to remote center which includes fail control information of cluster level and collected data from sensors.

(8) Failure detector of system-level does high-level modeling and controls the fail through sending feedback to local relay.

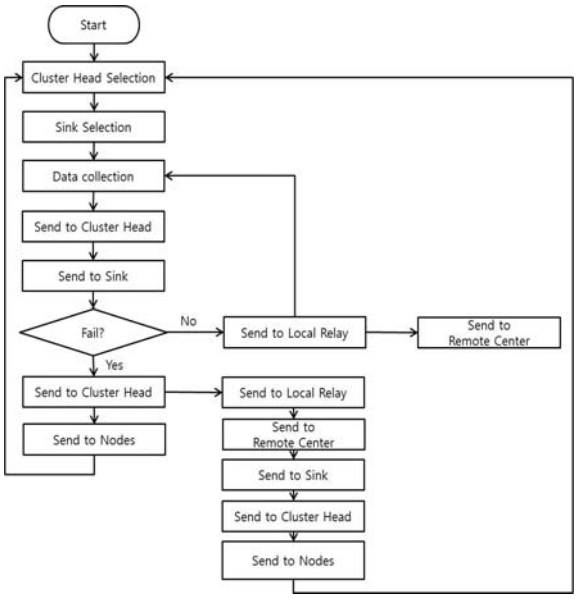


Figure 2. Two-tier Failure detector flow chart

3.3 Cluster Head Selection

(1) **INIT:**

(2) Set Mesh Routing;

(3) Calculate Avg Link Quality(node to node);

$$\{ \sum_{k=1}^{N-1} Link\ Quality / N-1 \}$$

(4) Calculate Node Priority;

$$\{ NP_k = \alpha_{CH} * Avg\ Link\ Quality + \beta_{CH} * Buffer\ capacity + \delta_{CH} * Residual\ Energy, \alpha_{CH} + \beta_{CH} + \delta_{CH} = 1 \}$$

(5) Exchange Node Priority Information;

(6) CH= NP_m

FOR NP_1 to NP_{N-1}

IF CH< NP_k **THEN**

CH= NP_k

k=k+1

ELSE

k=k+1

END IF

END FOR

We assume that a cluster has the number of N nodes and priority grades of parameters are provided. NP_k means node priority of Kth node. NP_m means priority of node itself. α_{CH} denotes weight of Avg link quality with other nodes, β_{CH} denotes weight of node's buffer capacity, δ_{CH} denotes weight of node's residual energy.

3.4 Sink Selection

(1) **INIT:**

(2) Calculate RTT from Local Center to Sink

(3) Receive pilot message from CHs

(3) Calculate Avg Link Quality(CHS to sink);

$$\{ \sum_{k=1}^{L} Link\ Quality / L \ \}$$

(4) Calculate Sink Priority;

$$SP_k = \alpha_S * RTT + \beta_S * Avg\ Link\ Quality + \delta_S * Buffer\ capacity + \gamma_S * Residual\ Energy, \quad \alpha_S + \beta_S + \delta_S + \gamma_S = 1$$

(5) Exchange Sink Priority Information;

(6) Sink= SP_m

FOR SP_1 to SP_L

IF Sink< SP_k **THEN**

Sink= SP_k

k=k+1

ELSE

k=k+1

END IF

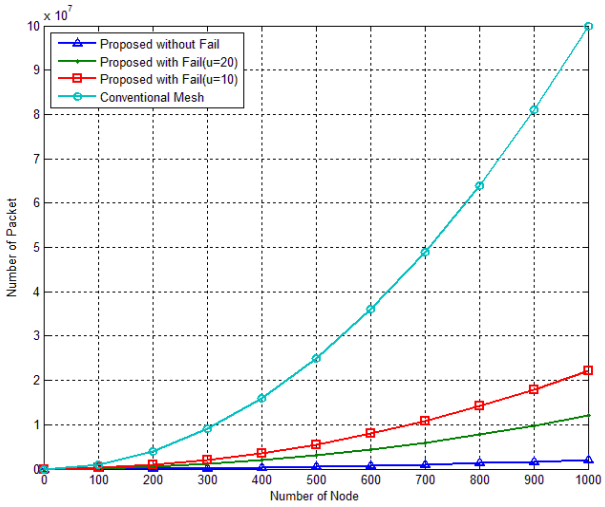
END FOR

4. PERFORMANCE EVALUATION

Table 1. Performance Evaluation Parameters

Parameters	Value
Network topology	Wireless Mesh Networks
Data Period(t)	10s
Measured Time(t)	100s
Power to run the transmitter circuitry(nJ/bit)	50
Power for the transmit amplifier to achieve an acceptable SNR (pJ/bit/m^2)	100
Distance(node to node)(m)	10
Packet Size(bit)	1000

Figure 3 shows the number of packet per the number of nodes. In large scale wireless sensor networks, conventional mesh topology increases the number of sending packets sharply by increasing the number of nodes. In this paper, we reduce the number of sending packets due to cluster head. Our proposed should update cluster head and sink selection per updated period and when the failures are detected. So the shorter update period or the more failures, the more power consumption.



. **Figure 3. Number of Packets as a function of Number of nodes**

Figure 4 shows the network power consumption per the number of nodes. Our proposed reduce the number of sending packets and buffer overflow. Therefore the number of sending and receiving is decreased. As mentioned above, the shorter update period or the more failures, the more power consumption.

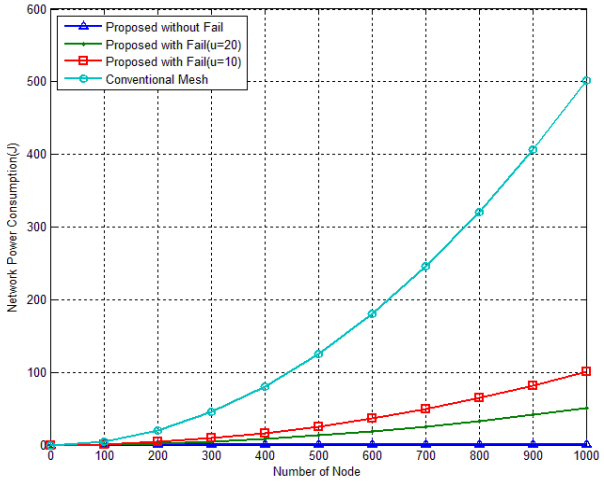


Figure 4. Network Power consumption as a function of Number of nodes

5. CONCLUSION

In this paper we propose the structure and cluster head/sink selection algorithm of Large Scale Disaster Wireless Sensor Networks for CPS. Through our proposed we show how the failure detector is applied in specific practical environments. As a result our proposed adapt the network situation by considering network parameters. It reduce collision and improve network lifetime by reduced the number of sending packets. Also it provide situational awareness actuation in real-time by considering RTT, link quality, buffer capacity, residual energy.

Our Two-tier failure detector provides prompt actuation and high level modeling in the event of a disaster. Therefore it is very efficient to both rescuers and savers in time critical environment such as the disaster areas.

6. ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation by Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A3012227).

7. REFERENCES

[1] Y.X. Yang, W.B. Yao, and Z. Chen, 2010. *Review of Disaster Backup and Recovery Technology of Information System*, Journal of Beijing University of Posts and Telecommunications, Vol. 33(2), Apr. pp. 1-6.

[2] R. A. Cortez, 2009. *A Cyber-Physical System for Situation Awareness Following a Disaster Situation*, IEEE Realtime Systems Symp.

[3] T.D. Chandra and S. Toueg, 1996. *Unreliable Failure Detectors for Reliable Distributed Systems*, Journal of the ACM, Vol. 43(2),pp. 225-267.

[4] W. Chen, S. Toueg, and M.K. Aguilera, 2002. *On the Quality of Service of Failure Detectors*, IEEE Trans.on Computers, Vol. 51(5), pp. 561-580.

[5] M. Bertier, O. Marin, and P. Sens, 2002. *Implementation and Performance Evaluation of an Adaptable Failure Detector*, Intl. Conf. on Dependable Systems and Networks, Washington DC, USA, 2002, pp. 354-363

[6] N. Hayashibara, X. Defago, R. Yared, and Katayama.T. 2004 ,*The ϕ Accrual Failure Detector*, Proc. 23rd IEEE Intl. Symp. On Reliable Distributed Systems, Florianpolis, Brazil, pp.66-78.

[7] N.X. Xiong, V. Athanasios, T. Laurence, etc, 2009. *Comparative analysis of Quality of Service and Memory Usage for Adaptive Failure Detectiors in Healthcare Systems*, IEEE Journal on selected areas in communication, vol. 27(4), May.

[8] He Jun, Wang Tao, Tang Lei, Wen Chuan-hua, 2010. *Research on Failure Detector in Remote Disaster-tolerant System*, icic, vol. 1, pp.199-202, 2010 Third International Conference on Information and Computing.

[9] Xiuqing Mao, Xingyuan Chen, Yingjie Yang, Junfeng Li, 2011. *An Improved Framework of Disaster-Tolerance Oriented Adaptive Failure Detection*, iscid, vol. 1, pp.228-231, 2011 Fourth International Symposium on Computational Intelligence and Design

LSIST: LEARNING STYLE BASED INFORMATION SEEKING TOOL

Nor Liyana Mohd Shuib
University of Malaya
Kuala Lumpur, Malaysia
liyanashuib@gmail.com

Rukaini Abdullah
University of Malaya
Kuala Lumpur, Malaysia
rukaini@um.edu.my

ABSTRACT
Learning can be enhanced when its activities are aligned with students' learning styles (LS). An important component of student learning activities is searching and retrieving reading materials. Research has shown that students have problems finding suitable reading materials due to the mismatch of the different attributes of the reading materials and learning styles. Existing information seeking tools are inadequate as these tools do not consider learning style in their query search. Hence there is a need to develop an information seeking tool that uses learning style to retrieve suitable reading materials. This paper presents the architecture for LSIST, a learning style based information seeking tool. In particular, we explained how LS is used in the LS based search module that allows the retrieval of reading materials that match students' LS.

Categories and Subject Descriptors
D.2.2 [**Software Engineering**]: Design Tools and Techniques – *modules and interfaces.*

General Terms
Design.

Keywords
Learning style, information seeking tool, information retrieval, reading material

1. INTRODUCTION
Finding suitable reading material that can be understood by students is very crucial in learning process. However, students have problems finding suitable reading material due to the mismatch of the different attributes of the reading materials and learning style. Students learning can be enhanced through the presentation of reading materials that are consistent with a student's particular learning style [11].

Students have different learning style. Some students may learn best by reading by text, others by reading from visual and others by reading examples. Therefore it is important to take into account students' learning style while developing information seeking tool and reading material. To date, none of the existing information seeking tools has the function that can match reading materials with students' learning styles. Hence there is a need for the development of information seeking tool with learning style consideration.

2. LITERATURE REVIEW
Reading materials are written document to be read. They contain data that are organized in the form of meaningful information. Students need reading materials to carry out activities such as solving problems, making decisions, reducing uncertainties, resolving conflicts, answering questions and satisfying curiosities. It helps them understand their courses.

Information in reading material can be presented in various forms such as text, tables, pictures, flow charts, drawings, maps, figures and mathematical expressions. These forms of presentation illustrate, explain or demonstrate information in the reading material [9]. Appropriate form of presentation in reading material is important because it can enhance understanding and help students encode the information more effectively in order to help them understand the information [3, 26, 27].

Most of the reading materials especially in education domain can now be accessed and easily found online, since many organizations have digitized the printed reading materials. Also publishers today publish reading material in print and online forms, making the information more accessible. With the changes in information accessibility, students can find many reading materials by using various information seeking tools. However, the question is whether the retrieved materials are suitable to the students.

2.1 INFORMATION SEEKING TOOL
Information seeking tool is an important element in information seeking process. Students use information seeking tool to retrieve additional reading material. Information seeking tool is a system that informs the user about the existence or non existence of the document related to the user request [24]. It is developed to facilitate information seeking.

Information seeking tools contains a set of components which are interrelated to facilitate searching process. The basic process of information seeking tool is shown in Figure 1 [4]. A file of potential search term is produced from organized information.

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.9

When a user submits a query, a comparison is made between the file and the query. A set of documents are then retrieved.

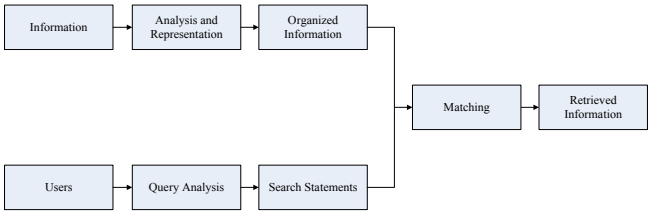


Figure 1. Basic Process of Information Seeking Tool [4]

Information seeking tool helps student to find information by organizing the information in order to make the information easier to search, identified and achieved. There are many information seeking tools available on the Internet. Students can choose available tools to retrieve the desired materials. Students may access and use these various information seeking tools for diverse reasons [6]. Information seeking tools differ in structure, in the way they function and utilize different methods and techniques for storing and retrieving information. A few examples of information seeking tools are described below:

- Online public access catalogues (OPAC)

OPAC is a computerized catalogue containing bibliographic records of items in a library [1]. Students usually used OPAC to find books from library online before borrowing it. By using OPAC, students can access bibliographic records of all type of materials in the library. They can search more than one collection within the same library or in different libraries. OPAC provides a simple search interface that enable users to browse and search the collections. Search facilities provided by OPAC are browse and search, keyword and phrase search, Boolean and proximity search (limited to keyword search option), subject heading such as LCSH search, and records search through selected keys such as author, title, ISBN or call number [5].

- Internet search engine

Internet search engine consists of a web page, images and files. It uses web crawlers or spiders to automatically retrieve information from millions of web pages on the web and then indexes the information. This makes Internet search engine the most comprehensive coverage of the web. An example of an Internet search engine is Google. Each Internet search engine has its own specific set of retrieval features. Search facilities provided by most Internet search engines are word, phrase and natural language search options, special options for image, audio and video search, multilingual search and basic search facilities such as Boolean search. They can restrict and rank searches with different criteria.

- Subject directory

Subject directories are created manually by assigning the submitted sites to a suitable subject category by the directory developers. Search facilities provided by most subject directories are browse and search options, word, phrase and subject search options and ability to rank result by relevance option. Yahoo and Britainica are examples of subject directories.

- Online database

Online database is a collection of scholarly journals that can be accessed online. It provides access to remote database through a database vendor or service provider. The format is similar to the traditional printed journals and can be in the form of HTML, PDF or both. Among the search facilities provided are browse and search, keyword and phrase search, common search such as Boolean search, truncation, field search, limiting search and range search. Browsing can be done on the entire collection or by issues. Examples of online databases are Elsevier, IEEE and ACM.

- Digital library

Digital library (DL) is a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network [2]. It provides a high quality resource within a particular subject area that is of interest to a specified audience [6]. It is filtered by library professionals and subject experts and added manually. Search facilities provided by most DLs are browse and search options, word, phrase and subject search options, and common search facilities such as Boolean search. DL search functions are limited because it is designed to focus more on content. Examples of digital library are Dspace@UM and DigiLibraries.com.

2.2 Difficulties in Seeking Suitable Reading Material

With the amount of information and the technology that is available today, it is easier to find information on any topic. However, the information obtained may not necessarily help students in their learning process. Information in this case can be defined as reading material that is used for learning. One of the problems is information overload. For example, when a student seeks for information on a particular subject, he looks at all possible information sources such as books, articles, and journals and ends up with a mountain of information, making it difficult to choose the most relevant and appropriate sources [6, 8].

Students do not necessarily find information that suits their style of learning. They usually evaluate information based on content or keyword that they used. As long as the information contains the keywords, they tend to assume that it is the right information for them. Such students need help from their lecturers to find suitable reading materials. Lecturers can help students find the reading material on the topic of their study at the appropriate level of knowledge but not all students can understand the same form of information. This situation is called one size fits all. For a standard learning process, lecturers normally provide the same form of information for all students. Students have to adapt to the information given. However, this does not work in practice. They end up not comprehending the information given [30].

Furthermore, reading materials are not categorized according to the way information is presented. This does not help students in choosing the right reading material. It is difficult or even impossible for student to learn when the material is not presented in their preferred learning style [17]. This means even though the information retrieved is relevant to their topic and level of knowledge; it may not contribute to better understanding of the subject.

2.3 The Need of Learning Style

Each individual learns in different ways. This is influenced by individual characteristics such as prior knowledge, education level, past experience, level of literacy, motivation, task confidence, aptitudes, and learning styles [10, 18, 20]. These differences affect the learning activities and outcome. One of the most important factors that affect the outcomes of learning is learning style [15].

Learning style is important in a learning process. It can help students enhance their learning capabilities [11]. For example, if students know and understand their learning style, they can choose reading materials that are suitable to their learning style that can assist them in understanding the subject better. This could enhance their learning capabilities.

2.4 Learning Style

Learning style is defined as the preference or predisposition of an individual to perceive and process information in a particular way or combination of ways [31]. There are many learning style models in the literature. In this study, we focus on learning style models that have intellectual approach to assimilating information and adaptable by learning strategies. These models focus on the processes by which information is obtained, sorted, stored and utilized. It includes Gregorc's Style Delineator [17], Kolb's Learning Style Inventory [22], Index of Learning Style (ILS) [12], Honey and Mumford's Learning Style [19], Gardner' Multiple Intelligence [16], and VARK learning style model [14].

While most of the models above discuss individual differences based on the way information is comprehended and arranged, Fleming and Mill [13, 14] introduced a learning style model that uses sensory modality called VARK which is an acronym for Visual, Aural, Read/Write and Kinesthetic. It emphasizes the preferences of taking in and taking out information through sensory channels. VARK has four preferences as shown below:

- Visual (V) - Visual learners prefer the use of diagrams, graphs and flow charts to represent printed information.
- Aural (A) - This mode describes a preference for information that is heard or spoken.
- Read/write (R) – Read/write learners prefer printed words and text as a means of taking in information
- Kinesthetic (K) - Kinesthetic preference refers to learning through the use of experience and practice.

A sensory modality is a combination of perception and memory – in other words, how the mind receives and stores information. Individuals have different sensory modality preferences when internalizing information. Sensory modality is one of the more practical and popular ways of defining and assessing learning style that one prefers when learning [8]. In this research, VARK learning style model [14] is chosen because its categorization of preferences is applicable to reading materials.

2.5 Limitation of Existing Information Seeking Tools

Research shows that none of the existing information seeking tools has a function that matches reading materials to students' learning styles [25]. Table 1 illustrates the differences between several information seeking tools in terms of the search facilities provided.

Table 1. Comparison of Information Seeking Tools Search Facilities

Search Facilities	OPAC	Internet Search Engine	Subject Directory	Online Database	DL
Browse and search	x	x	x		x
Keyword and phrase	x	x		x	x
Subject headings	x		x		x
Natural language		x			
Level of Expertise		x		x	
Ranked search result		x	x	x	x
Learning Style					

3. THE ARCHITECTURE OF LSIST

To develop an information seeking tool that retrieves reading materials that match students learning style, we propose LSIST. Learning style is incorporated in the tool's search process. This section presents the architecture of the tool. The main components of this tool are :

- Students' Learning Style Profile Component - To identify students' learning style preference based on learning style test.
- Reading Material Classification Component - To extract and categorize learning style attributes in reading materials.
- Matching Component - To match reading materials with students' learning style preference.

The proposed architecture as shown in Figure 2 involves two types of users; administrators and students. Administrators provide reading materials to the tool.

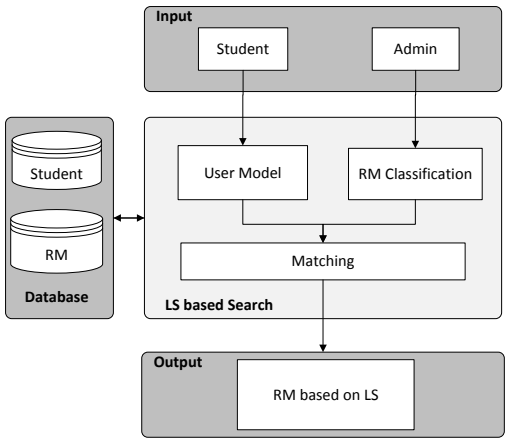


Figure 2. LSIST Architecture

LSIST consists of four main modules;

- The input module - Receive inputs from students and administrators.

- The LS based search module – this is a process module that contains three main components; User Model, Reading Material (RM) Classification and Matching. This module use information from Input module to categorize and match reading material with students' learning style.
- The database module - this module stores student information and reading materials provided by the administrator.
- The output module - This module presents the results from the LS based search module.

3.1 Input Module

The input module receives students' profile such as learning style preferences and search query. These inputs are used to develop a user model in the LS based search module and stored in the student database. This module also receives reading material profiles such as title, author, and learning style value from the administrator.

3.2 LS based Search Module

The LS based search module retrieves reading material based on keyword and learning style preference. This module contains three main components; User Model, RM Classification and Matching.

3.2.1 User Model

This model receives information from the input module and stores it in the database module. This is where students are categorized based on learning style preference. The filtering is based on learning style test and the results are then stored in the student database module.

3.2.2 RM Classification

This component extracts and categorizes learning styles' value from the reading material and stores the information in RM Database. Reading materials used in this study are limited to PDF type only. To extract and categorize learning styles' value in reading material, the content in the reading material is extracted using PDF extraction tool, iText. The content extracted is pre-processed to remove outliers and noisy data such as non-word characters, i.e. any character excluding a-z, A-Z, 0-9 and underscore (_) character.

Feature extraction is used to transform the input data (reading material) into a reduced representation set of features called feature vector. In this work, there are three feature vectors representing the three learning style preference consisting of Visual, Read/write and Kinesthetic. Audio preference is ignored because the reading materials considered are not in audio form.

Identifiers representing each feature vector as shown in Table 1 are identified and calculated from the content. For the purpose of standardization, all identifiers are coded from IEEE Learning Object Metadata [21].

Table 1: Identifiers based on LOM

Feature Vectors	Identifier
Visual	Figure, diagram, map, chart, graph, flowchart, arrow, circle, hierarchy, hierarchies, picture, table, equation, notation, formula, histogram, scatter plot, screenshot
Read/write	All words except words describing Visual and Kinesthetic preference
Kinesthetic	Example, practice, case study, exercise, simulation, experiment, self-assessment, application

If an identifier for a feature vector is found on a page then the value of that feature vector is increased according to the weight which is 1 or 0.5. If the value of the feature vector is already 1 then the next identifier of the same feature vector found on the same page is ignored. For each feature vector, the maximum value for each page is 1.

The values of the feature vectors for the whole document are used instead of the number of identifiers found in the pages. The results are shown in the form of percentages instead of categorizing each reading material onto a specific LS preference. This is to ensure any combination of learning style preference can be catered. The total value for all feature vectors is calculated

$$t = \sum v + \sum r + \sum k$$

Where;
t = total of all feature vectors
v =value of *visual feature vector* on each page
r = value of *read/write feature vector* on each page
k = value of *kinesthetic feature vector* on each page

The calculation for the various feature vectors is given below:

Value for *Visual* feature vector, *F(v)*,

$$F(v) = \left(\frac{\sum v}{t}\right) \times 100$$

Value for *Read/write* feature vector, *F(r)*,

$$F(r) = \left(\frac{\sum r}{t}\right) \times 100$$

Value for *Kinesthetic* feature vector, *F(k)*,

$$F(k) = \left(\frac{\sum k}{t}\right) \times 100$$

3.2.3 Matching

The matching component uses information stored in the database module to execute the searching and matching process. The reading material and student features need to be standardized because the values for the reading material feature vectors are in percentages, $[0,100]$ whilst the values for the student feature are in the range of $[1,16]$. The value for the student vector is identified using VARK learning style test from the VARK website which consists of 16 multiple choice questions with four answer selections corresponding to the four preferences [28].

To standardize these two values they are normalized to $[0,1]$. Once standardization is performed on a set of features, the range and scale should be similar, providing the distributions of raw feature values are alike. For the normalization, we use Softmax formula to ensure all of the output values are between 0 and 1 and that their sum is 1.

The process of matching reading material onto students' learning style is executed using k-nearest neighbor (k-NN) classification method. K-NN is a method for classifying objects based on closest relations in the feature space. This method is used because we want to consider all feature vector values without categorizing each document to only a specific LS preference. This reduces information loss. The k-NN classifier is based on the Euclidean distance. Euclidean distance is used to calculate the distance between students' learning style (*S*) and reading material learning style (*R*) value. It is chosen because it is sufficiently accurate to calculate the distance between multiple features vector to get the relationship [7].

Distance between feature vectors from one to another is calculated to check the closest relationship. In this case, how close LS feature vector of *S* and *R* is related. Feature vectors for both features are visual (*v)*, read/write (r) and kinesthetic (*k)* values. Each of feature vectors value has been normalize to $[0,1]$. We plot the feature vector of *S* and *R* as coordinates in Cartesian coordinate. Relation $d(S,R)$ between *S* and *R* LS value is calculated from the distance value of LS feature vectors.

The lower value means the closest relation.

$$d(S,R) = \sqrt{(S_v - R_v)^2 + (S_r - R_r)^2 + (S_k - R_k)^2}$$

Where;

d = Distance
S = Student
R = RM
v = Visual feature vector
r = Read/write feature vector
k = Kinesthetic feature vector

The process of matching reading material onto students' learning style is executed when students enter keywords in their query. The process for searching suitable reading material is presented as follow:

Step1: Query keyword of reading material

Step 2: For each RM with similar keyword query,

1. Calculate learning style value distance between reading material and student
2. Save distance and reading material to database
3. Sort the result from lowest to highest distance

Step 3: Show reading material from the database

3.3 Database Module

This module comprises two databases which are Student Database and RM Database:

- Student Database stores students' information such as gender, level of study and learning style preference.
- RM Database stores reading materials' information such as title, author and learning style value.

3.4 Output Module

The recommended reading material that resulted from the matching component is displayed in Output module. Students are asked to give feedback regarding the reading material retrieved from LSIST.

4. LSIST Development

The prototype for LSIST is designed using Structured Systems Analysis and Design Method (SSADM) [29] and developed using C#. PDF Library iTextSharp which is a complement of iText library is used to handle and manage PDF files. 77 reading materials in PDF type with various topics such as e-learning, data mining, artificial intelligence, research, system development, and knowledge management were uploaded into the prototype.

5. Conclusion

In this paper, we highlighted that existing information seeking tools are inadequate to address students' difficulty in retrieving suitable reading materials that match their learning styles. We then proposed LSIST, a learning style based information seeking tool and presented its architecture. In particular we described how learning style is used in the search module that enables the tool to retrieve reading materials that match students' learning styles. The usefulness and accuracy of this tool will be evaluated in the future.

6. REFERENCES

[1] Ariyapala, P.G. 2002. *Use of the University of Malaya's Library OPAC by foreign postgraduate students* . University of Malaya.

[2] Arms, W.Y. 2000. Digital Libraries. Cambridge, Mass., USA.

[3] Budhu, M. 2002. Interactive web-based learning using interactive multimedia simulations. *International Conference on Engineering Education (ICEE 2002)* (Manchester, UK, 2002).

[4] Chowdhury, G.. 1999. *Introduction to modern information retrieval*. Library Association Publishing.

[5] Chowdhury, G.G. and Chowdhury, S. 2003. *Introduction to Digital Libraries*. Facet Pub.

[6] Cooke, A. 2001. *A guide to finding quality information on the Internet: selection and evaluation strategies*. Library Association.

[7] Danielsson, P.-E. 1980. Euclidean distance mapping. *Computer Graphics and image processing*. 14, 3 (Nov. 1980), 227–248.

[8] Dobson, J.L. 2009. Learning style preferences and course performance in an undergraduate physiology class. *Advances in Physiology Education*. 33, (2009), 308–314.

[9] Dori, D. et al. 1997. The Representation of Document Structure: A Generic Object-Process. *Handbook on Optical Character Recognition and Document Image Analysis*. P.S.P. Wang and H. Bunke, eds. World Scientific Publishing Company.

[10] Evans, C. and Sadler-Smith, E. 2006. Learning styles in education and training: problems, politicisation and potential.

[11] Felder, R.M. 1995. A Longitudinal Study of Engineering Student Performance and Retention. IV. Instructional Methods and Student Responses to Them. *J. Engr. Education*. 84, 4, 361–367.

[12] Felder, R.M. and Silverman, L.K. 1988. Learning and Teaching Styles in Engineering Education. *Engineering Education*. 78, 7, 674–681.

[13] Fleming, N. and Baume, D. 2006. Learning Styles Again: VARKing up the right tree!! *Educational Developments*. 7,4, 4–7.

[14] Fleming, N.D. and Mills, C. 1992. Not Another Inventory , Rather a Catalyst for Reflection. *To Improve the Academy*. 11, (1992), 137.

[15] Gagné, F. 2004. A differentiated model of giftedness and talent (DMGT), pp.1–4. Available at: http://www.curriculumsupport.education.nsw.gov.au/policies/gats/assets/pdf/poldmgt2000rtcl.pdf. Accessed: January 20, 2013.

[16] Gardner, H. 1993. *Multiple intelligences: The theory in practice*. Basic Books.

[17] Gregorc, A.F. 1985. *Inside styles, beyond the basics*. Gregorc Associates.

[18] Atkins, H. et al., 2001. Learning Style Theory and Computer Mediated Communication. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications. Chesapeake, VA: AACE, pp. 71–75.

[19] Honey, P. and Mumford, A. 1992. *The Manual of Learning Styles*. Peter Honey Publications.

[20] Howles, L. 2006. Learning Styles: What the Research Says and How to Apply it to Designing E-Learning.

[21] IEEE 2002. Draft standard for learning object metadata. *IEEE Learning Technology Standards*. July, 1–44.

[22] Kolb, D.A. 1984. *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

[23] Kuhlthau, C.C. 1997. Learning in Digital Libraries: An Information Search Process Approach.

[24] Lancaster, F.W. 1986. *Vocabulary control for information retrieval. 2nd. ed*. Information Resources Press.

[25] Nor Liyana, M.S. et al. 2010. The Use of Information Retrieval Tools: a Study of Computer Science Postgraduate Students. *International Conference on Science and Social Research (CSSR 2010)* (Seri Pacific Hotel, Kuala Lumpur, Malaysia, 2010), 5–7 December.

[26] Pen˜a, C.I. et al. 2002. Intelligent agents in a teaching and learning environment on the Web. *ICALT 2002*.

[27] Stash, N. et al. 2004. Authoring of learning styles in adaptive hypermedia: problems and solutions. *The 13th International Conference on World Wide Web*.

[28] The VARK Questionnaire: 2010. *http://www.vark-learn.com/english/page.asp?p=questionnaire*. Accessed: 2010-01-01.

[29] Weaver, P.L. et al. 1998. *Practical SSADM 4+*. Pitman.

[30] Yang, Y. and Wu, C. 2009. An attribute-based ant colony system for adaptive learning object recommendation. *Expert Systems with Applications 28*. 36, 2, 3034–3047.

[31] Zapalska, A. and Brozik, D. 2006. Learning styles and online education. *Campus-Wide Information System*. 23, 5.

Harumanis Mango Flowering Stem Prediction using Machine Learning Techniques

R.S.M. Farook[a,*], H. Ali[b], A. Harun [c], Ndzi. D. L.[d], A. Y. M. Shakaff [c], Mahmad Nor Jaafar[e] , Z. Husin[a] , A.H.A. Aziz[a]

School of Computer and Communication Engineering, University Malaysia Perlis (UNIMAP), Perlis, Malaysia[a],
Block A, Kompleks Pusat Pengajian Seberang Ramai
No. 12 & 14, Jalan Satu, Taman Seberang Jaya Fasa 3
Seberang Ramai, 02000 Kuala Perlis
Perlis Darul Sunnah
Telephone Number : 604-9851654
Email: rohani@unimap.edu.my
Fax Number : 604-9851695

Information Technology Department, Polytechnic Tuanku Syed Sirajuddin, (PTSS), Perlis Malaysia[b]
School of Mechatronic Engineering, University Malaysia Perlis (UNIMAP), Perlis, Malaysia[c]
School of Engineering, University of Portsmouth, Portsmouth, UK[d]
Agrotechnology Research Station, University Malaysia Perlis (UNIMAP), Perlis, Malaysia[e]
rohassanie@yahoo.com, David.ndzi@port.ac.uk
{ zulhusin, hallis, aliyeon, mahmad }@unimap.edu.my

ABSTRACT

Harumanis Mango (M*angifera indica*) is known as one of the best table tropical fruit, due to its aroma and sweetness. Harumanis mango cultivar is included in the national agenda as a specialty fruit from Perlis, Malaysia for the world. Despite its overwhelming local demand in Malaysia and also internationally, the fruit supply never meets the demand. Mango flowering stem prediction is important as one of the factors to predict mango yield in order to implement effective forward marketing. Forward marketing is a co ntract that is signed between supplier and client based on the amount of delivery and the price of delivery in future, based on the predicted yield. In this paper, machine learning techniques are used to perform prediction of the flowering tree branches that could be used to predict yield in mango trees. Results shows that machine learning techniques could be used to predict the flowering branches.

Categories and Subject Descriptors

I.5.2 [**Computing Methodologies**]: Classifier Design and Analysis, Pattern Analysis.

General Terms

Algorithms, Performance.

Keywords

Machine learning; mango flowering stem prediction; soft computing

1. INTRODUCTION

Harumanis mango is one of the fruit that has high economic demand and potential for Malaysia export business especially the Perlis State in Malaysia. Perlis exported 3.1 metric tons of Harumanis mango to Japan in 2010 and has targeted the export demand to increase to 100 metric tons by 2020. Harumanis mango tree is a yearly fruit bearing trees and reproductive phase of the mango trees often starts from January and ends nearly on June. This type of mango is highly sensitive to the climate and only grows in Perlis and part of Surabaya in Indonesia. It requires a significant dry weather period to initial flowering and the productive phase can be

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.10

significantly affected by change in weather. Although it does grow on Surabaya, the variety that grows in Perlis is often highly valued for export and therefore attracts high foreign earnings. The demand outstrips supply most of the time and there is a need to study and understand the yield cycle in order to accurately predict supply.

Harumanis Mango growth and reproductive phases are illustrated in Figure 1 that depicts the growth and reproductive phases of Harumanis Mango in Perlis associated with the period of months. The vegetative growth period is approximately from July to December. December and January are considered as Pre-Flowering phase where the flower induction process is stimulated. January to February months is the period when the flowers grow and bloom. Fruit bearing occurs during March to April that leads to harvest from May to June.

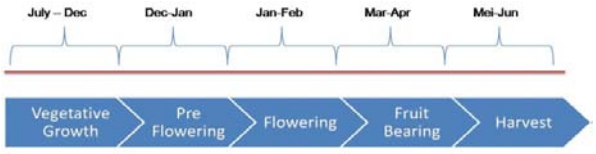


Figure 1. Harumanis Mango Growth and Reproductive Phases

The pre flowering and flowering phases are identified as important stages in the plant reproductive physiology. The Harumanis mango flowering induction event can be influenced by a few factors such as pruning, defoliation and nitrogen fertilizer application [3,4,10]. Since Harumanis mango tree grow in Malaysia, a tropic country, flowering induction is influenced by the climatic factors[2,6,18], and biotic factors[10,12,15]. In tropics climate countries, it is an important factor that the terminal stems of the trees are allowed to rest after the previous vegetative growth, to be able to produce reproductive shoots [11]. This resting time is necessary to provide the stem ample time to mature and grow sufficient whorl, length and diameter.

The mango trees yield predictions are essential to enable forward marketing signed between Malaysia and mango importing countries such as Japan. The forward marketing contract includes the specific mango quantity in tons to be delivered at specific times. A Harumanis mango exporter needs to know the approximate yield before agreeing to terms of the contract. The yield can be approximated once the farmer performs the possible flowering stems prediction.

Machine learning techniques application in agriculture is a relatively new approach for classification and yield prediction in agriculture. There are a few research studies that include machine learning techniques in agricultural domains for yield prediction [1,5], crop classification, crop disease detection [19], management and advisory expert system [7,8,9,13,14,20] .

Machine learning is about learning the structures from data. Machine learning techniques can be used to perform classification and prediction for future observation. Classification is a task of assigning objects to one of several predefined classes to create a classification model, whereas prediction is where the classification model is used to predict the new observations. A classification technique such as k-NN classifier, rule-based classifier and naïve Bayes classifiers employs a learning algorithm to find a model that best suits the relationship between the attributes and the classification categories.

A training set consists of data whose class are known and is used to build a classification model. The classification model is later applied to a test data set, to test the classification accuracy of the model. Evaluation of the classification model is based on the counts of test records correctly and incorrectly classified by the model. The model performance can be displayed using performance metrics determine the level of accuracy.

The performance of the model can be evaluated using several methods such as Hold out Method, Random Subsampling, Cross Validation and Bootstrap [16] . A Cross Validation technique is used to evaluate the performance of the classification models.

In this paper, machine learning techniques are applied to identify the possible flowering tree branches using biotic factors. Five different classifiers performance used to predict the possible flowering branches are compared. The classifiers used are k-Nearest Neighbour (k-NN), Naives Bayes, Support Vector Machine (SVM), Classification Trees (CAT) and Random Forest (RF).

The rest of the paper is outlined as follows; Section 2 discusses the data sets descriptions and method. Results and discussion are presented in Section 3 and the paper ends with the conclusion in Section 4.

2. METHODS

In this paper, the data from Harumanis Greenhouse at the Institute of Agrotechnology, University Malaysia Perlis are used. The biotic data consists of 254 stems of generative and vegetative flushes. The attributes and the data types are given in Table 1:

Table 1. The Attributes and the Data Type of the Biotic Factors of Harumanis Branches

Attributes	Data Type
Lysimeter	Nominal
Length of first whorl	Continous
Length of second whorl	Continous
Length of Third Whorl	Continous
The Diameter of the Third Whorl	Continous
Stem State	Categorical

There are 5 attributes that have been used in this research. These are lysimeter, that describe the type of lysimeter that the trees are planted in, length of the first, second and the third stem whorl (length in mm of the 3 whorl branches) and their diameters.

Lysimeter feature is assigned a v alue between 1 and 3 which represent the root zones of the mango trees. The mango trees are planted in 3 different lysimeter sizes which are micro-lysimeter (1), lysimeter (2) and unrestricted root zones (3). The micro-lysimeter has dimension of 50 cm in deep and a diameter of 50 c m while the other lysimeters size with dimensions of 0.75 m deep and 1.5 m in diameter. Mango trees with unrestricted root zones are planted directly into the ground.

Stem State describes the state of the stem. It is assign a state value of "flowering" or "vegetative". Stem State is the class that will be used by the machine learning techniques to learn while in the training process and build an appropriate model. The test data set is used to validate the models developed to predict the Stem State.

Figure 2 displays the features from the mango stem that are measured and used to classify the flowering stems. The stems whorls length are measured and recorded accordingly.

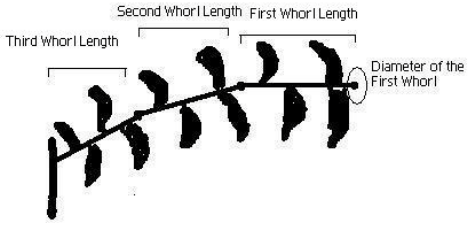


Figure 2. The typical mango terminal stems displaying the three whorls where the measurements are taken from. The whorls are representing the termination of each previous flush of vegetative growth.

Machine learning techniques that were used in this work are k-NN, Naïve Bayes, SVM, CAT and RF. k-NN algorithm is a technique used to classify the new observations based on the closest neighbor's labels. In this technique the distance (similarity) between the test set and the training set is used to determine the nearest-neighbor list. The appropriate number of k (the number of neighbor instances to be compared) is important to be determined. A small k value might cause the classifier to be susceptible to over fitting because of noise in the training data. On the other hand a large k value might cause misclassification because neighbors that are located far from the neighborhood will also be included in the classification decision.

The k- NN models uses Euclidean distance metrics as shown in Equation 1 to get the nearest neighbors

$$d = \sqrt{\sum (x_i - y_i)^2}$$

(Equation 1)

Algorithm for k-NN algorithm

Data : Training Samples D= { $\mathbf{x}_{1:N}$, $\mathbf{c}_{1:N}$}, Test Point **x***.
Result : Class of new point c*

1: Let k be the number of nearest neighbors.
2: for i = 1 **to** N **do**
3: calculate distance d($\mathbf{x}_i$, **x***) between x* and every sample in D;
4: find $\mathbf{x}_j$, for which the distance the smallest, the set of k closest training samples to x*;
5: $y' = \arg\max_{v} \sum_{(x_i, y_i) \in D_{x^*}} I(v = y_i),$
6: **end**

The naïve bayesian classifier algorithm is one of the classification algorithms where an instance is described with n, attributes $a_i (i = 1 \; to \; n)$ and the classified class v from the set of possible classes V is described as follows:

$$v = \arg \max_{v_i \in V} P(v_j) \prod_{i=1}^{n} P(a_i / v_j)$$

(Equation 2)

Compute a vector of elements

$$p_j = P(v_j) \prod_{i=1}^{n} P(a_i / v_j)$$

(Equation 3)

which, after normalization so that the sum of p_j is equal to 1, represents class probabilities. The class probabilities and conditional probabilities (priors) in the above formulae are estimates from the training data: class probability is equal to the relative class frequency, while the conditional probability of attribute value given class is computed by figuring out the proportion of instances with a value of i-th attribute equal to a_i among instances that from class v_j.

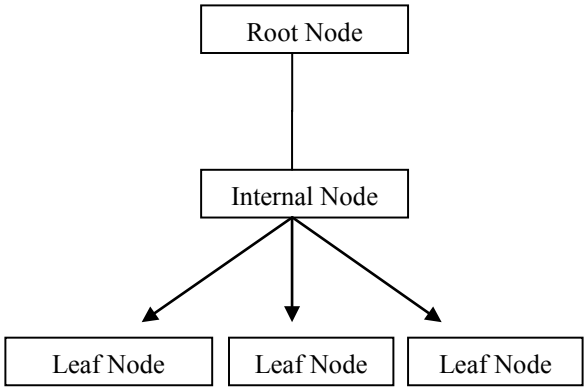
Relative frequency is used when computing prior conditional probabilities. So the total number of training examples is n and nc is the number of training example that has the specific condition. The relative frequency corresponding to the probability would be

$$P = \frac{n_c}{n}$$

(Equation 4)

SVM [17] is one of the techniques that is widely used in classification problems. SVM separates the classes independently with the hyperplane that maximizes the distance from a hyperplane separating the classes to the nearest point in the data set.

CAT is an algorithm that splits the training instances accordingly and builds a tree that consists of root node, internal nodes and leaf nodes as a model to be used to classify the test examples.



RF builds several classification trees to a data set and combines the predictions from all the trees. A classification tree is fitted to each bootstrap samples from the data. At each node, a small number of randomly selected variables are made available for binary partitioning. In random forest each variable that is importance is measured.

The classifier models using machine learning techniques have been built and tested using 10 fold cross validation tests. The k-NN technique has been applied to build a model to classify the flowering and vegetative branches. The Euclidean metrics has been used as learning metrics. The classification accuracy with values from 0 to 1, where 0 is the worst classification and 1 is the best classification has been recorded. Two learner metrics in k-NN technique performance are compared to find the better learner technique that could be used in the classification model. The Learner metrics are Euclidean and Hamming metrics. The results are displayed in Figure 4. Since the Euclidean metrics show better learning ability, this metrics has been used through out the training and testing of the data.

The other techniques, Naïve Bayes, SVM, CAT and RF have also been used to build the classifier models, tested and compared to report the best technique that could predict the flowering stems. The results are displayed and discussed in the following section.

3. RESULT AND DISCUSSION

Figure 3 displays the result of k-NN technique classification accuracy which varies the number of neighbors from 1 to 10, using 10 fold cross validation testing. The highest classification accuracy is achieved using Euclidean metrics for k value from 1 to 7 which is

0.6794. The accuracy decreases for k from 8 to 10, with values from 0.6711 to 0.6589. On the other hand the classification accuracy of Hamming learning metric is lower than that of Euclidean metrics. The highest accuracy is achieved at k =9 where the accuracy is 0.6534.

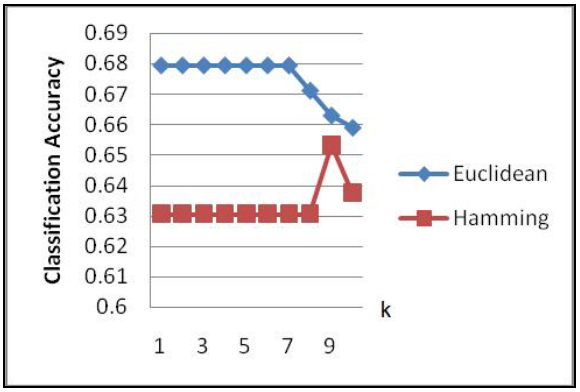


Figure 3. The Classification Accuracy using k-NN (k=1to10) and Different Leaning Metrics

Table 2 displays the classification accuracy for the various classification models in predicting flowering branches. Classification Trees classifier model outperforms the others with 77.95% accuracy rate followed by the SVM classifier model. Naives Bayes classifier model also achieve 72.43% accuracy rate. k-NN and Random Forest achieve less than 70% accuracy.

Table 2. Classififation Accuracy, Sensitivity and Specificity of the Classifier Models

Classifier Model	Classification Accuracy (%)	Sensitivity (%)	Specificity (%)
k-NN	67.94	29.79	82.50
Naives Bayes	72.43	65.96	76.25
SVM	74.86	85.11	68.75
Classification Trees	77.95	76.6	78.75
Random Forest	66.17	26.6	89.38

The sensitivity rates are also given in Table 2 where SVM outperforms the other techniques in classifying or predicting the flowering branches with 85.11% accuracy followed by the Classification Trees at 76.6%. Naives Bayes model achieves a lower Sensitivity rate at 65.96%. The k-NN and Random Forest sensitivity is very low at 29.79% and 26.6%, respectively.

SVM and Classification Trees classifier models outperform the other methods applied in this study in predicting the flowering stems and also the vegetative stem with accuracy levels of more than 70%.

4. CONCLUSION

The development of accurate yield prediction methods is invaluable both to suppliers and the importers. This is more critical in high value crops that can help in communities and government to predict and plan expenditure. The results presented in this paper show that SVM and Classification trees classifier models outperform other methods tested in this study in predicting the flowering stems. The results also demonstrate that the machine learning technique could be used to perform classification on flowering and non flowering stems. This classification algorithm can be used in Decision support system that could predict tree yield every season using biotic factors.

5. ACKNOWLEDGMENT

Special thanks to the UniMAP Agricultural Station, for providing the samples for data collection. This project is funded by UniMAP Short Grant (9001-00423), University Malaysia Perlis. Rohani S. Mohamed Farook acknowledges the sponsorship provided by UniMAP.

6. REFERENCES

[1] Basso, B. Spatial validation of crop models for precision agriculture. Agricultural Systems 68, 2 (2001), 97–112.

[2] Chacko, E.K. Physiology of vegetative and reproductive in mango (Mangifera indica L.) trees. Proc. 1st Australian Mango Research Workshop, (1986), 54–70.

[3] Davenport, T.L., Ying, Z., Kulkarni, V., and White, T.L. Evidence for a translocatable florigenic promoter in mango. Scientia Horticulturae 110, 2 (2006), 150–159.

[4] Davenport, T.L. Reproductive physiology. In: Litz RE (ed), The Mango, Botany, Production and Uses, 2008, 69–146.

[5] Elwell, D.L., Curry, R.B., and Keener, M.E. Determination of potential yield-limiting factors of soybeans using SOYMOD/OARDC. Agricultural Systems 24, 3 (1987), 221–242.

[6] Farook, R.S.M., Aziz, A.H.A., Harun, A., et al. Data Mining on Climatic Factors for Harumanis Mango Yield Prediction. Intelligent Systems, Modelling and Simulation, International Conference on 0, (2012), 115–119.

[7] Fukuda, S., Spreer, W., Yasunaga, E., Yuge, K., Sardsud, V., and Müller, J. Random Forests modelling for the

estimation of mango (Mangifera indica L. cv. Chok Anan) fruit yields under different irrigation regimes. Agricultural Water Management, (2012).

[8] Hernández-Sánchez, C., Luis, G., Moreno, I., et al. Differentiation of mangoes (Magnifera indica L.) conventional and organically cultivated according to their mineral content by using support vector machines. Talanta 97, (2012), 325–30.

[9] Hu, J., Li, D., Duan, Q., Han, Y., Chen, G., and Si, X. Fish species classification by color, texture and multi-class support vector machine using computer vision. Computers and Electronics in Agriculture 88, (2012), 133–140.

[10] Núñez-Elisea, R. and Davenport, T.L. Requirements for mature leaves during floral induction and floral transition in developing shoots of mango. Acta Hortic. 296, (1992), 33–37.

[11] Núñez-Elisea, R. and Davenport, T.L. Flowering of mango trees in containers as influenced by seasonal temperature and water stress. Scientia Horticulturae 58, 1-2 (1994), 57–66.

[12] Núñez-Elisea, R. Effect of leaf age, duration of cool temperature treatment, and photoperiod on bud dormancy release and floral initiation in mango. Scientia Horticulturae 62, 1-2 (1995), 63–73.

[13] Papageorgiou, E.I., Markinos, A.T., and Gemtos, T.A. Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application. Applied Soft Computing Journal 11, 4 (2011), 3643–3657.

[14] Pomar, J. and Pomar, C. A knowledge-based decision support system to improve sow farm productivity. Expert Systems with Applications 29, 1 (2005), 33–40.

[15] Ramírez, F. and Davenport, T.L. Mango (Mangifera indica L.) flowering physiology. Scientia Horticulturae 126, 2 (2010), 65–72.

[16] Tan, P.-N., Steinbach, M., and Kumar, V. Introduction to Data Mining. Addison Wesley, 2006.

[17] Vapnik, V.N. Statistical Learning Theory. John Wiley & Sons, Inc., 1998.

[18] Whiley, A.W. Environmental effects on phenology and physiology of mango-a review. Acta Hortic. 341, (1993), 168–176.

[19] Yang, C.-C., Prasher, S.O., Landry, J.-A., and Ramaswamy, H.S. Development of a herbicide application map using artificial neural networks and fuzzy logic. Agricultural Systems 76, 2 (2003), 561–574.

[20] Zheng, H. and Lu, H. A least-squares support vector machine (LS-SVM) based on fractal analysis and CIELab parameters for the detection of browning degree on mango (Mangifera indica L.). Computers and Electronics in Agriculture 83, (2012), 47–51.

A Text-Oriented Chinese Word Clustering Method Using Latent Dirichlet Allocation

Qiu Lin
School of Computer and Control Engineering
University of Chinese Academy of Sciences
No. 80 East Zhongguancun Road, Haidian District, 100190 Beijing, China
qiulin10@mails.ucas.ac.cn

Xu Jungang
School of Computer and Control Engineering
University of Chinese Academy of Sciences
No. 80 East Zhongguancun Road, Haidian District, 100190 Beijing, China
xujg@ucas.ac.cn

ABSTRACT
Word clustering is a trendy research issue in the field of natural language processing, which is widely used in applications such as document clustering, machine translation, and text classification. In this paper, a text-oriented Chinese word clustering method is proposed. Firstly, Latent Dirichlet Allocation algorithm is used to extract the topics (clusters) from nouns among each sentence in the text. And then the highest probability noun of each topic is selected as the centroids of the k-means algorithm. Based on these selected initial cluster centriods, k-means algorithm is used to calculate the word cluster. Finally, Chinese word similarity calculation method based on HowNet is used to calculate the similarity between each two words in the text. Experimental results on the People's Daily editorial data set show that our proposed method outperforms the graph-based word clustering baseline algorithms using a Web search engine on average similarity.

Categories and Subject Descriptors
I.2.7 [**Artificial Intelligence**]: Natural Language Processing – *language models, language parsing and understanding, text analysis.*

General Terms
Algorithms, Measurement, Experimentation

Keywords
Word clustering, Latent Dirichlet Allocation, K-means, Word similarity

1. INTRODUCTION
In the field of natural language processing, word clustering is a widely studied subject [1]. And it is important for automatic thesaurus construction, text classification, and word sense disambiguation [2]. A typical word clustering task is described as follows: given a set of words (nouns), cluster words into groups so that the similar words are in the same group (cluster). Let us take an example, assume a set of words is given: printer, print, InterLaser, ink, TV, Aquos and Sharp[1]. Apparently, the first four words are related to a printer, and the last three words are related to a TV [3].

Currently, there are a lot of researches on word clustering. The text-oriented word clustering has been widely studied and used in document clustering [4], document classification [5], and large-scale class-based language modeling in machine translation [6].

There are three categories of word clustering: (1) using various heuristic measures to obtain the distance (or similarity) of elements in the clustering process; (2) using statistical model (such as the likelihood function) to obtain the distance between words and the total number of clusters; (3) besides using the statistical model, adding a certain measure (such as perplexity) to control the increasing or reducing number of clustering process [7]. Due to the lack of high-quality Chinese data sets and the ineffectiveness of Chinese automatic segment algorithms, the results of the Chinese word clustering methods based on statistical model are not very satisfactory.

In this paper, we propose a Chinese word clustering method that combines the k-means algorithm and Latent Dirichlet Allocation (LDA) algorithm. K-means clustering [8] is traditionally viewed as an unsupervised method for data clustering, which does not guarantee unique clustering result because the initial clusters are chosen randomly. In order to solve this problem, the LDA algorithm is used to choose the initial *k* centroids and make the k-means more effective and independent on the initial clusters and instance order. LDA algorithm is a generative probabilistic model for collections of discrete data [9], which can mine the implicit topic model from the document or collection. Each sentence in the text is expressed as a word frequency vector, and thus the text message is transformed into one digital model. The topics are extracted from the nouns in each sentence through LDA algorithm, and then the nouns of the highest probability from each topic are chosen as the centroids of k-means. Finally, Chinese word similarity calculation method based on HowNet [10] is applied to calculate the similarity between each two words in the text. And in this paper, only Chinese words are considered.

The remainder of this paper is organized as follows. Section 2 describes the related work. Section 3 introduces the k-means algorithm, the LDA algorithm and Chinese word similarity calculation method based on HowNet. Section 4 explains one word clustering method which combines LDA algorithm and k-means algorithm. Section 5 presents the experimental results based on the People's Daily editorials data set. Finally, section 6

[1] InterLaser is laser printer made by Epson Corp. Aquos is a liquid crystal TV made by Sharp Corp.

concludes the work.

2. RELATED WORK

Word clustering is an important branch of natural language processing techniques, and also has caused wide concerns. At present, the researchers in this field have made some achievements.

Corpus is an instance of natural language, basically in line with the law of the syntax of the language. The basic idea of word clustering method based on corpus is grammatical features that extracted from the target word in the context, and the words which have the similar grammatical features can be in the same cluster [11]. According to this idea, Farhat et al [12] proposed a formal representation of the target word in the context, in which the target word is represented as a binary random variable, the Kullback-Leibler (KL) is used to calculate the distance between two words. Wang [13] proposed a word clustering method that is based on the bilingual parallel corpus word clustering, which combined statistical-based translation model with mutual information clustering algorithm used in single-language. A small-scale word clustering experiment using British-German bilingual parallel aligned corpus shows effective results, but this method based on corpus still has some defects including that suitable corpus is generally not readily available and the results relies on the size of the corpus.

The word clustering based on semantic features usually rely on a semantic knowledge base. These semantic resources are directly coded by language experts, so this method belongs to the rule-based method [11]. Hang Li [14] proposed a method that combining the Minimum Description Length (MDL) principle and the disambiguation method to derive a disambiguation method that makes use of both automatically constructed thesauruses and hand-made thesaurus. Wen Yang et al [15] proposed a bidirectional hierarchical clustering algorithm, which can cluster words in different categories simultaneously. Intuitively, the combination of the relationship between words occurs with a degree of regularity. They proposed the concepts of modifiable degree and modifiable distance in order to solve the problem caused by sparse data. The method based on semantic knowledge has some defects, such as poor results of compound word similarity.

Pragmatic feature refers to the feature of the word that shows in the specific application. The word clustering based on pragmatics feature is an application-oriented word clustering method that extracts the lexical features for a specific application [11]. Yutaka Matsuo [3] proposed an unsupervised algorithm for word clustering based on a word similarity measure by web counts. Each pair of words is queried to a search engine, which produces a co-occurrence matrix. By calculating the similarity of words, a word co-occurrence graph is obtained. A new kind of graph clustering algorithm called Newman clustering is applied for efficiently identifying word clusters.

Blei proposed the original LDA using EM estimation [9]. Griffiths and Steyvers applied Gibbs sampling to estimate LDA's parameters [16]. Since the inception of these works, many variations have been proposed, for example, LDA has previously been used to construct features for classification and reduce data dimension [17].

In this paper, a Chinese word clustering method is proposed, which combines the k-means algorithm and LDA algorithm. We extract the topics from nouns of each sentence through LDA algorithm, and then choose the highest probability noun of each topic as the centroids of k-means clustering algorithm. At last, we use k-means algorithm to cluster all the words in the text. And for the distance formula which in the k-means algorithm, we use the word similarity calculation method based on HowNet to calculate the similarity between each two word in the text.

3. THE EXISTING METHODS OF CHINESE WORD CLUSTERING AND SIMILARITY CALCULATION

3.1 K-means

Clustering is the process of partitioning or grouping a given set of patterns into disjoint clusters, which is done when patterns in the same cluster are alike and patterns belonging to two different clusters are different. The k-means algorithm has been shown to be effective in producing good clustering results for many practical applications.

In this section, we briefly describe the direct k-means algorithm [8]. For a given number of clusters k, k-means firstly selects k cluster centroids stochastically and partitions the objects to the nearest cluster centroid to form a cluster according to the Nearest-Neighbor rule, then computes the mean value of each cluster and makes it the new cluster centroid. The process described above is iterative continually and will be terminated by the error rule function convergence.

The k-means algorithm finds locally optimal solutions for minimizing the sum of distance between each data point and its nearest cluster centroid, which is equivalent to maximizing the likelihood given in the assumptions listed above. Note that k-means is defined over numeric (continuous-valued) data since it requires the ability to compute the mean. On the other hand, words cannot be represented with numeric data, so the distance between the words cannot be measured by Euclidean distance. In this paper, we use the results of LDA algorithm to get the centroids of k-means. And we use the method of Chinese word similarity calculation to measure the distance between the words, Chinese word similarity calculation method will be discussed in subsection 3.3.

K-means algorithm is one of the most widely used clustering algorithms in spatial clustering analysis. It is easy and efficient, but it also has some limitations: (1) it is sensitive to the initialization; (2) it does not perform well in global search and it is easy to get into local optimization. Random initial centroids could lead to different clustering results. And the more appropriate the initial centroids are chosen, the better the clustering results are [18]. So in this paper, Latent Dirichlet Allocation algorithm is used to get the initial seeds to perform semi-supervised learning. Through the preliminary clustering of LDA, we can get a better selection of initial centroids, which can improve the similarity of the clustering results.

3.2 Latent Dirchlet Allocation

Latent Dirichlet Allocation (LDA) [9] is a generative probabilistic model for processing collections of discrete data such as text corpus, which has quickly become one of the most popular

probabilistic text modeling techniques. LDA uses the bag of words model, which considers each document as a word frequency vector. The graphical model of LDA is shown in Figure 1.

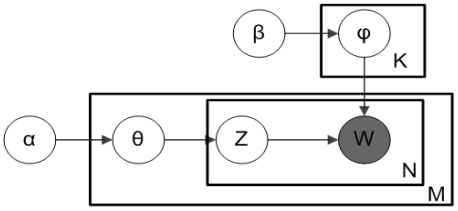


Figure 1. Graphical model of LDA

LDA could be described as follows [16]:

Word: a basic unit defined to be an item from a vocabulary of size W.

Document: a sequence of N words denoted by $d = (w_1, \ldots, w_n)$, where w_n is the n^{th} word in the sequence.

Corpus: a collection of M documents denoted by $D = \{d_1, \ldots, d_m\}$.

Given D documents is expressed over W unique words and T topics, LDA outputs the document-topic distribution θ and topic-word distribution ϕ. This distribution can be obtained by a probabilistic argument or by cancellation of terms in Equation 1:

$$p(z_i = j \mid z_{-i}, w) \propto \frac{n_{-i,j}^{(w_i)} + \beta}{\sum_{w'}^{W} n_{-i,j}^{(w')} + W\beta} \frac{n_{-i,j}^{(d_i)} + \alpha}{\sum_{j}^{T} n_{-i,j}^{(d_i)} + T\alpha} \quad (1)$$

where $\sum_{w'}^{W} n_{-i,j}^{(w')}$ is a count that does not include the current assignment of z_j. The first ratio denotes the probability of wi under topic j, and the second ratio denotes the probability of topic j in document di. Critically, these counts are the only necessary information for computing the full conditional distribution, which allow the algorithm to be implemented efficiently by caching the relatively small set of nonzero counts. After several iterations for all the words in all the documents, the distribution θ and distribution ϕ are finally estimated using Equation 2 and 3.

$$\phi_j^{(w_i)} = \frac{n_j^{(w_i)} + \beta}{\sum_{w'}^{W} n_j^{(w')} + W\beta} \quad (2)$$

$$\theta_j^{(d_i)} = \frac{n_j^{(d_i)} + \alpha}{\sum_{j}^{T} n_j^{(d_i)} + T\alpha} \quad (3)$$

Many applications are based on LDA algorithm, such as clustering, deduction, forecast and so on. LDA for clustering and k-means are both unsupervised clustering methods that do not require any training corpus. Therefore, this method gets rid of the negative effect of the low-quality corpus. And on the other hand, we add a dictionary for the data set into the original segment system, which improves the accuracy of segment results.

3.3 Similarity Calculation Based on HowNet

In order to calculate the distance between words in k-means algorithm, we use the word similarity instead of the distance. The higher the similarity between two words is, the closer the two words are.

HowNet [10] is an on-line common-sense knowledge base unveiling inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents. The concept can be divided into a number of sememes, which are the smallest basic semantic units that cannot be reduced further. Some sememes form a complex network structure. Because HowNet's words contain the semantic information, we can also take full account of its semantic information when we want to compute the similarity between Chinese words.

The similarity calculation relies on the similarity between two sememes. The calculation formula of the similarity between two sememes [19] is Equation 4.

$$Sim(p_1, p_2) = \frac{2 \times Spd(p_1, p_2)}{Dsd(p_1, p_2) + 2 \times Spd(p_1, p_2)} \quad (4)$$

$Spd(p_1, p_2)$ refers to the path length of the parent node which the two sememes p_1 and p_2 shared in the sememes hierarchy system. $Dsd(p_1, p_2)$ describes the length of the shortest path where p_1 and p_2 move upward gradually along parent nodes until they reach the second shared node. And the concept similarity formula [18] between two concepts C_1 and C_2 is Equation 5.

$$Sim(C_1, C_2) = \beta_1 Sim_1(C_1, C_2) + \sum_{i=2}^{4} \beta_1 \beta_i Sim_i(C_1, C_2) \quad (5)$$

$\beta_i (1 \le i \le 4)$ denotes the similarity of the first basic sememes, the other basic sememes, the relational sememes and the signal semems respectively, which are the adjustment parameters limited by $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1, \beta_1 \ge \beta_2 \ge \beta_3 \ge \beta_4 > 0$.

4. ONE IMPROVED CHINESE WORD CLUSTERING METHOD

4.1 The Formulation of the Algorithm

In this paper, we mainly discuss the Chinese similar word clustering in one text. As everyone knows, the traditional k-means clustering algorithm is a supervised learning algorithm, and the different choice of initial centers will bring great impact on the results, so a new method that combines the Latent Dirichlet Allocation (LDA) algorithm and k-means clustering algorithm is proposed. The main steps of this algorithm are listed in Figure 2.

Algorithm: Improved similar Chinese word clustering method

1: **for** each d_i in document set **do**
2: SS←sentence segment
3: WS←word segments for each sentence
4: NV←vectors expressed by nouns in each sentence
5: NL←noun list including the different nouns in the text
6: SeedSet ←extract the topics by LDA algorithm (NV_{i1}, NV_{i2}, … , NV_{in})
7: Cluster←k-means (centroids initialized by SeedSet, NL_i)
8: end **for**

Figure 2. The main steps of the algorithm

For each document in the document set, some pre-process work need to be done firstly (Line 2 - Line 5). Secondly, and then some topics are extracted by LDA algorithm and the highest probability nouns of each topic are chosen as the centroids of k-means (Line 6). At last, k-means algorithm is used to cluster all the words in the text (Line 7).

4.2 Baseline Algorithms

Recently, a series of effective graph clustering methods has been proposed. Pioneering work that specifically emphasizes edge betweenness was done by Girvan and Newman [3], we call the method as GN algorithm.

Yutaka Matsuo [3] proposed an unsupervised clustering algorithm for word clustering based on a word similarity measure by web counts. Each pair of words is queried to a search engine, which results in a co-occurrence matrix. By calculating the similarity of words, a word co-occurrence graph is created. GN algorithm is a new king of graph clustering algorithm. GN algorithm emphasizes betweenness of an edge and identifies densely connected sub-graphs.

The method to measure semantic similarity between words or entities using Web search engines2 has been introduced by many papers. Web search engines provide an efficient interface to this vast information. Page counts for the query P and Q, can be considered as an approximation of co-occurrence of two words (or multi-word phrases) P and Q on the Web [19]. And the most common methods of semantic similarity between words based on Web search engines are PMI, Jaccard, Overlap and Dice, which are defined as Equation 6-9 [20]. And H(P) stands for the page counts for the query P in a search engine.

$$PMI(P,Q)=\begin{cases}0 & if\ H(P\cap Q)\le c\\ \log_2\left(\dfrac{\frac{H(P\cap Q)}{N}}{\frac{H(P)}{N}\frac{H(Q)}{N}}\right) & otherwise\end{cases} \quad (6)$$

$$Jaccard(P,Q)=\begin{cases}0 & if\ H(P\cap Q)\le c\\ \dfrac{H(P\cap Q)}{H(P)+H(Q)-H(P\cap Q)} & otherwise\end{cases} \quad (7)$$

$$Overlap(P,Q)=\begin{cases}0 & if\ H(P\cap Q)\le c\\ \dfrac{H(P\cap Q)}{\min(H(P),H(Q))} & otherwise\end{cases} \quad (8)$$

$$Dice(P,Q)=\begin{cases}0 & if\ H(P\cap Q)\le c\\ \dfrac{2H(P\cap Q)}{H(P)+H(Q)} & otherwise\end{cases} \quad (9)$$

[2] *http://www.yahoo.cn* is used.

Here, N is the number of documents indexed by the search engine. In this paper, we set $N=10^{10}$ according to the number of indexed pages reported by Google. And we set $c=500{,}000$ in our experiments. In this paper, these four algorithms and k-means algorithm are chosen as baseline algorithms.

4.3 Text Preprocessing

In order to get the input of LDA algorithm, some preprocessing work should be done for the text.

Firstly, segment one text into some sentences, and segment each sentence into some words. Secondly, extract the nouns in the sentence according to the part-of-speech of these words. Finally, construct one vector with the nouns in each sentence and noun list including the different nouns in the text.

4.4 Improved Chinese Word Clustering Method

LDA can be used to convert word dimension of document into topic dimension. So LDA could be re-defined as follows:

Word: a basic unit defined to be an item from a vocabulary of size W.

Sentence: a sequence of N words denoted by $s=(w_1,...,w_n)$, where w_n is the n^{th} noun in the sentence.

Document: a collection of M sentences denoted by $D=\{s_1,...,s_m\}$.

LDA includes a process of generating the topics in each document, which greatly reduces the number of parameters to be learned and provides a clearly-defined probability for arbitrary documents. In collapsed Gibbs sampling, only z_{ij} is sampled, and the sampling is done conditioned on α, β and the topic assignments of other words z_{ij}.

As the output of LDA algorithm, k topics for the text can be got, and then the highest probability nouns of each topic are chosen as the initial cluster centroids of k-means algorithm.

5. EXPERIMENTATION

5.1 Data Set

In this paper, we adopt 516 People's Daily editorials from year 2008 to 2010 as the experimental date set. We manually divided these 516 editorials into five categories: politics (183 editorials), economy (114 editorials), culture (56 editorials), people's livelihood (91 editorials), science and technology (72 editorials).

5.2 Experimental Design

Due to the enormous amount of data, we just labeled 10% of the data sets, which includes 18 editorials of politics, 11 editorials of economy, 5 editorials of culture, 9 editorials of people's livelihood and 7 editorials of science and technology. We choose k (topic number) as valued from 5 to 15, and obtain the precision for different k, which is shown in Figure 3. From the results, we find that the best precision can reach 61.2% when k is 8. And the values of α and β are set as 0.2 and 0.1 respectively.

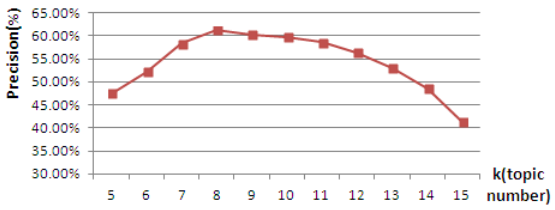


Figure 3. The precision for different value *k*

According to the results given by two methods of the word clustering, we will use the HowNet to calculate the average similarity of every cluster. And the average similarity of clusters is defined as Equation 10, where n stands for the number of clusters.

$$AveSim = \frac{\sum_{i=1}^{n} average\ similarity\ of\ cluster\ i}{n} \quad (10)$$

And the average similarity calculation expression of cluster i is defined as Equation 11, where m stands for the number of words.

$$ASC_i = \frac{\sum_{i=1}^{m}\sum_{j=i+1}^{m} similarity\ between\ word_i\ and\ word_j}{m*(m-1)/2} \quad (11)$$

In order to calculate the similarity between clusters, we choose the centriods that represent every cluster in IKL and k-means. And for GN algorithms, we choose a word (we also call the word as centroid) from every cluster and guarantee that the average similarity between the centroid and the other words in the same cluster is the highest. The calculation of the similarity between clusters is simplified to the calculation of the similarity between centroids. And the average similarity between clusters is defined as Equation 12, where n stands for the number of clusters.

$$ASBC = \frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n} similarity\ between\ centroid_i\ and\ centroid_j}{n*(n-1)/2} \quad (12)$$

5.3 Exprimental Results

Table 1 shows the average similarity of the method that combines k-means and LDA (we call the method as IKL), k-means and GN algorithms used in a Web search engine.

Table 1. The average similarity of IKL, k-means and GN algorithms

	politics	economy	culture	people's livelihood	science and technology
IKL	0.349	0.412	0.392	0.380	0.404
K-means	0.271	0.328	0.317	0.230	0.351
GN and PMI	0.255	0.346	0.311	0.258	0.360
GN and Jaccard	0.300	0.396	0.345	0.293	0.400
GN and Overlap	0.293	0.374	0.316	0.317	0.377
GN and Dice	0.281	0.382	0.32	0.304	0.382

The comparison results of six algorithms are shown in Figure 4.

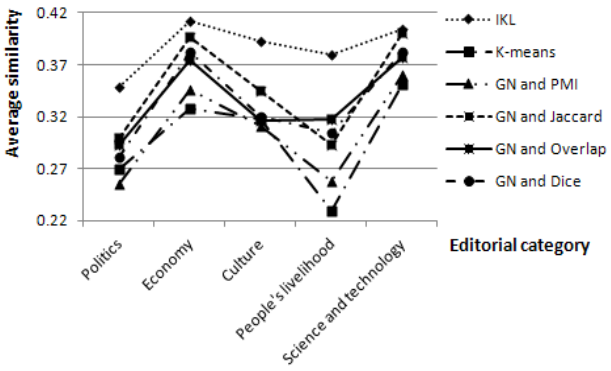


Figure 4. The comparison resluts of six algorithms

From Figure 4, we can see that the average similarity calculated by IKL is higher than the other algorithms, which means that using IKL can get better clustering effect within the cluster than the other algorithms.

The result of the average similarity between the clusters (centroids) is shown in the Figure 5.

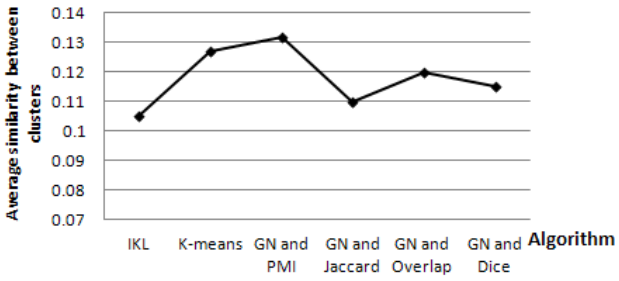


Figure 5. The average similarity between the clusters

From Figure 5, we can see that the average similarity between clusters of IKL is lower than the other algorithms, which means that using IKL can get better discrimination effect between the clusters than the other algorithms.

Also, the time consumption of GN algorithm using Web search engine is enormous. We set the nouns after pre-processing on a server[3] to obtain the web counts for each word and the web counts for pairs of words using a search engine. It takes about ten days to get all web counts. And on the other hand, IKL takes only about 4 minutes to get the result.

6. CONCLUSIONS

In this paper, one Chinese word clustering method, we call the IKL method, is proposed. We re-define the concept of word, document and corpus as word, sentence and document in LDA algorithm. LDA algorithm is used to choose the initial centroids of k-means, which is used to extract the topics (clusters) from nouns of each sentence in the text. It can make up the defect that k-means randomly selects k of the objects. And the result of the

[3] The server's configuration: 2.03 GHz E5606 Intel(R) Xenon(R) processor, 4GB DDR3 RAM and 500 GB SATA Hard Disk.

centroids of cluste rs obtained by LDA algorithm is relatively close to the final re sult of k-means clustering, thus increasing the Based on the People's Da ily editorial data set, experimental results show that both the average similarity of our propos ed algorithm and the time consumption are better than the k-means algorithm and four graph-based word clustering algorithms.

7. ACKNOWLEDGMENTS

This work is supported in part by the National Key Technology R&D Program of China under Contract No. 2012BAH23B03.

8. REFERENCES

[1] Sun, M.S., Zuo, Z.P., and Tsou, B.K. 2000. Part- of-speech identification for unknown Chines e words based on K-Nearest-Neighbors strategy. *Chinese Journal of Compu ters.* 23, 2 (Feb. 2000), 166-170.

[2] Khaled, A., David, M., and Nathan, S. 2002. An effcient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intellgence.* 24, 7 (Jul. 2002), 43-48.

[3] Yutaka, M. and Takeshi, S. 2006. Graph-based word clustering using a web search engine. In *Proceedings of the 2006 Conference on Empiric al Methods in Natur al Language Processing* (Sydney, Australia, July 22-23, 2006), 542-550.

[4] Noam, S. and Naftali, T . 2000. Document clus tering using word clusters via the information bottleneck method. In *Proceeding of the 23rd Annual ACM SIGIR conference* (Athens, Greece, July 24-28, 2000), 208-215.

[5] Li, H. and Naoki, A. 1998. Word clustering and disambiguation based on co-occurrence data. In *Proceedings of the 17th Inter national Conference on Computational Linguistics* (Montreal, Quebec, Canada, August 10-14, 1998), 749-755.

[6] Wen, Y., Yuan, C.F., and Huang, C.N. 2000. Clus tering of Chinese adjectives-Nouns based on c ompositional pairs. *Journal of Chinese Information Processing*. 14, 6 (Jun. 2000), 45-50.

[7] Thomas, L.G. and Mark, S. 2004. Finding scientific topics. In *Proceedings of the National Academy of Sciences* , 101 (Apr. 2004), 5228-5235.

[8] Li, F.F. and Pietro, P. 2005. A bayesian hierarchical model for learning natural scene categories. In *Proceedings of IEEE Computer Society Conference on Compter Vision and Pattern Recognition* (San Diego, CA, USA, June 20-26, 2005), 524–531.

[9] Zhang, Z., Zhang, J.X., and Xu e, H.F. 2008. Improved k-means clustering algorithm. In *Proceedings of 2008 International Congress on Image and Signal Pro cessing* (Sanya, Hainan, China, May 27-30, 2008), 169-172.

[10] Liu, Q. and Li, S.J. 2002. Word similarity computing based on Hownet. *Computational Linguistics and Chines e Language*. 7,2 (Aug. 2002) , 59-76.

[11] Danushka, B., Yutaka, M., and Mitsuru, I. 2007. Measuring semantic similarity between words using web search engines. In *Proceedings of the 16th International Conference on World Wide Web* (Banff, Alberta, Canada, May 08-12, 2007), 757-766.

[12] Baker, L.D. and Andrew, K.M. 1998. Distributional clustering of words for text classification. In *Proceedings of the 21st Annual ACM SIGIR Conference* (Melbourne, Australia, August 24-28, 1998), 96-103.

[13] Jakob, U. and Thorsten, B. 2008. Distributed word clustering for large scale class-based language modeling in machine translation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics* (Columbus, Ohio, USA, June 16-17, 2008), 755-762.

[14] Hu, H.P., Zeng, Q.R., and Lu, S.F . 2006. Research on Chinese word clustering. *Computer Engineering and Science*. 28,1 (Jan. 2006), 122-124.

[15] Hartigan, J.A. and Wong, M.A. 1979. A k-means clustering algorithm. *Journal of the Royal Statistical Society*. 28, 1 (Jan. 1979), 100-108.

[16] David, M.B., Andre w,Y.N., and Michael, I.J. 2003. Latent dirichlet allocation. *Journal of Machine Learning*. 3 (Mar. 2003), 993-1022.

[17] Dong, Z.D. and Dong, Q. 2011. Hownet. http://www.keenage.com/, 2011-04-10.

[18] Guo, H.E., Zhu, L.J., and Xu,S. 201 0. A survey on word clustering technique. *Digital Library Forum*. 5 (May. 2010), 14-18.

[19] Farhat, A., Isabelle, J.F., and O'Shaughnessy, D. 1996. Clustering words for statistical language models based on contextual word similarity. In *Proceedings of the Fourth International Congerence on Spoken Languge Pro cessing* (Philadelphia, USA, October 03-06, 1996), 180-183.

[20] Wang, Y.Y., Lafferty , J., and Waibel, A. 1996. Word clustering with parallel s poken language corpora. In *Proceedings of the Fourth International Conference on Spoken Language* (Philadelphia, USA, October 03-06, 1996), 2364-2367.

An XMPP Based Service Framework with A Telecare Application

Feng-Cheng Chang
Dept. of Innovative Information and Technology
Tamkang University, TAIWAN
135170@mail.tku.edu.tw

Hsiang-Cheh Huang
Department of Electrical Engineering
National University of Kaohsiung, TAIWAN
hch.nuk@gmail.com

ABSTRACT
Telecare is the term for providing remote care to elderly people, babies, and/or less able people. Different implementation approaches would be used for deploying a telecare application. In this paper, we design a software framework that is suitable for prototyping a telecare application. The framework is programming language independent. We may develop application modules using different languages and make them work together. The framework is also a realization of the service oriented architecture (SOA). Therefore we may integrate distributed components into an application. At the end of the paper, we choose a telecare scenario as the demonstration. The analysis, design, and deployment structure are described to show the effectiveness and feasibility of the framework.

Categories and Subject Descriptors
H.4 [**Information Systems Applications**]: Miscellaneous

General Terms
Design

Keywords
Extensible Messaging and Presence Protocol, XMPP, service-oriented architecture, SOA, telecare.

1. INTRODUCTION
The trend of prolonging human life expectancy indicates that there are more and more elderly people. Elderly people may need different assistant technologies to make their life better. In addition to the utilities that helps them physically, monitoring technology around them would reduce the safety concerns. Without any monitoring device, the caretaker stays nearby the elderly people, keeps an eye on them, and inevitably interferes their daily life. A solution to this situation is to adopt the telecare technology. In addition to the elderly people, telecare can be applied to a wider range. It is the concept of offering remote care to help old and physically less able people [14]. There have been different researches related to this concept. Some of them focus on the network [5], some on the data analysis [9], some on the software model [3] [1], etc.

In a telecare system, sensors may be integrated to monitor the environment or the biomedical conditions. Depending on the deployment constraints, a telecare system may be implemented in different forms. For example, it would be a dedicated hardware that detects fall event and automatically sends short messages to the caretaker's mobile phone; or a computer software that continuously analyzes the frames from the video recorder and sends fall alerts to the caretaker's messenger program. If we would like to build a prototype telecare application, using the commonly available computing devices and networks is a better approach in that we can easily modify and deploy the experimental components.

According to the properties of telecare applications, distributed communications among components are frequently encountered. Therefore, a telecare application can be viewed as the integration of necessary distributed service components. This is a good match to the concept of service-oriented architecture (SOA). We will discuss an SOA implementation based on the extensible messaging and presence protocol (XMPP).

The paper is organized as follows. In Sec. 2, we briefly describe the idea of prototyping a telecare application as distributed services. In Sec. 3, we describe the detailed design of the XMPP-based framework. Then we use a telecare scenario for demonstrating how to develop the application with this framework in Sec. 4. At the end, we conclude our work in Sec. 5.

2. MOTIVATION
As mentioned in Sec. 1, telecare technology is important for elderly people. Generally speaking, we may deploy different kinds of sensors that detect different events. When a potentially dangerous event occurs to the elderly people, the system notifies the caretaker to handle the event. By the help of telecare system, different caretakers would be involved: a professional person, a family member, or another helping system. No matter they reside locally or remotely, they are associated with the elderly people by the system.

Commercial telecare systems are usually implemented with proprietary network and sensor hardware. Sometimes such a high-precision or high-quality telecare system is not affordable. Therefore, we need an alternative architecture for prototyping the applications and/or reducing the costs. Fortunately, PC, wired and wireless LAN, and hand-held devices are commonly available. What we need is a software framework for developing the applications. The framework should be a platform for hosting distributed services (including sensors), and it is better if the framework is independent of programming languages. In our previous study [2], we proposed that a protocol centric framework is neutral to the underlying implementation. When the proper protocol is chosen, a service-oriented architecture can be

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.12

deployed without concerning much about the programming languages. In the following sections, we will explain the detailed framework design and how to use it as the infrastructure of a telecare application.

3. FRAMEWORK

In this section, we describe the design of the XMPP based service framework. First of all, we briefly introduce the general concepts of SOA and XMPP in Sec. 3.1. In Sec 3.2, we discuss the design of the fundamental service which is shared among all the derived services.

3.1 SOA and XMPP

As described by I. Sommerville, "*Service-oriented architectures (SOAs) are a way of developing distributed systems where the system components are stand-alone services, executing on geographically distributed computers* [13]." In terms of software design, it is the concept that an application is composed by several smaller pieces called services. It emerges as a development approach as a result of network technology improvement and wide availability. Generally speaking, a service is a standalone software component that exchanges information with the other services to accomplish a certain task. An application in this sense includes the algorithm to locate all the necessary services, to associate services together, and start the data or event flow to accomplish the task.

SOA can be implemented in different forms, such as the mature Common Object Request Broker Architecture (CORBA) [11] or web-based services. To integrate services, a broker for centralized management or agent-based technology for distributed management is necessary. The service management provides the functionality for registering and discovering a service. When dealing with distributed services, some standard service description specification should be used. For example, Web Services Description Language (WSDL) [8] is an XML based document for describing the features of a service. The matching algorithm of a service is not standardized, and depends on the design of the application.

Extensible Messaging and Presence Protocol (XMPP) [12] is an open specification for delivering messages and multi-casting presence status. It was developed as a messenger protocol called jabber. In an XMPP network, a client connects to a server and uses an e-mail like string (called bare JID) for authentication. The specification allows multiple logins of the same client, and thus the client may suggest an additional resource string for distinguishing different logins. Once the login process succeeds, the server responds the complete identifier (full JID) to the client. Note that the suggested resource string may be overridden by the server as long as the result full JID is unique. The formal identifier of a client is the string `<client_id>@<server>/<resource>`. When sending a message, the flow of the message goes through the source client, source client's server, the target client's server, and the target client. The connections among servers form a server federation. An XMPP server is required to deliver the message as soon as possible, and this is sometimes referred to be the real-time property of the XMPP.

When an incoming message is addressed by a bare JID, the server chooses the default login instance as the target. To determine the default message receiver, a priority is associated with each login. The integer values from 0 to 127 are the potential receiver, and the login with the highest one wins. The integer values from -1 to -128 can only be addressed by full JID. It is possible to change the priority at run-time, which implies that we can dynamically switch the default message receiver if necessary.

3.2 The Fundamental Service

As the open-source community and some software vendors adopted the XMPP core specification, more and more extensions have been added to support various functionalities, such as binary object transfer, service discovery, audio/video streaming, etc. In our previous work [2], we found that XMPP can be used as the underlying communication channel for SOA. In this section, we revise part of the design according to the practical implementation experiences. Because services vary from domain to domain, we only describe the fundamental service that is shared by all the derived services.

A service can be implemented in three forms: as a normal client (called a bot), as a component, and as a server extension (plugin). If performance is critical, the choice is server extension. However, it is not flexible because a server extension is based on the server implementation. Our goal is to develop a framework that allows integration of standalone services. Therefore, the former two approaches are better for this purpose. A client and a component are similar in that they should be authenticated before being available to the others. Most of the data exchange mechanisms, such as messaging and events, are handled in the same way. One of the differences is that a component has its own sub-domain and do not have a roster (buddy list). With a sub-domain on a server, a component becomes more accessible by the other services. A component is also trusted by the server, and has permissions to control restricted resources. Therefore, it is suggested that a service is developed as client-based and later migrated to component-based when necessary. In the following paragraphs, we focus on the client-based design because it meets the core requirements of the framework.

We need to enable the discovery mechanism for every service because it should be discoverable by the others. The XMPP Extension Protocol 0030 (XEP-0030) provides the functionality of discovery protocol. A client may request a list of available entities on the server, and queries the features announced by each entity. A feature is a free-format string, and we can use a feature to describe one or more capabilities of the service. An identifier to locate the run-time service instance is required. Although the full JID is the natural service identifier in an XMPP network, the resource string is determined by the server, not the application. In other words, a client may not have the same full JID for every session. A reliable way is to use a feature string as the application-specified identifier of the service instance, and then determine its full JID after the discovery process.

Below we summarize the procedure to instantiate a fundamental service:

1. Take the login JID and password to connect to the server.
2. Turn auto-subscription on to make its presence automatically available to the others.
3. Enable the XEP-0030 discovery mechanism.
4. Announce the bot identifier as a feature.
5. Wait for the session start event.

6. Send the initial presence with priority -1 so that it never receives messages addressed by the bare JID.
7. Fetch the initial roster so that it is able to discover features and receives presence events from the other entities.

To extend the fundamental service to a specific service, we may inherit or compose the fundamental functionalities into the specific service. After the initial setup process, the channel to the XMPP server is established. The service invokes the discovery functionality to identify all the necessary services, and collaborates with them using the supported messages. The detailed implementation of a service depends on the XMPP library we choose. Therefore the issues, such as event handling and message delivery, are out of the scope of this paper.

4. APPLICATION EXAMPLE

Based on the service framework described in Sec. 3, we may develop various kinds of services that communicate over the XMPP network. In this section, we describe how a telecare application is developed on the framework from scenario to implementation.

4.1 Scenario

Suppose we would like to build a telecare environment prototype without dedicated hardware. To achieve the goal, we need to write software that communicates with commonly available computer peripherals and electronic devices. The telecare environment helps the caretaker (e.g., a family at work) to monitor and remind the elderly people at home. The application provides the following functionalities:

- Fall detection in the bathroom and the living room.
- Reminder for taking medicine.
- Notify the caretaker when a fall is detected or the medicine is not taken.

We assume that the caretaker may choose a PC or a smart phone to receive notifications. Therefore we have limited computation capability on the receiver side.

4.2 Analysis and Design

Based on the aforementioned scenario, we retrieve the system requirements and design the software architecture. The functionality descriptions can be divided into four pieces: fall detection, reminder, notification sender, and notification receiver (the caretaker).

For fall detection, a component that monitors human body movement is required. There are many detection algorithms depending on different kinds of sensors. For instance, some algorithms are based on accelerometer [6], some are based on video analysis [7], and some are based on image analysis [10]. In this example, we choose image analysis approach for demonstration purpose. By this approach, one of the issues is when to take a shot. We think a sensor that detects abrupt movements is suitable for elders' fall event. When a possible fall event occurs, it triggers the camera to take a shot for analysis. If the analysis result is positive, it triggers the notification sender to notify the caretaker.

For the reminder part, we need a scheduler that emits an event at the given time. The message for registering a timed event contains the following fields:

- `((<start time>,<interval>,<stop time>), <message>)`
  - start time (inclusive): string `<YYYY-MM-DD hh:mm:ss>`. If `<start time>` is `None`, it indicates the current time.
  - interval: string `<dd:hh:mm:ss>`, where dd represents the number of days. If `<interval>` is `None`, it indicates that the message will be sent once.
  - stop time (exclusive): string `<YYYY-MM-DD hh:mm:ss>`. If `<stop time>` is `None`, it indicates that there is no limit on the number of times the message should be sent.

The event is actually a kind of activation signal. Therefore we may re-use the notification mechanism to activate the reminder and render the message. After the medicine is taken, the reminder message should be stopped. The issue here is that how to detect the take-medicine condition. We assume that the medicine is stored in a box, and use the open-close of the lid to represent the condition. If the medicine is not taken for a given timeout, the reminder message is stopped and a notification is sent to the caretaker.

The notification sender and receiver can be accomplished using the popular publisher-subscriber pattern. The sender accepts the subscriptions from the clients. A subscription contains a tuple (event, subscriber) as the request. When the given event is ready, the sender multi-casts the event information to all the subscribers. The notification sender also has an event-receiving interface for other components to deliver events.

4.3 Integration

In this example, we need to integrate pure-software components and sensors. To make a sensor accessible in the service framework, we connect the physical device to a computer and write a wrapper service that communicates with the sensor to obtain the readings. The connection between a sensor and the computer would be wired or wireless, depending on the deployment requirements. Here we assume the connections are wired and it is sufficient for demonstration purpose. In the implementation, we connect the sensors to the Arduino Mega 2560 board. The program uploaded to Arduino board translates the sensor readings to a value recognized by the corresponding wrapper. The wrapper then communicates with the board through the USB (emulated RS232) port. The software services are implemented with the Python SleekXMPP [4] library.

The implemented service components are list below:

- Movement service: detects abrupt body movement.
- Image capture service: captures an image on-demand.
- Fall image analysis service: fetches an image from the webcam, analyze it, and sends a notification if fall is detected.
- Scheduler service: sends a message to the registered service at a specific time.

- Reminder service: produces beeps on request and sends notification if timeout.
- Button service: detects box lid open-close and sends messages to stop the reminder.
- Notification service: receives an event and multi-casts it to the subscribers.

To form an XMPP network, at least one server should exist to serve the message delivery. Instead of setting up our own server, we use the google-talk server in this example. The deployment of the services could be as shown in Fig. 1. Due to the limit of the picture view, we only show the configuration of the body movement sensor with the webcam and the taken shot in Fig. 2.

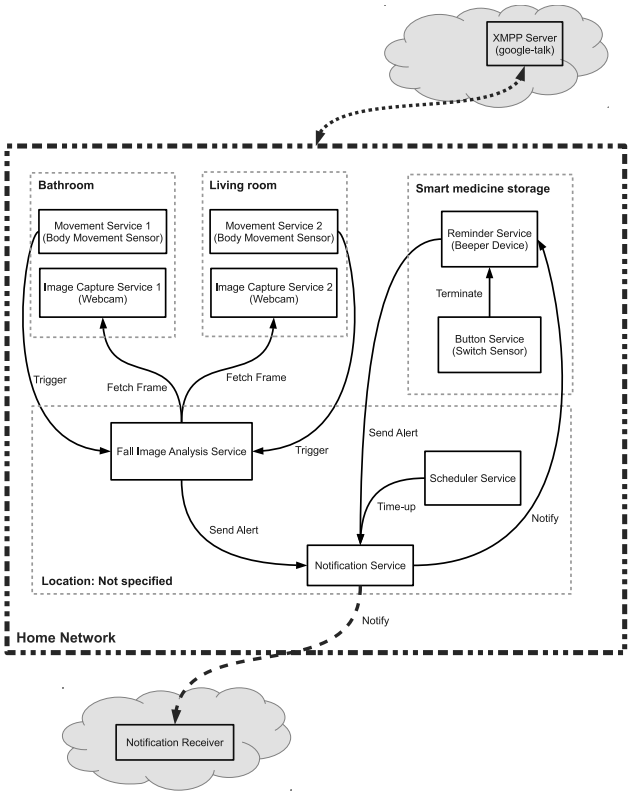


Figure 1: Deployment of the telecare application

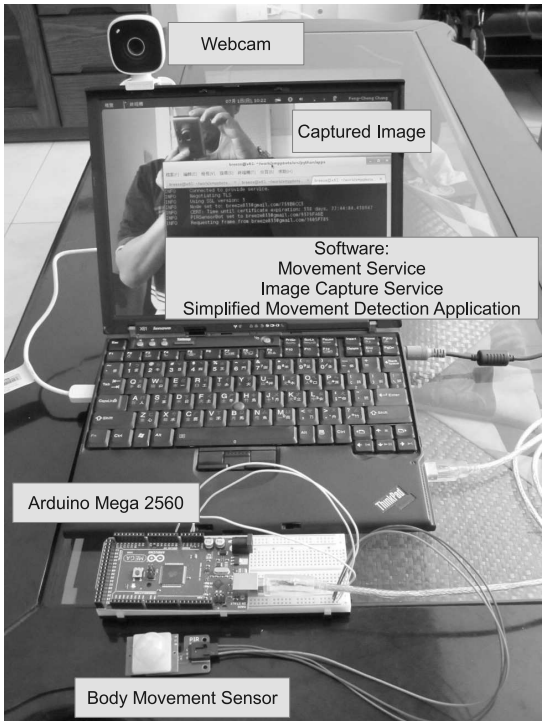


Figure 2: Configuration of simplified movement detection application

5. CONCLUSION

The XMPP specifications are open and can be used to develop efficient message exchanging programs. The SOA concept becomes more and more popular because many modern software applications integrate distributed components. By implementing SOA using XMPP, we get a flexible and efficient framework for developing distributed applications.

In this paper, we revised and extended the design of our previous study. A detailed fundamental service implementation for the XMPP based framework is described and implemented. To show the effectiveness of the framework, we chose a telecare application scenario, analyzed the requirements, identified the necessary services, and implemented the hardware/software design. A deployment structure and a physical implementation picture were also provided to demonstrate the feasibility of the design. In the telecare application, a few services were included: a fall detection function that is composed by a movement service, an image capture service, and a fall analysis service; a reminder function that integrates scheduler and beeper services; and some other assistant services such as the notification service. The flexibility is shown in two aspects: the service implementation is independent of programming languages; and the important XMPP network entity, the server(s), could a third-party service such as google-talk.

6. ACKNOWLEDGMENTS

This work was partially supported by the NSC, Taiwan, under Grants NSC 99-2221-E-032-050 and NSC 101-2221-E-032-052. The hardware devices used in the projects were partially supported by Tamkang University, Taiwan, under grant of the subsidiary project for Technology and Society (year 2011).

7. REFERENCES

[1] D. Berian, V. Gomoi, and V. Topac. A hybrid solution for a telecare system server. In *Applied Computational Intelligence and Informatics (SACI), 2011 6th IEEE International Symposium on*, pages 589-592, May 2011.

[2] F.-C. Chang and D.-K. Chen. The design of an XMPP-based service integration scheme. In *IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2011)*, pages 33-36, Dalian, China, Oct. 2011.

[3] C.-H. Chen, W.-T. Huang, Y.-Y. Chen, and Y.-J. Chang. An integrated service model for telecare system. In *Asia-Pacific Services Computing Conference, 2008. APSCC '08. IEEE*, pages 712-717, Dec. 2008.

[4] N. Fritz. SleekXMPP. http://sleekxmpp.com/. [Online; Access: 23-Jan-2013].

[5] S.-J. Hsu, H.-H. Wu, S.-W. Chen, T.-C. Liu, W.-T. Huang, Y.-J. Chang, C.-H. Chen, and Y.-Y. Chen. Development of telemedicine and telecare over wireless sensor network. In *Multimedia and Ubiquitous Engineering, 2008. MUE 2008. International Conference on*, pages 597-604, April 2008.

[6] H.-Y. Kung, C.-Y. Ou, S.-D. Li, C.-H. Lin, H.-J. Chen, Y.-L. Hsu, M.-H. Chang, and C.-I. Wu. Efficient movement detection for human actions using triaxial accelerometer. In *Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on*, pages 113-114, Jan. 2010.

[7] C.-W. Lin and Z.-H. Ling. Automatic fall incident detection in compressed video for intelligent homecare. In *Computer Communications and Networks, 2007. ICCCN 2007. Proceedings of 16th International Conference on*, pages 1172-1177, Aug. 2007.

[8] C. K. Liu and D. Booth. Web services description language (WSDL) version 2.0 part 0: Primer. W3C recommendation, W3C, June 2007. http://www.w3.org/TR/2007/REC-wsdl20-primer-20070626.

[9] R. Martinez-Lopez, D. Millan-Ruiz, A. Martin-Dominguez, and M. Toro-Escudero. An architecture for next-generation of telecare systems using ontologies, rules engines and data mining. In *Computational Intelligence for Modelling Control Automation, 2008 International Conference on*, pages 31-36, Dec. 2008.

[10] T. Matsubara, H. Satoh, and F. Takeda. Proposal of an awaking detection system adopting neural network in hospital use. In *World Automation Congress, 2008. WAC 2008*, pages 1-6, 28 2008-Oct. 2 2008.

[11] OMG. Common object request broker architecture (CORBA/IIOP).v3.1. Technical report, OMG, Jan. 2008. http://www.omg.org/spec/CORBA/3.1/.

[12] P. Saint-Andre, K. Smith, and R. Tron↑con. XMPP: The Definitive Guide. O'Reilly, April 2009.

[13] I. Sommerville. Software Engineering. Addison-Wesley, September edition, 2011.

[14] Wikipedia. Telecare. http://en.wikipedia.org/wiki/Telecare, March 2012. [Online; access 28-June-2012].

Exploring the Relationships among Corporate Governance, Social Responsibility and Financial Performance: A Perspective of the Enterprises' Selection of Successors

Keh-Luh Wang
Graduate Institute of Finance,
National Chiao Tung University,
Hsinchu, Taiwan, ROC
+886-3-5712121ext 57081
lkwang@mail.nctu.edu.tw

Chi Chiang
Department of Management Science
National Chiao Tung University,
Hsinchu, Taiwan, ROC
+886-3-5712121ext57166
cchiang@mail.nctu.edu.tw

Chiu-Mei Tung
Department of Management Science
National Chiao Tung University,
Hsinchu, Taiwan, ROC
+886-988262176
dcm@ebas.gov.tw

ABSTRACT
Corporate chief executive officers (CEOs) play critical roles in corporate governance (CG) and the financial performance (FP) of firms. Corporate social responsibility (CSR) is the universal issue faced by all enterprises. How to extend the development of CG, CSR and FP would be an important concern for the enterprises' selection of successor. Previous studies mostly focus on the individual or pair relationship among CG, CSR and FP, while seldom adopt the perspective of the enterprises' selection of successors. This study aimed to explore the dynamic causal relationship among the enterprises' selection of successors, CG, CSR and FP using system dynamics (SD). The findings of the study were as follows: 1. the enterprises' selection of successors and CSR positively influenced CG and FP; 2. FP influenced the selection of successors and CG; 3. morality and human caring positively influenced CG, CSR and selection of successors; 4. selection of successor played a vital role in firm's CG mechanism. The findings suggest that morality and human caring is the drive for the dynamic causal relationship among CG, CSR and FP; thus, when selecting the enterprises' successors, morality and human caring should be treated as the important indicators.

Categories and Subject Descriptors
A.0 General–Conference proceedings.

General Terms
Management

Keywords
Corporate governance, financial performance, corporate social responsibility, selection of successors.

INTRODUCTION
The Asian financial crisis of 1997-1998 and the corporate scandals, such as the events of Enron and WorldCom in the United States and the scandal of Parmalat in Europe have highlighted the importance of effective CG systems. Those scandalous events have shown a compelling need for an effective CG mechanism in developed countries as well as other emerging market countries. However, the effectiveness of CG in public corporations is typically restricted due to informative asymmetry and managerial self-interest incentives. As a result, a vital acceleration of CG activities occurs. Previous studies have indicated that an effective CG can resolve the conflicts of interest between shareholders and non-investing stakeholders. Along with the tremendous development of the CG issue, one of a significant issues regarding CG in recent years is the growth of CSR.

The issues related to what formed the best CG mechanisms and why firms were engaged in CSR have been discussed over the past years. According to the World Commission on Environment and Development, sustainable development comprises major challenges associated with three areas: environment, economy, and social issues. Business organizations are seen as being especially well-placed to tackle sustainability problems via their CSR. The CSR concept has evolved since the 1950s, formalized in the 1960s, and proliferated in the 1970s [1]. Based on previous studies, CSR can be broadly considered as the activities and the responsiveness to the needs of the organizations' stakeholders beyond organizations' self-interests and law regulations.

Although an enormous body of literature has emerged concerning the relations between CG and CSR, between CG and FP, and between CSR and FP, and the interrelations among CG, CSR and FP, there is no universally agreed rationale relation among CG, CSR and FP. Several studies have claimed that CSR initiatives are a waste of valuable resources and do not maximize firm value. For example, Jamili et al.[2] suggested that there is a discernable overlap between CSR engagement and CG mechanisms. Barnea and Rubin [3] proposed an overinvestment hypothesis, which considered CSR engagement as a principal-agent relation between managers and shareholders. Therefore, Barnea and Rubin [3] argued that insiders have an interest in overinvesting in CSR in order to acquire private benefits of reputation building as good social citizens.

In contrast, some studies have found a positive relation between CSR and FP, supporting the conflict-resolution hypothesis, that is based on the stakeholder theory. In particular,

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.13

Jo and Harjoto [4] proposed that the lag of CG variables positively affects firms' CSR engagement, in turn, the CSR engagement positively influences FP. Accordingly, firms' CSR engagement with the community, environment, diversity, and employees plays a significantly positive role to enhance FP. Furthermore, Margolis and Walsh [5] also portrayed a generally positive association between investing in socially responsible activities and financial performance.

Previous studies have shown that CEOs influence the quality of the information available for the board of directors and investors. The key success factor of the enterprise of the founder/CEO duality is the founder/CEO extremely familiar with the industry; thus, the founder/CEO can recognize the key reforms of the enterprises. One of the motives of hard-working bosses in many enterprises is that the personal wealth of the founder/CEO duality is closely associated with the corporate stock prices.

Founders and CEOs have strong sense of responsibility toward CG and FP and even demonstrate their CSR for stockholders by not receiving annual salary. However, they will be retired one day and how new successors extend the development of CG, CSR and FP will be the important concern of the enterprises' selection of successors. Corporate CEOs apparently play critical roles in CG and FP; CSR is the universal issue faced by all enterprises. How to extend the development of CG, FP and CSR will be an important concern for enterprises to select their successor. While most of the literature has focused on the individual or pair relationships among CG, FP and CSR, this study adopted the perspective of enterprises' selection of successors to probe the dynamic causal relationships among CG, CSR and FP using SD. Furthermore, this study was attempted to find the key factors of the enterprises' selection of successors. The main finding of this study indicated that morality and human caring were the drive factor for the dynamic causal relationships among CG, CSR and FP. The findings may offer some evidences when selecting enterprises' successors, the morality and human caring should be considered an important indicator.

2. CG, CSR and FP

Prior studies mostly focus on the individual or pair relationship among CG, CSR and FP, while seldom analyze the dynamic causal relationship by the perspective of SD. Literature review is presented below for further analysis:

2.1 CG and CSR

The issues of CG and CSR have been discussed for decades. The major responsibility of the managers was to increase shareholders' wealth; thus, managers and even executives can viewed as the employees of the stockholders. However, if the goals of principal (shareholders) and those of the agent (managers) are not identical, the agency problem may occur. The CG means to make sure that the managers' actions match the important stakeholders' benefits by governance management mechanism. One of the goals of the enterprises is to provide stockholders with good return on investment. However, in most of the limited companies, stockholders authorize the corporate control and strategic decision-making to the managers. Thus, managers become the agents of stockholders and they should develop the strategy of maximizing long-term return on investment for stockholders.

The CSR image can be explained as the consumers' association of public affairs, social art and social welfare contributed by the enterprises. The public mostly identifies with the public welfare activities and regard them as ethical and generous behavior.

If the enterprises regard the ethical responsibility as responsibility instead of obligation, they may obtain potential profits. The reason is that by practicing social responsibility, the enterprises can communicate with investors, stockholders, customers, the employees or supports and be identified by them. In other words, the enterprises continuously promote CSR strategy to stakeholders who will accept and strategy and support the enterprises.

The enterprises can construct trust and increase the reputation for morality. It will attract the consumers, the employees, suppliers and distributors and acquire the reputation. Secondly, in the era with immediate and globalized communication and media, immoral behavior will be spread negatively and cause the increase of fines and lawsuit cost. Based on the above, the enterprises can promote the CSR activities by globalized media and information to draw the attention of the consumers to enhance the reputation.

2.2 CG and FP

The effect of CG mechanism on corporate performance or stockholders' wealth or capital would treated as an issue, and good CG mechanism will increase corporate operational performance, enhance stockholders' wealth or reduce the cost. CG mechanism can enhance corporate operational performance and reduce the cost since good CG mechanism directly influences FP. Will it result in positive CSR and influence FP? There are few related studies.

CG mechanism aims to manage the senior managers' behavior by their turnover intention. There is a negative relationship between senior managers' turnover and stock price performance of the previous period. After senior managers leave the company, the corporate performance is enhanced. However, if new senior managers are from external world, the corporate performance will be significantly enhanced. With effective corporate internal monitoring mechanism, the senior managers with inferior performance encounter the threat of turnover. However, senior managers' shareholdings might influence the effectiveness of the internal monitoring mechanism. There is a negative relationship between senior managers' shareholding, corporate performance and senior managers' turnover. Although Jensen and Meckling [6] indicated that managers' ownership leads to the consistency between managers' and stockholders' benefits, according to recent studies, managers' ownership will make it more difficult to dismiss the managers with inferior performance. Therefore, the enterprises' selection of successor is significantly related to CG and FP.

2.3 CSR and FP

Most of the companies with rapid growth share the same experience. For sustainable operation, they must satisfy three kinds of stakeholders: customers, the employees and stockholders. Thus, the stakeholders are influential. The implementation of the strategy directly influences operational performance. Will the enterprise' friendliness, participation in public welfare and behavior of CSR influence FP? It has been a disputable question. There is no unified theory behind CSR engagement, and there are two main alternative explanations.

On the one hand, some scholars suggested that the enterprises' undertaking of social responsibility will result in high cost, restrict the product development and reduce the competitiveness. On the other hand, some scholars suggest that

facing highly competitive market environment, companies should use CSR as a strategic tool to respond to expectations of various stakeholders and thus create a favorable corporate image. In reality, companies have regarded CSR activities as a necessity, thus urging managers to contemplate how to implement CSR activities consistently with their business strategy. If the enterprises can satisfy the expectation of stakeholders at different levels, they will positively enhance the financial performance. Higher social performance will result in better financial performance. The statement is called social impact hypothesis.

3. THE ENTERPRISES' SELECTION OF SUCCESSORS

Ryan et al. [7] indicated that executive values and motivation lie at the heart of ethics research in corporate govrnance. Goel and Thakor [8] also stated that CEOs attributes affect various corporate decisions, the CEO selection process can affect the firm's investment policy as well as the efficacy of any corporate governance mechanism.

For the enterprises, with a good successor system, the organizations will predict future manpower demand, recognize the key talents' learning process and train them at the strategic positions. For the employees, a good successor system allows them to receive the training to develop their potential. The individuals' career development will be systematic, and they will be properly promoted. Thus, the employees' sense of achievement and their loyalty increase.

Ideal corporate successor planning should be based on logic, political manipulation and emotion. Therefore, the concerns for scheduling, candidates' qualification and assessment should be more thorough. Scheduling of successor planning, the leaders and progress are usually the questions of the enterprises. Generally speaking, although complete successor planning is led by incumbent leaders, the directors should also monitor the process in order to guarantee the implementation. Regarding the scheduling, first, it must confirm the qualification of new CEO and assessment. Secondly, the enterprises select the appropriate internal candidates, set up leader training plan and accomplish regular evaluation by sufficient time. The most important characteristic of CEOs is the intention to invest in time and efforts and develop the common consensus between high-rank management and directors regarding the corporate strategies and organizational management. Upon such condition, they can predict the professional capability, experience and personality traits of successors according to the operational vision.

Because of different organizational cultures, the enterprises usually have different successor cultures. For instance, in Asia, including Taiwan, the succession in family businesses is considerably different from the resolution of the board of directors in U.S. and Europe. Thus, Friedman and Olk [9] suggested that according to the guidance of incumbent CEOs, candidates or criteria planned, the succession can be divided into the following four types: crown heir, horse race, coup d'Etat and comprehensive search.

In a word, in the enterprises' selection of successors, the successors' work at basic level, knowledge of different departments, outstanding personality traits and different backgrounds and concepts will influence the performance at the position. It will be critical for the enterprises' future success.

4. RESULTS

The SD is characteristic of understanding the behavior of complex systems; therefore, the study clarified the dynamic relationships presented in the following sections among the successor selection, CG, CSR and FP using SD.

4.1 CG plays a pivotal role

As shown in Figure 1, the selection of successors had a positive effect on CG. CSR also relatively positively influenced CG; in modern society with transparent information, by respecting stakeholders' opinions to fulfill CSR, the enterprises would lead to positive effects; CG also positively influenced FP. Conse-quently, many literatures have suggested that FP is the important issue of CG. CG thus plays a pivotal role in the dynamic relationships among selection of successors, CG, CSR and FP.

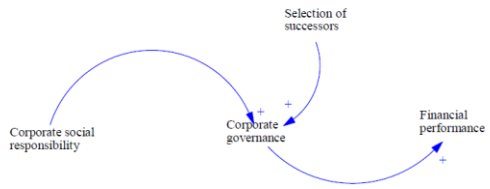


Figure 1. Corporate governance as the pivot

4.2 FP feedback influences selection of successors and CG

As shown in Figure 2, the influential factors of successor selection not only include CG, but also refer to FP. Accordingly, FP will positively influenced selection of successors as well as CG due to the effect of feedback. There was a positive loop among FP, selection of successors and CG, and the loop presented their causal relationship.

In modern society with the rise of morality, corporate leaders should identify with the public's opinions and cannot violate the justice of morality to harm the rights of the public. Morality and human caring both positively influenced CSR, CG and selection of successors and they were the key factors. When the enterprises selected the successors, they should concern the important personality traits or the spirit of business culture.

In modern society with advances of information, the enterprises' accomplishment of CSR and CG can be immediately absorbed by the public and it would influence the enterprises' reputation. They both positively influenced the reputation. In addition, when the enterprises fulfill CSR, the expenditure should increase and it should be a burden. Thus, it is the main reason that some scholars or enterprisers suggest that the enterprises should not undertake CSR.

The enterprises' fulfillment of CSR can enhance corporate reputation, and it positively influenced the corporate operational efficacy; however, it should also increase the expenditure of CSR due to the socially responsible actions and negatively influenced the corporate operational efficacy. Therefore, positive and negative effects of CSR on the corporate operational efficacy formed a dynamic relationship through the reputation and the expenditure of CSR. When the enterprises tried to accomplish CSR, they should also measure their own condition of resources in order to acquire a balance. The corporate operational efficacy also influenced FP and the selection of successors and it formed dynamic causal feedback loop.

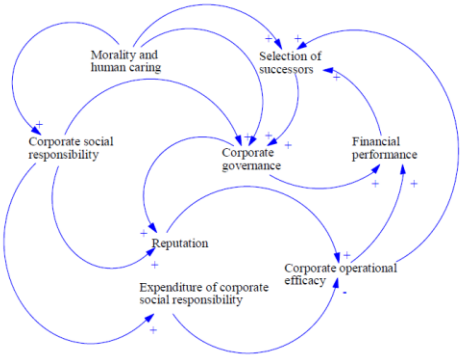


Figure 2. The dynamic causal relationships among the enterprises' selection of successors, CG, CSR and FP

5. CONCLUSIONS

By literature review and causal feedback loop analysis of System Dynamics (see Figure 2), this study found that among corporate governance, corporate social responsibility and financial performance, corporate governance is the pivot and their relationships are shown below: corporate governance is the pivot. Financial performance feedback influences the selection of successors and corporate governance, morality and human caring are the key factors. Corporate social responsibility and corporate governance positively influence reputation. There are positive and negative effects of corporate social responsibility on corporate operational efficacy. The main contribution of this study is to demonstrate morality and human caring as the key factors since morality and human caring both influence the selection of successors, corporate governance and corporate social responsibility and they become the critical indicators to evaluate the enterprises' successors.

In Asian countries, including Taiwan, in family enterprises, listed or non-listed, the directors, supervisors and managers usually have blood or marriage relations. By intersect holdings, they usually show decisive votes. Due to asymmetric information, the management and supervision of family business are not transparent. The rights of small stockholders are usually neglected and the investors will try to avoid the shares of family business. Thus, the traditional family operational model becomes difficult and the capital market cannot be enhanced. Due to marriage relations, in family business, the employment is only for the relatives and excellent talents cannot obtain the positions in management. Incompetent family members hold the key positions. Because of ambiguous property rights, the ownership and operation rights in family business is not separated. It is the point which is criticized the most.

"The sons inherit fathers' business" is the traditional Chinese concept. In short time, family businesses still exist in the market. However, it does not mean all family businesses follow the same way. A vigorous enterprise should change and create the organizations to match the change of environment with their own development. Family business should break through the limitation. When cultivating the family members to be the successor, they should properly introduce the external talents. Thus, they will not stand out in fair competition.

According to traditional economic model, the enterprises should pursue maximum profits. It can be feasible in the past and it matches corporate ethics. However, nowadays, since the public's expectation toward the enterprises and the change of business environments, the enterprises should not treat traditional economic model as the principle of corporate governance, and they must accept the view in social responsibility model to include social responsibility in corporate governance. They not only pursue reasonable profits, but also fulfill social responsibility. When selecting the successors, the enterprises should include morality and human caring as the most important concerns. The concept and implementation of morality and human caring is associated with the fulfillment of corporate social responsibility, corporate governance and financial performance. They will be the common factors of successor selection in eastern and western enterprises.

6. REFERENCES

[1] Carroll, A. 1999. Corporate social responsibility. *Business & Society*, 38, 3, 268.

[2] Jamali, D., Safieddine, A., and Rabbath, M. 2008. Corporate governance and corporate social responsibility synergies and interrelationships. Corporate Governance: *An International Review*, 16(5), 443–459..

[3] Barnea, A. and Rubin, A. 2010. "Corporate social responsibility as a conflict between shareholders," *Journal of Business Ethics*, 97, 71–86.

[4] Jo, H. and Harjoto, M. A. 2012. "The Causal Effect of Corporate Governance on Corporate Social Responsibility," *Journal of Business Ethics*, 106, 53–72.

[5] Margolis, J. and Walsh, J. P. 2003. Misery loves companies: Rethinking social initiatives by business. Administrative Science Quarterly, 48, 268–305.

[6] Jensen, M. C. and Meckling, W. H. 1976. "Theory of the Firm: Managerial Behavior, Agency Cost and Ownership Structure," *Journal of Financial Economics*, 3, 305-360.

[7] Ryan, L. V., Buchholtz, A. K. and Kolb, R.W. 2010. "New Directions in Corporate Governance and Finance: Implications for Business Ethics Research," *Business Ethics Quarterly* 20(4), 673-694.

[8] Goel, A. M. and Thakor, A. V. 2008. "Overconfidence, CEO Selection, and Corporate Governance," The Journal of Finance, 63(6), 2737-2784.

[9] Friedman**,** S**.** and Olk, P. 2004. "Varieties of CEO succession," *Global CEO*, March, 41-51.

A New Distance Metric for Formation of Non-Uniformly Distributed Incremental Clusters

A.M.Sowjanya
Dept. of CS & SE
College of Engineering
Andhra University
91-09666657782
sowmaa@yahoo.com

M.Shashi
Dept. of CS & SE
College of Engineering
Andhra University
91-09949072880
smogalla2000@yahoo.com

ABSTRACT
Conventional clustering algorithms aim at identification of groups of entities having high cohesion and separation with an inherent characteristic of nearly uniform distribution within a cluster. However, non-uniformly distributed clusters are also significant in applications like modeling the growth pattern of neighborhoods in urban areas, modeling the protection mechanism against infections due to injuries in a human body etc. This paper presents an algorithm for the formation of non-uniformly distributed clusters and their incremental growth. A non-uniformly distributed cluster demands a different distance metric to simulate its growth pattern which is different from the growth pattern of a uniformly distributed cluster. The authors proposed a new proximity metric called 'Inverse Proximity Estimate' (IPE) which considers the proximity of a data point to a cluster representative as well as its proximity to a farthest point in its vicinity to determine the membership of a data point into a cluster. The performance of this incremental algorithm was tested on benchmark datasets available in UCI repository.

Categories and Subject Descriptors
H.2.8 [**Database Applications**]: *Data mining*

General Terms
Algorithms

Keywords
Data mining, Clustering, Incremental clustering, non-uniformly distributed clusters, Inverse Proximity Estimate, k-means, Cluster Feature, Farthest neighbor points, Clustering accuracy, CFICA.

1. INTRODUCTION
Data clustering is one of the important data mining activities used extensively in many domains like statistics, biology, geology, image processing, pattern recognition, information retrieval etc. Since clustering discovers patterns from a wide variety of domain data many clustering algorithms were developed by researchers. The main problem with the conventional clustering algorithms is that, they mine the static database and generate a set of patterns in the form of clusters. Numerous applications maintain their data in large databases or data warehouses and thus many real life databases keep growing incrementally. New data may be added periodically either on a daily or weekly basis. For such databases, the patterns extracted from the original database become obsolete. Conventional clustering algorithms handle this problem by repeating the process of clustering on the entire database whenever a significant set of data items are added. The process of re-running the clustering algorithm on the entire dataset is inefficient and time-consuming. Thus most of the conventional clustering algorithms are unsuitable for incremental databases.

An incremental database requires the design of new clustering algorithms [1,2,3,4] . A solution to handle this problem is to integrate a clustering algorithm that functions incrementally [5]. Incremental clustering algorithms permit a single or a few passes over the whole dataset to put the updated item into the cluster. With respect to the size of the set of objects, algorithms and number of attributes, incremental clustering algorithms are of scalable nature [6]. The model of incremental algorithms for data clustering is necessitated by realistic applications where the demand sequence is not known in advance and the algorithm should keep a constantly fine clustering using a restricted set of operations resulting in a solution of hierarchical structure [7].

Data points in a cluster can be either uniformly distributed or non-uniformly distributed. The shape of a uniformly distributed cluster is nearly globular and its centroid is located in the middle (geometrical middle). Non-uniformly distributed clusters have their centroid located in the midst of dense area, especially if there is a clear variation in the density of data points among dense and sparse areas of the cluster. The shape of such clusters is not spherical and the farthest points of a non-uniformly distributed cluster are generally located in the sparse areas.

For clusters with uniformly distributed points, the usual distance measures like Euclidean distance hold good for deciding the membership of a data point into a cluster. But

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.14

there exist applications where clusters have non-uniform distribution of data points. In such cases, the Inverse Proximity Estimate (IPE) proposed in CFICA [8] will be useful as it better recognizes the discontinuities in data space while extending the cluster boundaries.

Specifically, while modeling the growth patterns of neighborhoods in an urban area the shape of the cluster must be non-globular as a neighborhood consists of a smaller densely populated area (down town) and a larger sparsely populated area (posh localities). As the city grows new colonies are expected to develop adjacent to the sparser side of the city while vertical growth or infiltration into sparser area is expected on the denser side to keep pace with additional population in the city. As a consequence of this, the growth of the city appears to be bounded on the denser side and open towards the sparser side. This paper presents a new algorithm for the formation and maintenance of incremental clusters to simulate such situations. Normal Euclidean distance metric will not discriminate the data points existing on denser side from the sparser side. So, the authors have proposed a new proximity metric, IPE which is capable of discriminating the data points as per the requirement of simulating the growth pattern of a non-uniformly distributed cluster.

2. CFICA

Cluster Feature Based Incremental Clustering Approach (CFICA) is capable of clustering incremental databases starting from scratch. However, during the initial stages refreshing the cluster solution happens very often as the size of the initial clusters is very small. Hence for efficiency reasons a partitional clustering algorithm is applied on the initial collection of data points to form clusters. We have used k-means algorithm for initial clustering and the original CFICA algorithm has been retained. Specifically changes were made to the structure of the cluster feature to increase efficiency. It has two important steps namely initial clustering of the static database and handling of incremental data points.

2.1 Initial Clustering of the Static Database

We used the k-means clustering algorithm here for initial clustering to obtain k number of clusters as it can be easily implemented and is suitable for clustering datasets with numerical attributes because it uses mean as cluster representative. Also k-means is relatively scalable and efficient in processing large datasets with computational complexity O(nkt) where n is the total number of objects, k is the number of clusters and t represents the number of iterations.

2.2 Clustering of Incremental Database

After initial clustering of the original static database using k-means algorithm, the clustering solution is obtained in the form of cluster features (CF's). Now, using CFICA, the incremental database is clustered.

2.2.1 Computation of Cluster Feature (CF)

The concept of cluster feature for clustering the incremental database has been adopted from BIRCH [9] as it supports incremental and dynamic clustering of incoming objects. As CFICA handles partitional clusters as against hierarchical clusters handled by BIRCH, the original structure of cluster feature went through appropriate modifications to make it suitable for partitional clustering.

The Cluster Feature (CF) is computed for every cluster c_i obtained from the k-means algorithm. In CFICA, the Cluster Feature is denoted as,

$$CF_i = \{n_i, \vec{m}_i, \vec{m}_i, Q_i, \vec{ss}_i\}$$

where n_i → number of data points;

$\vec{m}_i$ → mean vector of the cluster C_i with respect to which farthest points are calculated; $\vec{m}_i$→ new mean vector of the cluster C_i that changes due to incremental updates; Q_i → p-farthest points of cluster C_i ; $\vec{ss}_i$ → squared sum vector that changes during incremental updates.

A Cluster Feature is aimed to provide all essential information of a cluster in the most concise manner. The first two components n_i and $\vec{m}_i$ are essential to represent the cluster prototype in a dynamic environment. The Q_i , set of p-farthest points of cluster C_i from its existing mean $\vec{m}_i$, are used to handle unevenly distributed and hence irregularly shaped clusters; ; $\vec{m}_i$, the new mean is essential to keep track of dynamically changing nature / concept – drift occurring in the cluster while it is growing. $\vec{ss}_i$, squared sum is essential for estimating the quality of cluster in terms of variance of data points from its mean.

2.2.2 Insertion of a New Data point

Whenever there is a new data point Δy finding the appropriate cluster for that data point is important. If the incoming data point Δy cannot be included into any of the existing clusters, then it separately forms a new singleton cluster.

The results produced by standard partitional clustering algorithms like K-means are not in concurrence with this natural expectation as they rely upon Euclidean distance metric (ED) for discriminating data points while determining their membership into a cluster. So, the author suggests a modification to the Euclidean distance (ED) by adding a Bias factor. Bias is the increment added to the conventional distance metric in view of formation of more natural clusters and better detection of outliers. It considers the unevenness / shape of the cluster reflected through a set of p- farthest points to estimate the proximity of new data points to the cluster.

$$IPE_{\Delta y}^{(i)} = ED(\vec{m_i}, \Delta y) + B$$

where, $IPE_{\Delta y}^{(i)}$ → Inverse Proximity Estimate (IPE)

$ED(\vec{m_i}, \Delta y)$ → Euclidean distance between the centroid, $\vec{m_i}$ and the incoming data point and Δy of the cluster C_i.

B → Bias factor

One of the factors that assess the bias is that it increases with the distance of the new point to a farther point. So, bias is proportional to the Euclidean distance, ED between the farthest point (q_i) and the incoming new data point (Δy).

$$\therefore \quad B \;\alpha\; ED(q_i, \Delta y)$$

As the distance between the centroid and the particular farthest point in the vicinity of the new point increases, elongation of the cluster with respect to that farthest point should be discouraged.

Bias also increases with the distance of a particular farthest point to its centroid. Therefore, bias is proportional to the Euclidean distance, ED between the centroid ($\overrightarrow{m_i}$) and the incoming new data point (Δy).

$$\therefore \quad B \;\alpha\; ED\,(\overrightarrow{m_i}, q_i)$$

Hence, bias is estimated as a product of $ED\ (q_i, \Delta y)$ and $ED\ (\overrightarrow{m_i}, q_i)$ mathematically. Therefore bias is expressed as,

$$B = [\ ED\ (q_i, \Delta y)\ * ED\ (\overrightarrow{m_i}, q_i)\ ]$$

Bias is introduced into the distance metric and the modified distance metric ($IPE^{(i)}_{\Delta y}$) is calculated as follows:

$$IPE^{(i)}_{\Delta y} = ED\ (\overrightarrow{m_i}, \Delta y) + B$$

i.e.,

$$IPE^{(i)}_{\Delta y} = ED\ (\overrightarrow{m_i}, \Delta y) + [\ ED\ (q_i, \Delta y)\ * \mathrm{ED}\ (\overrightarrow{m_i}, q_i)]$$

2.2.3 Incremental Clustering Approach

The Inverse Proximity Estimate $IPE^{(i)}_{\Delta y}$, used by CFICA is used for effectively identifying the appropriate cluster of an incoming data point Δy. In other words, it estimates the proximity of the incoming point to a cluster based on the cluster centroid (mean m_i), farthest point in the vicinity of the incoming point (Q_i) and the incoming data point (Δy).

For each cluster c_i, the Euclidean distance E_D is calculated for the following set of points: centroid and incoming point($\overrightarrow{m_i}, \Delta y$), farthest point in the vicinity of the incoming point and incoming point $(q_i, \Delta y)$, centroid and farthest point in the vicinity of the incoming point $(\overrightarrow{m_i}, q_i)$. Upon the arrival of a new data point Δy to the existing database S_D which is already clustered into C = {C_1, C_2, ……., C_k} clusters, its distance, $IPE^{(i)}_{\Delta y}$ to i[th] cluster for all i = 1 to k is calculated using the following equation.

$$IPE^{(i)}_{\Delta y} = ED\ (\overrightarrow{m_i}, \Delta y) + [\ ED\ (q_i, \Delta y)\ * \mathrm{ED}\ (\overrightarrow{m_i}, q_i)]$$

where,

$ED\ (\overrightarrow{m_i}, \Delta y)$ → Euclidean distance between $\overrightarrow{m_i}$ and Δy

$ED\ (q_i, \Delta y)$→ Euclidean distance between q_i and Δy

$\mathrm{ED}\ (\overrightarrow{m_i}, q_i)$→ Euclidean distance between $\overrightarrow{m_i}$ and q_i

2.2.4 Finding the farthest point in the vicinity of incoming data point Δy

The set of p - farthest points (Q_i) of the i[th] cluster, c_i are calculated as follows: First Euclidean distances are calculated between the data points within cluster c_i and the mean of the corresponding cluster, $\overrightarrow{m_i}$. Then, the data points are arranged in descending order with the help of the measured Euclidean distances. Subsequently, the top p -farthest neighbor points for every cluster are chosen from the sorted list and these points are known as the p - farthest points of the cluster c_i with respect to the mean value, $\overrightarrow{m_i}$. Thus a list of p - farthest points are maintained for every cluster, c_i. These p - farthest points are subsequently used for identifying the farthest point, q_i.

In order to find the farthest point q_i (data point in Q_i) which is in the vicinity of the incoming data point Δy, the Euclidean distance is calculated for that data point Δy with each of the p -farthest points of that cluster. Then, these p -farthest neighbor points are sorted based on the Inverse Proximity Estimate $IPE^{(i)}_{\Delta y}$ and thereby, for each cluster, the data point having minimum distance is taken as the farthest point q_i.

2.2.5 Finding the appropriate cluster

The Inverse Proximity Estimate is used, to find the appropriate cluster for the new data point, Δy. The new data point, Δy is assigned to the closest cluster only if the calculated distance measure, $IPE^{(i)}_{\Delta y}$ is less than the predefined threshold value, λ. Otherwise, the data point Δy is not included in any of the existing clusters, but it separately forms a new singleton cluster. In such a case, the number of clusters is incremented by one.

2.2.6 Updating of Cluster Feature

Whenever a new data point Δy is added to the existing database, that new data point may be included into any of the existing clusters or it may form a new cluster. So after the new point gets inserted, updating of CF is important for further processing.

When a new data point, Δy is included into an already existing cluster C_i , its cluster feature (CF_i) is updated without requiring the original data points of C_i and hence supports incremental update of the clustering solution. In particular, the n_i, $\overrightarrow{m_i}$ and ss_i fields of CF_i are updated upon the arrival of a new data point Δy into the cluster C_i . However, the Q_i representing the p-farthest points and the centroid of the previous snapshot $\overrightarrow{m_i}$ were kept without any changes until the next periodical refresh.

Case 1: Inclusion of new data point into any existing cluster

The addition of a new data point into the cluster C_i, naturally results in change of mean ($\overrightarrow{m_i}$ to $\overrightarrow{m_i}$). So, a new set of p-farthest points have to be computed for the incremented cluster C_i'. Calculating p-farthest points again, every time the cluster gets updated, is not an easy task as it involves recalculation of the Euclidean distance, ED for every data point in the incremented cluster C_i' with the updated mean, $\overrightarrow{m_i}$.

For pragmatic reasons, it was suggested to refresh the CF_i , only in case it deviates significantly from its original value indicating concept drift [10]. So, deviation in mean is calculated.

$$\text{Deviation in mean} = \frac{\vec{m_i} - \vec{m_i}}{\vec{m_i}}$$

If the deviation in mean is greater than δ, which is user defined then the new set of p-farthest points have to be computed for the incremented cluster. Otherwise, the same set of p-farthest points along with the old mean value, m_i is maintained for the incremented cluster also. It may be noted that m_i in the CF_i always represents the centroid based on which the p-farthest points are identified. Hence needs to be changed whenever new set of p-farthest points are identified.

Case 2: New data point forms a singleton cluster

If the new data point forms a new cluster separately, the cluster feature (CF_i) has to be computed for the new cluster containing the data point Δy . CF_i for the new cluster contains the following information: Number of data points becomes1, as it is a singleton cluster, In this case, $\vec{m_i} = \vec{m_i} = \Delta y$, list of p-farthest points is null and squared sum is zero.

Finding the appropriate cluster to incorporate the new data point Δy and updating of the cluster feature after adding it to the existing cluster or forming a separate cluster are iteratively performed for all the data points in the incremental database ΔS_D .

2.2.7 Merging of Closest Cluster Pair

Once the incremental database ΔS_D is processed with CFICA, a merging strategy is used to maintain reasonable number of clusters with high quality. Merging is performed when the number of clusters increases beyond 'k' (k in k-means) while ensuring that increase in variance which indicates error is minimum due to merging. It is intuitive to expect an increase in the error with the decrease in the number of clusters. A closest cluster pair is considered for merging if only the Euclidean distance between the centroids of the pair of clusters is smaller than user defined merging threshold (θ).

The procedure used for the merging process is described below:

Step 1: Calculate the Euclidean distance ED between every pair of cluster centroids ($\vec{m_i}$).

Step 2 : For every cluster pair, with Euclidean distance, ED less than the merging threshold value,θ (ED $\ll \theta$), find increase in variance (σ^2).

Step 3 : Identify the cluster pair with minimum increase in variance and merge that cluster pair.

Step 4 : Recalculate the mean for the new merged resultant cluster, C_R.

Step 5 : Repeat steps 1 to 4 until no cluster pair is merged or until the value of 'k' is adjusted.

After merging the closest cluster pair, now the Cluster Feature (CF) has to be computed for the merged cluster. Hence the Cluster Feature of the new cluster 'k' which is formed by merging two existing clusters 'i' and 'j' is determined as a function of CF_i and CF_j

$$CF_k = f(CF_i, CF_j)$$

Thus Cluster Feature provides the essence of the clustered data points thereby avoiding explicit referencing of individual objects of the clusters which may be maintained in the external memory space.

3. RESULTS

CFICA has been implemented using the Iris, wine and yeast datasets from the UCI machine learning repository. Datasets with instances belonging to more number of classes are prone to clustering solutions with less accuracy. *Iris dataset* [11] comprises 150 instances distributed among 3 classes and each class corresponds to a type of iris plant. *Wine dataset* [12] comprises 178 instances describing 3 types of wines in terms of 13 constituents/attributes. Yeast dataset [13] contains 1484 instances distributed into 10 classes and has 9 attributes. It predicts the localization site of protein. It can be seen that wine and yeast datasets contain a considerable number of attributes. After applying attribute reduction the number of attributes came down to 3 for wine and to 5 attributes for yeast.

These datasets are processed dynamically by dividing the data points in each dataset into chunks. For example, the Iris dataset is divided into 4 chunks (Iris1 consisting of 75 instances, Iris2, Iris3 and Iris4 consisting of 25 instances each). Initially, the first chunk of data points is given as input to the k-means algorithm for initial clustering i.e Iris1. It generates k number of clusters. Then, the cluster feature is computed for those k initial clusters. The next chunk of data points is input to the new approach incrementally. For each data point from the second chunk, the Inverse Proximity Estimate (IPE) is computed and the data points are assigned to the corresponding cluster if the calculated distance measure is less than the predefined threshold value, λ. Otherwise, it forms as a separate cluster. Subsequently, the cluster feature is updated for each data point. Once the whole chunk of data points is processed, the merging process is done if only the Euclidean distance between the centroids of the pair of clusters is smaller than user defined merging threshold, θ. Here, λ = 10 and θ = 4.Finally, the set of resultant clusters are obtained from the merging process. Similarly, the cluster solution is updated incrementally upon receiving the later chunks of data Iris3 and Iris4.

The performance of CFICA is evaluated on the above datasets. The evaluation metric [14, 15] used for estimating the performance in terms of clustering accuracy (CA) is given below,

$$\text{Clustering Accuracy, } CA = \frac{1}{N}\sum_{i=1}^{T} X_i$$

where, N → Number of data points in the dataset

T → Number of resultant cluster

X_i → Number of data points of majority class in cluster i

The Clustering Accuracy of CFICA and BIRCH are shown in figures 1 to 3. Clustering accuracy of the resultant clusters is calculated by changing the k-value (order of initial clustering).

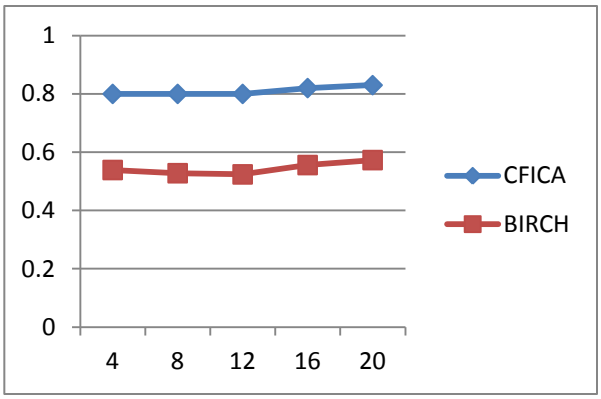


Figure 1. CA vs. number of clusters (k) for Iris dataset.

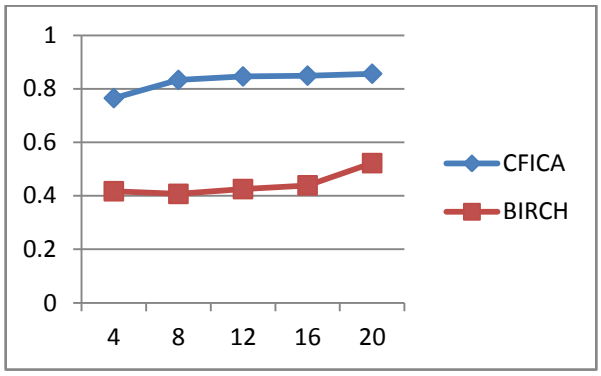


Figure 2. CA vs. number of clusters (k) for Wine dataset.

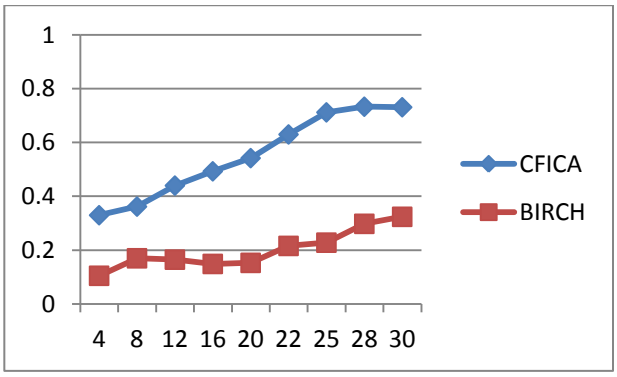


Figure 3. CA vs. number of clusters (k) for Yeast dataset.

The above results demonstrate that CFICA performs better than BIRCH in terms of Clustering Accuracy (CA). BIRCH algorithm is unable to deliver satisfactory clustering quality if the clusters are not spherical in shape because it employs the notion of radius or diameter to control the boundary of a cluster.

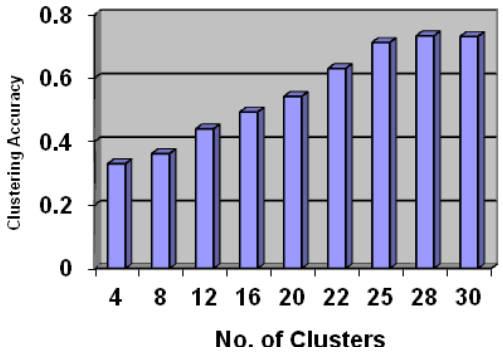


Figure 4. Clustering Accuracy vs. number of clusters (k) for Yeast dataset.

From the above figure 4 it can also be seen that as the number of clusters increase the clustering accuracy also increases.

4. CONCLUSIONS

In the context of incremental clustering while adopting the existing patterns or clusters to the enhanced data when a significant chunk of data points arrives, it is often required to elongate the existing cluster boundaries in order to accept new data points if there is no loss of cluster cohesion. It can be observed that the Euclidean distance between single point cluster representative and the data point will not suffice for deciding the membership of the data point into the cluster except for uniformly distributed clusters. Instead, the set of farthest points of a cluster represent the data spread within a cluster and hence can be considered for formation of natural clusters. So, CFICA makes use of a new proximity metric, Inverse Proximity Estimate (IPE) which considers the proximity of a data point to a cluster representative as well as its proximity to a farthest point in its vicinity to determine the membership of a data point into a cluster.

5. ACKNOWLEDGMENTS

Our thanks to R.A.Fisher, creator of the Iris datset [16]; Forina, M. et al, the original owners of Wine dataset [17] and Kenta Nakai, the creator and maintainer of Yeast dataset [18] for using the above datasets from the UCI Machine Learning Repository.

6. REFERENCES

[1] D. Fisher, "Knowledge acquisition via incremental conceptual clustering," Machine Learning, vol. 2, 1987, pp.139-172.

[2] J. Gennary, P. Langley, and D. Fisher,"Model of Incremental Concept Formation," Artificial Intelligence Journal, vol. 40, 1989, pp. 11-61.

[3] M. Charikar, C. Chekuri, T. Feder, and R. Motwani, "Incremental clustering and dynamic information retrieval," 29th Symposium on Theory of Computing, 1997, pp. 626 - 635.

[4] M. Ester, H. P. Kriegel, J. Sander, M. Wimmer, and X. Xu, X, "Incremental clustering for mining in a Data Warehousing environment," Proc. of the 24th Int. Conf. on Very Large Databases (VLDB'98), New York, USA, 1998, pp. 323-333

[5] Chien-Yu Chen, Shien-Ching Hwang, and Yen-Jen Oyang, "An Incremental Hierarchical Data Clustering Algorithm Based on Gravity Theory", Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, 2002, pp. 237 – 250.

[6] Dan Simovici, Namita Singla, "Metric Incremental Clustering of Nominal Data", ICDM, 2004, pp: 523-526.

[7] Dimitris Fotakis, "Incremental algorithms for Facility Location and k-Median", Theoretical Computer Science, Vol: 361, No: 2-3, 2006, pp: 275-313.

[8] A.M. Sowjanya, M. Shashi, "Cluster Feature-based Clustering Approach (CFICA) for numerical data", International Journal of Computer Science and Network Security, 2010, Vol.10, No.9, pp.73-79.

[9] T. Zhang, R. Ramakrishnan, M. Linvy, "*BIRCH: An Efficient Data Clustering Method for Very Large Databases"*, Proceedings ACM SIGMOD International Conference on Management of Data, 1996, pp.103-114.

[10] Hung-Leng Chen, Ming-Syan Chen, Su-Chen Lin, "Catchimg the Trend: A Framework for Clustering Concept-Drifting Categorical Data", IEEE Transactions on Knowledge and Data Engineering, 2009, Vol. 21, No.5, pp.652 - 665.

[11] Iris dataset : http://archive.ics.uci.edu/ml/datasets/Iris

[12] Wine dataset : http://archive.ics.uci.edu/ml/datasets/Wine

[13] Yeast dataset : http://archive.ics.uci.edu/ml/datasets/Yeast

[14] Zengyou He, Xiaofei Xu, Shengchun Deng, "Clustering mixed numeric and categorical data: A cluster ensemble approach", abs/cs/0509011.

[15] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values", International Conference on Data Mining and Knowledge Discovery, Vol. 2, No.3, September 1998, pp. 283-304.

[16] Fisher,R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936); also in "Contributions to Mathematical Statistics" (John Wiley, NY, 1950).

[17] Forina, M. et al, PARVUS - An Extendible Package for Data Exploration, Classification and Correlation. Institute of Pharmaceutical and Food Analysis and Technologies, Via Brigata Salerno, 16147 Genoa, Italy.

[18] Kenta Nakai and Minoru Kanehisa, "Expert Sytem for Predicting Protein Localization Sites in Gram-Negative Bacteria", PROTEINS: Structure, Function, and Genetics 11:95-110, 1991.

Knowledge-driven Computer Aided Process Innovation Method

Junhao Geng Gangfeng Wang Xitian Tian
School of Mechanical Engineering, Northwestern Polytechnical University
127 Youyi Xilu, Xi'an 710072, China
86-29-88460462
gengjunhao@nwpu.edu.cn wgf8998@163.com tianxt@nwpu.edu.cn

ABSTRACT
The process innovation is difficult, c ostly and time-consuming because it mainly depends on limited knowle dge and random inspiration of individual. T his paper presents a knowle dge-driven computer aided process innovation method in order to change this situation. In this method, proc ess innovation kno wledge is described with formal format, organized into process innovation knowledge neural network. The knowledge can be accumulated with bilayer social WIKI network. When a process problem is input into as activating signal, the knowledge neurons process and transmit the signal in the knowledge neural network, and finally an innovative solution is generated. So, the process problem is solved based on proce ss innovation knowle dge with structured procedure. The knowledge-driven c omputer aided process innovation method can implement th e systematic and structured process innovation and improve the innovation efficiency and quality. A welding pr ocess innovation instance was shown to confirm this method.

Categories and Subject Descriptors
J.1 [**Administrative data processing**]: Manufacturing.

J.6 [**Computer-aided engineering**].

I.2.4 [**Knowledge Representation Formalisms and Methods**]: Relation systems.

General Terms
Management, Design, Theory.

Keywords
Process innovation, innovation knowledge management, knowledge neural network, knowledge accumulation, social WIKI network, problem solving.

1. INTRODUCTION
Process innovation is a creative practice which cre ates new technical principle and production mode; improves manufacturing capacity and effic iency with mod ern scientific knowledge [1]. However, manufacturing proces s innovation m ainly depends on the experience and inspiration of a few of engineers. Meanwhile, the knowledge can't be a ccumulated from innovation instances; the successful and failing experiences can't be used for reference effectively. The dependence on limited individual knowledge and random innovative techniques (for example, brainstorming or trial and error method) directly leads to the low efficiency, low success rate and serious waste of materials.

The concept of proce ss innovation is proposed by J. Schump eter [1]. But, almost all of researches focus on t he concept, management regulations and policies of process innovation, how to implement the process innovation with technical method is still absent. J.H. Geng proposed the concept of computer aided process innovation (CAPI) firstly [2].

CAPI is a branch of computer aided innovation (CAI), process innovation knowledge (PIK) is the basis to implement the process innovation activity. In the last decade, CAI technology has evolved into CAI 2.0 and Enterprise 3.0 [3, 4], but it focuses on product innovation, and does n't meet the requirements of process innovation. Process innovation is different from product innovation [5]. Product innovation giv es attention to virtual product, but process innovation must pay attention to final product, in-process product and manufacturing process. At the same time, process innovation activity must soak into laborers, producer material, manufacturing objects and their combining mode, and is restricted by enterprise manufacturing capability and other real factors. In general, process innovation has more broad technology domain, more complex and long procedure and needs more mass, fuzzy and discrete innovation knowledge than product innovation.

So, we consult the concept of CAI and the technology of product innovation, focus on the characteristics and requirements of process innovation, research how to implement process innovation knowledge management a nd process problem resolving from technical level for knowledge-driven c omputer aided process innovation (KCAPI) in order to improve the quality and success rate of process innovation.

2. FRAMEWORK OF KCAPI
There are several remarkable c haracteristics for CAI s uch as knowledge-driven innovation process, application of some innovative theories, including methods and tools and clos e link with IT [6]. So, we define CAPI as flowing: CAPI is a system engineering which is based on innovativ e methodology and knowledge management technology, aims at creating or reforming manufacturing technologies, builds structured process innovation theory and methods with computer aided technology, supports all involved people to participate in the activity of process innovation knowledge accumulation and process problem solving.

In mass customization production mode, changeable products need steady and mature manufacturing process. The lifecycle of process covers the lif ecycles of m ultiple products. So, process innovation has been the main component of technical innovation [7]. CAPI provides innovative technology to s olve process problems occurring in different phases of product lifecy cle, so CAPI system is an im portant and organic element of manufacturing informationization system engineering as shown in Figure 1.

CAPI is a system engineering whose system frame includes knowledge management domain a nd problem solving domain. Knowledge management do main provides knowledge base for process innovation activity. Problem solving domain uses PIK to resolve process problem and generates process solution. The two domains are divided into three levels: method level, tool level and procedure level. CAPI is based on TRIZ, ontology and other basic theories [3, 6, 8]. Meanwhile, the adoption and diffusion of process innovation mainly depends on individual but not organization, so CAPI emphasizes full participation and collective intelligence. The position of CAPI is shown as Figure 2.

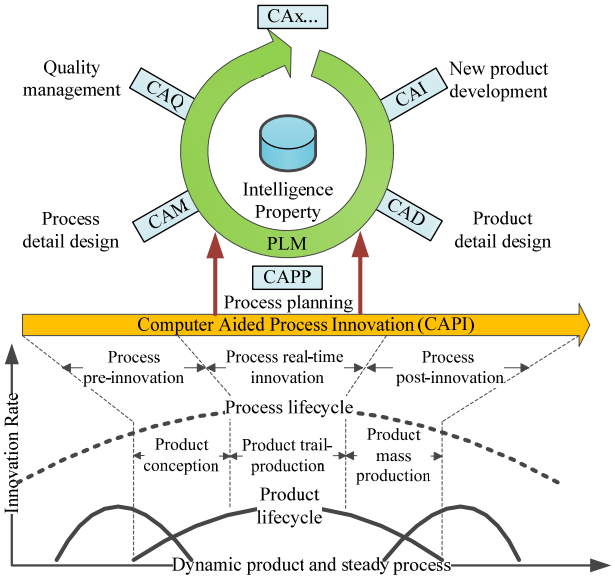


Figure 1. Position of CAPI

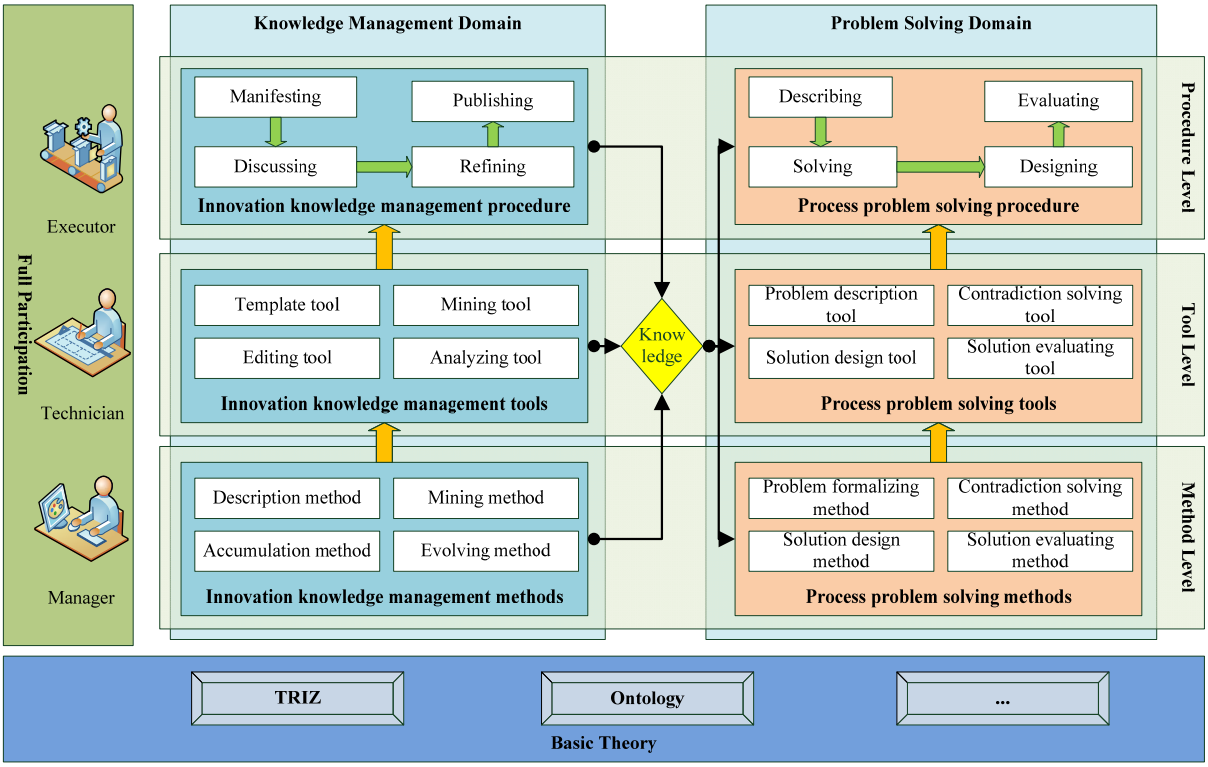


Figure 2. Framework of KCAPI

3. FORMALIZED BUILDING OF PIK NETWORK

PIK supports to implement pro cess innovation activity correctly resulting in the generation of new process knowledge. Five kinds of PIK are used in process innovation activity which is de scribed in Table 1.

Process innovation depends on a mass of PIK. Only after the PIK are connected together from discrete knowledge unit to knowledge network, the PIK can play the biggest role. Proces s innovation knowledge network is highly similar to neural network . Neural network consists of l arge numbers of neuron, accepts external stimulation and outputs control action through the interaction between the neurons. PIK network consists of la rge numbers of PIK unit, accepts the stimulation of process problem and outputs

innovative solution through the interaction between PIK units. So, we proposed a building method of PIK network by the imitation of neural network.

Table 1. Classification of PIK

Kind	Concept	Function
Problem description Template	The formal description format of process problem with the manufacturing domain glossary and semantics.	Descript process problem formally and uncover process contradiction.
Process contradiction matrix	The permutation and combination of process contradiction parameters and the mapping relation between process contradiction and innovative principle.	Solve process contradiction and indicate the innovation direction.
Science effect	The basic scientific principle and its typical implementary structure of multidisciplinary theory.	Map innovation direction to bas ic implementation structure.
Innovative solution instance	The formal description of typical process innovative solution or technical patent.	Provide reference solution for detail design of process innovation solution.
Manufacturing capability description	The formal description of enterprise manufacturing capability such as machining capability of equipment.	Evaluate the manufacturability of process solution for selecting optimized solution.

Process innovation depends on a mass of PIK. Only after the PIK are connected together from discrete knowledge unit to knowledge network, the PIK can play the bigges t role. Process innovation knowledge network is highly similar to neural network. Neural network consists of large nu mbers of neuron, accepts external stimulation and outputs control action through the interaction between the neurons. PIK network consists of larg e numbers of PIK unit, accepts the stimulation of process problem and outputs innovative solution through the interaction between PIK units. So, we proposed a building method of PIK network by the imitation of neural network.

We call PIK unit as process innovation knowledge neuron (PIKN) and build IKN by imitating the structure of neuron. The PIKNs connects each other through knowledge interface in order to make knowledge neural network. PIKN has knowledge parameter input interface and knowledge re sult output interface respectively corresponding to dendrite and axon of neuron. The encapsulation space of PIKN is mapped to cell membrane of neuron. The knowledge attribute is mapped to cytoplasm of neuron. The core handling process is mapped to nucleus of neuron. The structure of PIKN is described as shown in Figure 3.

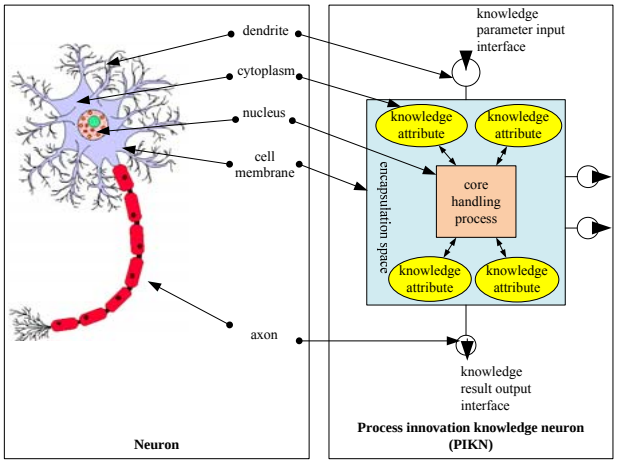


Figure 3. Structure of PIKN

According to function, PIKN c an be divided into description PIKN (including Problem description template knowledge, called as description neuron (DN)), solving PIKN (including Process contradiction matrix knowledge and science effect knowledge, called as contradiction neuron (CN) and e ffect neuron (EN) respectively), solution PIKN (inclu ding Innovative s olution instance knowledge, called as solution neuron (S N)) and evaluation PIKN (including manufacturing capability description knowledge, called as capability neuron (AN)) . Abundant and discrete IKNs are stored in knowledge base. IKNs accept the input of process problem as stimulatory signal. Then the signal is passed into knowledge encapsulation space, is dealt with knowledge core handling process with knowledge attributes. The processing result will be transmitted to other PIKNs to solve problem iteratively, so the process innovation knowledge neural network (PIKNN) is generated. Finally the process solution is output and process problem is solved. The structure of PIKNN is shown as Figure 4.

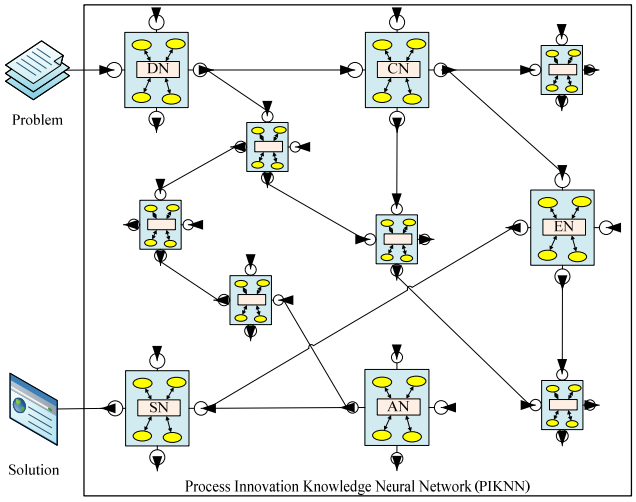


Figure 4. Structure of PIKNN

4. COLLABORATIVE ACCUMULATION OF PIK

Process innovation activity involves large numbers of people, but their knowledge levels are uneven, and their knowledge domains

are different. So, the proces s of PIK accumulation should be a process where all people can participate, each people can unleash personal strengths, collective intelligence can be integrated together, and the PIK can be turned from recessive to dominant, from rough to refined and from discrete to correlative.

Social network provides a knowle dge exchanging, s haring and manifesting platform based on relationship network and interest topic beyond background and specialty. WIKI network provides a knowledge refining and linking platform through page lock and collaborative editing [9]. So, bas ed on the advantages of social network and WIKI network, oriented at PIK unit building phase and PIK network building phas e, we p roposed PIK accumulation method based on bilayer social WIKI network as shown in Figure 5.

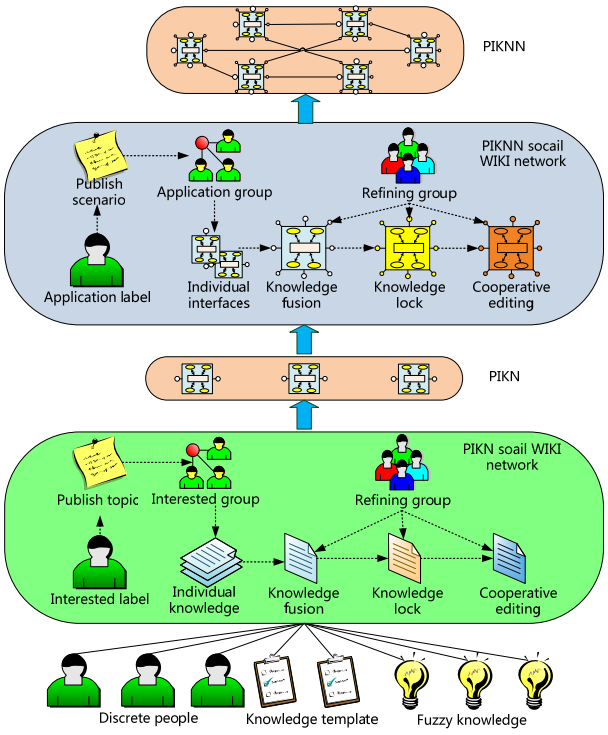


Figure 5. Bilayer social WIKI network

The first layer social WIKI network is used to accumulate PIKNs. Discrete technicians publish PIKN to pics. The interested technicians for some topic are gathered into a group through social relationship. Then they discuss this topic and manifest this knowledge using knowledge template from the point of view of individual specialty and experience. Some of mos t rational individual knowledge will be merged into a rough PIKN. Then, this rough PIKN will be locked and be refined editing through WIKI technology by knowledge refined group. Finally, the high quality PIKNs will be accumulated.

The second layer social WIKI network is used to accumulate PIKNN. PIKN needs specific interfaces and parameters for specific application domain and scenario in order to form self-organizing knowledge network. Its accumulation process is similar to PIKN accumulation. Firstly, discrete technicians publish application scenarios. Then, some interested technicians for some scenario discuss through social relationship, each of them can add specific interfaces and parameters individually for PI KNs. Then, the different interfaces and parameters of the same PIKN will be merged. Finally the PIKNs will be locked, cooperatively edited and refined. The PIKNs with specific interface and parameters can accept specific problem input, self-organize and self-associate to form the PIKNN and solve the process problem.

5. STRUCTURED SOLVING OF PROCESS PROBLEM

When the PIKNN is accumulated, the process problem can be solved with structured procedure in the PIKNN. In fact, knowledge-driven process innovation is a procedure where PIKNN accepts process problem input and outputs process innovative solution. The basic steps are described as Figure 6.

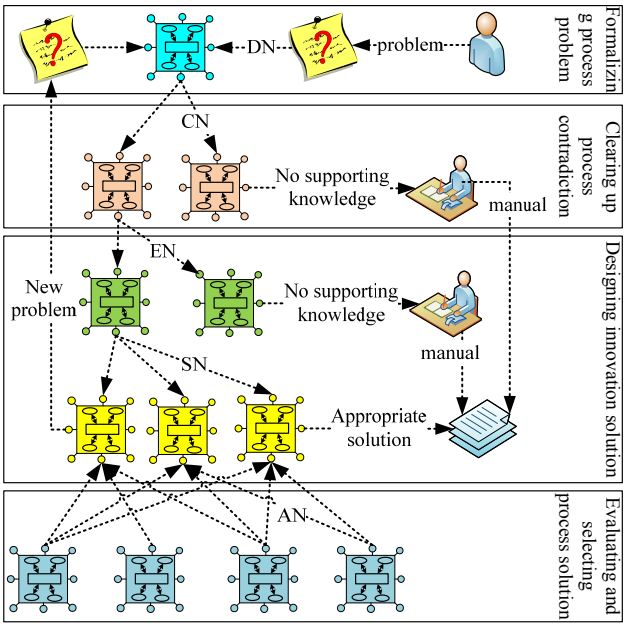


Figure 6. Solving process of process problem

Step 1: after accept process problem input, appropriate DNs formalize this problem, and output process contradictions.

Step 2: suited CNs input process contradictions, resolve it and get specific innovative principles, and the principles will be turned to function requirements.

Step 3: right E Ns input function requirements and give the basic solution structure according to the relative effects.

Step 4: according to the similar structure and application scenario, SNs are matched and be edited in detail to be detail solutions.

Step 5: the detail solutions are evaluated with ANs based on the used manufacturing resource, and finally output the solution with good manufacturability.

In the s olving process, if there are no appropriate PIKNs to support solving, the manual design is needed. If the s olution of process problem results in new process problem, this new process problem will be solv ed iteratively until the process contradiction is eliminated completely or some compromise is recognized.

Hereon, we use a process innovation instance of pressure transmitter to s how the validit y of knowledge-driven compute r

aided process innovation. An enterprise developed a new pressure transmitter. But the layout of the components is special, this result in more missing welding and incomplete weld with traditional weld technology. The solving process is described as follows:

1) Formalizing process problem. The generated DN like this:

(S)Product requirement→(V)change→(O)welding position→(P)15 °	CAUSE	(S)Component layout →(V)increase→(O)weld defect→(P)30%.

2) Clearing up process contradiction. The CN is got from formal description of process problem as: welding position→weld defect, and the innovative principle is got: solder substitute and separation.

3) Designing innovation solution. The EN is got by using above CN and a basic solution structure is generated: Non-uniform hot melt welding effect. Then, a similar SN is matched and a detail solution is proposed: using the two alloy solders with different melting point and non-contact Reflow Oven method.

4) Evaluating and selecting process solution. The used manufacturing equipments, tools and other resource is linked to ANs to evaluate this solution, finally this solution is adopted because its manufacturability is acceptable.

Using the new manufacturing method, this qualified rate of this pressure transmitter is improved from 84% to 97%.

6. CONCLUSION

The knowledge-driven computer aided process innovation method uses PIKN and PIKNN to formalize process innovation knowledge, accumulates innovation with bilayer social WIKI network, solves process problem with a structured procedure. This method can meet the requirements of process innovation knowledge, decrease the arbitrariness, randomness and cost of process innovation activity, increase the success rate of process innovation activity. In the future, We will focus on some automated method to improve the automation degree of process innovation accumulation and process problem solving, such as process contradiction matrix mining method oriented technical patent and iterative solving method for complex problem.

7. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China under grant number 51105313.

8. REFERENCES

[1] J. Schumpeter. 1990. *The Theory of Economic Development*. Commercial Press, Beijing, China.

[2] J.H. Geng, X.T. Tian. 2010. Knowledge-based Computer Aided Process Innovation Method. *Adv. Mater. Res.* 97-101, (Mar. 2010), 3299-3302. DOI= 10.4028/www.scientific.net/AMR.97-101.3299.

[3] S. Husig, S. Kohn. 2011. "Open CAI 2.0'' – Computer Aided Innovation in the era of open innovation and Web 2.0. *Comput. Ind.* 62, (2011), 407-413. DOI=10.1016/j.compind.2010.12.003.

[4] F. Carbone, J. Contreras, J. Hernández and J. M. Gomez-Perez. Open Innovation in an Enterprise 3.0 framework: Three case studies. *Expert Syst. Appl.* 39, (2012), 8929 – 8939. DOI=10.1016/j.eswa.2012.02.015.

[5] R. Simonetti, D. Archibugi, R. Evangelista. Product and process innovations: how are they defined? How are they quantified?. *Scientometr.* 32, 1(1995), 77-89.

[6] S. Husig, S. Kohn. 2009. Computer aided innovation—State of the art from a new product development perspective. *Comput. Ind.* 60, (2009), 551-562. DOI= doi:10.1016/j.compind.2009.05.011.

[7] Bi Ke-xin, Huang Ping and Shi Fang-fang. 2012. A study on the process and model for process innovation in manufacturing enterprises based on knowledge management. *J. Syst. Manag.* 21, 4(Jul. 2012), 478-484.

[8] C. Zanni-Merk, D. Cavallucci and F. Rousselot. 2009. An ontological basis for computer aided innovation. *Comput. Ind.* 60, (2009), 563-574. DOI= 10.1016/j.compind.2009.05.012.

[9] J. Baumeister, J. Reutelshoefer and F. Puppe. 2011. KnowWE: a Semantic Wiki for knowledge engineering. *Appl. Intell.* 35, (2011), 323–344. DOI=10.1007/s10489-010-0224-5.

Plant Chili Disease Detection using the RGB Color Model

Zulkifli Bin Husin
School of Computer and Communication Engineering
Universiti Malaysia Perlis
Perlis, Malaysia
zulhusin@unimap.edu.my

Abdul Hallis Bin Abdul Aziz
School of Computer and Communication Engineering
Universiti Malaysia Perlis
Perlis, Malaysia
abdulhallis@unimap.edu.my

Ali Yeon Bin Md Shakaff
School of Mechatronic Engineering
Universiti Malaysia Perlis
Perlis, Malaysia
aliyeon@unimap.edu.my

Rohani Binti S Mohamed Farook
School of Computer and Communication Engineering
Universiti Malaysia Perlis
Perlis, Malaysia
rohani@unimap.edu.my

ABSTRACT
Nowadays, chili is an important and high value product that able to give higher returns to farmers. However, chili plant fruitfulness should be given priority so that it's not damaged by pets and diseases. There are a few diseases that could attack the chili plant through the leaves. This research paper describes an image processing technique that identifies the visual symptoms of chili plant diseases using an analysis of colored images. This project proposed the design of software program that recognizes the color and shape of the chili leaf image. A few problems and constraints had to be identified before starting the project such as the different color of chili leaf, shape of chili leaf taken in different angle and distance, the group of the chili leaf and the resolution of the image captured. LABVIEW software is used to capture the image of chili plant in RGB color model and MATLAB software is used to enable a recognition process to determine the chili plant disease through the leaf images. The image recognition processes include the threshold, complementation, edging, segmentation, colors comparison and colors recognition. The recognition result of this research is about 93.3% from 120 images of chili plant. The proposed method in recognizing chili plant disease is demonstrated by experiments.

Categories and Subject Descriptors
I.4.6 [**Image Processing and Computer Vision**]: Segmentation: Edge and feature detection

General Terms
Algorithms, Experimentation

Keywords
Chili disease; leaf image; image processing techniques

1. INTRODUCTION
Chili is included in the main horticultural commodities. At certain times, it becomes a very high demand in the market because supply is limited. Business chili indeed belongs in the high-risk plants. Therefore, strategies and technical knowledge and the field became an important matter to be mastered. The systematic and structured should be developing so that it will use by operators to increase the overall reduction. Many farmers refused to cultivate chili in the rainy season due to an increase of the disease are very high so the output is high risk and cannot guarantee the quality control and productivity is good.

Chili plant is threatened by a wide variety of plant diseases and pests. These can damage leaf, seedling, chili fruit and tree-trunk and wipe out entire harvests. About 42% of the world's total agricultural crop is destroyed yearly by diseases and pests. Farmers often must contend with more than one pest or disease and new pesticide-resistant pathogenic strains attacking the crop. The traditional method of identifying plant pathogens is through visual examination. This is often possible when after the only major damage has already been done to the crops. Therefore, any treatments will be not use at this time. To save the chili plant from irreparable damage by louse, farmers should be able to identify any infection even before it becomes too visible. An attack by disease causing organisms generates a complex immune response in a chili plant, resulting in the production of disease specific proteins involved in plant defense and in limiting the spread of infection. Louse also produce proteins and toxins to facilitate their infection, before disease symptoms appear such as the leaf color will change from their attributes. These leaf colors play vital role in the development of plant disease detection. Advances in vision technology, software technology, and biotechnology have made the development of such this disease detection is possible. These projects are designed to detect chili plant diseases early, either by identifying the presence of the louse in the plant (by testing for the presence of louse movement) or the color and shape produced by either the louse or the plant during infection. These techniques require minimal processing time and are more accurate in identifying louse. And while some require equipment and training, other procedures can be performed and monitor on site by automatically or by a person with no special training. So far, this research has been designed only to detect diseases of chili plant through leaf image.

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.16

As a result, producing chili is a daunting task as plants are exposed to various attacks from micro-organisms and diseases to bugs and pests. Influence the next phase of the attack is usually seen through the leaves stems or fruit inspection. To solve this potential problem, early identification and diagnosis of disease to determine the precise and rapid implementation of preventive measures should provide operators to solve the above problem before seriously damage to the whole of chili plant. It could be more details explanation in chapter 2 and chapter 3.

2. PROBLEMS DESCRIPTION AND CHALLENGES

Fertility of crops comes from plants that are free from pest and diseases and there is no interference in the environment, and always cut to stabilize growth. Figure 1 shows the result of plants chili being attacked by the disease. Figure 2 shows the larva, nymph and adult mites attack the leaves of plants chili with liquid smoke bud and fruit. The attack caused the leaf to drop. Leaves that are attacked will be curly and curved downward.

Among the major vector of attack is a lice Trip. To control this pest, it will used plastic mulch and make sure the seed is given treatment before planted. The Trip can be eliminated by using poison spray and how much it can be use it depends on how many percent of leaves damage. Therefore, it becomes problems to the farmers to control and to manage their sprays activities for certain periods of time.

Figure 1. Samples of plant chili disease [5]

Figure 2. Pets and lice Trip [5]

3. MATERIALS AND METHODS

3.1 Data Acquisition

Image processing is traditionally concerned with preprocessing operations such as Fourier filtering, edge detection and morphological operations. Computer vision extends the image processing paradigm included with understanding of scene content and objects classification methods. Therefore, this research paper demonstrates the use of image processing techniques to detect the chili plant disease through the leave image. The system consists of two major parts (*see* Figure 3) such as the image captured by the camera using LabVIEW software tools as a Graphical User Interface (GUI) and image processing techniques using MATLAB software tools.

The first part of the project is to capture the image of chili leaf. The images of the chili leaf are fixed with 7.1M pixels sizes were captured by camera (*see* Table 1). MATLAB software is chosen to perform the image processing techniques using RGB color model. Image processing techniques requires numerous standard procedures and steps to identify and to recognize the color of chili plant image. Figure 3 illustrates the block diagram of plant chili disease detection system. The photo images prepared for the experiment sample have some fixed criteria as mention before. Both of the healthy and diseased leaf samples were used for the experimental purpose of this research. For better result, the leaves sample should be in good condition and follow their fixed criteria. Throughout the photo capturing section, the distance of the camera and the leaf was adjustable in order to get a clear shot of leaf pattern.

The input photo image is JPG file image and the size of resolution is 3872 x 2592 pixels.

```
a=imread('A(1).JPG');
A=imresize(a, [800 536]);
```

The *imread* function is used to read the image from graphic file. The *imresize* function is to returns an image of the size specified by [m-rows n-cols]. Images are resized for easier image processing. Figure 4 shows the block diagram of capturing the chili plant image [10][11].

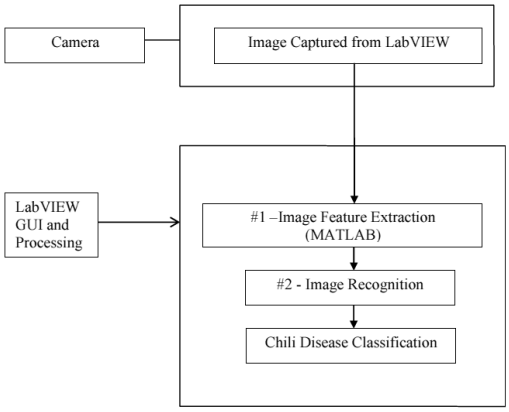


Figure 3. The diagram of plant chili disease detection system

Table 1. The photo image characteristic

Type	Leaf sample image	WebCAM image
Format	JPG	PNG
Resolution [m,n]	800 x 536 pixels	640 x 480 pixels
Columns, m	800	600
Rows, n	536	480

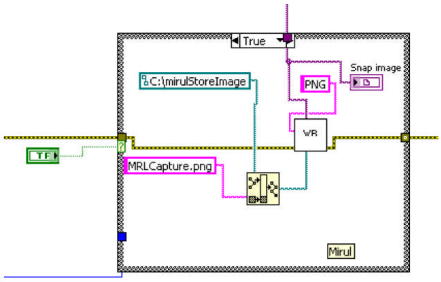


Figure 4. The Image captured using LABVIEW IMAQ Vision

3.2 Image Processing Methods

Basically, the colors of humans and some other animal perceive in an object are determined by the nature of light reflected from the object. With the absorption characteristics of the human eye, colors are seen as variable combinations of the primary colors such as red (R), green (G), and blue (B) (*see* Figure 5). The purpose of type color model is to facilitate the specification of colors in some standard for easier process to recognize the images. In essence, color model is a specification of a coordinate system where each color is represented by a single point. In digital image processing, the hardware-oriented models most commonly used in practice for example the RGB (red, green, blue) model for color monitors and a broad class of color video cameras; the CYM (cyan, yellow, magenta) and CMYK (cyan, magenta, yellow, black) models for color printing and the HSI (hue, saturation, intensity) model, which corresponds closely with the way of humans describe and interpret color. In this research, the RGB model is used to determine the disease of plant chili through leaf images.

In the RGB model [6], each color appears in its primary spectral components of red, green, and blue. This model is based on a Cartesian coordinate system. RGB primary values are at three corners; the secondary colors cyan, magenta, and yellow are at three other corners; black is at the origin; and white is at the corner farthest from the origin (*see* Figure 6). In this model, the gray scale extends from black to white along the line joining these two points. The different colors in this model are points on or inside the cube, and are defined by vectors extending from the origin [7].

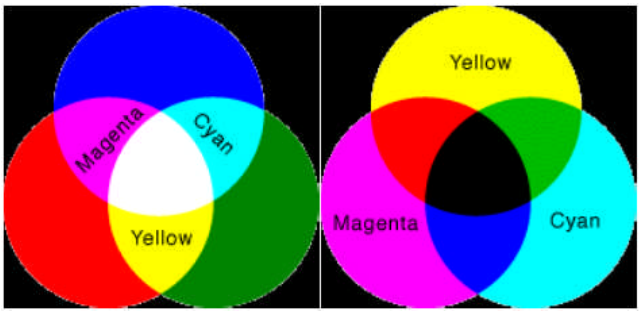


Figure 5. Primary and secondary colors of light and pigments [4][12]

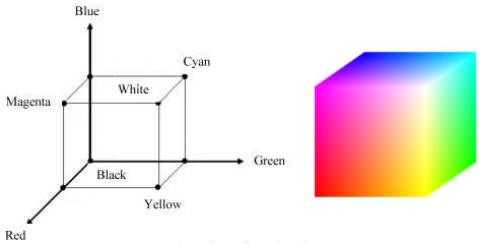


Figure 6. The RGB colors cube [4][12]

3.3 Feature Extraction

The input image is enhanced to preserve information of the affected pixels before extracting chili leaf image from the background [1][2]. The color model respectively is used to reduce effect of illumination and distinguish between chili and non-chili leaf color efficiently [3]. The resulting color pixels are clustered to obtain groups of colors in the image is shown in Figure 7.

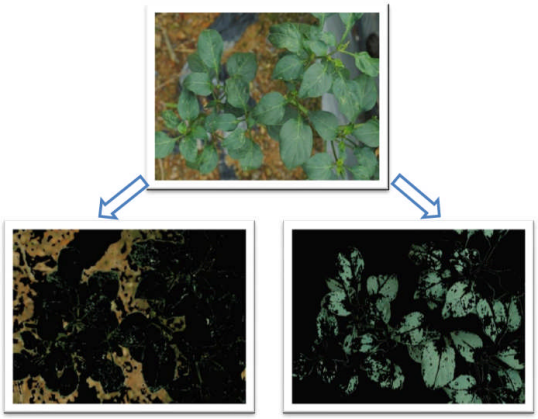


Figure 7. The result of color clustering

3.4 Image Recognition

Extracts the color features of image, which can be used for color matching or other applications related to the color information such as color identification and color image segmentation [8][9][10]. Figure 8 and Figure 9 shows the samples of chili plant image with healthy and diseased condition. Figure 10 shows the samples result of histogram graph after the complete process of clustering methods is done.

Figure 8. The healthy of plant chili

Figure 9. The plant chili disease

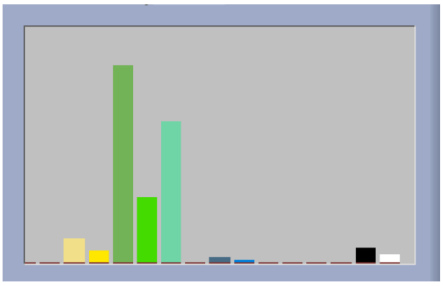


Figure 10. The result of histogram graph

4. RESULTS AND DISCUSSION

The system was tested with image of 800x536 pixels. The total samples are 120. The graphical user interface (GUI) result of healthy and disease plant chili is shown in Figure 11 and Figure 12. All the plant chili samples are tested. Table 2 illustrated the samples result of plant chili healthy and disease that was implemented in this paper. Figure 13 shows the result of healthy and risky classification of plant chili disease.

The method implemented in this research paper is effective and fastest method in detection of plant chili disease. The overall result was satisfying (93.3%) and is considered as a successful project. This paper has introduced a number of techniques in image processing for image and color recognition of an image photo.

As a conclusion, this research strongly recommends to be use for early detection of plant chili disease through leaf inspection. Leaves images captured are processed to determine the healthiness of each chili plant. By using leaf samples in RGB color model technique, it will identify the potential problems to the chili plants before its goes seriously damage for all chili field. With this method, the use of harmful chemicals on plants can be reduced and hence to ensure a healthier environment. It also may be even lowering the production cost of the maintenance and producing a high quality of chili

Figure 11. The result of healthy plant chili

Figure 12. The result of plant chili disease

Table 2. The samples of plant chili healthy and risky result

Sample	Yellow	Cyan	Green	Healthy/Disease
1	5.304	0	0	High Risky
2	2.597	0	0	Risky
3	0.035	10.347	0	Healthy
4	2.917	13.511	0	Risky
5	0.413	14.474	0	Healthy
6	0	18.116	0	Healthy
7	0.19	18.519	0	Healthy
8	0	24.711	0.05	Healthy
9	0.003	11.199	0.06	Healthy
10	4.336	12.911	0.293	Risky
11	0	28.316	0.35	Healthy
12	0	17.073	0.608	Healthy
13	4.022	3.234	0.639	Risky

14	0	2.102	0.79	Healthy
15	0.04	1.454	0.793	Healthy
16	8.365	0	0.835	High Risky
17	0.139	0	0.858	Healthy
18	6.928	0	0.936	High Risky
19	2.104	9.507	1.006	Risky
20	3.698	0.417	1.088	Risky
21	3.151	0	1.137	Risky
22	1.802	3.719	1.266	Risky
23	1.141	0	1.646	Risky
24	3.696	0	1.668	Risky
25	2.715	10.842	2.031	Risky
26	2.009	0	2.246	Risky
27	1.067	15.543	2.349	Risky
28	1.737	7.758	2.524	Risky
29	3.786	0	2.595	Risky
30	3.601	0	2.848	Risky
31	2.743	7.614	3.367	Risky
32	1.694	0	3.706	Risky
33	0.03	0.023	3.772	Healthy
34	2.539	0	3.789	Risky
35	1.192	0	3.917	Risky
36	3.946	0	3.922	Risky
37	6.074	0	4.008	High Risky
38	4.454	2.5	4.161	Risky
39	8.966	0	4.166	High Risky
40	0	0	4.259	Healthy
41	4.066	0	4.595	Risky
42	2.55	0	4.733	Risky
43	1.154	0	5.405	Risky
44	0	19.67	5.455	Healthy
45	3.661	0	5.57	Risky
46	1.146	0	5.594	Risky
47	0	0.462	6.128	Healthy
48	1.044	0.25	6.338	Risky
49	4.558	1.927	6.556	Risky
50	1.405	0	6.675	Risky
51	1.308	0	6.748	Risky
52	4.323	0	7.029	Risky
53	7.115	0	7.302	High Risky
54	9.809	0.002	7.415	High Risky
55	2.198	0	7.439	Risky
56	6.351	0	7.828	High Risky
57	2.017	0	7.919	Risky
58	0.226	0.443	7.926	Healthy
59	4.117	0	8.305	Risky
60	4.225	0	8.33	Risky
61	4.523	0	8.46	Risky
62	5.173	0.536	8.886	Risky
63	7.743	0	9.139	High Risky
64	1.309	1.831	9.266	Risky
65	0.067	12.605	9.351	Healthy
66	3.542	0	9.456	Risky
67	12.621	0	9.689	High Risky
68	3.337	1.377	9.814	Risky
69	1.515	0	9.974	Risky
70	3.379	0	10.29	Risky
71	1.453	0	10.385	Risky
72	3.351	0.44	10.491	Risky
73	1.038	0	11.057	Risky
74	3.121	28.41	11.15	Risky
75	16.973	0	11.181	High Risky
76	6.739	0	11.402	High Risky
77	13.251	0	11.557	High Risky
78	3.606	0	12.1	Risky
79	7.044	0.521	12.209	High Risky
80	3.774	0	12.555	Risky
81	1.386	0.357	12.717	Risky
82	1.212	6.084	12.857	Risky
83	5.362	0	13.402	Risky
84	0.216	0.279	14.049	Healthy
85	1.107	24.883	14.129	Risky
86	2.488	0	14.232	Risky
87	1.085	0	14.458	Risky
88	3.97	0	14.689	Risky
89	1.828	0	15.077	Risky
90	2.451	0.138	15.92	Risky
91	1.126	3.059	16	Risky

92	2.375	0.264	16.418	Risky
93	12.183	0	16.837	High Risky
94	1.825	0.001	17.16	Risky
95	1.269	0	17.21	Risky
96	3.775	0	17.293	Risky
97	1.317	0	17.315	Risky
98	1.289	0	17.391	Risky
99	2.373	0.171	17.61	Risky
100	2.656	0	17.863	Risky
101	0.519	0.09	18.351	Healthy
102	0.731	0.52	18.581	Healthy
103	2.356	0	18.839	Risky
104	2.932	0	19.905	Risky
105	0.828	0.06	20.417	Healthy
106	0.828	0.06	20.417	Healthy
107	0.914	0.067	20.432	Healthy
108	0.83	0.029	20.93	Healthy
109	0.739	0.163	21.218	Healthy
110	0.835	0.067	21.426	Healthy
111	1.248	14.954	21.441	Risky
112	0.014	0.007	22.916	Healthy
113	0.624	0.16	23.127	Healthy
114	5.312	0.934	26.58	Risky
115	1.688	0	26.718	Risky
116	4.078	0.011	27.854	Risky
117	9.653	2.382	27.91	High Risky
118	12.981	0	28.157	High Risky
119	8.397	0	29.625	High Risky
120	0.521	0	31.742	Healthy

5. ACKNOWLEDGMENTS

This study was supported by the Ministry of Science, Technology & Innovation (MOSTI) and Universiti Malaysia Perlis.

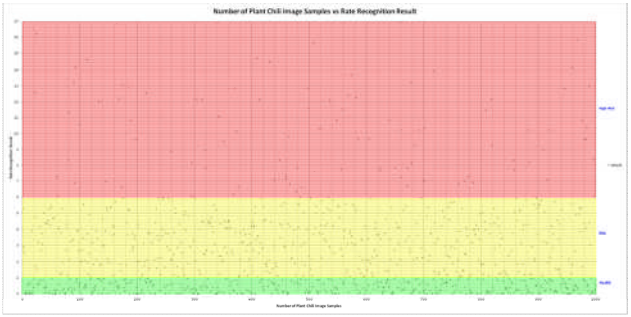


Figure 13. The Samples of Plant Chili Recognition of Healthy and Risky Result

6. REFERENCES

[1] J.-X. Du, X.-F. Wang, and G.-J. Zhang, "Leaf shape based plant species recognition," Applied Mathematics and Computation, vol. 185, 2007

[2] H. Fu and Z. Chi, "Combined thresholding and neural network approach for vein pattern extraction from leaf images," IEEE Proceedings-Vision, Image and Signal Processing, vol. 153, no. 6, December 2006

[3] J. Du, D. Huang, X. Wang, and X. Gu, "Shape recognition based on radial basis probabilistic neural network and application to plant species identification" in Proceedings of 2005 International Symposium of Neural Networks, ser. LNCS 3497. Springer, 2005

[4] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, Digital Image Processing Using MATLAB. Prentice Hall, 2004

[5] Norhashim Kamisan, "Penanaman Cili: Panduan Lengkap Kaedah Fertigasi" Grupbuku Karangkraf, 2011

[6] B. C. Heymans, J. P. Onema, and J. O. Kuti, "A neural network for opuntia leaf-form recognition," in Proceedings of IEEE International Joint Conference on Neural Networks, 1991

[7] Amnon Shashua and Tammy Riplin-Raviv (2001). The Quotient Image: Class-Based Re-Rendering and Recognition with Varying Illuminations, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, pp. 129-139

[8] Ingeborg Tastl and Gunther Raidl, *Transforming an analytically defined color space to match psychophysically gained color distances,* the SPIE's 10th Int. Symposium on Electronic Imaging: Science and Technology (San Jose, CA), vol. 3300, 1998, pp. 98–106

[9] A.L Yuille, D. Snow, and M. Nitzberg, *Signfinder : Using color to detect, localize and identify informational signs,* Int. Conf. on Computer Vision ICCV98, 1998, pp. 629–633

[10] B. Funt and G. Finlayson, *Color constant color indexing,* IEEE Trans. On Pattern Analysis and Image Processing **17** (1995), 522–529

[11] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley, *Color transfer between images,* IEEE Computer Graphics and Applications September/ October (2001)

[12] Maria Petrou and Panagiota Bosdogianni (2003). *Image Processing: The Fundamentals.* 3rd Edition, John Wiley & Sons, LTD, England.

Reversible Data Hiding with Prediction-based Histogram Alteration

Hsiang-Cheh Huang
Nat'l Univ. of Kaohsiung
700 University Road,
Kaohsiung 811, Taiwan
hchuang@nuk.edu.tw

Feng-Cheng Chang
Tamkang University
180 Linwei Road,
Ilan 262, Taiwan
135170@mail.tku.edu.tw

Wai-Chi Fang
Nat'l Chiao Tung Univ.
1001 Ta-Hsueh Road
Hsinchu 300, Taiwan
wfang@nctu.edu.tw

Sheng-Hong Li
Nat'l Univ. of Kaohsiung
700 University Road,
Kaohsiung 811, Taiwan
hch.nuk@gmail.com

ABSTRACT
In this paper, we propos e a new algorithm in reversible data hiding, with the applications of copyright and privacy protection for natural image s. Security issues for digital images have long been an important topic in research and practical applications, and reversible data hiding has attracted more and more attention. Unlike conventional schemes, we take the prediction of luminance values and their spatia l coordinates into consideration in order to look for better performances. W e take the inherent characteristics of original image into account, with the prediction-based difference histogram alteration. By doing so, the larg er embedding capacity with similar output image quality in reversible data hiding can be obtained. Simulation results demonstrate the sup eriority over exis ting schemes, and the effectiveness for practical applications for digital images.

Categories and Subject Descriptors
H.5 [**INFORMATION INTERFACES AND PRESENTATION**]: H.5.1 Multimedia Information Systems

General Terms
Algorithms.

Keywords
Reversible data hiding, histogram, prediction, authentication.

1. INTRODUCTION
Data hiding is one of the im portant schemes in the applications of digital rights management (DRM) [1], including content authentication, covert communi cation, and copyright protection. Conventional watermarking techniques aim at examining authenticity between embedded watermark and extrac ted one. Under this scenario, at the encoder, the watermark is embedded with the algorithm designed by researchers, and irreversible degradation of original multimedia content can be expected. A t the decoder, the watermark is extracted, and compared with the embedded one to examine the authenticity . Only the e xtracted watermark should be examined, and the image is ignored.

The concepts and algorithms of reversible data hiding have emerged in early 2000's [2][3]. At the encoder, secret information is embedded into m ultimedia contents, mostly images, by algorithms developed b y researchers. At the decoder, unlike conventional watermarking, not only the embedded watermark should be extracted, but the original image sho uld also be recovered. More importantly, the extracted watermark and the recovered image should be identical to their counterparts at the encoder. With these requirements, practical implementati ons for reversible data hiding can roughly be classified into categories , one is to modify the histogram of original image from the global point of view [4], and the other is to alter the relationships among neighboring pixels from the standpoint of spatial locality [2].

Performance evaluations relating to reversible data hiding algorithms can be briefly stated as follows [5][6].

- Reversibility. This is the most important criterion of our data hiding algorithm. After reception of marked image, at the decoder, with reasonable amount of s ide information, both the hidd en information and the original image should be recovered.
- Embedding capacity. This is the number of bits o f secret for hiding into the image, which is an important factor in algorithm design. For the larger capacity, it implies that more secret data can be hidden into original images.
- Output image quality, also named imperceptibility, should be as resemble as its original counterpart.
- Side information, which is the reas onable amount of auxiliary information for the decoder, should be necessary for enhanced security.

In this paper, we focus on reversible data hiding algorithm with the prediction of luminance values bas ed on spatial locality of original image. Difference values between predicted image and its original counterpart are utilized for data hiding. The histogram of difference values is produced, a nd intentional alteration of the histogram, where no c alculation is needed, can make reversible data hiding possible. With the mea sures listed above, simulation results exhibit the better performances over existing algorithms.

This paper is organized as follows. In Section 2 we discuss about conventional algorithms in reversible data hiding. In Section 3 we then describe the propos ed algorithm based on prediction techniques. Simulation results are demonstrated in Section 4. Finally, we conclude this paper in Section 5.

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.17

2. CONVENTIONAL SCHEMES

Conventional schemes in reversible data hiding can be classified into two categories, one is by modifying the difference value between neighboring pixels [2][3], and the other is by altering the histogram of original image [4]. For the former one, it takes the local characteristics into account, and for the latter one, it considers the slightest modification of histogram based on the global statistics of original image.

2.1 The Difference-Expansion-Based Scheme

The difference expansion (DE) method is one of the earliest schemes for reversible data hiding [2][3]. It follows the concepts from wavelet transforms with three steps. First, turn the pair of spatial pixel values at neighboring positions into two frequency coefficients, that is, high and low frequencies. Then, modify the high-frequency part and keep the low-frequency part intact. Data embedding is accomplished by doubling the high-frequency part, and adding the secret bit into the multiplied difference. This is the origin of the term of "difference expansion". And finally, turn frequency coefficient back to modified spatial pixel values.

The DE-based scheme is famous for the high capacity for data embedding. However, by modifying the high-frequency part, the overflow problem may be emerged because the modified pixel values may lie outside 0 and 255. Coordinates of such locations, named "location map", should be recorded as side information for decoding. The size of location map would decrease the effective capacity of the scheme.

2.2 The Histogram-Based Scheme with Luminance and Its Difference

The histogram-based method [4] is famous for its ease of implementation and few overhead generated. By intentionally increase the luminance values that are larger than the luminance with the maximal occurrence in the histogram, data embedding can be accomplished. Capacity is limited by maximal occurrence, which is the major drawback with this type of scheme.

It is easily observed that luminance values between neighboring pixels are similar for ordinary images. Thus, by taking difference values and producing corresponding histogram, conventional histogram-based method [4] can be utilized for data hiding, while the increased capacity can be expected. By following this concept, new data hiding scheme can be proposed in Sec. 3.

3. PREDICTION-BASED SCHEME FOR REVERSIBLE DATA HIDING

Here we describe the prediction-based reversible data hiding algorithm in this paper. Suppose the size of original image is 512×512. We first divide the original images into two types: (a) 4096 blocks with the size of 8×8, or (b) 1024 blocks with the size of 16×16. The standard deviation σ_k of each block is calculated, where $0\le k\le 4095$ for 8×8 blocks, and $0\le k\le 1023$ for 16×16 blocks. For blocks with smaller standard deviation, they can be classified as the smooth blocks. Hence, they are expected to hide more secret information because difference values tend to be concentrated around zero. On the contrary, blocks with larger standard deviation are known as active blocks. In order to ease the classification of blocks, we set the integer-valued threshold σ_{th} with the conditions that

$$\text{block}_k=\begin{cases}\text{active}, & \text{if } \sigma_k\ge\sigma_{\text{th}};\\ \text{smooth}, & \text{if } \sigma_k<\sigma_{\text{th}}.\end{cases} \tag{1}$$

We use the 1024- or 4096-bit block map for representing the classification of blocks; if the block is active, it is denoted by bit '1', otherwise it is denoted by bit '0'. The block map is served as the side information for the decoder. To balance the classification result, we adjust σ_{th} to obtain around half of smooth blocks, and half of active blocks in the original image.

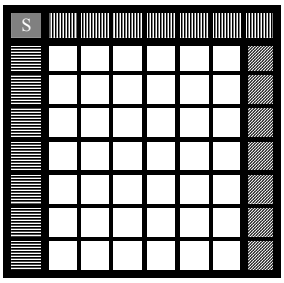


Figure 1. The 8×8 original image is classified into four types.

Data embedding procedures can be described as follows.

Step 1. Generate the predicted image. We take the block X with the size of 8×8 in the original image as an instance in Fig. 1. Prediction of data starts from the pixel at upper-left corner, denoted by 'S' as the seed. The predicted image is denoted by X_p. The seed locates at $X(0,0)$. Next, luminance prediction of the pixels in the first row, shown in vertical lines, is calculated.

$$X_p(0,j)=X(0,j-1),\quad j=1,\cdots,7. \tag{2}$$

Then, luminance prediction of the pixels in the first column, shown in horizontal lines, is calculated based on the weighting factors in Fig. 2.

$$X_p(i,0)=\text{round}\left(\frac{n_{\text{N}}\cdot X(i-1,0)+n_{\text{NE}}\cdot X(i-1,1)}{n_{\text{N}}+n_{NE}}\right),\quad i=1,\cdots,7. \tag{3}$$

Next, luminance prediction of the pixels in the last column, shown in diagonal lines, is calculated.

$$X_p(i,7)=\begin{cases}\min(X(i-1,7),X(i,6)), & X(i,7)\ge\max(X(i-1,7),X(i,6))\\ \max(X(i-1,7),X(i,6)), & X(i,7)\le\min(X(i-1,7),X(i,6))\\ X(i-1,7)+X(i,6)-X(i-1,6), & \text{otherwise}\end{cases} \tag{4}$$

For the remaining pixels, $i=1,\ldots,7,\quad j=1,\ldots,6$, predicted pixels can be calculated with pixels in the upper-left (northwest direction), upper (north direction), upper-right (northeast direction), and left (west direction), with the properly selected weighting factors

$$X_p(i,j)=\text{round}\left(\frac{n_{\text{NW}}\cdot X(i-1,j-1)+n_{\text{N}}\cdot X(i-1,j)+n_{\text{NE}}\cdot X(i-1,j+1)+n_{\text{W}}\cdot X(i,j-1)}{n_{\text{NW}}+n_{\text{N}}+n_{\text{NE}}+n_{\text{W}}}\right) \tag{5}$$

Weighting factors are represented with coordinates in Fig. 2, where the subscripts denote the direction of each factor, and subscript c implies the current position. With Eq. (3) and Eq. (5), by following raster scanning, we are going to adjust n_{NW}, n_{N}, n_{NE}, and n_{W}, for obtaining the larger capacity with the better output image quality.

Step 2. After the prediction procedure, the difference between original and predicted images are calculated by

$$d(i,j)=X_p(i,j)-X(i,j). \tag{6}$$

n_{NW}	n_N	n_{NE}
n_W	n_C	n_E
n_{SW}	n_S	n_{SE}

Figure 2. Weighting factors for data prediction. Subscripts imply the direction relationships to the central factor n_C.

The histogram of difference $d(i, j)$, denoted by D, is generated.

Step 3. We set the integer-valued embedding level, EL, and the embedding round index, $r_{em} = \lfloor \frac{1}{2} \cdot \mathrm{EL} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function, for data embedding. EL means the number of bins in difference histogram around 0, and r_{em} operates as separator between positive and negative values.

Data embedding schemes for odd-valued and even-valued EL are somewhat different. For odd-valued EL, if EL =3, $r_{em} = 1$, we take -1, 0, and 1 in the difference histogram for data embedding. The sum of the heights of the three implies the embedding capacity. On the other hand, for even-valued EL, if EL=4, $r_{em} = 2$, we take -1, 0, 1, and 2 in the difference histogram for data embedding. Again, the sum of the heights of the four implies the embedding capacity. Larger EL brings about larger capacity.

The odd- or even- valued r_{em} also influences the data embedding procedure. For odd-valued EL,

$$\tilde{D} = \begin{cases} D + r_{em}, & \text{if } D > r_{em}; \\ D - r_{em} - 1, & \text{if } D < -r_{em}; \\ D, & \text{else.} \end{cases} \quad (7)$$

For even-valued EL,

$$\tilde{D} = \begin{cases} D + r_{em}, & \text{if } D > r_{em}; \\ D - r_{em}, & \text{if } D < -r_{em} + 1; \\ D, & \text{else.} \end{cases} \quad (8)$$

Step 4. Suppose the difference value after the embedding of secret data bits, w, becomes D'. For an odd-valued r_{em}, if $r_{em} \geq 1$ data embedding is performed by

$$D' = \begin{cases} \tilde{D} + r_{em} - 1 + w, & \text{if } \tilde{D} = r_{em}; \\ \tilde{D} - r_{em} - w, & \text{if } \tilde{D} = -r_{em}; \\ \tilde{D}, & \text{else.} \end{cases} \quad (9)$$

After this, r_{em} is decreased by 1. After performing recursively until $r_{em} = 0$, data embedding is performed by

$$D' = \begin{cases} \tilde{D} - w, & \text{if } \tilde{D} = r_{em}; \\ \tilde{D}, & \text{else.} \end{cases} \quad (10)$$

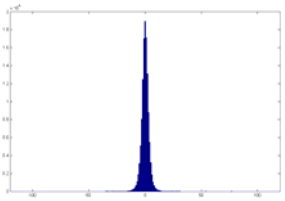
On the other hand, for an even-valued r_{em},

$$D' = \begin{cases} \tilde{D} + r_{em} - 1 + w, & \text{if } \tilde{D} = r_{em}; \\ \tilde{D} - r_{em} - 1 - w, & \text{if } \tilde{D} = -r_{em} + 1; \\ \tilde{D}, & \text{else.} \end{cases} \quad (11)$$

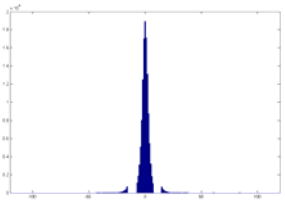
After this, r_{em} is decreased by 1. Eq. (11) is performed recursively until $r_{em} = 0$.

Step 5. Reconstruct the output image by adding the difference value back. By using raster scan, the modified difference value after data embedding, D', can be turned into $d'(i, j)$, and marked image $X_w(i, j)$ is obtained by

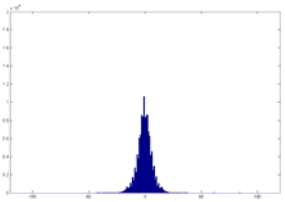
$$X_w(i, j) = X_p(i, j) - d'(i, j). \quad (12)$$



(a) Histogram of difference values for Lena.



(b) Emptying selected difference values in (a) for EL = 15. Values between –7 and 7 are ready for data embedding.



(c) Histogram of difference value after data embedding.

Figure 3. Comparisons of difference histograms. (a) Original difference histogram. (b) Emptying regions in Fig. 3(a) for EL=15. (c) Difference histogram after data embedding.

Extraction of secret data and recovery of original image are the reverse procedures to the embedding counterpart. At the beginning, the marked image $X_w(i, j)$, the embedding level, EL, and the four coefficients for performing prediction in Fig. 2, n_{NW}, n_N, n_{NE}, and n_W, are received. Pixels at the first column in Fig. 1 are recovered first based on Eq. (3), then the first row and the last column are predicted by Eq. (2) and Eq. (4), respectively, and finally, the rest of the pixels are calculated with Eq. (5). Difference values are then produced. By following procedures in reverse order, original image and hidden secret can be recovered.

We make comparisons of histograms of differences Fig. 3. By careful selection of the EL value, secret data can be reversibly hidden into original image with good output image quality and acceptable capacity. We easily observe that EL serves as the side information for the decoder. In comparison with conventional histogram-based scheme [4], or difference expansion (DE) scheme [2][3], the amount of side information is relatively fewer than its corresponding counterparts.

4. SIMULATION RESULTS

We have conducted experiments with several test images to examine the effectiveness of proposed algorithm. Performances with relating algorithm in [5] are also compared.

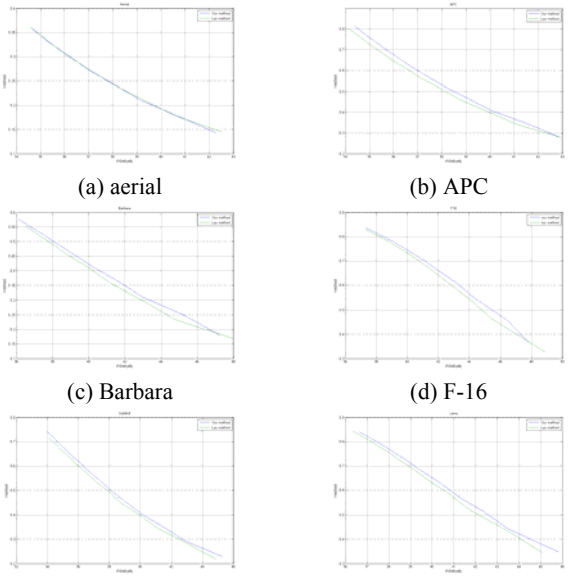
Both output image quality and embedding capacity are observed. Side information for decoding is also addressed. Due to limited space, we take the Lena test image and present the parameters in more detail. For the Lena image with size of 512×512, it is divided into 16×16 blocks, and the standard deviations of all the 1024 blocks are calculated. Next, suppose around half of the blocks are active, and the other half are smooth in Eq. (1), we choose $\sigma_{th} = 8$, which leads to 536 active blocks, denoted by '1', and 488 smooth ones, denoted by '0', in the block map. The 1024-bit block map is served as the beginning of secret information for embedding. Then, in Eq. (2) to Eq. (5), the four parameters for image prediction, n_{NW}, n_N, n_{NE}, and n_W, are trained to reach better results in the data hiding. Considering practical implementations, these four parameters are integer values between 1 and 8. After training, suitable parameter values are $(n_{NW}, n_N, n_{NE}, n_W) = (1, 8, 7, 8)$, the difference values can be produced in Eq. (6), and with the provided embedding level, EL, data can be embedded subsequently with Eq. (7) to Eq. (12). For EL=16, 119,997 and 108,267 bits can be embedded into smooth and active blocks, respectively, which leads to a total of 228,264 bits, or 0.87 bit/pixel (bpp).

Next, performance comparisons for six test images, including aerial, APC, Barbara, F-16, goldhill, and Lena, are depicted in Fig. 4. We can easily observe that results with our algorithm, shown in blue curves, outperform those with [5], shown in green curves. With the same amount of capacity, the output image quality in PSNR with our algorithm performs better than those with [5]. Regarding to the side information, In [5], both the parameters for prediction, or $(n_{NW}, n_N, n_{NE}, n_W) = (1, 2, 1, 2)$, which are fixed for different test images, and the selected embedding level EL, are the side information for decoding. For our algorithm, both the four parameters and the EL values need also be provided at decoder, with the flexibility that the four parameters are adjustable for enhanced performances.

Finally, we perform the authentication test for our algorithm. The four values of n_{NW}, n_N, n_{NE}, and n_W can also be served as the secret key for correct decoding and authentication at the decoder. If the four values are received erroneously, even the correct EL value is utilized, after decoding, neither the original image nor the hidden secret would be correctly recoverable. With our algorithm, it is extendable to data authentication algorithm, while the algorithm with [5] may have limited flexibility in this application.

5. CONCLUSIONS

In this paper, we propose an effective algorithm in reversible data hiding by utilizing the prediction of pixel values, and by altering the difference values between predicted image and original one to make data hiding possible. Considering the output image quality, the embedding capacity, and the side information for decoding, our results outperforms relating algorithms in literature. Simulation results have also pointed out the performance assessments under a variety of tests, with practical applicability and ease of implementation of proposed algorithm.



(a) aerial (b) APC (c) Barbara (d) F-16 (e) goldhill (f) Lena

Figure 4. Comparisons of capacity and image quality between our method and the scheme in [5] for six test images.

6. ACKNOWLEDGMENTS

The authors would like to thank National Science Council (Taiwan, R.O.C) for supporting this paper under Grant No. NSC101-2220-E-390-002.

7. REFERENCES

[1] Huang, H. C., and Fang, W. C. 2010. Metadata-based image watermarking for copyright protection. *Simulation Modelling Practice and Theory* 18, 4 (Apr. 2010), 436-445. DOI = http://dx.doi.org/10.1016/j.simpat.2009.09.004.

[2] Alattar, A. M. 2004. Reversible watermark using the difference expansion of a generalized integer transform. *IEEE Trans. Image Process.* 13, 8 (Aug. 2004), 1147-1156. DOI= http://dx.doi.org/10.1109/TIP.2004.828418.

[3] Kim, H. J., Sachnev, V., Shi, Y. Q., Nam, J., and Choo, H. G. 2008. A novel difference expansion transform for reversible data embedding. *IEEE Trans. Information Forensics and Security* 3, 3 (Sep. 2008), 456-465. DOI = http://dx.doi.org/10.1109/TIFS.2008.924600.

[4] Ni, Z., Shi, Y. Q., Ansari, N., and Su, W. 2006. Reversible data hiding. *IEEE Trans. Circuits Syst, Video Technol.* 16, 3 (Mar. 2006), 354-362. DOI = http://dx.doi.org/10.1109/TCSVT.2006.869964.

[5] Luo, H., Yu, F. X., Huang, Z. L., Chen, H., and Lu, Z. M. 2012. Reversible data hiding based on hybrid prediction and interleaving histogram modification with single seed pixel recovery. *Signal, Image and Video Processing* 2012 online. DOI = http://dx.doi.org/10.1007/s11760-012-0306-4

[6] Huang, H. C., and Fang, W. C. 2011. Authenticity preservation with histogram-based reversible data hiding and quadtree concepts. *Sensors* 11, 10 (Oct. 2011), 9717-9731. DOI = http://dx.doi.org/10.3390/s111009717.

An Online Software Upgrade Method for Intelligent power Distribution Terminal

Zhiyuan Xie
North China Electric Power University
North China Electric Power University,
Baoding, Hebei, China
+8613703365092
zhiyuanxie@263.net

Shengxiang Deng
North China Electric Power University
North China Electric Power University,
Baoding, Hebei, China
+8613730203515
dengshengxiang@163.com

Fenfen Dong
North China Electric Power University
North China Electric Power University,
Baoding, Hebei, China
+8615930481553
dongfenfen1990@126.com

ABSTRACT
In order to realize intelligent power distribution terminal to modify and improve software in operation continuously, stably and reliably, this paper designs an online upgrade method for intelligent power distribution terminal to renew the software based on the Flash memory of LPC2138. And the online upgrade solutions, the design of communication protocol, division of program storage space, online updating algorithm and security algorithm are discussed in detail. The experimental results show that the design achieves the performance required. It can complete the intelligent power distribution terminal online upgrade safely and reliably for practical projects.

Categories and Subject Descriptors
D.1.2 [**Programming Techniques**]: Automatic Programming.

General Terms
Algorithms, Design.

Keywords
Intelligent power distribution terminal; software upgrade; In-application programming (IAP)

1. INTRODUCTION
Intelligent power distribution terminal is a very important device in electric power system, which can realize power network working safely, reliably and effectively. According to the collected data, it is able to control the network equipment intelligently. It need to be modified and perfected constantly for application loopholes are inevitable. Once there were some bugs in the software, we could only exchange the devices or update the software with the programming interface. But software update of power system control equipment must be finished in the operation process. So that a method for software upgrade online applied in the intelligent power distribution terminal has an active effect with low cost and convenient operating.

Nowadays, there are two commonly used software update method for the intelligent power distribution terminal: In-System programming (ISP) and In-Application programming (IAP), which many MCUs' Flash memories support. For the ISP, it needs the condition of external trigger [1]. When upgrading, the equipment must be back to factory or taken down for technician to operate, which is neither convenient but also to the disadvantage of the power grid operation, nor causing the costs [2]. IAP is more agile to upgrade software, through the special design of boot program memory. It can program the Flash by calling the IAP function to upgrade the software, which means it can program while the device is operating.

This paper puts forward an online software upgrade method for intelligent power distribution terminal based on wireless communication, which can be convenient, reliable to update software. Abnormalities won't cause the system crash when software updating.

2. THE DESIGN OF THE INTELLIGENT POWER DISTRIBUTION TERMINAL TO UPGRADE SYSTEM

2.1 IAP Technology
The system of this paper makes use of LPC2138 as its MCU [3] [4].The LPC2138 is based on a 32 bit ARM7TDMIS-S CPU. It has 32kb RAM and 512kb Flash memory, which supports ISP or IAP via on-chip bootloader software. In-Application programming (IAP) is the performing of erasing or writing operations on the on-chip Flash memory as directed by the end-user application code. Single Flash sector or fll chip erase is 400ms and programming of 256 bytes is 1 ms. Besides, it has multiple serial interfaces including two UARTs, two Fast I2C-bus (400 kbits), SPI and SSP with buffering andvariable data length capabilities. IAP code of LPC2138 is in BootBlock, written by manufacture before leave the factory [5]. The Flash boot loader provides the interface for programming the Flash memory. User shall not modify the IAP code, but can call it. IAP program is the Thumb code, locating at address 0x7ffffff0. It is not only to realize the jump but also to realize transition status, when calling IAP function. Because IAP needs to use the top 32 bytes of RAM, users have to change the stack space position [6] [7]. The IAP function could be called in the following way using C. This section is taken from the User Manual.

Define the IAP location entry point.

#define IAP_LCOATION 0x7ffffff1

Define data structure or pointers to pass IAP command table and result table to the IAP function.

unsigned int command[5];

unsigned int result[2];

Define pointer to function type, which takes two parameters and returns void. Note the IAP returns the result with the base address of the table residing in R1.

typedef void (*IAP) (unsigned int [],unsigned int []);

IAP iap_entry;

Setting function pointer

iap_entry= (IAP) Iap_location.

Call IAP routine.

iap_entry (command, result);

As seen above, iap_entry does not have to be defined anywhere in the application, as the linker now knows it is been defined in the image residing in Flash through the symdefs file.

Fig. 1 shows the general operation steps of how to use IAP to program Flash [8]. There are four aspects in the IAP calling process as follows:

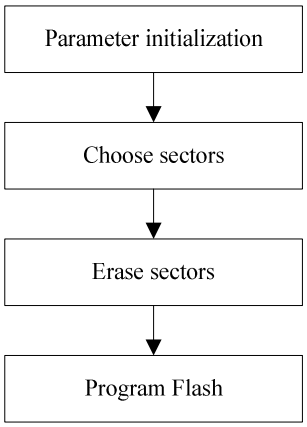


Figure 1.The flow chart of program Flash by IAP

- Parameter initialization: It needs to define some parameters before IAP, such as the system clock, IAP entry address, input parameters and the output parameters.
- Choose sectors: It must first select operation sector before operate flash. It can select multiple sectors at one time. The command code of choosing sector is 50.
- Erase sectors: It must erase operation sector before write flash. It can erase multiple sectors at one time. The command code of erasing sector is 52.
- Program Flash: In this stage, the IAP function will write date from RAM to Flash. The command code of programing Flash is 51.

2.2 Online Software Upgrade Solutions

The online software upgrade uses two serial interfaces, including UART and SPI. UART is connected with wireless module, which uses mobile network. The SPI is connected with EEPROM, which saves data which is mainly the new ad used programs. Fig. 2 shows the design solution of online upgrade for intelligent power distribution terminal. The entire system is composed of four parts, including one hand terminal equipment, two wireless equipment and one intelligent power distribution terminal. The hand terminal equipment is made up of the ARM11, which is running the WinCE. The intelligent power distribution terminal system makes use of LPC2138 as its MCU. First of all, hand terminal equipment, as the source host of new program files, sends update command through the wireless module. Then hand terminal equipment sends the new code to intelligent power distribution terminal after receiving the responding answer. Intelligent power distribution terminal will response the update command through the interrupt mode by verifying its own version number and judge whether to start files downloading or not, then it will receive and check the upgrade data to realize online upgrade. If the upgrade process fail or wait for overtime, intelligent power distribution terminal will give up the upgrade task automatically, and go back to run the original code.

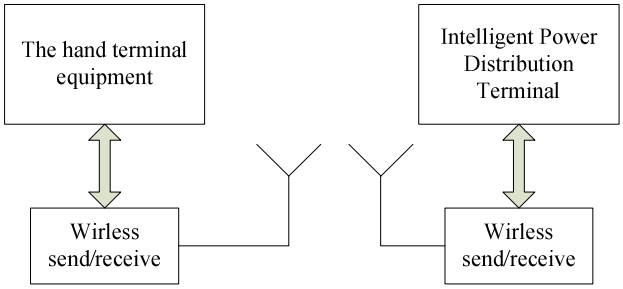


Figure 2.The design scheme of online software upgrade

According to the software structure, online upgrade system includes three kinds of working mode and three program module, as shown in Fig. 3.Guiding code, running on management mode, is in charge of the working mode of the intelligent power distribution terminal and guide the system into the application mode or upgrade mode. It can complete the task of complicated program loading and guiding. Upgrade code running on upgrade mode, is responsible for updating new code. In upgrade mode, all of the interrupt is closed. The user code runs on application mode, which is the application program of intelligent power distribution terminal. It is responsible for the daily work.

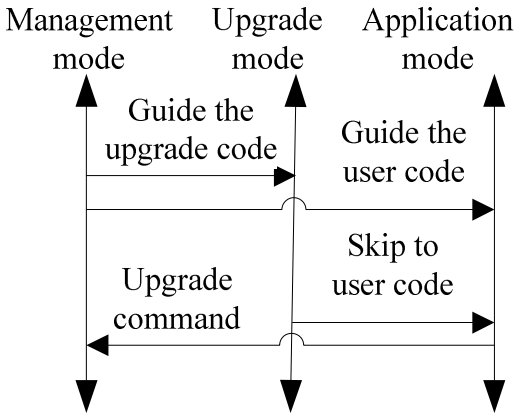


Figure 3.The software structure.

3. THE DESIGN OF ONLINE UPGRADE ALGORITHM

3.1 Communication Protocol

According to the reliability requirements of the design, the paper designs the corresponding communication protocol, as shown in Fig.4. This protocol divides upgrade code into frame as the unit for transmission. Each frame includes the head character, control domain, mark domain, the data length domain, data domain, check domain and end domain. And each domain contains some certain bytes. In the frame format, the starting character means the beginning of a frame of information. Control domain means the position of the frame in the total data and shows whether it is the end frame. Mark domain records the frame data retransmission times. End domain indicates the end of a frame of information.

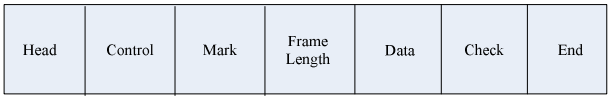


Figure 4.The communication protocol.

3.2 Program storage Space

In order to realize the online software upgrade and fault protection, the system divides the Flash of the MCU into four parts consisting of one Boot storage area, one current user storage area, and two user code storage areas to make the memory's structure logical and avoid the codes covering with each other. Fig. 5 shows the division. As the figure shows, the Boot area, loading the Bootloader of system, is in charge of the software upgrade. The user code is divided into two areas consisting of A and B: one is used to save the operating code; another is used to save the new software code. For safety consideration, the upgrade code is written in the spare area, but not covering the original one, and stores the new code when checksum is right. Then renew the therapy program entry pointer in the information storage area. Even if the upgrade is unsuccessful or breaks off, the last edition code is still stored in another area, to which the programming entry can return to point; the device can still work with the original code. When the current operational code is located at A, the system will upgrade to B. On the other hand, it can upgrade to A when the current operational code is located at A. The area of current user sign stores the currently operational code area and some constants needing backup, which store the beginning address of the operating code.

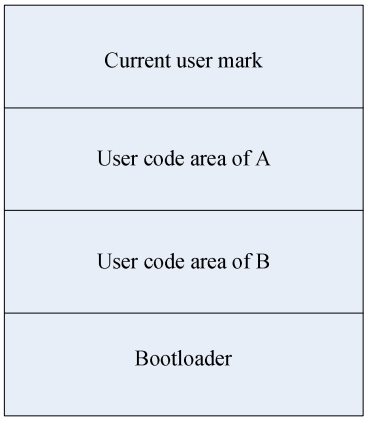


Figure 5.The flow chart of upgrade code.

3.3 Upgrade Algorithm

Because the software update process is automatic, this paper designs the upgrade code specially, written in C language. The starting address of the update is 0x00000000. The device should test some special associated conditions and then starts the upgrade mode. Firstly, the users should start the MCU's reset mode by an update command. The user code will jump to management mode after receiving update command, and then enter the upgrade mode. This kind of design can guarantee priority of software upgrade. And the system keeps reading the serial communication port's buffer to receive the upgrade data and check whether the data is correct. If all data is correct, the system will send response of receiving success and start software upgrade. The MCU erases the specified segments in the Flash, and then writes upgrade data to the Flash. Then the system skips to the new application program and updates the current user mark. The following is the code to ski to the application program:

```
int main(void)
{       …
    UserProgram = (void (*)()) user;
        if (*FlagPoint == A){UProgram = (void (*)()) (A);}
        else
        {
                if (*FlagPoint == B)
                {UProgram = (void (*)()) (B);}
                else{…}
        }
        (*UProgram)();
}
```

The entire system's operating principle of the IAP wireless high-speed configuration and downloading technology for the upgrade systerm is showm in Fig. 6.

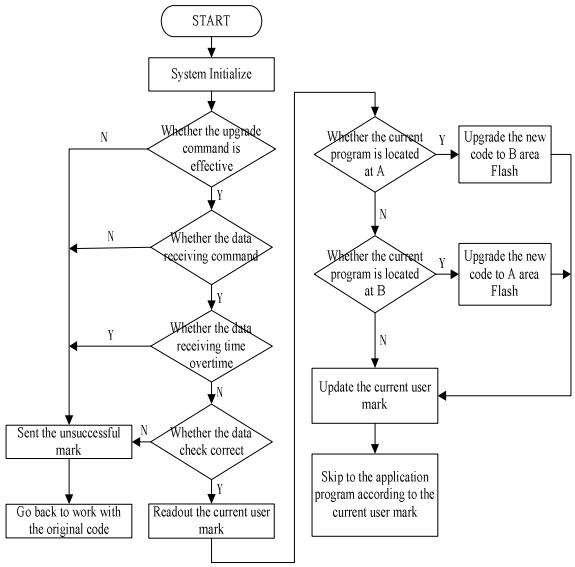


Figure 6.The flow chart of upgrade code.

When we want to use ADS1.2 software to write Bootloader and application program separately, it uses the scatter, which is used to specify the ARM connector how to distribute the storage address of RO, RW and ZI data to loading the code into the division of Flash just as follows:

```
ROM_LOAD 0x0
{
; IAP Bootloader area
   SYSTEM 0x00000000
   {
      Startup.o (vectors, +First)
      main.o(+RO)
      target.o(+RO)
      *(+RO)
   }
}
; Usercode area
ROM_LOAD2 0x8000
{
   USER 0x8000 FIXED
   {
      User.o (+RO, +First)
      SSP.o(+RO)
   }
   IRAM 0x40000000
   {
      Startup.o (MyStacks)
      * (+RW,+ZI)
   }
   HEAP +0 UNINIT
   {
      Startup.o (Heap)
   }
   STACKS 0x40002000 UNINIT
   {
      Startup.o (Stacks)
   }
}
```

The starting address of IROM is 0x00000000, and the first 0x8000 are used to store the upgrade code. And the user code is locating at the starting address of 0x00008000. It costs 0x00040000. The starting address of IRAM is 0x40000000. It is used to store the variables.

3.4 Security Algorithm

In order to avoid the system breaking off during the software upgrade process which will cause the upgrade task failure and the system crash, this paper formulates the corresponding security algorithm, according to the types of accident. It can ensure the intelligent power distribution terminal continuly update the software online during the stably and reliably operating.

- Before upgrading, firstly, the software will check whether there is legal upgrade command. Then the user sends password and entering orders to get the operating right. If the command is to go back, the system will readout the used program from the used application program area of Flash directly, and wite the used program to the application program area and execute.
- For data, it uses the data validation method, not only for each frame, but also for the entire packet, after the end of the entire packet data transmission, to ensure that all data is corrected received. And during the process, the host will wait for the answer from the device after sending each frame to insure the reliability.
- This design starts the timing function of timer. If the upgrade waits for the overtime, the system will return to the original user code.
- If exceptions occur during the upgrade process, the system will often reset and the reset mark will increase by opening watchdog function. Then the system will enter the management mode to give up this upgrade task. The last edition code is still stored in another area, to which the programming entry can return to point, and the system can still go back to work with the original code.

4. EVALUATION

The design applying the method for upgrading the software of intelligent power distribution terminal is tested in detail after the design and the program are completely finished. The test result shows that the design has good feasibility in the application. The code is built with ADS1.2, analyzed and sent by the hand terminal equipment. More than 30 times of test are carried out in the actual and complex environment, and all upgrades are completed successfully in 3 minutes. As a result, the design is feasible and has a high successful rate at the effective distance. But as a matter of fact, the IAP mode programming to the memory (Flash) needs another absolute complex code stored in the Flash to analysis the upgrade file commonly that will get the code more and more complex. There are some factors affecting the communicating capability. For example, the code size decides the time. The small file will reduce the time obviously. The data in each frame also enhances the efficiency of system. You should better balance the efficiency and reliability to design the protocol of the each frame. Increasing the amount of data in each frame will enhance the efficiency but the more data in a frame, the less reliable it would be. The protocol should balance the efficiency and reliability.

5. CONCLUSION

According to the requirements of the intelligent power distribution terminal, such as continuous, stable and reliable operation, this paper proposes an online software upgrade method, which offers several novel and original features. This method uses a different software frame structure with the common way. It can realize the online upgrade function and keeping the high

reliability and simplifying the system of intelligent power distribution terminal. Even if the upgrade is unsuccessful, the system can still return original status.

6. ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (Grant NO. 61172075). I deeply appreciate my advisor, Dr. Zhiyuan Xie, for his Careful guidance and advice.

7. REFERENCES

[1] Naixuan Shi, Wei Feng, Jian Wang, Xiaoyong Ji. 2010. *Design and Implementation of Software On-line Update Mechanism Base on TMS320VC55x DSP*. Communications Technology.2010 7(223), 236-238.

[2] Yang, C, Hongwei, H, Lin, X, Luming, L, Bozhi, M. 2006. A Software Upgrade Method for Micro-electronics Medical Implants. In *Proceedings of the 28th IEEE EMBS Annual International Conference*. (Aug 30-Sept 3, 2006). New York City, USA, NY 5009-5012.

[3] ZLG, Zhang Hua. 2005. *Thinking in ARM-LPC213x/LPC214x*. Beijing University of Aeronautics and Astronautics Press.

[4] Philips Corporation. 2005. *LPC213x User Manual Rev.01*. Koninklijke Philips Electronivs N.V.

[5] Peng Jinghua. 2008. *Romote On-line Upgrade of Embedded System's Software Based on GPRS*. Journal of Computer Applications, 2008 2,519-521.

[6] Philips Corporation.2004. *AN10256*. Koninklijke Philips Electronivs N.V.

[7] Shen Jianhua. 2005. *ARM System Developer's Guide: Designing and Optimizing System Software (In Chinese)*. Beijing University of Aeronautics and Astronautics Press.

[8] ZLG.2007. *AN070701*. GuangZhou Zhiyuan ELECTRONICS CO., LTD.

.

A Mobile Live Multi-Sports Events Recommendation System

Zhenchen Wang
Queen Mary, University of London
Mile End Road, E1 4NS
zhenchen.wang@eecs.qmul.ac.uk

ABSTRACT
This paper describes a work to support the live multi-sports events recommendation on ubicomp devices including smart phones and laptops. Content recommendation systems are in mainstream use enabling consumers to filter video content to match personal preferences. To date the main focus on the use of video recommender systems has been on video retrieval and simple live video channel selection. The use of recommender systems to personalise live multiple concurrent events represents somewhat different challenges. A personalisation approach is needed that handles: recommendations for concurrent live multi-sports events that are constrained by a dynamic schedule; multi-valued dynamic user preferences (that may change as events progress); a characterisation of domain objects and user selections that reflects multiple semantics rather than a universal semantics. Hence, the critical tasks within a recommendation system including user preference modelling, user group modelling, and recommendation generation process modelling, need to be modified to handle the limitations of existing approaches. In this paper, a recommendation model concerning these issues is presented and evaluated. It is found that the proposed model can be flexibly implemented on different ubiquitous devices and generate accurate recommendations.

Categories and Subject Descriptors
H.1.2 User/Machine Systems H.2.1 Logical Design

General Terms
Algorithms, Design

Keywords
Group Recommendation, data processing, data mining, sports programs recommendation, machine learning

1. INTRODUCTION
Sport is one of the most popular genres for viewing broadcast video worldwide. It is estimated that about one in six or one billion people worldwide watch World Cup football, and that about three quarters of the world's population watch some of the Olympics, with typically up to a quarter of this audience viewing these events live. And more and more people are used to viewing them on mobile devices. Watching an event live as opposed to watching a recording event is a different user experience primarily because it is the present time, when viewers may strongly empathise with the competition participants and outcome but the outcome is not yet determined. The vast majority of personalisation approaches for sports video have been designed to enable viewers to more easily retrieve previously recorded content that matches users' interests rather than for personalizing live event viewing. Personalising the viewing of live events appears to be simple as it is akin to personalising the selection of general TV entertainment channels, displaying only those channels that match users' preferences. For sports, a sub-genre classification system, represented as metadata, is often used to characterise specific sports events and to characterise viewer preferences. Typically, for general entertainment TV viewing, an on-line guide can be accessed that defines the event schedule as a list of sports discipline event instance descriptions, e.g., long Jump Preliminary Heat 'A', etc. The channels on offer can be matched to viewer preferences, by viewers manually or automatically using algorithms such as types of recommendation system. Personalising the viewing of live multi-event sports events such as the Olympics introduces new challenges compared to personalizing sports video retrieval and general entertainment TV broadcasts.

Human viewers have one focus of attention to view multiple events even if a multi-view screen is used. For live multi-sports events, an individual viewers schedule is semi-deterministic and is dynamic. A viewer can decide to subjectively switch between events because of several reasons such as the current view is not captivating, the preferred athletes are not successful as expected, because of specific event incidents or because of the score status, etc. Note also that live multi-view sports channels are dynamic reflecting the event schedule. There may be no fixed channels as in TV entertainment systems so one can't always identify a specific sports view by number. Hence, video recommender systems tend to recommend channels by programme content genre rather than by channel number. In some research cases, the content and preferences can be represented semantically. Semantic descriptions may alleviate the issue that users must know a priori the content provider descriptions, rather they can use their own semantics for their preferences and match these to the semantics of the content provided. Note also, that for video retrieval, quite detailed explicit user feedback can be gathered from users during the interaction. However, explicit feedback from users during viewing of live events needs to be minimised else the usability of the system is reduced as it detracts viewers from the immersion of following an event.

Another issue remains that affects the personalisation for live multi-sports events, scalability. Generally, sports events tend to be viewed in high definition with the same views being broadcast to all viewers in a region. Multiple cameras may be used for each

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.19

event instance, of the order of five hundred cameras may be used in an Olympic event, however a human director interleaves and orchestrates these multi-camera views into a single stream for that event. It is too costly to transmit live sports video content to match very many different individual user preferences. This introduces the need for more personalised view channels [18]. The main approach here is to personalise content delivery with respect to user groups rather than to individual users. To summarise, a personalisation approach is needed that handles: Multi-valued dynamic user preferences (that may change as events progress) and dynamic user group classification.

The remainder of this paper is organized as follows: existing approaches regarding user group and preference modelling and recommendation generation are discussed in section 2. The proposed recommendation model is presented in section 3. Section 4 describes the model in the context of a live sports event broadcast service. Section 5 presents the evaluation of the developed group recommender system and analyses the evaluation results. Section 6 presents the conclusions of this work.

2. RELATED WORK

As stated, previously, current personalisation approaches to the video domain focus more usually on video retrieval, such as in [18] and [5]. Personalisation approaches that are used to select live event schedules are primarily static and make use of coarse-grained content descriptions such as a single entertainment genre is used for a static channel, e.g., in [15], [3], [11], [20]. Some approaches support the use of more finely-grained content metadata, e.g., using a semantic representation, to match to user preferences [16], [2]. Both of these approaches are not suited to the live approach because they tend not to support dynamic, multi-valued user preferences that are time-constrained. In order to recommend items to a group of users sharing similar preferences, target user groups must be clearly defined and the recommended items have to match the shared preferences. In this section, the existing research works is considered in more detail with respect to three main sub-processes to support recommender systems.

In order to recommend items to a group of users, target user groups must be clearly defined. In most studies, groups are often seen as a number of users sharing the similar preferences. Existing user grouping approaches largely rely on individual preferences which are usually collected via either an explicit or an implicit approach or both. Explicit preferences can be retrieved via explicit user interface as user preference input [9]. Implicit preferences are often obtained via the monitoring of user interaction with the system such as in [4]. As these approaches link user preferences directly to the user group definition.

User preferences are normally formalized with respect to some predefined metrics in order to enable a recommender system to deal with users through considering them to share similar preferences. A rating system such as in [19], [10], [1], is often used to present users' preferences. Nevertheless, it is normally difficult to calibrate these subjective ratings among a group of users. In [14], the user personality is the determinant of the preferences; the weakness of this approach is that it may partially cluster users according to some personality stereotypes that are quite subjective. In [7], the recommended item tags and metadata are used to present user preferences. One constraint of this approach is that it requires a large amount of user input tags that can be associated with recommended items. Because these user preference modelling approaches are very much user feedback centred and lack of analysis on multi-character of the domain objects.

According to [8] and [21], an aggregation function is often used to generate the group recommendations. In [8], the preference aggregation approaches are specifically summarized as: merging of sets of recommendations; aggregation of individuals' ratings for particular items; and construction of group preference models. The first approach could reduce the incentive for group recommendation as it needs to obtain individuals' recommendations beforehand. The second approach could easily generate partial recommendations due to a user's subjective rating standard. The last approach as in [1] and [21], aggregates the preferences of individual group members to form a model of the preferences of the group as a whole. This approach can effectively avoid repetitive computations for users for the same recommended items. However, user preferences can still be ill-defined and the user preferences and recommended items definition are heavily relying on users' feedback.

3. GROUP RECOMMENDATION MODEL

In this section, three modelling tasks of user group modelling, user preference modelling and recommendation generation modelling are proposed which enables the proposed model to be applicable within the context of live multi-sports events.

3.1 Recommendation Model Overview

The benefits of this model structure (see Figure 1) are threefold. First, it frees the user preference modelling from user feedback. Second, it separates user group modelling from user preference modelling and thus makes these two tasks less dependent. Third, Items definition is not dependent upon user preference definition. Instead, this defines user preferences that enables them to be characterise in terms of domain objects. The main advantage of the proposed structure is that it relieves the constraints imposed by user feedback on user preference modelling and item definition which supports a more flexible approach to the design and implementation of recommendation generation.

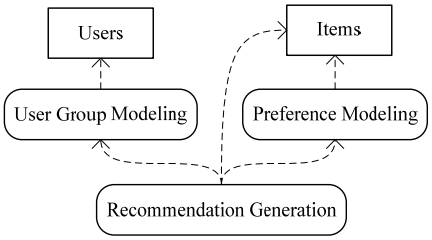


Figure 1 Proposed tasks orchestration structure

3.2 User Group Modelling

The user group modelling, here, is independent of the user preference modelling in the sense that the user information is used to cluster users rather than their preferences.

Two types of user information are used to describe a group of users, termed session invariant user information and session variant user information. Session invariant user information can describe users in a general way and is relatively static across multiple user sessions, e.g., demographic information. Session variant user information describes session specific information, e.g. a user session duration.

The advantage of using session invariant user information to group users is twofold. First, once users are grouped, groups can be reused across user sessions and across different access systems.

Second, no system inference process is required as users can be explicitly asked to provide this information. As a result, the user group can be expressed as

$$Group_i = Agg\{\{ui_{1j}, \dots, ui_{1l}\} \in ui_1, \dots, \{ui_{ij}, \dots, ui_{im}\} \in ui_i\} \quad (1)$$

ui_i denotes the ith user information and ui_{ij} denotes the jth information value of ui_i.

3.3 User Preference Modelling

User preferences can be described as a function of the likeness degree of an item. The preference only exists when there are counterpart items that can be chosen. i.e.$Pref(x) = f_{likeness}(x, L_x)$, where x denotes the item and L_x denotes the item list where x is chosen from. The likeness degree can vary between items based upon the value of a preference function.

A list of items can be described as an aggregation of a set of attributes with predefined discrete values, e.g. a sports event with competition type=team, individual, stadium size=large, small. i.e.

$$L_x = Agg\{\{a_{1i}, .., a_{1k}\} \in a_1, \dots \{a_{ni}, .., a_{nk}\} \in a_n\} \quad (2)$$

a_n denotes the *nth* attribute and a_{ni} denotes the *ith* value of the attribute. Therefore, the user preference of an item attribute value can be a function of a likeness degree of a predefined discrete attribute value, i.e.

$$Pref(a_{ni}) = f_{likeness}(a_{ni} \in a_n) \quad (3)$$

The user preference of an item attribute thus can be described as a sort function:

$$Pref(a_n) = Sort(Pref(a_{n1}), .., Pref(a_{ni})) \quad (4)$$

With the same token, a list of items can also be further defined as a sort function of the item attributes in the form of:

$$Pref(L_x) = Sort(Pref(a_i), .., Pref(a_n)) \quad (5)$$

Based upon the sorted preference of each attribute value the preferred items can be arranged in order from the item list L_x and enables a top-k group recommendation. e.g., if the $Pref(L_x) =$ team>small, then teamwork events will be initially selected and ranked before individual events and eventually the teamwork events in small stadium will be ranked before the teamwork events in large stadium.

3.4 Recommendation Generation Modelling

The recommendation generation task is critical as it associates a user group and a preference model. It also is the key task to produce the recommended items to new users.

The association between groups and items could be either strong or weak. Therefore, according to equations (1) and (2), the group-item association degree function is defined as a sort function of common user information to discrete item attribute value association degrees.

$$f_{association}(Group_i, L_x) = Sort(f_{association}(Group_i \in \{ui_1, \dots, ui_i\}, \{a_1, \dots, a_n\})) \quad (6)$$

As a strong association between the user information and a particular item attribute value will indicate the preference of that attribute, hence

$$Pref(Group_i, a_{ni}) = f_{association}(Group_i, a_{ni}) \quad (7)$$

Due to the fact that the recommended item is comprised of a set of attributes, therefore the preference of an ideal item can be expressed as the ranked preference of its attributes as

$$Pref(Group_i, Item_{ideal}) = Sort(Pref(Group_i, a_{n1}), .., Pref(Group_i, a_{ki})) \quad (8)$$

In order to recommend the top-k items to a user usr_i belonging to a group, equation (6) can be re-defined as a sort function

$$Pref(usr_i \in Group_i, L_x) = Pref(usr_i, L_x | Item_{ideal}) \quad (9)$$

$Pref(usr_i, L_x | Item_{ideal})$ denotes the user group preference of a list of recommended items given an ideal item.

4. REFICATION OF THE MODEL

In this section, a group recommender system is reified. The target users are live sports events online viewers and the recommended items are Olympic sports.

4.1 User Group

Two categories of common user information are used. The first category is the user demographic information including age, gender and race. The second category is the sports behaviour related information including the Watch Sports Events Frequency. Table 1 lists the information used and the corresponding predefined discrete values from which a number of groups of users.

Table 1 User information with predefined values

Common User Information	Predefined Discrete values
Age	Young (≤35), Middle Age(>35 & ≤60), Old (>60)
Gender	Male, Female
Race	American, African, Asian, European
Watch Sports Frequency	Daily, weekly, monthly, seldom

4.2 User Preferences

Sports are defined with six attributes based upon the method used in [17]. The likeness of a particular attribute values a_{nk} can be expressed in a binary form, which can be further formalised in equation (10), where the TRUE indicates a large quantity condition, e.g. large size competition area, large number of in-game sessions etc. whereas FALSE indicates a small quantity condition. Therefore, if let a_{nk} denote the 'large number of players', then a Boolean value of TRUE can be assigned to the a_{nk} of current viewing event when it satisfies the TRUE condition in equation (10).

$$f_{likeness}(a_{nk} \in a_n) = a_{nk} > \frac{1}{m}\sum_{j=1}^{m} a_{nkj} : True? False, \quad (10)$$

$$where\ m\ denotes\ m_{th}\ live\ sports\ events$$

4.3 Recommendation Generation

An association function can be defined in terms of existing machine learning techniques. In this work, three well known techniques are employed alternatively, the Decision Tree (DT), Bayesian network (BN) and Bayesian Point Machine (BPM).

4.3.1 Decision Tree

DT is composed of three elements including a decision node, an edge and a leaf. The decision node in this work can be the common user information attribute value whereas the leaf can be the likeness of a sports attribute value in a form of binary values, e.g. '1'indicates like '0'indicates not like.

In order to choose the best attribute as the root of decision tree or sub decision tree, information gain based upon the Shannon entropy is often used to discriminate each decision node. The information gain between an invariant user information attribute value and the likeness of a sports attribute value thus can be expressed as

$$Info\ Gain(u_i, a_{nk}) = Info(u_i) - Info_{a_{nk}}(u_i) \quad (11)$$

$$Info(u_i) = -\sum_{i=1}^{m} \frac{fq(Group_n, u_i)}{|u_i|} log_2 \frac{fq(Group_n, u_i)}{|u_i|} \quad (12)$$

$$Info_{a_{nk}}(u_i) = \sum_{a_{nki} \in [0,1]} \frac{\left|u_i^{a_{nk}}_{a_{nki}}\right|}{|u_i|} Info\left(u_i^{a_{nk}}_{a_{nki}}\right) \quad (13)$$

$Group_n$ denotes user group set, $fq(Group_n, u_i)$ denotes the number of u_i type users in the user group class. $u_i^{a_{nk}}_{a_{nki}}$ denotes the user information attribute value for which a value of either of 0 (FALSE) or 1(TRUE) for the sports attribute a_{nk}as expressed in a likeness function. When the information gain ratio is required as in the DT algorithm C4.5 [13], so called split information can be used.

$$Split\ Info(u_{i,} a_{nk}) = -\sum_{i=1, a_{nki} \in [0,1]}^{m} \frac{\left|u_i^{a_{nk}}_{a_{nki}}\right|}{|u_i|} log_2 \frac{\left|u_i^{a_{nk}}_{a_{nki}}\right|}{|u_i|} \quad (14)$$

$$Gain\ Ratio(u_{i,} a_{nk}) = \frac{Info\ Gain(u_i, a_{nk})}{Split\ Info(u_{i,} a_{nk})} \quad (15)$$

The tree built allows each of the user information attribute values to be associated with the likeness of sports attribute values. In order to recommend the top-k items to a new user usr_i belonging to a group, three additional processes are required namely classification, preference ranking, and recommendation.

A classification process enables the system to find out the decisions *D* on likeness (i.e. preference) of each sports attribute value corresponding to a set of user information attribute values. A most-fit strategy can be used for the case that not all user information attributes values can be classified using the tree built. A preference $Pref(usr_{i,} a_{nk})$ can be obtained by examining reduced user information given a user information reduction function f_r, e.g. ignore one user information attribute value. The reduction function is iterated until a decision is reached.

$$Pref(usr_{i,} a_{nk}) = f_r(D) \quad (16)$$

A preference ranking process allows the system to assign a weight ω to the likeness of each sports attribute value. The ω will enable a ranking process for each attribute value for an attribute. The attribute value with largest weight will be chosen as the decisive attribute value. A further ranking process will rank the attributes according to their decisive attributes' weights. ω will be the decision accuracy and the sort function in equation (8) can be based upon the following equation to support descending sorting.

$$\omega(usr_{i,} a_{nk}) = \frac{Pref_{True}(usr_{i,} a_{nk})}{Pref_{Total}(usr_{i,} a_{nk})} \quad (17)$$

$Decision_{Total}$ denotes the total number of existing data with the required user information attribute values in the classification process. $Decision_{True}$ denotes the number of correct decisions in the training data.

Finally, the recommendation process generates the top-k list of recommended sports based upon result of equation (8) that gives the user group's preference for an ideal recommended sport. The preference of a list of recommended sports can be obtained through iterative comparisons between recommended sports based upon the ideal sport. This can be expressed as:

$$Pref(usr_i, L_x | Sp_{ideal}) = Iterative\ Sort\left(a_{nk}^i \in Sp_i, \dots, a_{nk}^t \in Sp_t\right) \quad (18)$$

a_{nk}^i denotes the *i*th sport's decisive attribute valuea_{nk}, $Sp_{ideal} = \{a_{nk}, \dots, a_{ot}\}$ denotes the ideal recommendation with a particular order of attribute values with the first attribute value has the highest preference priority. The iterative sorting direction, i.e. ascending or descending, depends on the indicated condition for each sports attribute value, e.g. if a_{nk} represents 'more players', descending sorting will be used.

4.3.2 Bayesian network

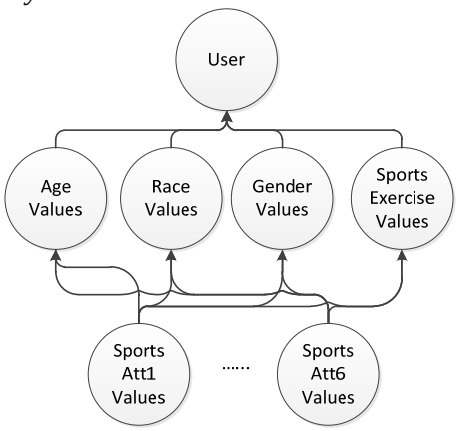


Figure 2 Bayesian Network

A Bayesian network is a probabilistic graphical model which can be used to handle uncertain information. The graphical model usually consists of two components. The graphical component which is a directed acyclic graph (DAG) links the nodes with edges. The numerical component represents the conditional probability distribution of nodes in terms of parent nodes. Naïve Bayes is a simple form of Bayesian network which has one root node (the unobserved node) and assumes child nodes (the observed nodes) are independent to one another.

In this work, the graph model can be structured as shown in Figure 2. The conditional probability between a parent node and a child node can be calculated in terms of the given training data set. This can be expressed as:

$$P(N_{Child} \mid T) = \frac{P(T \mid N_{Child})\, P(N_{Child})}{P(T)} \quad (19)$$

N_{Child} denotes a predefined discrete value of either a user information attribute or a sports attribute. T denotes the total evidence for a child node in the training set.

Once the Bayesian network is quantified, it will be able to recommend the top-k sports to the new user. The recommendation process is pretty much the same as for the DT approach. The only difference lies in the first step, in which a combined user information probability for each sports attribute value is obtained and the probability values are used to represent the $Pref(usr_{i,} a_{nk})$ in terms of a preference probability for each sports attribute value.

$$Pref(usr_{i,}a_{nk}) := P(a_{nk}|\, usr_i \in Group_i) = \frac{P(u_1 \mid a_{nk}) \ldots P(u_k \mid a_{nk})}{P(usr_i)} \quad (20)$$

4.3.3 *Bayesian Point Machine*

Bayes point machine [6] is a learning algorithm for kernel classifiers which approximates the Bayes-optimal decision by the centre of mass of a version space.

Given the hypothesis space H and the user groups training set$G$, the version space can be defined as

$$V(G) = \{h \in H \mid h(Group_i) = a_{nk}\} \quad (21)$$

In order to classify a group of users to each sports attribute value, i.e. either 0 or 1, the Bayes classification strategy is used to obtain a loss incurred by each hypothesis h applied to $Group_i$ and to weight it according to its posterior probability $P_{H|G}(h)$. The tested sports attribute value with the minimum expected loss will be chosen as the user group preferred sports attribute value. The Bayesian point algorithm thus can be defined as:

$$A_{bp}(G) = MinP_G\,[P_{H|G}\left(los\big(h(G), H(G)\big)\right)] \quad (22)$$

$A_{bp}(G)$ denotes the Bayes point which is the classifier$h_{bp} := A_{bp}(G) \in H$.

Once the group preferences are classified in terms of sports attributes values, the top-k sports recommendation can be performed following the same steps as those used in the DT approach except that in BMP approach the $Pref(usr_{i,}a_{nk})$ will be linked to the posterior probability $P_{H|G}(h)$, i.e.

$$Pref(usr_{i,}a_{nk}) = P_{H|G}(h) \quad (23)$$

5. RECOMMENDATION MODEL EVALUATION

In this section, the group recommender system is evaluated with a system which has both a back end system and a front end mobile Web TV system which is developed on Windows Phone platform as shown in Figure 3.

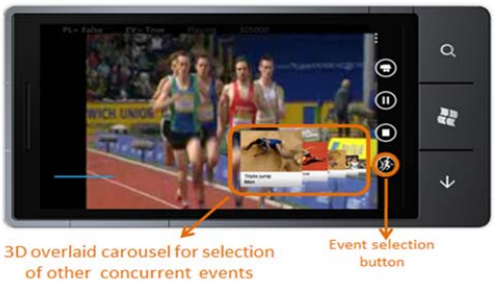


Figure 3 Live sports viewing system Windows Phone front end

5.1 System Setup

Figure 4 shows the system setup, a data acquisition system is attached to retrieve both the inputs and outputs of the recommender system. The inputs include the live feed of the sports, i.e. recommended sports, the current viewing sports of users and the user information for each user. The outputs are the top-k recommendation sports for each user group. For evaluation, the test system front end has a lab trial mechanism which is used to automatically generate a virtual user profile and the system is also able to automatically select the events in every ten seconds according to the generated user profile. The purpose of using this approach here is to simulate the real use scenario assuming user information and viewing preference can be range in diversity.

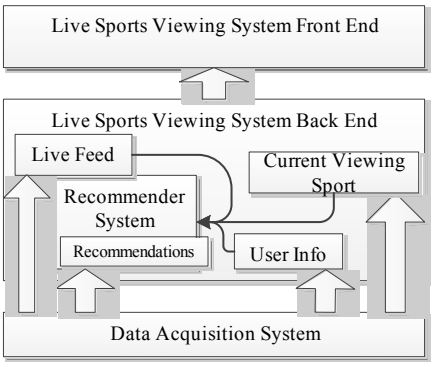


Figure 4 Evaluation system setup

5.2 Evaluation Scenario and Metric

The evaluation scenario can be summarized as: To recommend the top-9 sports to 10 users who belong to different user groups based upon the existing 90 users. Table 2 lists the elements of evaluation scenario.

Table 2 Evaluation scenario elements

Scenario Elements	Values
Total Generated User Number	100
Training Sample	90
Recommended Sports	Football, Basketball, 100m, 400m swimming, 5000m, Beach Volleyball, Long jump, Javelin, High jump

A hundred virtual users are generated from test system front end instances. The current viewing sports of the testing users represent the user selected sports, i.e. the ground truth, which can be compared against the recommended sports. Recommendation accuracy is used as the evaluation metric of the model.

5.3 Recommendation Accuracy

In this work, three well known techniques are employed alternatively, i.e. the Decision Tree (DT C4.5), Bayesian network (BN) and Bayesian Point Machine (BPM) are used to generate the top-k recommendation respectively where 'k' is the number of available sports from a live feed. A random recommendation is generated to represent a system without a recommendation function. The efficiency of the system is evaluated in terms of the mean recommendation accuracy which can be expressed as:

$$\varepsilon = \frac{1}{n}\sum_{j=1}^{n} 1 - \frac{Position_i}{k} \quad (24)$$

Where n denotes the number of recommended users, $Position_i$ denotes the position of the user selected sports i in the recommended sports list. A 10-fold cross validation approach is used to test the recommendation accuracy with 90 users in the training set and 10 users in the test set.

6. EVALUATION RESULTS

The results indicate that a group recommendation system is functioning with decent recommendation accuracy. Among the three implemented recommendation generation methods, the decision tree has the highest recommendation accuracy (mean =87%, median = 86%). The BPM has a slightly higher

recommendation accuracy (mean =73%, median = 71%) than BN ((mean =68%, median = 68%)).

7. CONCLUSION

In this work, a group recommendation model has been proposed for live multi-sports events viewing on mobile devices such as smart phones and laptops. The model is able to address the challenges including Multi-valued dynamic user preferences (that may change as events progress) and dynamic user group classification. The model is implemented within the context of live sports broadcast services with the purpose to recommend online users the top-k sports. The performance of the developed recommender system for the proposed evaluation scenario shows that it outperforms a system without recommendation system. Based upon the machine leaning techniques used, a mean recommendation accuracy ranging from 68% to 87 % can be achieved. In the future, we plan to investigate how to extend the recommendation model to recommend users views when event views using multiple cameras are made available within each individual sports event.

8. REFERENCES

[1] Boratto L., Carta S., Chessa A., Agelli M., and Clemente L., 2009. Group Recommendation with Automatic Identification of Users Communities. In *Proceedings of the 2009 IEEE/WIC/ACM international Joint Conference on Web intelligence and intelligent Agent Technology*,Washington, DC, pp.547-550.

[2] Cavallaro A., Steiger O., Ebrahimi T., 2005. Semantic Video Analysis for Adaptive Content Delivery and Automatic Description. *IEEE Transactions On Circuits And Systems For Video Technology* , Vol. 15, No. 10, pp. 1200-1209.

[3] Chao L., Changsheng X., Hanqing L. 2010. Personalized Sports Video Customization Using Content and Context Analysis. *Int. J. Digital Multimedia Broadcasting*.

[4] De Á. and Zorzo S., 2009. A personalized TV guide system: an approach to interactive digital television. In *Proceedings of the 2009 IEEE international Conference on Systems, Man and Cybernetics* pp.11-14.

[5] Gibbon D., and Liu Z., 2009. Large scale content analysis engine. In *Proceedings of the First ACM Workshop on Large-Scale Multimedia Retrieval and Mining* . LS-MMRM '09. ACM, New York, NY, pp.97-104.

[6] Herbrich R., Graepel T. and Campbell C., 2001. Bayes point machines. *J. Mach. Learn. Res.* 1, pp.245-279.

[7] Hölbling G., Thalhammer A, and Kosch, H. 2010. Content-based tag generation to enable a tag-based collaborative tv-recommendation system. In *Proceedings of the 8th international interactive Conference on interactive Tv&Video*. EuroITV '10. ACM, New York, NY, pp.273-282.

[8] Jameson A. and Smyth B., 2007. Recommendation to groups. *The Adaptive Web*, LNCS, pp. 4321:596–627.

[9] Jameson A., S. Baldes, T. Kleinbauer, 2004. Two methods for enhancing mutual awareness in a group recommender system. In: Proceedings of the International Working Conference on *Advanced Visual Interfaces*, pp.447–449,

[10] Masthoff J., and Gatt A., 2006. In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. *User Modelling and User-Adapted Interaction*, 16(3-4): pp.281–319.

[11] Ming L., Yantao Z., Shi-Yong N., et al., 2008. Personalized event-based news video retrieval with dynamic user-log. IEEE Int. Conf. on *Multimedia and Expo*, pp.1157 – 1160.

[12] Poslad S., Pnevmatikakis A., Nunes M. et al. 2009. Directing Your Own Live and Interactive Sports Channel. 10th Int. Workshop on *Image Analysis for Multimedia Interactive Services*, WIAIMS'09, pp. 275 – 279.

[13] Quinlan J., C4.5, Programs for machine learning. Morgan Kaufmann San Mateo Ca, 1993.

[14] Recio-Garcia J., Jimenez-Diaz G., Sanchez-Ruiz A., and Diaz-Agudo B., 2009. Personality aware recommendations to groups. In *Proceedings of the Third ACM Conference on Recommender Systems* RecSys '09. ACM, New York, NY, pp.325-328.

[15] Shumeet B., Rohan S., Sivakumar D., et al., 2008. Video suggestion and discovery for youtube: taking random walks through the view graph. Proc. *17th international conference on World Wide Web*, pp. 895-904.

[16] Tsinaraki C., Polydoros, P., Christodoulakis, S. , 2004. Interoperability Support for Ontology-Based Video Retrieval Applications. *Lecture Notes in Computer Science.* Vol. 3115, pp. 582-591.

[17] Wang Z., Poslad, S. Patrikakis C., Pearmain A, 2009. "Personalised Live Sports Event Viewing on Mobile Devices," *Mobile Ubiquitous Computing, Systems, Services and Technologies, 2009. UBICOMM '09. Third International Conference on* , pp.59-64.

[18] Wei Y., Bhandarkar S., and Li K., 2006. Client-centered multimedia content adaptation. *ACM Trans. Multimedia Comput. Commun. Appl.* pp.1-26.

[19] Xu J., Zhang L., Lu H. and Li Y., 2002. "The Development and Prospect of Personalized TV Program*", Proceedings of the IEEE 4th Internacional Symposium on Multimedia Software Engineering* (MSE'02), California, USA.

[20] Yifan Z., Xiaoyu Z., Changsheng X., 2007. Personalized retrieval of sports video. Proc. Int. workshop on Workshop on *multimedia information retrieval*, pp. 313-322.

[21] Yu Z., Zhou X., Hao Y., Gu J., 2006. TV program recommendation for multiple viewers based on user profile merging. *User Modelling and User-Adapted Interaction* 16(1) pp.63–82.

The Study to the Evaluation Model for the Performance Audit of Government Informatization Projects

Jinyu Tian
School of Business and Administration
North China Electric Power University, Baoding, China
+8615933966466,071003
306669351@qq.com

Yuting Liang
School of Business and Administration
North China Electric Power University, Baoding, China
+8615933966466,071003
306669351@qq.com

ABSTRACT
Modern information and communication technology are applied to the government informatization projects, which integrates management and service through network technology, and contributes to the changes from the organs of government departments by the control type to the service authority. In order to increase the benefit of informatization investment and also ensure its continuous development establishing an informatization projects performance audit evaluation system is crucial. In accordance with the evaluation model for the performance audit of government informatization projects, fuzzy hierarchy model, quantitative and qualitative analysis method are used to build an effective evaluation system to evaluate, standardize and guide the construction of government informatization projects.

Categories and Subject Descriptors
K.6.4 [**Management of Computing and Information Systems**]: System Management –Management audit

General Terms
Management, Design, Theory.

Keywords
Informatization Projects; Performance Audit; Evaluation Model

1. INTRODUCTION
With the increasing development of the era of knowledge economy, as well as the management of the national economy of information technology, information technology has become one of the most potential productivity in contemporary society, and information resource provides strategic resource and core competence for national economic and social development, which promote the regional development and modernization in the city. After 20 years of development government informatization projects in China, while the transition from exploring to full application has basically come true, accumulates much informatization assets to become the focus in the management of government assets. How to effectively strengthen the management and supervision of the government informatization projects and to improve the performance of the use of funds has become a top priority [1].

2. OVERIEW ABOUT THE GOVERNMENT INFORMATIZATION PROJECTS PERFORMANCE AUDIT

2.1 Government Performance Audit
The performance audit is defined as an independent supervisory activity that in accordance with relevant regulations and standards, an independent auditing agency or personnel use certain audit procedures and methods to supervise, evaluate and verify economy, efficiency and effectiveness of units or specific projects audited, and make recommendations for improvement and promotion of enterprise management, in order to improve the effectiveness[2].

Refer to the common practice in the world, in terms of subject and object from the performance audit, the subjects of performance audit are government audit authorities. The object of the performance audit is mainly the public sector, public works projects and public utilities.

Government performance audit is defined that based on the evaluation criteria, the national audit authorities use a comprehensive evaluation index system and methods to examine and evaluate economy, efficiency and effectiveness of economic activities of the government with its subordinate departments, and project units, as well as the construction of public works projects, and to correct the error, and then make reasonable suggestions for the improvement of their work.

2.2 Performance Audit of Informatization Projects
Performance audit of informatization projects is a new direction for performance audit. With the continuous coverage of the information technology and information technology continues to improve, the role of performance audit of informatization projects has become increasingly important.From the narrow and broad understanding of the concept of informatization projects performance audit: First, the narrow sense is effect of acceptance after the completion of the construction of informatization projects, the operation of the process of using the project's cost-effectiveness of the audit and evaluation; generalized sense is performance audit evaluation of a series of processes throughout the entire life cycle of informatization projects, its main purpose is to improve the overall operating efficiency of the system, to meet organizational development requirement much better.

Of course, the narrow concept of informatization projects performance audit in the audit practice is relatively easy to perform, but the generalized one is on behalf of the future trend.

Performance audit on informatization projects must evaluate the performance of informatization projects and then audit according

Research Notes in Information Science (RNIS)
Volume13, 2013
doi:10.4156/rnis.vol13.20

to the results of the evaluation of informatization projects by comparing with the construction phase of plan, goal, investment of informatization projects audit.

2.3 The Objectives of Informatization Projects Performance Audit

The objective of informatization projects performance audit is to review the informatization projects of audited entity, as shown in Table 1.

Table 1 Table of informatization projects performance audit objectives listed

Objectives	Content
Economy	To obtain a certain number and quality of output at the lowest resources cost
Efficiency	To strive to achieve the maximum output with resources invested on informatization projects
Effectiveness	The relationship between the expected results and actual one. The task of the audit is to check whether the target is expected to achieve

3. THE ISSUES ABOUT GOVERNMENT INFORMATIZATION PROJECTS PERFORMANCE AUDIT

The government carries out informatization projects performance audit practice facing some difficulties, as shown in Table 2 below.

Table 2 Table analyzing problems of government informatization projects performance audit

Problems	Content Analysis
Lack of reasonable informatization projects performance audit system	a. Investment returns of the department is unable to effectively assess, while informatization projects continues to deepen and advance. b. Thus phenomenon that is contrary to the essential characteristics
Lack of quantitative reasonable performance audit	a. Audit focuses on the process of informatization projects construction, that project and acceptance. b. Functioning performance, the sustainable development capacity and benefits given to the users are lack of attention.
Assessment of operational mechanisms are inadequate	a. Subjects of assessment are often superior administrative organ, most of the assessment forms are “campaign-style” “competition style", "surprise". b. While it lack persistent determination on informatization projects performance; as to management process, it is closed, and lack transparency and openness.

Therefore, to build a scientific informatization projects performance audit system has become a key link in the government informatization construction.

4. THE GOVERNMENT INFORMATIZATION PROJECTS PERFORMANCE AUDIT EVALUATION MODEL

4.1 Model Design Thoughts

The government informatization projects performance audit is a process that through the study of informatization projects performance audit evaluation indicators, models, methods we explore to establish a scientific evaluation index system and management mechanisms and gradually focusing on the strategic orientation of the benefits of construction projects, which improves the overall level of construction applications and management of informatization projects.

The main evaluation content is shown in Figure 1.

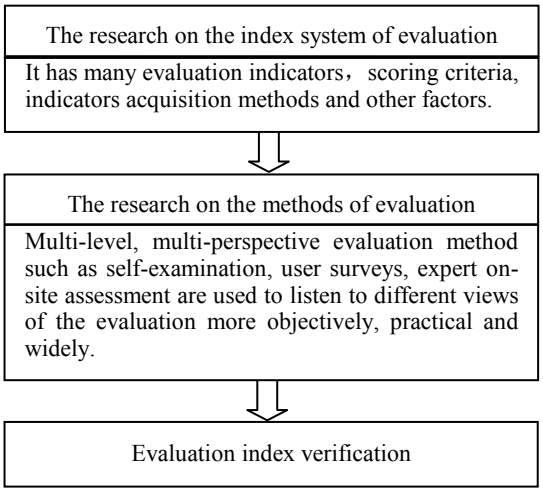


Figure 1. Figure of performance audit evaluation

4.2 Design of Comprehensive Evaluation Index System

The informatization projects is an important part of the national "Gold Faith" project, informatization projects performance cannot be measured directly with the the increase of economic benefits or reduction of administrative costs, it is indirectly reflected in the ability and level of services to enhance government departments, to promote government departments the performance of their duties in place and other aspects.

Indicators of informatization projects performance audit should include aspects of system construction, operation and maintenance quantitative indicators as well as qualitative indicators reflecting the project outputs. Therefore, in order to audit the effectiveness of the government informatization projects objectively and effectively, we propose a comprehensive evaluation index system, establish a comprehensive evaluation model based on the

principle of fuzzy hierarchical comprehensive evaluation used for performance audit evaluation on the government informatization projects. We regard informatization projects performance audit as the overall goal of evaluation, and then further refine the main indicators.

We set comprehensive evaluation index system in accordance with the grading. Compared with the goal of the aforementioned government informatization projects performance audit, we set the seven criteria layer, and then refine into 20 indicators, the index system is shown in Table 3.

Table 3 Evaluation index system

Target layer	Criteria layer	Index layer
Informatization projects performance audit E	Completion of Construction B1	a. Pass rate of project acceptance b. Installation rate of safely equipment c. Business coverage,
	Level of system operation and maintenance B2	a. System trouble free rate b. average daily number of fault maintenance c. Fault response rate d. Troubleshooting rate e. Patency rate of core network f. Success rate of data backup g. Users' evaluation on system stability and improvement of business efficiency
	Achievement degree of government functions B3	a. Improvement of work quality and administrative law enforcement capability b. Effect on business innovation
	Improvement of social service capability B4	a. Quality of public services b. Response to emergency c. Resources sharing degree
	Improvement of administrative management B5	a. Improvement of administrative efficiency and quality b. Improvement of inner control capability
	Financial indicators B6	a. Rate of actual in place b. Utilization of funds and budget completion
	Organization and implementation of project B7	a. Legality and compliance of financial management b. Quality of accounting information

4.3 The Analysis of Fuzzy Hierarchy Model

4.3.1 The procedures of analytic hierarchy process

Step 1 is modeling the hierarchical structure. This is the most important step in analytic hierarchy process. At the top is the target level. The level in the middle is called criterion level. The measures layer is at the bottom.

Step 2 is building up the judgment matrix. After modeling the hierarchical structure, the next is to compare the elements of each item from each level. In this problem, use 1~9 scale to quantify the different conditions. We put the meaning of each scale in the table 4:

Table4 Meaning of each scale in judgment matrix

scale	meaning
1	An element is equal importance to the other
3	An element is slightly important than the other
5	An element is obviously important than the other
7	An element is great important than the other
9	An element is extremely important than the other
2,4,6,8	2,4,6,8 indicates the mid-value of 1~3,3~5,5~7,7~9 respectively
reciprocal	If the value of element i compare to j is b_{ij}, then the value of j compare to i is $b_{ij}=1/b_{ij}$

Suppose the elements of a level have relations with the elements of the lower level. The elements in the lower level compare with each other and we record the results as bij. Then we define B the the results as b_{ij}. Then we define B the judgment matrix.

$$B=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}$$

Step 3 is single hierarchical sort. For the judgment matrix B mentioned above, we calculate the characteristic root and feature vector meet formula $\mathrm{BW}=\lambda_{\max}W$.

The component of W is weight of each element which conduct single hierarchical sort. The root method is introduced as the following:

Firstly, for each line of the judgment matrix, we calculate the product of all elements from the same line, record it as $\mathrm{W_i}: W_i=\prod_{j=1}^{n} b_{ij}$.

Secondly, we calculate the Nth root of W_i , and name it as $\overline{\mathrm{W_i}}$: $\overline{W_i}=\sqrt[n]{W_i}$.

Thirdly, we normalize the vector $\overline{W} = [\overline{W_1}, \overline{W_2}, ..., \overline{W_n}]^T$, that is $a_i = \frac{\overline{W_i}}{(\sum_{j=1}^{n} \overline{W_j})}$, then $A = [a_1, a_2, ..., a_n]^T$ is the characteristic root we need, which is the weight.

Forthly, we calculate the maximum characteristic root $\lambda_{\max}$ of judgment matrix.

Step 4 is checking consistency for judgment matrix. To test the consistency of judgment matrix, we need calculate the consistency index CI, and $\mathrm{CI} = \frac{\lambda_{\max} - n}{n-1}$ .If the value of CI is zero, then it has full consistency.

The larger is CI, the worse is consistency. Reversely, the smaller is CI, then the judgment matrix is closer to full consistency.

Comparing CI to the average random consistency index RI, the RI value of each order is in the table 5:

Table5 RI values of low order

order	1	2	3	4
RI	0.00	0.00	0.58	0.9
order	5	6	7	8
RI	1.12	1.24	1.32	1.41

When the order is larger then 2, radio of CI and RI is called the consistency radio, CI and RI are under the same order, and named the consistency radio as CR, $\mathrm{CR} = \mathrm{CI/RI}$.

In generally, when $CR < 0.1$, the judgment matrix has good consistency, otherwise, values of judgment matrix need to be adjusted.

4.3.2 The procedures of fuzzy comprehensive evaluation

The fuzzy comprehensive evaluation is mainly according to each evaluation factor, evaluation standard and important degree of all relative element to set up the fuzzy comprehensive evaluation model, then evaluate each object. The procedures are in the following:

Step 1 is determining the set of evaluation index. We build up the domain set made up by various main indexes influencing on objects to be evaluated.

Step 2 is determining the set of weight of evaluation index. According to the importance of each index, the related analysis method is used to define the weight as a_j, and these weights make up the set of weight we need.

Record it as $\mathrm{A} = \{a_1, a_2, \cdots, a_n\}$ and $\sum \mathrm{a}_j = 1 (a_j \geq 0; i = 1,2,\cdots,n)$.

Step 3 is determining the set of rating level. We assume the set as $\mathrm{V} = \{v_1, v_2, \cdots, v_n\}$.

In generally, V_j is the membership of each object to be evaluated according to different rating levels.

To obtain these evaluation results by target comparison method, we use questionnaire survey method and the expert evaluation method.

In the process of evaluation, for each qualitative index from each layer of index, we design a group of fuzzy evaluation values, such as the grade evaluation values {excellent, good, fair, poor}.

Every quantitative index is changed to qualitative index based on its characteristics and the grade evaluation values mentioned above.

For example, if the success rate of data backup is ninety-nine percent, then all questionnaire survey deem the index as excellent.

Step 4 is building up the fuzzy matrix of single factor evaluation. The fuzzy subset of single factor is $\mathrm{R_i} = \{r_{i1}, r_{i2}, \cdots, r_m\}$, it is that for each evaluation index, the evaluation value of each object to be evaluated according to the set of rating level, which is usually expressed as a percentage.

Step 5 is decision on fuzzy synthetic evaluation. We consider the distribution of weights in the case of multi-factor and set the decision model of fuzzy synthetic evaluation of fuzzy synthetic evaluation as: $\mathrm{B} = \mathrm{A} \cdot \mathrm{R}$, $\mathrm{B} = \{\mathrm{b}_1, \mathrm{b}_2, \cdots, \mathrm{b}_n\}$; $\mathrm{b_j} = \cup \{a_i \cap r_{ij}\}\ i = 1,2,\cdots,n$.

The symbols " $\cap$ ", " $\cup$ " respectively represent the minimum value and the maximum value. The meaning of b_j is the membership of the objects to be evaluated to the rating level of j in rating level set when considering all factors.

Then, a decision is made based on the value of membership. For an evaluation issue, the set of evaluation factor is $\mathrm{U} = \{\mathrm{u}_1, \mathrm{u}_2, \cdots, \mathrm{u}_n\}$, the evaluation set is $\mathrm{V} = \{\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_n\}$, the weight of each evaluation index is $\mathrm{W}_1, \cdots, \mathrm{W}_m$, then the problem of fuzzy synthetic evaluation can be described as calculating the fuzzy multiplied $U \cdot V$.

So the government informatization projects performance can be evaluated and audited more effectively.

5. ClOSING

With the implementation of the national informatization strategy, the government informatization projects construction is imminent. In this paper, the theory of comprehensive evaluation is applied to build a multi-level evaluation model set suitable for China's national conditions and the informatization projects characteristics. In the study, the conclusions are obtained as the following:

Firstly, through the collection and analysis of the existing literature, our specific environmental factors are added to the index system to realize dynamic model.

Secondly, construction of informatization projects can improve the level of government information, optimize departmental processes and improve the working environment and the state of all levels of management and operations staff.

Thirdly, informatization projects performance audit evaluation model is study to the results of evaluation of informatization projects construction based on the theory of performance audit. The national audit institutions can investigate the level of information technology through the model of this stage, to verify the importance of vigorous development of information technology projects.

6. ACKNOWLEDGMENTS

It is my grateful thanks for journal to supply the templates which help me to modify my paper combining with reality. And this paper is supported by the subject of performance audit of government investment projects (project number: 201201117).

7. REFERENCES

[1] Sheng Yingxian. 2012. *Try to talk about the government investment informatization construction projects performance audit. Chinese Agricultural Accounting.*

[2] Shang Zizhong. *Performance audit is an effective means to improve the operation efficiency*. *Western Forum*(Dec. 2007).

[3] Richard E. Brown, James. Pyers. 1988. *Putting Teeth into the Efficiency and Effectiveness of Public Services. Public Administration Review* 735-742.

[4] Liao Xiangwu. 2009.*Design of Performance valuation System for Government informatization projects. Informatization Research* (Vol.35, No.9, Sep. 2009).

[5] Yang Guanbiao, Gao Yingyi. 2005. *Theory and application of fuzzy math. Guangzhou: South China University of Technology press*, 2005.

[6] UNPAN. *UN Global E-Government Readiness Report 2004* [R]. 2004.

[7] Egon Berghout, Theo-Jan Renkema. *Methodologies for IT Investment Evaluation: A Review and Assessmen ， In Grembergen W. V. Information Techonlogy Evaluation Methods and Management*[M]. London: Idea Group Publishing, 2001, (8).

Lightweight Web Methodology with Practical Applications to Medical Fields

Atsushi Togashi
Graduate School of Project Design, Miyagi University *
togashi@myu.ac.jp

Yu Kitano
Graduate School of Project Design, Miyagi University
kitano@00index.co.jp

Hiroki Suguri
Graduate School of Project Design, Miyagi University
suguri@myu.ac.jp

ABSTRACT

We have identified four typical problems in building web applications: High learning cost; difficulty of distributed development; scope creep; and coupling of user interface and business logic. To solve these problems, we have developed lightweight methodology and Perl-based framework for web applications. In this paper, we discuss effectiveness of our approach by describing our lightweight methodology and lightweight framework. We have successfully applied these techniques in developing several business applications, which illustrate real-world usefulness of the proposed methodology and framework.

Categories and Subject Descriptors

D.2.10 [Design]: Methodologies

General Terms

Design

Keywords

Lightweight Web Methodology, Web API Framework, Medical System

1. INTRODUCTION

General-purpose nature of the web makes it the most popular usage of computer system today. To develop web applications effectively, following four challenges must be overcome: (1) High learning cost; (2) Difficulty of distributed development; (3) Scope creep; and (4) Coupling of user interface and business logic.

1.1 High Learning Cost

Learning cost to start developing web application is expensive because wide range of expertise is necessary. In the case of development in Java, it is indispensable to understand HTML, JavaScript, web server administration, JSP/Servlet, database, and typical design patterns in addition to Java language itself. Session management and security issues specific to web application [1] [2] also increase learning cost. Usually, knowledge about frameworks such as Struts and Hibernate is additionally required.

A framework is a set of functionalities commonly required to build applications. Many frameworks arrange large number of components to build feature-rich applications. However, the initial cost of learning to start using such heavyweight framework is huge.

1.2 Difficulty of Distributed Development

Division of work and integration to system is necessary in developing web application with a team of developers. Multiple technologies are involved in web application development, such as server-side programming, database middleware, and client-side page design. Specialized engineer is required for each technical domain, which is complicated by itself. In addition, integrating them into a system is much more difficult.

MVC pattern is a traditional approach to logically divide the system into three components of Model, View, and Controller. Each component is assigned a specific task and communication between the components drives the system. MVC approach is popular in modeling and developing web applications [3] [4] [5]. However, the model must be carefully designed and fixed at first. Modeling is an intensive and time-consuming task that becomes bottleneck early in the development process before distributing the work to page designers and programmers.

1.3 Scope Creep

System requirements are getting more and more complicated today. Since web is bleeding edge of business, it is common to begin developing systems without fixing the requirements. Even in the middle of the development process, requirements keep changing and the system must quickly embrace the changes. Traditional waterfall project management technique is no longer effective in such scope creep situations.

Many web applications are developed using application frameworks that are based on MVC approach, such as Struts and Ruby on Rails [6]. It is true that MVC approach is said to resilient to changes. However, it is only true when the model is kept intact. If the model is forced to be altered due to requirement changes, system-wide modification of the program is likely to occur.

*1-1 Gakuen, Taiwa-cho, Kurokawa-gun, Miyagi 981-3298 Japan

1.4 Coupling of User Interface and Business Logic

Web application is roughly divided into HTML page design rendered in client browser, and background business logic programmed in server side. Design of user interface and program of business logic must be strictly distinguished, allowing page designer and programmer to work separately. However, in many existing web application frameworks, programming logic such as field repetition and input validation must be coded in page layout along with HTML. Therefore, changing user interface design requires programmer's involvement as well as page designer's work.

The purpose of the paper is to propose lightweight methodology and framework to solve these problems. In section 2, we describe proposed lightweight methodology and application framework in detail. Application examples are reported in section 3. Section 4 concludes the paper by summarizing the outcomes and mentioning future work.

2. PROPOSED LIGHTWEIGHT METHODOLOGY AND FRAMEWORK

This section explains proposed lightweight methodology and framework to solve the four problems discussed in the introduction. The techniques are targeted for small- and medium-sized applications. Suggested number of pages is less than thirty. With such applications, large framework such as J2EE is unnecessary. Larger system can be built by integrating such medium-sized applications.

Our approach focuses on rapid prototyping. It is most effective for developing web application that specifications cannot be fixed at early stage of the project. In other words, the application framework implementing our methodology is not quite efficient when the specifications are determined at the beginning and kept intact during the whole process of the project. In such cases, traditional heavyweight application framework is sufficient, combined with waterfall project management methodology.

We also set the target users of the framework as novice developers. It is ideal to employ highly skilled engineers to develop applications. However, doing so is often impossible due to many reasons. It is sometimes neglected but very important task to assure productivity, quality and security of the system built by group of inexperienced programmers.

2.1 Lightweight Methodology

Before explaining design and implementation of the proposed framework, we discuss the lightweight methodology, which is a foundation of the framework.

2.1.1 Programming Language

The programming language must be productive and easy to learn. We chose Perl. Perl is one of so-called lightweight languages along with Ruby and Python. These lightweight languages offer rich set of functionalities in simple syntax.

2.1.2 Development Environment

For the programming environment, we wanted to avoid using heavyweight IDE such as Eclipse and NetBeans. These feature-rich tools are difficult to learn for novice developers. Difference between local debugging environment and remote deployment environment is also problematic. Instead, we incorporated browser-based development environment (i.e., file browser, text editor, and debugger) inside the framework, which handles server files directly.

2.1.3 Lightweight MVC

Web application is divided into two main components. One is client page design comprising HTML, CSS, JavaScript, and media files. The other is server program and database. The client side and the server side can be developed asynchronously. In many cases, end users want to see client page design before server programming has been completed. To meet these requirements, our lightweight methodology abstracts web application as a set of page transitions.

In traditional MVC approach, Controllers manage Models and Views. Multiple Models can be represented by a View through a Controller. A Model must cooperate with multiple Controllers (Figure 1). This complicated structure requires precise modeling and is vulnerable to requirement changes.

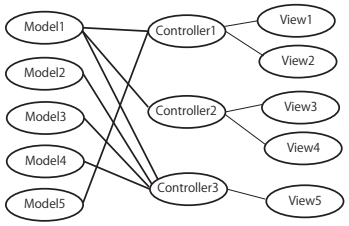


Figure 1: Traditional MVC.

In our approach, MVC is redefined as follows: Model is a Perl program module that corresponds to a set of pages, View is HTML template discussed later, and Controller is the web application framework itself. A Model can handle multiple Views through the Controller, and a View must belong to a Model. We call it lightweight MVC (Figure 2).

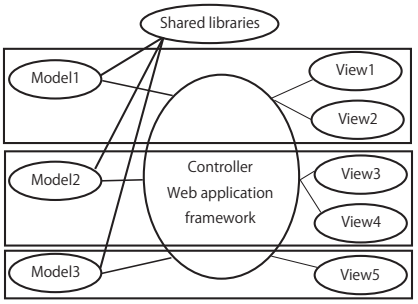


Figure 2: Lightweight MVC.

Typically, a programmer is responsible for a Model. Views that belong to the Model are related pages, among which page transition occurs. For example, a Model represents registered users. Views that belong to the Model are registration page, modification page, deletion page, and summary page of the users. Thus, the programmer's scope of responsibilities is clearly defined. When shared libraries are required among multiple models, the libraries can be programmed in ordinary object-oriented fashion.

2.2 Lightweight Framework

Table 1 shows server environment supported by the lightweight framework.

Table 1: Server environment.

Operating system	Linux and Windows Server
Web Server	Apache
Database	PostgreSQL and SQLite
Language	Perl

Figure 3 depicts overview of the structure of lightweight framework.

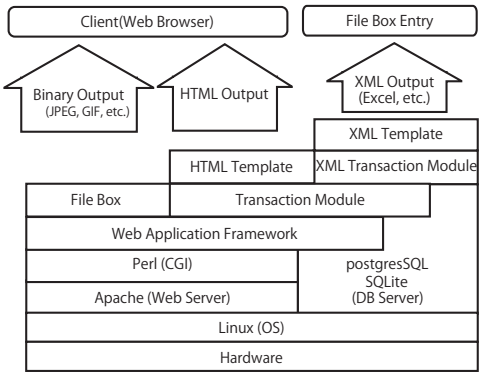


Figure 3: Structure of lightweight framework.

As illustrated in Figure 4, the framework consists of five Perl modules. Four modules (kwdebug.cgi, kw.cgi, kwInputAssist.cgi, and kwfile.cgi) are accessed by web browser. API module provides application programming interface.

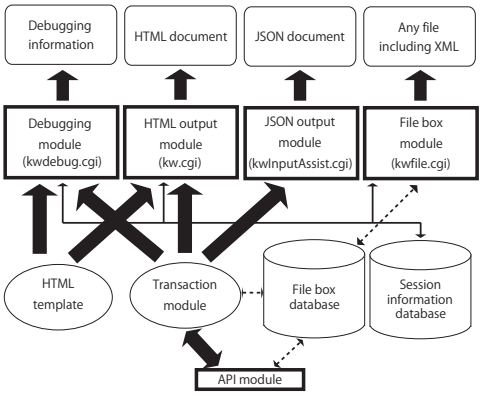


Figure 4: Perl modules.

Module kw.cgi is the core of the framework. It calls back transaction module supplied by developer to generate HTML output from HTML template, which is also supplied by developer. Module kwfile.cgi is responsible for binary file output from file box. Module kwdebug.cgi is a superset of kw.cgi for debugging purposes. Module kwInputAssist.cgi interacts with JavaScript program running on the browser by sending JSON data.

2.2.1 HTML Output

HTML template embeds data items in page layout along with regular HTML description. The data items are defined in the transaction module. In the HTML template, designer can use variable tags to be replaced with run-time value of the variable.

In MVC terms, kw.cgi corresponds to Controller, HTML template corresponds to View, and transaction module corresponds to Model. Page transition of the web application controlled by kw.cgi is handled by HTML form parameters sm and ss. The parameter sm specifies transaction module that is executed after page transition. The parameter ss designates HTML template used by the transaction module. If ss is null, default HTML template is used, which is specified in the transaction module. In other words, next transaction module after pressing submit button is determined by the value of sm, while HTML template is chosen by the values of sm and ss. Figure 5 represents the relationship between the components.

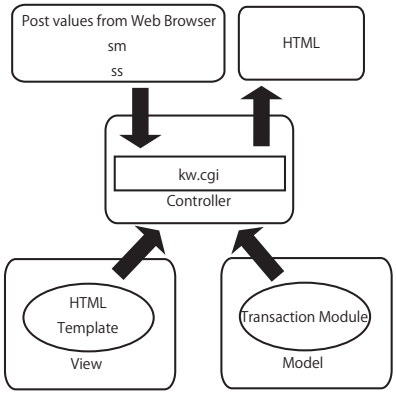


Figure 5: kw.cgi.

If transaction module is empty, empty application is generated that only performs page transitions defined by HTML template. This is useful for checking the page transitions before getting to the programming. It is recommended that HTML template (View) is designed and implemented at first, before transaction module (Model) is designed and implemented.

2.2.2 Control Flow of tag substitution

Figures 6 and 7 respectively show top level and substitution loop of control flow to generate HTML output from HTML template and transaction module. Module kw.cgi, which generates HTML output, executes function do_HTMLtemplatename() in the transaction module that is specified in form parameter sm, where HTMLtemplatename is the name of HTML template specified in form parameter ss. The purpose of function do_HTMLtemplatename() is to process page-specific business logic before substitution loop of tags begins.

The substitution loop recursively substitutes tags found in the HTML template with run-time value of the variable or execution result of Perl script to generate final HTML output. The substitution loop is a function that takes loopname as an argument, where loopname is a name of the loop spec-

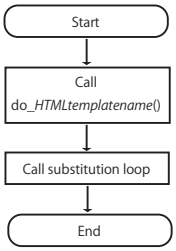


Figure 6: Top level of control flow.

ified in Begin Loop tag (@~loopname~ {~@) of the HTML template.

The substitution loop executes function loopname_loadArray() in the transaction module at first. (This is skipped if the loop is the top level of the recursion.) Programmer writes code in that function to initialize the loop. The return value of the function is number of iterations or -1 if undetermined. Module kw.cgi set the number of iterations in hash variable $kw::out_param{loopname$MAX}. A hash variable is an associative array of key-value pairs where key can be non-integer. The value can be used in the HTML template to adjust the number of rows or columns of the repeated field.

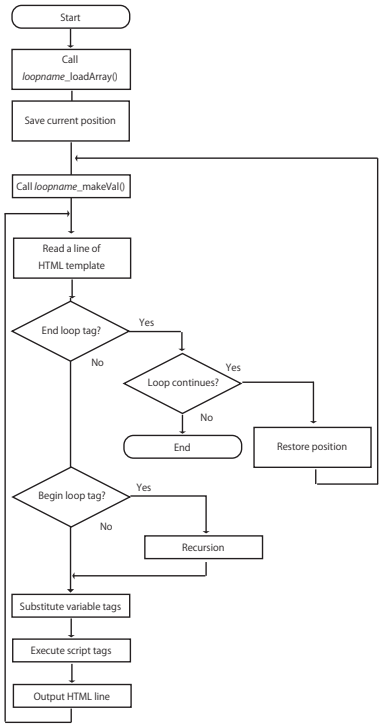


Figure 7: Substitution loop.

The Controller kw.cgi saves memento of current position in the HTML template. Then function loopname_makeVal() in the transaction module is called back. (In the case of top level of recursion, function makeVal() is called.) In this function, substitution values are stored in %kw::out_param.

Then, current line of the HTML template is copied to variable $line. If $line contains End Loop tag (@~}~@) and the loop breaks, the substitution loop function returns. If $line contains End Loop tag (@~}~@) and the loop continues, memento of saved position in the HTML template is restored and function loopname_makeVal() is called back again to process the next iteration of the loop.

If $line contains Begin Loop tag (@~loopname~{~@), the loop is recursively processed. Then, variable tags ($~variable~$) in $line are substituted with the value of corresponding variable found in hash $kw::out_param{variable}. Basically, kw.cgi performs substitution shown in Figure 8. Sanitization is omitted due to simplicity.

```
$line=~ s/\$\~(.*?)\~\$/$out_param{$1}/eg;
```

Figure 8: Variable substitution.

If $line contains Perl script tag (!~code~!), the code is executed and the result substitutes the tag. At this point, $line is substituted HTML line, which is sent to the browser. The next line of HTML template is processed likewise.

Table 2 shows major tags used in the HTML template. Table 3 lists typical callback functions programmed in the transaction module.

Table 2: HTML template tags.

Tag	Description
!~Code~!	Replaced with execution result of Code as Perl script.
$~variable~$	Replaced with value of variable.
@~loopname~ {~@	Begin loop identified by loopname.
@~} ~@	End loop.

Table 3: Callback functions.

Function	Description
loopname_loadArray()	Called at the beginning of tag substitution for loopname. Returns number of iterations.
loopname_makeVal()	Called at the beginning of repetition for loopname. Sets %kw::out_param for loopname.
makeVal()	Called at the beginning of top level. Sets %kw::out_param for non-loop.

2.2.3 Page Transition

Page transition is usually specified by ss in POST or GET parameter of the View as shown in Figure 9. In addition, transaction module can explicitly specify sm and ss to transit to the desired page, which is illustrated in Figure 10

If sm and ss are not specified when kw.cgi is accessed by browser, default transaction module and default HTML template are used, which are defined in initialization file.

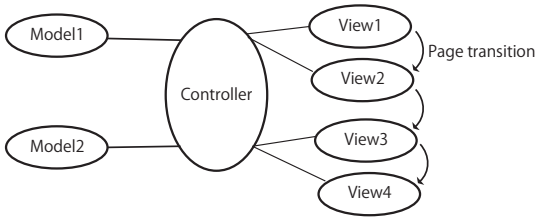


Figure 9: Page transition by ss.

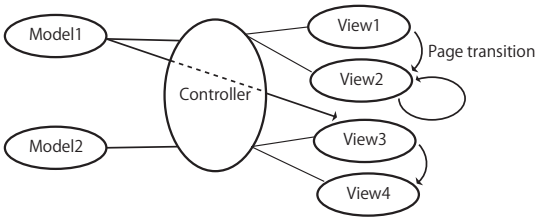


Figure 10: Page transition by transaction module.

Usually, the default is a login page. When user credentials are insufficient or combination of sm and ss is illegal, error module and error page is invoked.

2.2.4 File Box

It is common that web application must deal with image files, PDF files and Microsoft Office documents. To output these non-HTML files, file box may contain arbitrary binary files, as shown in Figure 11.

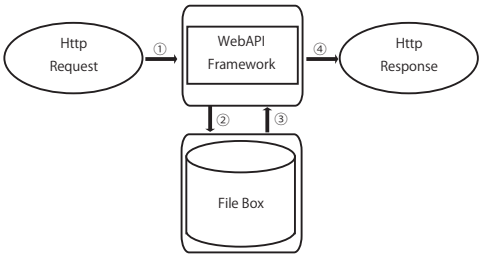


Figure 11: File box.

Transaction module may register non-HTML files with file box. These files are sent to HTTP response by specifying the file identifiers to kwfile.cgi. It is also possible to enter URL in the browser that specifies the file identifier.

XML files can be generated and registered as easily as HTML output by using XML template, which is similar to HTML template described above. This is useful when the web application produces XML data files that are used by client applications such as Microsoft Excel.

XML template can be edited by Microsoft Excel as shown in Figure 12. Figure 13 shows generated XML output, which is opened by Microsoft Excel.

2.2.5 Debugging and Miscellaneous functions

Debugging module kwdebug.cgi displays run-time values of variables including %kw::out_param set by callback functions

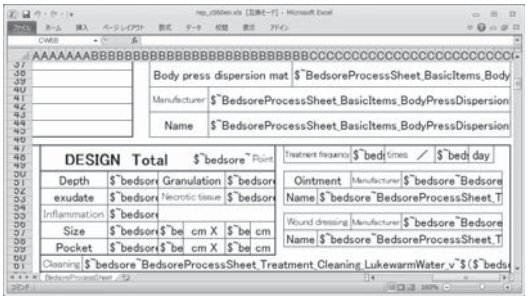


Figure 12: XML Template.

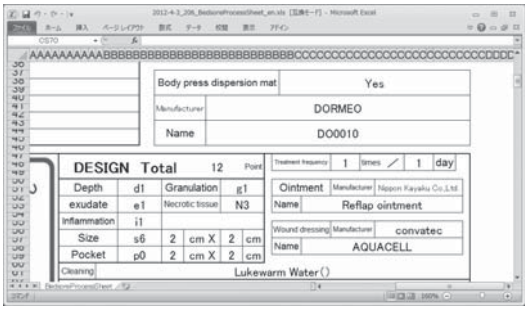


Figure 13: Microsoft Excel Output.

of transaction module. It also offers Adobe Flash-based text editor for client browser to edit server files without using local editor and file transfer tools. This simplifies the development of web application. Especially in Japan, character code set on server (UTF-8) is different from that of Windows client (Shift JIS). Developers must always be very careful to convert the character set properly, which is heavy stress. We solved the problem by offering server-based text editor.

In addition, the framework provides miscellaneous functions such as numeric formatting and Japanese Imperial Calendar processing.

3. APPLICATIONS TO MEDICAL FIELDS

In this section, we discuss applications to the medical fields built on top of the proposed lightweight framework.

Miyagi Cardiovascular and Respiratory Center and Miyagi University jointly developed Nutrition Support Team (NST) and Bedsore Management System [7] [8] using the proposed methodology and framework in 2008. NST is a team comprising doctors, nurses and nutritionists to comprehensively support patients. The team medical care was introduced in Japan in 2007 by Ministry of Health, Labor and Welfare. Therefore, no existing system can be referred to develop the new application. It was impossible to fix requirement specifications at initial stage of development. We adopted agile project management and rapid prototyping using the framework to redesign, rebuild and refactor the application again and again.

Miyagi regional clinical cooperation passport is a medical information system by which hospitals, clinics, and nursing stations can securely share patient information. Parties can view, edit, add, and delete medical records on bedsores,

gastrostomy, and oral cavity care by PC using standard web browser or by iPad using dedicated iOS application. Android client application is under development. The system started in year 2009 linking six hospitals and clinics for taking care of bedsores of patients who move between the medical facilities. As of August 2012, more than fifty health care institutions have joined the network. As the number of users increased, specification changes also increased. We handled enhancement requests successfully utilizing the framework. Currently, total number of pages is thirty and more than ten Microsoft Excel output (via XML) are generated.

The company at which first author works developed human resource management system using the framework. Total number of page is approximately 200. This large system was built by combining smaller subsystems, each comprising approximately 30 pages. The proposed lightweight framework has been proven to perform well in such a large application. The system was released to customers in 2008. Even now, new features are added and requirement changes are implemented.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed lightweight methodology and lightweight framework for developing web application. We have actually developed real-world applications that are actively utilized and constantly improved. To develop web application, orthodox method is to employ highly skilled engineers and to adopt heavyweight methodology, framework, IDE, and project management techniques. However, another approach is to use lightweight methodology and lightweight framework with unskilled developers. We have proven that web application can be developed with limited financial and human resources.

The outcome of the research is summarized to the following four points.

(1) The lightweight methodology and framework enabled cost-effective development of web application. Compared to Java-based framework and Ruby on Rails, learning cost is cheap. Even unskilled programmers could take part in the development of high-reliability large application like medical passport system. We have also developed simple applications simpler, like bus timetable optimization system.

(2) We made distribution and integration of workload effectively by defining scope of responsibilities for each page. Programmers and designers cooperated smoothly to finish the system in short time.

(3) The proposed framework allowed rapid redesigning and rebuilding of the system to satisfy requirement changes from users. In fact, during the development of bus information system, users frequently requested to change the functionality of the system. In the case of NST application, it took six months to finalize the requirements. During that period, even fundamental changes of workflow occurred, resulting complete redesigning of screen transitions. The proposed framework handled these changes successfully.

(4) The framework decoupled page design, application logic, and data model thoroughly. Our lightweight MVC separates View and Model distinctively, and framework itself becomes Controller. Therefore, the application does not suffer from "pseudo MVC" problem. The application is resilient to requirement changes.

After successful development and deployment of applications based on the framework, we also identified two future work. The first thing to do is to incorporate test tools in the framework. In unit testing framework such as PerlUnit, dedicated testing program must be coded. However, when specifications change, the testing program must be changed, too. This is not productive. We are planning to develop unit testing library as a part of our framework to solve this problem.

Second issue is version control. Tools like CVS, Subversion, and Git are popular among developers. However, unskilled programmers are not familiar with the notion and operation of version control. In many cases, simple versioning such as those found in Microsoft Office is sufficient for small- and medium-sized project. We are designing easy-to-use versioning tools to incorporate to our lightweight framework.

5. ACKNOWLEDGMENTS

The work was partly supported by Sugiura Foundation for Development of Community Care.

6. REFERENCES

[1] Martin Johns, SessionSafe Implementing XSS Immune Session Handling, Lecture Notes in Computer Science, Volume 4189, pp. 444-460, 2006.

[2] Martin Johns, Code-injection Vulnerabilities in Web Applications Exemplified at Cross-site Scripting, Information Technology, Vol. 53, No. 5, pp. 256-260, 2011.

[3] N. Mitsuda and N. Fukuyasu, Issues behind the Use of Web Application Frameworks, Computer Software, Vol. 27, No. 3, pp. 2-12, 2010.

[4] Avraham Leff, James T. Rayfield, Web-Application Development Using the Model/View/Controller Design Pattern, Proc. Enterprise Distributed Object Computing Conference, pp. 118-127, 2001.

[5] Wojciechowski, J., Sakowicz B., Dura K., Napieralski, A., MVC model, struts framework and file upload issues in web applications based on J2EE platform, Proceedings of the International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science, pp. 342-345, 2004.

[6] Ruby on Rails, http://rubyonrails.org/

[7] Soichi Shibata, Emi Asada, Takako Takahashi, Reyo Ikeda, Yu Kitano, and Atsushi Togashi, Development of Effective Patient Information Sharing System in Nutrition Management, The Japanese Journal of Clinical Nutrition, Vol. 117, No. 1, pp. 79-83, 2010.

[8] Atsushi Togashi, Yoshitsugu Takahashi, Hiroyuki Aoki, Ken Tomii, Tatsuya Uesaki, Ippei Miyauchi, Yu Kitano, Soichi Shibata, Kan Endo, and Reyo Ikeda, Research and Development of Regional Medical Information System, Proc. 70th National Convention of Information Processing Society of Japan (4), pp. 4-447-4-448, 2008.

[9] Satoshi Nakajima, "Pseudo-MVC" Problem in Ruby on Rails, http://satoshi.blogs.com/life/2009/10/rails_mvc.html

Service enhancement, performance and product focus: A moderating model

Haitao LI
School of management, Harbin Institute of technology,
Heilongjiang, China
+86-0451-86414042
haitaoli2007@163.com

Huashan LI
School of management, Harbin Institute of technology,
Heilongjiang, China
+86-0451-86414042
lihuashan2007@gmail.com

Yezhuang TIAN
School of management, Harbin Institute of technology,
Heilongjiang, China
+86-0451-86414042
tianyezhuang@hit.edu.cn

ABSTRACT
Service strategy-environment configuration is the core issue in manufacturing servitization research. In this paper, we take the "fit as moderation" perspective to exploit the different service enhancement practice choice under different status of product focus. The result show that 1) Service enhancement practice is positively related to both market performance and financial performance, and 2) the relationship between service enhancement practice and market performance is mediated by product focus. This paper provided an empirical evidence for the environment-strategy configuration literature.

Categories and Subject Descriptors
J.1 [**Administrative Data Processing**] Area of application*s* – *manufacturing, business.*

General terms
Management.

Keywords
Service enhancement; business performance; product focus

1. Introduction
No one doubts Darwin's natural selection theory that "survival of the fittest", like creatures acclimatize themselves to the environment, the enterprises also respond to the changes of the external environment. With the increasing competitive intensity and the need to exploit new growth potential, traditional product-manufacturing firms are extending their service business [1, 2]. While extending the service business seems the right way to escape the trap of decreasing product margins and ever more complex customer needs, there is a paucity of empirical research concerning the "service paradox", namely that it appears more difficult for firms to make incremental profits by adding services than might be expected[3, 4]. The reasons for "service paradox" are multi-sided, some scholars attribute to the high cost when providing service [5], but take from a fit perspective, and many problems are actually caused by the misfit between environment and strategy [2].

As suggested Oliva and Kallenberg, there is a transition continuum from pure product manufacturers to service providers, extending the service business is a step-wise transition process [6]. According to Gebauer, the choice of service strategy should fit the external environment [2]. Product focus is a dimension of an enterprise's external environment, which represent the market's demand for tangible products or the intangible service, there is also an continuum from tangible products to intangible service, from a strategic fit perspective, it's crucial to configure a firm's service to the product focus.

We are interested in how manufacturers acclimatize themselves to the requirement of the changing environment? In this paper, we take the "fit as moderation" perspective, to exploit the different service enhancement practice choice under different status of product focus. This paper provided an empirical evidence for the environment-strategy configuration literature. And also, this paper provides implications for managers.

The paper is organized as follows: The next sections review the literature on service enhancement and strategy-environment configuration. The paper then introduces the proposed hypothesis and a moderation model. The method and result are then explained. The paper concludes with suggestions for future research.

2. Literature review

2.1 Service enhancement
Increasingly, manufacturers are extending service business, Berger and Lester proposed the notion of "service enhancement" from a business strategy point of view, focuses on the impact of services on the selection of competitive strategy and the access of competitive advantage [7]. The thesis of service enhancement is that manufacturing companies enhance the competitiveness of their products and services as an important source of value creation through the product-based services [7]. This is can be viewed as a way to escape the trap of decreasing product margins and ever more complex customer needs, by providing customer ser-

Research Notes in Information Science (RNIS)
Volume13, May 2013
doi:10.4156/rnis.vol13.22

vice, so as to enhance the competitiveness of manufacturing enterprises and obtain a new value creation source [6].

Service enhancement is similar to the notion of servitization which depict the transition from product manufacturer to product-service integrate offering provider [8]. In the evolution of servitization, many manufacturing companies have moved dramatically into services and so caused the boundaries between products and services to become blurred. Today's customers are increasingly demand an integration of product and service offerings rather than single product, manufacturers should provide a bundle of service enhanced products rather than single tangible products. There has a strong complementarity between the development of manufacturing industry and high value-added services, "the service enhanced-product" provided a huge space for development.

2.2 Strategy-environment configuration

There are many kinds of service enhancement practices, manufacturers can be positioned on a product-service continuum ranging from products with services as an "add-on", to services with tangible goods as an "add-on" [6]. At the one extreme point of the continuum, firms achieve a competitive position as a product manufacturer. They produce essentially core products, with services purely as add-ons. Profits and revenue are generated mainly through the company's core products. At the other extreme point, products are merely an add-on to the services. Products represent only a small part of total value creation. The dominant share of total value creation stems from services.

According to the strategy–environment fit framework, external environment have a strong influence on the service strategy choice, manufacturers should choose their service strategy according to the external environment [2]. Moreover, organizational performance depends partly on the strategy–environment fit [9]. Product focus is a vital dimension of firm's external environment which depicts customer's preference ranges from the physical product to intangible service. In a physical product-preferred market, customers will emphasis the physical attribute of product, manufacturers should act as a product manufacturer. While in an intangible service-preferred market, manufacturers should act as a service provider.

3. Research hypothesis

3.1 Service enhancement and performance

For the following reasons, service enhancement practice was supposed to be positively related to business performance. First, higher profit margin and more revenue were mentioned as the main driver for manufacturers to extend service business, in some sectors, service revenues can be one or two orders of magnitude greater that new product sale [1]. Compared to several other strategic options such as fostering innovation and technology and product quality, competing through services enables product manufacturers to earn the highest potential margins[3].

Second, service was viewed as a differential factor, hence gaining competitive advantages for manufacturers. Furthermore, Competitive advantages achieved through services are often more sustainable since, being less visible and more labor dependent, services are more difficult to imitate [3, 6]. The value-add of services can enhance the customer value to the point, where, homogeneous physical products are perceived as customized. These increase barriers to competitors [10].

Third, services are also claimed to create customer loyalty to the point where the customer can become dependent on the supplier. Services tend to induce repeat sale and, by intensifying contact opportunities with the customer, can put the supplier in the right position to offer other products or services [8, 10].

Therefore, we respect a positive relationship between the service enhancement practice and business performance:

Hypothesis 1: Service enhancement practice is positively related to market performance.

Hypothesis 2: Service enhancement practice is positively related to financial performance.

3.2 The moderating effect of product focus

The importance of strategy-environment has been recognized in previous literature. Neu and Brown argue that "firms that successfully develop B2B service will align strategy with conditions of the service business unit's external environment and adapt several factors of organization to align with the newly formed [service] strategy" [11]. Gebauer categorized four different service strategies and pointed out that each strategy should be aligned with the external environment [2]. That means organizational performance of manufacturing companies moving along the transition line depends on the proper alignment between environment, strategy, and factors in organizational design. There-fore, the following hypothesis is proposed:

Hypothesis 3: Service enhancement practice will have a strong effect on market performance when customers emphasize service.

Hypothesis 4: Service enhancement practice will have a strong effect on financial performance when customers emphasize service.

The theoretical framework could be summarized as follows:

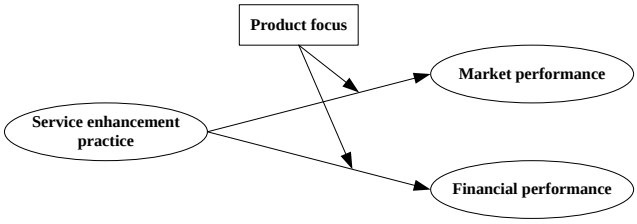


Figure 1 Theoretical framework

4. Methods

4.1 Sample

This study uses data from the 5th round International Manufacturing Strategy Study (IMSS-V). The IMSS was started by Voss and Linderberg in 1993, carried out for every four years, previous IMSS rounds were carried out in 1993, 1997, 2001, and 2005. It assesses the strategies, practices, and performance of manufacturing companies. Data from previous IMSS rounds were used in manufacturing strategy studies and supply chain management studies, published in journals such as Journal of Operations Management et. al.[12]. The questionnaire was based on Likert five point scales along with some objective measures.

The IMSS-V was carried out in 2009, including 719 samples from 23 countries and 7 industries. Due to the increasing trend of manufacturing servitization, some service issues were included in IMSS-V.

4.2 Measurement

According to previous studies, there are three type of service enhancement practice, in IMSS-V, related items were contained. Respondents were required to indicate the effort put into implementing the following action programs in the last three years within their company:

SEP1: Our company is actively engages in expanding the service offering to our customers;

SEP2: We are actively developing the skills in the organization needed to improve the service offering;

SEP3: We deliberately design products so that the after sales service is easier to manage/offer.

Among each question, respondents were asked to choose the degree of effort within a 1-5 Likert five point scale table, 1 stands for the program had never been implemented and 5 stands for the program were implemented at a high level. We calculated the average of SEP1, SEP2 and SEP3 to measure service enhancement practice.

We measured business performance by compare sales, market share, return of sales (ROS) and return of investment (ROI) to the main competitors in a five point Likert scale, which 1 stands for worse than competitors and 5 stands for better than competitors. These four items were clustered into market performance (Sales and Market share) and financial performance (ROS and ROI). Average was calculated to measure both market performance and financial performance.

We measured product focus by asking the respondents to choose the degree of product focus within a 1-5 Likert five point scale table, 1 stands for physical attribute and 5 stands for emphasize service.

The measures could be summarized as table 1. The Cronbach's Alpha of the measurement of service enhancement practice and business performance is 0.812 and 0.827, which is over the minimum level of 0.7, which means that service enhancement practice and business performance are well measured.

Table 1 measurement

Variables	Items	Mean	Std. error	Cronbach's α
Service enhancement practice (SEP)	SEP1	2.94	1.25	0.812
	SEP2	3.13	1.18	
	SEP3	3.08	1.28	
Business performance (BP)	Sales	3.36	0.88	0.827
	Market share	3.33	0.88	
	ROS	3.22	0.81	
	ROI	3.21	0.81	
Produce focus (PF)	PF	2.85	1.20	--

5. Result

Hierarchical regression is used to test the relationship between service enhancement practice and business performance, and also the moderating effect of product focus. Table 2 describes the regression coefficients. Model 1 and model 3 tested the relationship between SEP and business performance. SEP is positively related to market performance (β=0.126, p<0.01) and financial performance (β=0.128, p<0.01), hypothesis 1 and hypothesis 2 were validated. Model 2 and model 4 tested the moderating effect of product focus. The interaction between SEP and PF is positively related to market performance (β=-0.082, p<0.05), indicating that the relationship between SEP and performance is mediated by product focus. But the interaction between SEP and PF is not significantly related to financial performance. Hypothesis 3 is validated, while hypothesis 4 is not.

Following Zatzick et. al., in Figure 2 we plot the relationship between SEP and market performance (+/- 2 s.d.)[13], the graph shows support for hypothesis 3: Service enhancement practice is positively related to market performance.

Table 2 Regression of performance

Model	Model 1	Model 2	Model 3	Model 4
Dependent variables	Market performance		Financial performance	
SEP	0.126**	0.138***	0.128**	0.136**
PF	0.025	0.024	-0.004	-0.006
SEP×PF		0.082*		0.058
R^2	0.018	0.025	0.016	0.019
ΔR^2		0.007*		0.003
F	5.442**	4.978**	4.011*	3.223*

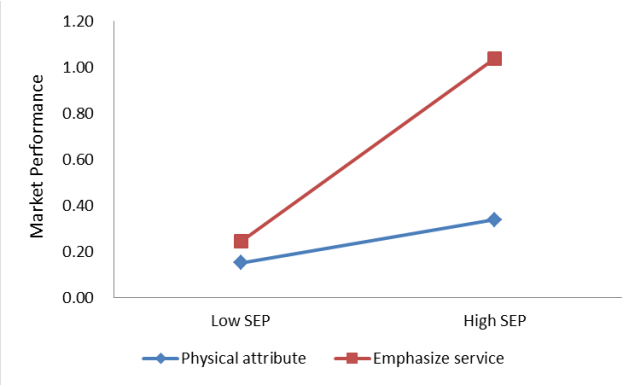


Figure 2 Moderating effect of product focus on the relationship between SEP and market performance

6. Discussion

"Service paradox" has long been an impassable issue for manufacturers in their transition from product producer to product-service offerings provider. In our research, we validated the positive relationship between service enhancement practice and business performance. And also, we tested the moderating effect of product focus on the relationship between service enhancement practice and business performance.

6.1 Theory implication

Our research contributes to the manufacturing servitization literature in two ways: First, we conducted an empirical research and validated the positive relationship between service enhancement practice and business performance. Prior study have proposed that

service is a value-added activity for manufacturers, but there is a paucity of empirical research concerning the phenomenon and that which does exist raises the question of a service paradox, namely that it appears more difficult for firms to make incremental profits by adding services than might be expected[3, 4]. Our research offered a positive insight.

Second, this paper contributes to the strategy-environment configuration literature by testing the moderating effect of product focus on the relationship between service enhancement practice and business performance. Prior research on strategic fit has proposed that the misfit between strategy and external environment is a source for business failure [2]. In this paper, we emphasized the importance of product focus on manufacturers' service strategy choice, namely, service enhancement is more effective if customers emphasize the service aspect of product.

6.2 Management implication

This paper also provides implications for operation managers. Service enhancement is a concept of significant potential value, providing routes for companies to move up the value chain and exploit higher value business activities. It is a way to get rid of the increasing competition in the product market. However, service enhancement does not however represent a panacea for all manufactures, it just provide an option, when take this option, manufacturers should take the external environment into consideration, especially the customers' need. The most important aspect of overcoming service paradox is choosing the right service strategy according to the external environment.

7. Conclusion

Two conclusions could be summarized: 1) Service enhancement practice is positively related to both market performance and financial performance, and 2) the relationship between service enhancement practice and market performance is mediated by product focus. Service enhancement practice is a useful way to help manufacturers to get rid of the ever increasingly intension competition in the product sector. But manufacturers have to configure their service strategy according to the external environment, especially customer's product focus.

References

[1] Oliveira, P. and Roth, A.V. 2012. Service orientation: the derivation of underlying constructs and measures. *International Journal of Operations & Production Management*, Vol. 32 Iss: 2 pp. 156 – 190. DOI= http://dx.doi.org/10.1108/01443571211208614

[2] Gebauer, H.2008. Identifying service strategies in product manufacturing companies by exploring environment–strategy configurations. *Industrial Marketing Management*, 37(3): p. 278-291. DOI= http://dx.doi.org/10.1016/j.indmarman.2007.05.018

[3] Gebauer, H., E. Fleisch, and Friedli, T. 2005. Overcoming the service paradox in manufacturing companies. *European Management Journal*. 23(1): p. 14-26. DOI = http://dx.doi.org/10.1016/j.e-mj.2004.12.006

[4] Ulaga, W. and Reinartz W. J. 2011. Hybrid Offerings: How Manufacturing Firms Combine Goods and Services Successfully. *Journal of Marketing*: Vol. 75, No. 6, pp. 5-23. DOI= http://dx.doi.org/10.1509/jm.09.0395

[5] Neely, A. 2008. Exploring the financial consequences of the servitization of manufacturing. *Operations Management Research*, 1(2): p. 103-118. DOI= http://dx.doi.org/10.1007/s12063-009-0015-5

[6] Oliva, R. and Kallenberg, R. 2003. Managing the transition from products to services. *International Journal of Service Industry Management*, 14(2): p. 160-172. DOI = http://dx.doi.org/10.1108/0-9564230310474138

[7] Berger, S. and Lester, R.K. 1997. Made by Hong Kong. *Oxford University Press*. Hong Kong.

[8] Vandermerwe, S. and Rada, J. 1989. Servitization of business: adding value by adding services. *European Management Journal*, 6(4): p. 314-324. DOI = http://dx.doi.org/10.1016/0263-2373(88)90033-3.

[9] Mintzberg, H., 1979. The structuring of organizations: A synthesis of the research. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship

[10] Mathieu, V. 2001. Product services: from a service supporting the product to a service supporting the client. *Journal of Business & Industrial Marketing*, 16(1): p. 39-61. DOI= http://dx.doi.org/10.1108/08858620110364873

[11] Neu, W.A. and Brown, S.W. 2005. Forming successful business-to-business services in goods-dominant firms. Journal of Service Research, 8(1): p. 3-17. DOI= http://dx.doi.org/10.1177/1094670505276619

[12] da Silveira, G.J.C. 2005. Market priorities, manufacturing configuration, and business performance: an empirical analysis of the order-winners framework. *Journal of Operations Management*, 23(6): p. 662-675. DOI= http://dx.doi.org/10.1016/j.jom.2005.01.005.

[13] Zatzick, C.D., Moliterno, T.P. and Fang, T. 2012. Strategic (MIS) FIT: The Implementation of TQM in Manufacturing Organizations. *Strategic Management Journal*, Forthcoming. DOI= http://dx.doi.org/10.1002/smj.1988

The Influence of Service Discrimination on the Second Visit of Non-shopping Customers

Dong LI
School of management, Harbin Institute of technology, Heilongjiang, China
+86-0451-86414042
sxlidong@yahoo.com.cn

Zhe LI
School of management, Harbin Institute of technology, Heilongjiang, China
+86-0451-86414042
lizhe.hit@gmail.com

ABSTRACT
With the constantly rising of the service industry, the importance of service quality and customer's satisfaction are approved by the academia. The academia had researched many results in the relationship between the customer's positive mood and the level of the satisfaction of customers, the customer's intentions of shopping again. But the research of the academia is very seldom in the relationship between the customer's negative mood and the level of the satisfaction of customers and the customer's intentions of shopping again. This thesis mainly use literature research method and experimental analogic method to analyze the problems above, and mainly use SPSS software to carry on the statistical analysis.

Categories and Subject Descriptors
J.1 [**Administrative Data Processing**] Type of application*s* – *marketing, business.*

General terms
Management, Performance.

Keywords
service discrimination, non-shopping customers, second visit

1. Introduction
Among the customers who are visiting stores (either traditional stores or online stores), only a small part of them will transact with the stores and become the real purchasers, while the vast majority are non-shopping customers who are just "looking around" without clear shopping intentions. As their relationship with merchants is not as close as that of the shoppers with merchants, this kind of customer relationship is not obvious, and difficult to build. However, they are important to the development of the stores, for their number is large and potential is strong. Therefore, in order to help the stores obtain broader market space, it is necessary to conduct good customer relationship management on them.

Under normal circumstances, researchers conduct analysis according to the data from the purchasers, but ignore to observe the groups who have not purchased (Siddhartha Chib et al, 2004) [1]. Some scholars call the people who have purchase interest or intentions but have not purchased products as potential customers (Kim, 2009 [2]; Luo Jing, 2009 [3]) for observation and study. However, the non-shopping customers are not equivalent to general potential customers. Strictly speaking, it refers to the customers who visit stores but do not buy anything, and a considerable number of them even do not have clear shopping intentions. The characteristics of this kind of customers still need in-depth study, not only in the theoretical study but also the practical application. In practice, many merchants tend to attach great importance to purchasers and take various measures to provide warm, thoughtful and meticulous service for them, but for non-shopping customers who have no purchase intentions or behaviors, they seem to ignore them and seldom offer services for them.

To solve the above problems, in-depth research is needed on the basis of existing service research achievements. Moreover, it is necessary to guide stores to take different and targeted service designs for shoppers and non-shopping customers, and especially offer right service for the non-shopping customers, which will have practical guiding significance on the service marketing of the stores.

2. Theoretical Background

2.1 Relevant research on service discrimination
Liu Zhibiao (2002) believes that the discrimination strategy in the banking industry is widely applied as an operating mode in the corporate practice in developed countries [4]. The study of Thomas S.F. Chan (1999) shows that in the U.S. housing insurance market, compared with the price guarantee, merchants prefer to make quality guarantee, for quality guarantee is abstract and not concrete [5]. After comparing the housing insurance quality provided in the two regions with a large number and a small number of minorities, it is found that the housing insurance quality provided in the region with a small number of minorities is higher than that with a large number of minorities.

2.2 Relevant research on non-shopping customers
Studies related to the services for non-shopping customers are mainly in the research of potential customers and pre-sales service (Jin Liyin, 2006 [6]; Kim et al., 2009 [2]). In the service encounter, the impacts of the service quality are different on existing cus-

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.23

tomers and potential customers. Cronin and Taylor (1992) think that in most cases, customers will not form accurate expectations for service quality before purchasing services. According to the concept of "expected satisfaction" proposed by Simonson et al. (1992), potential customers have expected satisfaction conviction before purchasing services [7].

2.3 Relevant research on service encounter

The narrow meaning of service encounter refers to the face-to-face interaction between service providers and recipients in service situation; while the generalized service encounter means all the interactions (including hardware facilities and all other tangible objects) between customers and service enterprises [8]. The study of Jin Liyin (2008) shows that in the service encounter, the improper language communication of employees may induce the negative emotions of customers, and other non-verbal communication factors like manners and auxiliary languages all have significant impacts on the positive and negative emotional reactions of customers. Moreover, the physical appearance, dress and clothing of employees are more like a "motivating factor" affecting the emotional reactions of customers. Though the lack of this aspect will not cause too much negative impact on the customer sentiment, it plays an active role in promoting the formation of positive customer sentiment [9].

3. Research hypothesis

Researches at home and abroad have proven that in order to win customer recognition, a sincere heart and friendly attitude is needed for customers. If a customer suffers language complaints for he only tries the products without buy anything in the store, the psychological dissatisfaction will prompt him to go shopping in the stores of its competitors, and he will not come to this store the next time. It is more likely that he will do "word-of-mouth publicity" later so as to vent his dissatisfaction while reducing the possibility of surrounding consumers to go to that store. In addition, verbal communication may induce the negative emotions of the customers. While some body language and nonverbal communication such as the appearance and clothing of the employees have important impact in promoting the positive emotions of the customers, although the lack of this aspect will not have a negative impact on customer emotions (Jin Liyin, 2008). Neat appearance as well as unified and decent clothing of employees is likely to leave a good impression to consumers [10]. In summary, hypothesis 1 can be concluded:

Hypothesis1: service discrimination reduces the "second visit" willingness of non-shopping customers.

The mental endurance and stress resistance of people varies with their different ages, genders and education degrees. Therefore, when consumers suffer discrimination service as non-shopping customers in the shopping process, some consumers will take it for granted, while others will be very angry. Thus, different consumers would have different reactions when suffered the same service discrimination; moreover, the same consumer would have different reactions when suffered different service discrimination. Hence hypothesis 2 can be concluded:

Hypothesis 2: different service discrimination has different impact on the "second visit" willingness of non-shopping customers.

4. Research Design

4.1 Classification of service discrimination

This paper takes college students as the research object, and conducts definition and classification of service discrimination through interviews. The main interview content is to take about the discriminative behaviors they encountered in the shopping process. After interviews, the interview contents can be sorted as in Table 1:

Table 1 Interview contents

Classification No.	Performances of service discrimination
1	Reluctant to allow customers to try products or try on clothes; After learning the consumers will not buy, they no longer let them try; Repellent to customers' trying; Customers are not allowed to touch products.
2	After learning the consumers just try will not buy, sellers' attitudes change; After learning the selling price is low, the enthusiasm of sellers falls; The attitude is too indifferent; Ignore potential customers; no attentive service; seldom answer your questions or help you find the products you are looking for; After learning that I'm just looking around, the attitude of sellers is bad and indifferent; Indifferent to the non-shopping customers
3	After learning the customers will not buy, the service attitude becomes indifferent, and sometimes slightly contemptuous; After learning the customers will not buy, sellers even rudely shout, "Don't come if you have no money" and alike.
4	After learning the customers will not buy, when I ask about prices, the salesman replies impatiently, or even does not answer me; When I inquiries, the answer is brief and rough.

Summarize the discriminative behaviors the students encountered in the shopping process and conclude the features of each classification, the service discrimination can be divided into the following four categories: repellent service discrimination, indifferent service discrimination, complaining service discrimination and preserving service discrimination.

4.2 Design of service scenes

Targeted for the above four discrimination services and combined with related specific performances, the following four service scenes are designed:

Scene design for repellent service discrimination:

Scene: you find a nice dress. When you want to take it over for further understanding, you find a sign on the clothes reading: "Don't touch if not buy it".

Scene design for indifferent service discrimination:

Scene: you walk into a store. The salesman asks enthusiastically: "What do you want to buy?" If you answer: "Nothing. I'm just looking around.", she will no longer care about you.

Scene design for complaining service discrimination:

Scene: After trying one or two nice-looking clothes, you do not buy it after you take it off. The salesman who serves you will complain: "Why did you try them since you don't buy them"

Scene design for preserving service discrimination:

Scene: you find a nice-looking dress. When you want to see more colors or styles, the salesman will reply: "Will you buy it today? If you do not buy it today, I'll show you when you plan to buy it." When you ask if the price can be lower, the salesman will reply: "Let's discuss about it when you plan to buy it."

The questions designed in the four scenes are the same. The customer satisfaction degree and second visit chance when consumers face the above four scenes of service discrimination, and apply 5-grade satisfaction questionnaire for the experimental survey.

4.3 Setting of experimental control group

In order to ensure the completeness and accuracy of experimental results as well as the reliability and authenticity of experimental data, this experiment designs experimental control groups. There are four sets of questionnaires in total (see attachment), comparing each other in pairs. Questionnaire 1 and questionnaire 3 as well as questionnaire 2 and questionnaire 4 are mutually experimental control groups. The service discrimination for non-shopping customers and the customer satisfaction degree of non-shopping customers who still receive warm service and the second visit chance. Questionnaires 2 and 4 are designed to solve sequence errors. For the convenience of investigation, the questions in questionnaires 2 and 4 just changed the question order rather than fully list the 24 items.

4.4 Selection of experimental objects

The survey mainly adopts the simulation method, relies on questionnaires, takes college students as the main survey and research objects, and applies posttest control group design method for grouping experimental investigation. Not all designs need to be measured before, but the posttest is a must to determine the effect of the experimental processing.

4.5 Experimental time control

Four experiments are needed, and the experimental time is focused between the end of October to the beginning of November in 2012. The number of experimental objects every time is 120. Simulate the scenes in questionnaire form and collect the questionnaires. Finally, 480 questionnaires are handed out and 480 copies are collected, 463 of which are valid questionnaires.

5. Data analysis

5.1 The relationship between service discrimination and the second visit willingness of non-shopping customers

The experimental group is marked as group 1, and the control group is marked as group 2. Independent samples T-Test is utilized to test the second visit willingness of customers received service discrimination and normal service. The results are shown in Table 2:

Table 2 Independent Samples Test

		Sig.	Sig. (2-tailed)	Std. Error Difference
Second visit willingness 1	Equal variances assumed	.236	.000	.10735
Second visit willingness 2	Equal variances assumed	.349	.000	.11854
Second visit willingness 3	Equal variances assumed	.144	.000	.09965
Second visit willingness 4	Equal variances assumed	.691	.000	.10045

As can be seen from the results of the above table, no matter variance homogeneity or not, the mean values between the two groups all have significant difference (sig value= 0.000). Then compare the mean values of each group as shown in Table 3:

Table 3 Group Statistics

	VAR00001	N	Mean	Std. Deviation	Std. Error Mean
Second visit willingness 1	1	228	2.3421	1.12504	.07451
	2	235	3.7191	1.18297	.07717
Second visit willingness 2	1	228	2.9342	1.36671	.09051
	2	235	3.7362	1.17968	.07695
Second visit willingness 3	1	228	1.6711	1.12310	.07438
	2	235	4.0809	1.02002	.06654
Second visit willingness 4	1	228	1.6886	1.09241	.07235
	2	235	3.3489	1.06896	.06973

As can be seen from the table, no matter in which scene, the mean values of the experimental groups are all significantly lower than those of control groups, that is, the second visit willingness of non-shopping customers who felt service discrimination will greatly reduce, which verifies the above hypothesis 1: service discrimination reduces the "second visit" willingness of non-shopping customers.

5.2 The influence relationship of different service discrimination on the second visit willingness of non-shopping customers

One-sample T-test way is applied (only for experimental groups), and the results are shown in Table 4 and Table 5.

Table 4 One-sample Statistics

	N	Mean	Std. Deviation	Std. error mean
Second visit willingness 1	228	2.3421	1.12504	.07451
Second visit willingness 2	228	2.9342	1.36671	.09051

Second visit willingness 3	228	1.6711	1.12310	.07438
Second visit willingness 4	228	1.6886	1.09241	.07235

Table 5 One-sample Test

	t	df	Sig. (2-tailed)	Mean Difference
Second visit willingness 1	18.013	227	.000	1.34211
Second visit willingness 2	21.370	227	.000	1.93421
Second visit willingness 3	9.022	227	.000	.67105
Second visit willingness 4	9.518	227	.000	.68860

As can be seen from the above two tables, the influence of different service discrimination on the second visit willingness of non-shopping customers is different. As for the latter two types of service discrimination (complaining service discrimination and preserving service discrimination), the second visit willingness of customers is lowest; for the first service discrimination (repellent service discrimination) , the second visit willingness of customers is in the middle; and for the second service discrimination (indifferent service discrimination), the second visit willingness of customers is highest. The willingness order of non-shopping customers is (from high to low):

indifferent service discrimination, repellent service discrimination, preserving service discrimination, and complaining service discrimination

It can be concluded: there is a positive relationship between customer satisfaction and the second visit willingness of non-shopping customers; it also verifies the above hypothesis 2: different service discrimination has different impact on the "second visit" willingness of non-shopping customers.

5.3 Experimental analysis of service discrimination on the psychological impact of non-shopping customers

When consumers feel unfair, they will get angry. In the experiments of this paper, this aspect is also investigated. One-sample T-test is adopted to analyze the experimental groups. The results are shown in Table 6 and Table 7:

Table 6 One-sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Anger degree 1	228	2.6228	1.30664	.08653
Anger degree 2	228	2.1886	1.26437	.08374
Anger degree 3	228	3.7851	1.25314	.08299
Anger degree 4	228	3.6535	1.25517	.08313

Table 7 One-sample Test

	t	df	Sig. (2-tailed)	Mean Difference
Anger degree 1	18.753	227	.000	1.62281
Anger degree 2	14.195	227	.000	1.18860
Anger degree 3	33.559	227	.000	2.78509
Anger degree 4	31.922	227	.000	2.65351

As can be seen from the above two tables, various service discrimination types all bring customers a certain degree of negative emotional reaction (anger). Among them, the reaction of the latter two service discrimination types (complaining service discrimination and preserving service discrimination) are strongest, the first service discrimination (repellent service discrimination) lies in the middle, and the second service discrimination (indifferent service discrimination) is the weakest.

5.4 Experimental analysis on the influence of customer negative emotions on the second visit willingness of non-shopping customers

According to the data obtained in this experiment, the regression analysis is conducted in the relationship between customer negative emotional reaction (anger) and the second visit willingness of non-shopping customers under the four types of scenes. The results are shown in Table 8:

Table 8 Regression analysis

Model		Beta	Sig.
1	Anger degree 1	-.364	.000
2	Anger degree 2	-.495	.000
3	Anger degree 3	-.266	.000
4	Anger degree 4	-.240	.000

a. dependent variable: second visit chance

As can be seen from the above table, the negative emotional reaction (angry) has a significant negative impact on the second visit willingness of the non-shopping customers. Therefore, it can be concluded that service discrimination will bring customer negative emotional reaction (anger), which will reduce the second visit willingness of customers. For the latter two types (complaining service discrimination and preserving service discrimination) where customer negative emotional reaction is the strongest, the second visit willingness of customers is also lowest. The first type (repellent service discrimination) lies in the middle, while in the second type (indifferent service discrimination), the customer negative reaction is the weakest, and their second visit willingness of customers is highest.

6. Suggestions for the non-shopping service encounter

6.1 Suggestions for repellent service discrimination

From the above analysis, it can be seen that repellent service discrimination is an approach mainly adopted by merchants for non-shopping customers. The customer satisfaction of non-shopping customers for repellent service discrimination ranks the second of all types, which indicates that some consumers are inclusive to such discrimination. The merchants may want to remind consumers for the protection of precious and fragile products, and its starting point is understandable. However, the practical effect of this practice is similar to keep people away. The world today is no longer the era of "good wine needs no bush". Merchants should take advantage of all reasonable legal means to attract consumers. For "Don't touch if not buy it" and "Try it before you buy", it is self-evident that which will be more favored by customers.

6.2 Suggestions for indifferent service discrimination

The customer satisfaction of non-shopping customers for indifferent service discrimination ranks the first of all types, which indicates consumers do not hold entirely negative attitude to such service discrimination. The mean value of second visit willingness of non-shopping customers for indifferent service discrimination in the experimental groups is 2.9342. For the corresponding questions in the control groups, the mean value of satisfaction is 3.7362. Although there is a gap, the difference is not huge. Therefore, it can be concluded that merchants need not spare much labor to serve non-shopping customers. When there are few people in the store, the salesman needs to observe the non-shopping consumers carefully, and conduct proper introduction for the products they are slightly interested in. If the salesman is able to catch the attention of consumers, he can do further recommendation. However, when there are many people in the store especially during holidays, salesmen shall pay no attention to the non-shopping customers, but mainly care the consumers have clear shopping intentions.

6.3 Suggestions for complaining service discrimination

The customer satisfaction of non-shopping customers for complaining service discrimination ranks the first of all types, which fully shows that consumers are antagonistic and discontent for complaining service discrimination. The mean value of second visit willingness of non-shopping customers for complaining service discrimination in the experimental groups is 1.6711. For the corresponding questions in the control groups, the mean value of satisfaction is 4.0809. It can be seen that the difference is very large, and analyzed that consumers hate complaining service discrimination. This phenomenon must be resolutely stopped. Therefore, merchants must strengthen the quality education of salesmen. While improving their business level, their quality level must be attached great importance. Although there will be mistakes inevitably, taking the initiative to admit mistakes can still be accepted by consumers.

6.4 Suggestions for preserving service discrimination

The customer satisfaction of non-shopping customers for preserving service discrimination ranks the third of all types, which indicates that customers are basically unacceptable for preserving service discrimination. The mean value of second visit willingness of non-shopping customers for preserving service discrimination in the experimental groups is 1.6886. For the corresponding questions in the control groups, the mean value of satisfaction is 3.3489. Therefore, stores should strengthen staff training not only for enhancing their business level, but also for improving their service awareness. Even if the customers have not bought anything, they shall not leave bad impressions to consumers. The stores can also set customer feedback forms or complaint phones. If there are unexpected contradictions, try to solve them in the stores.

7. Conclusion

Taking the documentary research and simulation experimental as the research approach and college students as the research object, this paper classifies service discrimination and conduct in-depth analysis of service discrimination influence on the second visit willingness and psychological factors of non-shopping customers. The research results show that: 1) service discrimination reduces the second visit willingness of non-shopping customers; 2) different service discrimination has different impact on the "second visit" willingness of non-shopping customers; 3) various service discrimination types will bring customers a certain degree of negative emotional reaction (anger); 4) there is a negative correlation among customer negative emotions, customer satisfaction degree and service quality evaluation.

References

[1] Xiaodong Liu, and Wanquan Liu. 2005, Credit Rating Analysis with AFS Fuzzy Logic. *Advances in Natural Computation,* Vol 3612:pp. 1198-1204. DOI= http://dx.doi.org/10.1007/11539902_152

[2] Kim, Hee-Woong, Gupta and Sumeet. 2009, A comparison of purchase decision calculus between potential and repeat customers of an online store. *Decision Support Systems*, 47 (4): p.477-487.DOI= http://dx.doi.org/10.1016/j.dss.2009.04.014

[3] Woo Gon Kim, Xiaojing Ma, and Dong Jin Kim. 2006, Determinants of Chinese hotel customers' e-satisfaction and purchase intentions. *Tourism Management*, 27(5): p. 890-900. DOI= http://dx.doi.org/10.1016/j.tourman.2005.05.010

[4] Angela V. Hausmana, and Jeffrey Sam Siekpe. 2009, The effect of web interface features on consumer online purchase intentions. *Journal of Business Research*, 62(1): p. 5-13. DOI= http://dx.doi.org/10.1016/j.jbusres.2008.01.018,

[5] Gregory D. Squires. 2003. Racial Profiling, Insurance Style: Insurance Redlining and the Uneven Development of Metropolitan Areas. *Journal of Urban Affairs*, 25(4): p. 391-410. DOI= http://dx.doi.org/10.1111/1467-9906.t01-1-00168

[6] Tor Wallin Andreassen, and Bodil Lindestad, 1998, Customer loyalty and complex services: The impact of corporate image on quality, customer satisfaction and loyalty for customers with varying degrees of service expertise. I*nterna-*

tional Journal of Service Industry Management, 9(1): p.7-23. DOI= http://dx.doi.org/10.1108/09564239810199923

[7] Lisa D. Ordóñez, Terry Connolly, 2000. Regret and Responsibility: A Reply to Zeelenberg et al. (1998). Qrganizational Behavior and Human Decision Processes, 81(1): p. 132-142. DOI= http://dx.doi.org/10.1006/obhd.1999.2834

[8] Simintiras Antonis, Adamantios Diamantopoulos, and Judith Ferriday. 1997, Pre - purchase Satisfaction First - Time Buyer Behavior : Some Preliminary Evidence. *European Journal of Marketing*, 31 (11/ 12):p. 857 - 872. DOI= http://dx.doi.org/10.1108/03090569710190578

[9] Cedric Hsi-Jui Wu. 2007, The impact of customer-to-customer interaction and customer homogeneity on customer satisfaction in tourism service—The service encounter prospective. Tourism Management. 28(6): p. 1518-1528. DOI= http://dx.doi.org/10.1016/j.tourman.2007.02.002

[10] Chung-Tzer Liu, Yi Maggie Guo, and Chia-Hui Lee. 2011, The effects of relationship quality and switching barriers on customer loyalty. *International Journal of Information Management*, 31(1): p. 71-79. DOI= http://dx.doi.org/10.1016/j.ijinfomgt.2010.05.008

CONCEPTUALISING THE EVALUATION OF HRIS IN PUBLIC SECTOR

Prof. Dr. Junaidah Hashim
Dept of Business Administration
KENMS
Inter'l Islamic Univ Malaysia
Jalan Gombak, 53100 K.L
MALAYSIA
junaidahh@iium.edu.my

Faizal Haji Zainuddin
Dept of Business Administration
KENMS
Inter'l Islamic Univ Malaysia
Jalan Gombak, 53100 K. L
MALAYSIA
nouvomas@hotmail.com

Prof. Dr. Saodah Wok
Dept of Communication
KIRKHS
Inter'l Islamic Univ Malaysia
Jalan Gombak, 53100 K.L
MALAYSIA
wsaodah@iium.edu.my

ABSTRACT
This conceptual paper proposes a theoretical framework in examining the reaction of employees in the Malaysian public sector towards HRIS, and to examine the relationship of HRIS use with the organization performance. In Malaysia, HRIS is one of the pilot projects under the Electronic Government Flagship. Pertinent literature is reviewed and a few hypotheses are proposed.

Categories and Subject Descriptors
[Human centered computing]: Human capital interaction - Collaborative and social computing systems and tools, Social networking sites

General Terms
Management

Keywords
HRIS, Malaysia, Public sector

1. Background of Study

The increasing importance of IT has led many organizations to integrate IT into their daily operations with such purposes as improving customer services, reducing operational costs, improving production quality, increasing profits, and expanding market share (Byrd *et al*., 2006; Dardan *et al*., 2006/2007; Liu & Tsai, 2007; Merono-Cerdan, 2008; Rivard *et al*., 2006). Organizations have invested heavily in IT for operational excellence, new products, services, and business models, customer and supplier intimacy, improved decision-making, competitive advantage, and last but not the least for survival. As a growing essential service, IT is an essential component to all structured organizations including all levels of government (Streib & Willoughby, 2005). Due to this, it also has driven a demand in technology consumption by the public sector organizations. One of the key methods is to adopt methods of service delivery utilizing IT in its HRM.

However many users in organizations are underutilizing information systems (computer applications implemented to improve business efficiency and productivity) or are not using information systems in the ways intended, resulting in failure of organizations to achieve optimal performance gains from investments in IT (Allison, 2003; Almutairi, 2007; Jasperson, Carter & Zmud, 2005). According to Al-Gahtani (2004), difficulties in transferring IT into practice have been experienced in many organizations, and which seem to be worse in developing countries.

The Government of Malaysia has invested a lot over the last 10 years transform the public sector towards achieving greater productivity and efficiency through the information system in the HRM. Thus, his study aims to examine the reaction of employees in the Malaysian public sector toward HRIS, and to examine the relationship of HRIS use with the organization performance.

1.1 HRIS in Malaysian Public Sector

In Malaysia, HRIS is one of the pilot projects under the Electronic Government Flagship. The HRIS is an important initiative undertaken by the Government. It has been developed using local resources and expertise and not based on any „off the shelf" human resource application packages available in the market. The project was finalized in 1998 when Multi-Media Resources Corporation Berhad (MMRCB) was selected after a process of multi-track negotiation with several short-listed companies and evaluation of their proposals.

HRIS objectives is to provide a single interface that enables officers and staff in government sectors ministries and agencies to perform their human resource functions more efficiently and productively within an integrated environment. The project covers functions that are knowledge-based as well as those related to operations and management. It encompasses all aspects of human resource development covering the entire spectrum from the time an officer is appointed into the Civil Service until he retires. Among the management functions covered under the system are organizational development, rightsizing, and formulation of schemes of services, human resource planning, service matters and employer-employee relations. With the introduction of HRIS, all human resource elements related to the Malaysian Civil Service can be planned in a systematic and comprehensive manner. This is because HRIS will provide easy access to up-to-date data and information. It also offers the use of

Research Notes in Information Science (RNIS)
Volume13, May 2013
doi:10.4156/rnis.vol13.24

several techniques to generate and evaluate various alternative human resource scenarios. HRIS will also assist public sector personnel in various aspects of their daily operation such as recruitment, deployment, confirmation of appointment, training, promotion, pensions and welfare of workers. Eventually, this will involve work processes related to salary, leave, loans, asset declaration, medical administration, disciplinary action, consultation, career guidance and trade union claims.

The full implementation of HRIS will ensure more effective distribution of manpower among agencies. HRIS will enable the Public Service Department (PSD) and other departments to monitor the utilization of human resource in every unit and agency more accurately. This will lead to effective planning whereby any excess manpower can be redeployed to agencies that are facing a shortage. The Public Service of Malaysia will be able to perform better when the capability and skills of its workforce meet the relevant job specification and requirements thus result in productivity. HRIS will also allow management to identify suitable personnel to fill vacancies by matching personnel profiles with job profiles thereby enabling public personnel to deliver their best for the nation.

2. Review of Literature

2.1 Technology Acceptance Model (TAM)

There is a growing body of research that seeks to determine the constructs that lead to the acceptance of IT and ultimately its use (Gefen *et al.,* 2003; Lai & Li, 2005; Legris *et al.,* 2003; Lu *et al*., 2005; Venkatesh *et al*., 2003; Wang *et al.,* 2003). Among the many models, TAM is the classical model and appears to be the most widely accepted among IS researchers (Davis, 1989; Davis *et al.,* 1989; Legris *et al.,* 2003) The TAM which was developed by Davis (1989) as a tool to assess user acceptance of technology, and this theoretical model continues to be widely used by researchers (Alshare, Freeze & Kwun, 2009; Baker-Eveleth, Eveleth, O'Neill, & Stone, 2006; Davis & Wong, 2007; Ha & Stoel, 2009; Lin, 2008; Lin & Chou, 2009).

TAM postulates that an individual‟s intention to engage in the use of an IT is influenced by perceived usefulness (PU) and perceived ease-of-use (PEOU) of the system (Davis, 1989; Davis *et al*., 1989; Legris *et al.,* 2003). Researchers have used TAM to examine the possible antecedents of PU and PEOU: such as CSE, perceived risk, training, prior use, and similar experiences (Chan & Lu, 2004; Gefen *et al.,* 2003; Legris *et al*.; Lu *et al.,* 2005; Venkatesh, 2000; Venkatesh & Davis, 2000; Wang *et al.,* 2003). TAM does not include TRA‟s subjective norm as a determinant of behavioral intention. As Fishbein and Ajzen (1975) acknowledged "subjective norm is one of the least understood aspects of TRA" (p. 304). Therefore, due to its uncertain theoretical and psychometric status, subjective norm was not included in TAM (Davis *et al.,* 1989).

While the classical TAM provides a solid framework for determining behavioral intentions to use IS, there are several limitations to TAM (Legris *et al*., 2003) as well as several variations and extensions to the model (Adams et al., 1992; Jackson *et al.,* 1997; Vankatesh & Davis, 1996). The majority of TAM studies involved students as participants using automation software or systems development applications and the measurements reflect the variance in self reported use (Legris *et al*., 2003). Researchers believe that better results could be realized if the TAM processes are carried out in a business environment using business professionals or real customers as participants as well as using business process applications (Legris *et al.,*2003*).* Davis *et al.* (1989), in a foundation paper on the TAM, noted that there are various external factors such as individual characteristics, situational constraints, and managerial controlled interventions that impinge behavior, and would therefore be determinants of perceived usefulness and perceived ease of use, which were both determined to be antecedents of intention to use IT. Many researchers have therefore developed several extended TAMs with variables relating to individual and organizational factors and used the extended theoretical models to assess the adoption and usage of various technology innovations.

2.2 Technology Acceptance Model 2 (TAM2)

TAM2 is an extension of the classical TAM and includes social influence processes and cognitive instrumental processes as determinants of PU and usage intentions (Legris *et al.*, 2003; Vankatesh & Davis, 2000). TAM2 reflects the impact of three social influence processes and four cognitive instrumental processes as impinging on PU and ultimately an individual‟s IU or reject a new IS. The three social influence processes are: subjective norms, voluntariness, and image; while the four cognitive instrumental processes are: job relevance, output quality, result demonstrability, and perceived ease of use. Figure 2 gives a graphic overview of TAM2.

TAM2 theorizes that, "in a computer usage context, the direct compliance-based effect of subjective norm on intention over and above PU and PEOU will occur in mandatory, but not voluntary, system usage settings" (Venkatesh & Davis, 2000, p. 188). In TAM2, voluntariness is, therefore, shown as a moderating variable. Venkatesh and Davis (2000) define image as "the degree to which use of an innovation is perceived to enhance one‟s status in one‟s social system" (p. 189). Individuals, therefore, will react in accordance to subjective norms in an effort to maintain a favorable image in that social group. TAM2 postulates that subjective norm will positively influence image (Vankatesh & Davis, 2000). Experience refers to an individual‟s knowledge and beliefs about a system and is developed through repeated use of that system. TAM2 theorizes that intentions to use IS will change over time based on direct experience acquired through ongoing usage. Additionally, experience can be viewed as a moderating variable in the TAM2 construct.

Cognitive instrumental processes in TAM2 represent a series of determinants of PU and include: job relevance, output quality, result demonstrability, and PEOU (Vankatesh & Davis, 2000). Job relevance refers to the degree to which an individual perceived the task as applicable to his or her job while output quality refers to an individual‟s perception of how well the system performs a specific task (Vankatesh & Davis, 2000). Result demonstrability looks at how tangible the results to job performance are in using an IS. PEOU examines how easy or effortless a system is to use. TAM2 postulates that all cognitive instrumental processes will positively impact PU and ultimately an individual‟s IU and IS (Venkatesh & Davis). Both TAM2 and the classical TAM have been frequently used to predict an individual‟s use of IS; however, the TAM2 and TAM explain only 40% of system‟s use (Legris *et al.,* 2003). Therefore, neither TAM nor TAM2 appears superior in explaining actual system‟s use (Legris *et al.,* 2003). This study then will be proposing that productivity and efficiency factors in TAM may yield a model

that allows for a more precise understanding of the role of these constructs in explaining HRIS use in HR departments/unit in the context of Malaysian Public Service. This is due to that IT researchers believe that to explain consistently more than 40% of system use, TAM should be modified to include other constructs and be extended to include other variables related to human and social change processes (Legris *et al.,* 2003*).*

2.3 Hypothesized Model

This study will be using variables from a model called TAM2 by Venkatesh and Davis (2000). Originally, TAM2 is a well-tested model for technology acceptance called the *technology acceptance model* (TAM) which was suggested by Davis (1989). Davis had three variables attached to TAM: two independent variables (Perceived ease of use, Perceived usefulness) and a dependent variable (Intention to use the system that leads to usage behavior). The two independent variables from the TAM model will be used as the principle independent variables for this study.

It was argued that intention to use the system leads to usage behavior. The behavior intention of the sample in this comparative study is to interact with the HRIS system. Hiltz and Johnson (1990) said that the interaction with systems affects the attitude of the user. Wixom and Todd (2005) said that the satisfaction is derived from the user‟s feelings or attitudes about the system characteristics. As Venkatesh and Davis (2000) presented a theoretical extension for TAM called TAM2, this extended theory uses social influence and cognitive instrumental process to explain perceived usefulness and usage intentions. Therefore, such extension is vital for this dissertation in order to understand comprehensively on the sample intention to use the HRIS system. The variables that explain social influential processes include subjective norm whereas variables that explain cognitive instrumental processes include job relevance, output quality, result demonstrability, and perceived ease of use.

2.3.1 Perceived Usefulness

Prior research has shown that perceived usefulness (PU) is a major determinant of user acceptance or it has a positive effect on system use (Agarwal & Karahanna, 2000; Davis e*t al.,* 1989; Gefen et *al.,* 2003; Straub, Limayem, & Karahanna- Evaristo, 1995). Perceived usefulness is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" (Davis, 1989, p.320). According to the result of Davis‟s study, perceived usefulness is the main determinant of user technology acceptance. Davis assessed that if an information system is high in user perceived usefulness, the system is one for which a user believes that using it could have positive influence in job performance.

According to the motivation theory, the influence of PU on system use can be explained in that an individual is inclined to accept a new information system when he perceives it to be instrumental for achieving valued outcomes. Most technology acceptance studies use perceived usefulness, as the major independent construct to determine users‟ intentions toward technology adoption. This is also later supported by Bandura (1982) posits that behavior is partially determined by ones' perception of the consequences of his or her actions. Studies have shown PU influenced adoption behaviors (Davis, 1989; Szajna, 1996; Gong *et al.,* 2004; Venkatesh & Davis, 2000). Besides that many empirical studies have found that perceived usefulness is an important determinant of intention to use and also of attitude (Colvin & Goh, 2005; Premkumar & Bhattacherjee, in press; Vankatesh & Davis, 2000). Therefore, perceive usefulness could be related to effectiveness on the job, to more productivity at work, such as consuming less time or money, and to relative motivation for usage of that particular technology. In this study, perceived usefulness refers to the concept that HRIS is perceived as useful. Usefulness has been tested relative to the system‟s ability to increase performance, productivity, and effectiveness. Thus, the following hypothesis was developed.

H1: Perceived usefulness will have a positive effect on intention of using HRIS

2.3.2 Perceived Ease of Use

According to Davis (1989), EOU is, "the degree to which a person believes that using a particular system would be free of effort" (p. 320). Davis's assumption is that, other factors being equal, "an application perceived to be easier to use than another is more likely to be accepted by others" (p. 320). EOU has two important roles within the TAM framework – it directly affects attitude and it impacts attitude and intention to use via PU. Researchers have shown EOU as an important factor in adoption behaviors (Venkatesh, 2000; Davis, 1989; Gong *et al.,* 2004; Davis *et al.,* 1989; Venkatesh & Morris, 2000). This easiness includes mental and physical effort, especially in the learning phase. In this study, this variable can be defined as HR practitioners‟ perceptions that their usage of the HRMIS is effort-free.

Consistent with TAM and later TAM2, perceived ease of use has an effect on both intention to use and perceived usefulness, though some studies found that perceived ease of use has no influence on intention to use, since they omitted the attitude factor in their models (Davis, 1989; Vankatesh & Davis, 2000).

H2A: Perceived Ease of Use has a positive influence on perceived usefulness of using HRIS

H2B: Perceived Ease of Use has a positive influence on intention of using HRIS

2.3.3 Result Demonstrability

Result demonstrability, defined by Moore and Benbasat (1991, p. 203) as the "tangibility of the results of using the innovation," is hypothesized to be cognitive instrumental processes affecting perceived usefulness. Venkatesh and Davis (2000) argued that people form perceived usefulness judgments in part by cognitively comparing what the system is capable of doing with what they can achieve from using the system. Users are expected to form more positive perceptions of the CRM system if the relationship between usage and positive results is readily discernable. In other words, "the more visible its [an innovation] advantages are…the more likely it is to be adopted" (Duncan, 1973, p. 39, as cited in Moore and Benbasat, 1991). It is assumed that in a small business environment, the effects from using CRM software are more readily visible to employees due to its simpler structure and fewer layers of hierarchy.

Empirically, Agarwal and Prasad (1997) applied Moore and Benbasat‟s (1991) theory of innovation adoption to study information technology acceptance and continued usage thereafter. Although result demonstrability is not a significant determinant of initial usage decision, it proves to be a major predictor of continued usage. The researchers concluded that "sustained use in the future is driven primarily by rational

consideration: that is, the benefits offered by an innovation to potential adopters as well as their ability to consciously recognize and articulate these benefits" (p. 570). Since this study examined HRIS system utilization after the initial usage decision, this construct is assumed to be of strong relevance.
H3: Result demonstrability will have relation on practitioners' perceived usefulness of using HRIS

2.3.4 Job Relevance
Job relevance is adopted from the extended TAM named TAM2 by Venkatesh and Davis (2000). It is defined as "an individual‟s perception regarding the degree to which the target system is applicable to his or her job" (p. 191). The four empirical tests of TAM2 demonstrated consistent effects of job relevance on users‟ perceptions of perceived usefulness.

Considering the restrained resources of small businesses, it is unlikely that they can afford the most advanced and sophisticated CRM package. Small businesses can end up implementing systems that fit the budgets rather than those that fit the job requirements of its users. Therefore, it is hypothesized that, for small businesses, job relevance has direct relationship on perceived usefulness. In addition, it is argued that in situations where usage is not voluntary, usefulness and utilization is more a result of how the technology fits than other attitudes of users toward using it (Goodhue & Thompson, 1995). Since the extended TAM focuses on HRIS usage in organizational settings where usage is a job function, the perception of the system‟s usefulness and the resulting level of utilization is assumed to be directly influenced by how the users view the system fits their job requirements or in other words, job relevance.
H4: Job relevance will have relation on practitioners' perceived usefulness of using HRIS

2.3.5 Output Quality
Output quality can be judged by evaluating the intermediate or end products of using a system (Davis *et al.*, 1992). Prior research has shown that the output quality has a significant effect on perceived usefulness (Davis *et al*., 1992). Several studies also examined perceptions of output quality as a potential determinant of information use (O‟Reilly, 1982; Zmud, 1978).

Venkatesh and Davis (2000) suggest that when a set of multiple relevant systems is available, one is likely to choose a system that delivers the highest output quality. All else being equal, increased output quality is likely to improve an individual‟s job performance, thus influencing his perception of usefulness. Although TAM2 does not imply a direct effect of output quality on PU, the relationship between output quality and PU are examined in the present study, considering possible influence of the quality of output produced by the system. With respect to output quality, in their TAM2 model, Venkatesh and Davis and quick access to information, the ability to produce a greater number and variety of HR-related reports, and reducing costs (Ngai and Wat, 2006) and hence result in the productivity of the human resources.

(2000) theorized that, over and above considerations of what tasks a system is capable of performing and the degree to which those tasks match their job goals (job relevance), people will take into consideration how well the system performs those tasks, referred to as *perceptions of output quality*. With respect to the HRMIS system, output quality is the quality of their performance in HRM related matters and works.
H5:Output Quality will have relation on HR practitioners' perceived usefulness of using HRIS

2.3.6 Demography Aspects
Since e-HRM mandatorily requires the inclusion of employees, qualification as well as acceptance of employees to individually adopt e-HRM is crucial. Following previous research, individual qualification and motivation seem to systematically vary with different demographical attributes, while in particular age, gender, and education may influence individual adoption. First, age is taken as relevant for individual adoption. In addition, gender seems to be of relevance, since research yields that females have less overall experience with IT and are more likely to have negative attitudes towards IT (Morris *et al.,* 2005; Zhang, 2005). Moreover, education is associated with central predictors of individual adoption such as IT anxiety, enjoyment and perceived usefulness (Zhang, 2005).
H6. The demography of an organization reveals an effect on the adoption of HRIS

2.3.7 HRIS as Enabler for Productivity and Efficiency in HRM
As human resource departments continue to move to internet or web-based technology, the fastest growing trend in the delivery of HR information is employee self-service. These applications give employees the ability to access and maintain their personal HR information via the web. Another growing trend is the adoption of managerial self-service (MSS) which provides managers access to a variety of HR tools and information via the web. Most manager HR-related tasks can be completed via MSS applications including pay administration/compensation, performance management, staffing, and employee development. Research has demonstrated the cost-effectiveness of using ESS, MSS, and HRIS. The common benefits of HRIS frequently cited in studies included improved accuracy, the provision of timely and quick access to information, and the saving of costs. Sadri and Chatterjee (2003) stated that when the HRIS function was computerized, faster decision making can be carried out on the development, planning, and administration of HR because data can be much easier to store, update, classify, and analyze. In addition, they noted that HRIS can strengthen an organization‟s character. Using advanced HRIS often leads to the provision of timely.

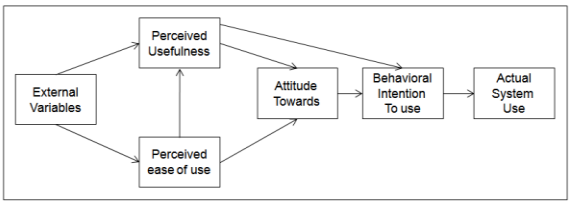


Figure 1: Original TAM (Davis, 1989)

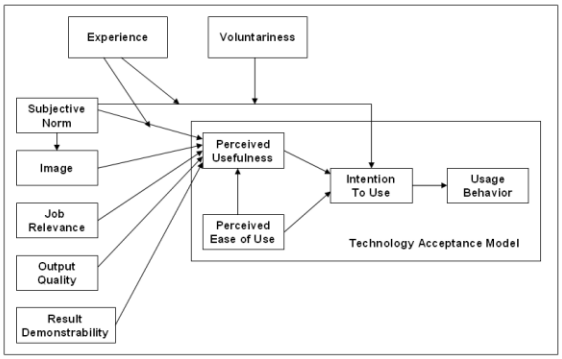


Figure 2: TAM2 (Vankatesh & Davis, 2000)

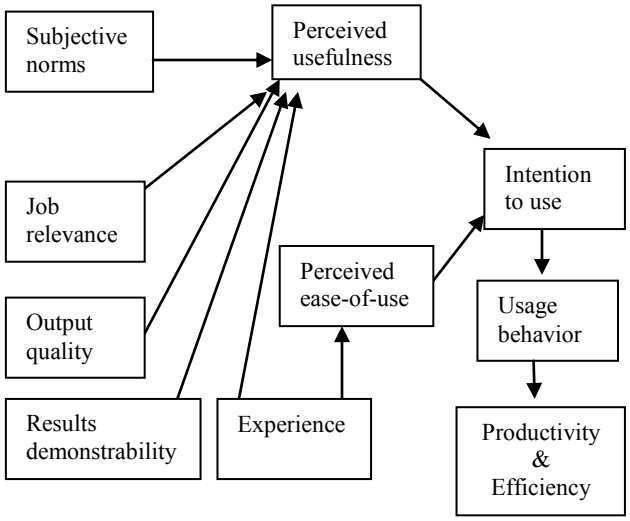


Figure 3: Hypothesized Model

References

[1] Adams, D. A., R. R. Nelson, P. A. Todd. 1992. Perceived usefulness, ease of use, and usage of information technology: Areplication. *MIS Quart.* 16(2) 227–250.

[2] Agarwal, R., and Prasad, J. (1997). The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. *Decision Sciences, 28*(3), 557-582.

[3] Agarwal, R., and Karahanna, E. (2000). Time flies when you're having fun:Cognitive absorption and beliefs about information technology usage. *MIS Quarterly, 24*(4), 665-694.

[4] Al-Gahtani, S. S. (2004). Computer technology acceptance success factors in Saudi Arabia: An exploratory study. *Journal of Global Information TechnologyManagement, 7*(1), 5-29.

[5] Allison, B. (2003). Are you maximizing your technology investments? *Orange CountyBusiness Journal, 26*(30), 5-6.

[6] Almutairi, H. (2007). Is the „Technology Acceptance Model" universally applicable?: The case of the Kuwaiti ministries. *Journal of Global Information Technology Management, 10*(2), 57-80.

[7] Alshare, K. A., and Alkhateeb, F. B. (2008). Predicting students usage of Internet emerging economies using an extended technology acceptance model (TAM*). Academy of Educational Leadership Journal, 12*(2), 109-128.

[8] Baker-Eveleth, L., Eveleth, D. M., O'Neill, M., and Stone, R.W. (2006). Enabling laptop exams using secure software: Applying the technology acceptance model. *Journal of Information Systems Education, 17*(4), 413-420.

[9] Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist,37*(2), 122-147.

[10] Benbasat, I. and Moore, G. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, 2(3), 192-222.

[11] Bhattacherjee, A., and Premkumar, G. (2004, June). Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. *MIS Quarterly*, *28*(2), 229–254.

[12] Byrd, T. A., Lewis, R. L., and Bradley, R. V. (2006). IS infrastructure: The influence of senior IT leadership and strategic information systems planning. *The Journal of Computer Information Systems, 47*(1), 101-113.

[13] Chan, S., and Lu, M. (2004). Understanding Internet banking adoption and use behavior: A Hong Kong perspective. *Journal of Global Information Management, 12*(3), 21-43.

[14] Colvin, C.A. and Goh, A. (2005). Validation of the technology acceptance model for police. *Journal of Criminal Justice*, 33, p. 89-95.

[15] Compeau, D., Higgins, C. A., and Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly, 23*(2), 145-158.

[16] Dardan, S., Stylianou, A., and Kumar, R. (2006). The impact of customer-related IT investments on customer satisfaction and shareholder returns. Journal of Computer Information Systems, 47(2), 100-111.

[17] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

[18] Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *ManagementScience, 35*(8), 982-1003.

[19] Davis, F. D., Bagozzi, R. P.,and Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology, 22*(14), 1111-1132.

[20] Davis, R., and Wong, D. (2007). Conceptualizing and measuring the optimal experience of the e-learning environment. *Decision Sciences Journal of Innovative Education,5*(1), 97-126.

[21] Fishbein, M., and Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

[22] Gefen, D., Karahanna, E., and Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly, 27*(1), 51-90.

[23] Goodhue, D., and Thompson, R. (1995). Task-technology fit and individual performance. *MIS Quarterly, 19(*2), 213-236.

[24] Gong, M., Xu, Y., and Yu, Y. (2004). An enhanced technology acceptance model for webbased learning. *Journal of Information Systems Education, 15*(4), 365-374.

[25] Ha, S., and Stoel, L. (2009). Consumer e-shopping acceptance: Antecedents in a technology acceptance model. *Journal of Business Research*, 62(5), 565-57.

[26] Hiltz, S., and Johnson, K. (1990). User satisfaction with computer mediated communication systems. *Management Science*, *36*(6), 739–764.

[27] Jackson, C., Chow. S., and Leitch, R. (1997). Toward an understanding of the behavioral intention to use an information system. *Decision Sciences, 28*(2), 357-389.

[28] Jasperson, J., Carter, P. E., and Zmud, R. W. (2005). A comprehensive conceptualization of post-adoptive behaviors associated with information technology enabled work systems. *MIS Quarterly, 29*(3), 525-557.

[29] Lai, V. S., and Li, H. (2005). Technology acceptance model for internet banking: An invariance analysis, *Information & Management*, *42*(1), 373-386.

[30] Legris, P., Ingham, J., and Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information and Management, 40,* 191-204.

[31] Lin, W. (2008). Construction of on-line consumer behavior models: a comparative study of industries in Taiwan. *International Journal of Commerce & Management,18*(2), 123-129.

[32] Liu, P., and Tsai, C. (2007). Effect of knowledge management systems on operating performance: An empirical study of Hi-Tech companies using the balanced scorecard approach. *International Journal of Management, 24*(4), 734-743.

[33] Lu, H., Hsu, C., and Hsu, H. (2005). An empirical study of the effect of perceived risk upon intention to use online applications. *Information Management & Computer Security, 13*(2/3), 106-120.

[34] Merono-Cerdan, A. L. (2008). Groupware uses and influence on performance in SMEs. *Journal of Computer Information systems, 48*(4), 87-96.

[35] Moore, G. C., and Benbasat, I. (1991). Developing of an instrument to measure the perception of adopting an information technology innovation. *Information Systems Research, 2*(3), 192-222.

[36] Morris, M.G., Venkatesh, V. and Ackerman, P.L. (2005), "Gender and age differences in employee decisions about new technology", IEEE Transactions on Engineering Management, Vol. 52 No. 1, pp. 69-84.

[37] O"Reilly, C. A. (1982). Variation in decision makers" use of information sources: The impact of quality and accessibility of information. *Academy of Management Journal, 25*, 756-771.

[38] Rivard, S., Raymond, L., and Verreault, D. (2006). Resource-based view and competitive strategy: An integrated model of the contribution of information technology to firm performance. *Journal of Strategic Information Systems, 15*(1), 29-50.

[39] Straub, D., Limayem, M., and Karahanna-Evaristo, E. (1995). Measuring system usage: Implications for IS theory testing. *Management Science, 41*(8), 1328-1342.

[40] Streib, G. D., and Willoughby, K. G. (2005, Spring). Local governments as e-governments: Meeting the implementation challenge. *Public Administration Quarterly, 29*(1) 78–110.

[41] Szajna, B. (1996). Empirical evaluation of the revised technology acceptance model. Management Science, 42(1), 85-92.

[42] Venkatesh, V., and Davis, F. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, *27*(3), 451-481.

[43] Venkatesh, V., and Davis, F. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science, 46*(2), 186-204.

[44] Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model.*Information Systems Research*, *11*(4), 342-365.

[45] Venkatesh, V. and Morris, M. (2000). Why don"t men ever stop to ask for directions? Gender, social influences and their role in technology acceptance and usage behavior. *MIS Quarterly*, *24*(1), 115-139.

[46] Venkatesh, V., Morris, M., Davis, G., and Davis, F. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425-478.

[47] Wang, Y., Wang, Y., Lin, H., and Tang, T. (2003). Determinants of user acceptance of Internet banking: An empirical study. *International Journal of Service Industry Management, 14*(5), 501-519.

[48] Wixom, B. H., and Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research, 16*(1), 85-102.

[49] Zmud, R. W. (1978). An empirical investigation of the dimensionality of the concept of information. *Decision Sciences, 9*(2), 187-195.

USE OF SOCIAL NETWORKS SITES (SNS), CAREER SUCCESS AND INFORMATION POWER

Assoc. Prof. Dr. Saodah Wok
Dept of Communication
KIRKHS
Inter'l Islamic Univ Malaysia
Jalan Gombak, 53100 K.L
MALAYSIA
wsaodah@iium.edu.my

Prof. Dr. Junaidah Hashim
Dept of Business Administration
KENMS
Inter'l Islamic Univ Malaysia
Jalan Gombak, 53100 K.L
MALAYSIA
junaidahh@iium.edu.my

Asst. Prof. Dr. Nurita Juhdi
Dept of Business Administration
KENMS
Inter'l Islamic Univ Malaysia
Jalan Gombak, 53100 K.L
MALAYSIA
nurita@iium.edu.my

ABSTRACT
This study aims to examine the relationship between the use of social network sites and career success. It further examines whether this relationship is mediated by information power of the networkers. This study used primary data collected from 308 employees of various sizes of organisation in Malaysia. The findings revealed that there is no relationship between the studied variables and information power does not mediate the relationship. The implications of the study are discussed.

Categories and Subject Descriptors
[Human computer interaction]: Collaborative and social computing systems and tools - *social networking sites*

General Terms
Management

Keywords
Social network sites, Career success, Information power

1. INTRODUCTION
1.1 Background Of The Study
The rapid widespread of internet and its innovation has attracted many researchers to use the internet implication as their focus of study. For example, many researchers are interested to study about Social Networks Sites (SNS). According to Ellison, Steinfeld and Lampe (2007), SNS is an online space where individuals are allowed to present themselves, articulate their social network, and establish or maintain relationship. Social networking is growing in popularity especially among the teenagers (Raine, 2008). In these recent years, SNS become popular among the adult and working people as well. Of all SNS, facebook and twitter are the most popular social networks. This is probably because the social networks' uses and features are user friendly, that make it easy for people to use it. Most of social networkers will log on into their account at least once a day (Bilton, 2010). The social networks are popular for many reasons, for example it manages to increase social capital and providing entertainment (Ray, 2007).

Besides for entertainment purposes, social network like facebook also provides information for job search strategy including LinkUp, business cards, inside Job, My LinkedIn Profile, and Work With Us (Osborn & LoFrisco, 2012). This information is very important for one future career.

According to NACE (2009), 35 000 students use online resources to find job. In a job search process, the employers often inspect the SNSs of student applicants to find the most suitable candidate (Giordani, 2006). The use of SNS has enabled the users to gain access to more information which non-users may not have. To what extent the use of SNS is able to help career development and success among working adults remain unknown. To date, there is limited studies done on SNS and career, and these studies focus on the use of SNS with recruitment activities, and they were studied from the employers' perspective. This is the uniqueness of the present study where it reduces the knowledge gap in the subject by studying on new dimension of SNS, where it studies on information power and career success.

2. REVIEW OF LITERATURE
2.1 Related Theory
Social capital consists of resources that are embedded within people's social network. Three main approaches have been made to conceptualise social capital. The first approach is the weak tie theory, focused on the strength of the social tie used by a person in the process of finding a job (Granovetter, 1973). Strong ties consist of frequent, close relationships, emotionally intense ties with friends, advisors and co-workers. The information possessed by any member of this circle is likely to be widely shared with the other members, being thus quickly redundant with the information possessed by the other members. Weak ties are viewed as a connection to densely knit network outside the individual's direct contact, are infrequent, not emotionally intense that provide non-redundant information (Barros & Elves, 2003; Hatala, 2009). Granovetter (1973) argued that it was more likely that weak ties rather than strong ties would provide a greater opportunity for new information about job leads.

The second approach was Burt's (1992; 1997) structural hole approach to social capital based in alters and in egos social network. A structural hole existed when two alters were not connected to each other. They are identified as a gap. Burt (1992,) hypothesized that low-density networks (i.e., networks where few of the members are mutual friends) result in better sources of valuable information. The theory postulates that individual who possess many structural holes within their network are in an advantageous position from a power position and with regard to upward mobility. The result is simple:

better-connected people do better (Burt, 2000). Structural holes provide an individual with three primary benefits: (1) greater bargaining power, (2) more unique and timely access to information, and (3) greater visibility and career opportunities through the social system. The test of this theory concluded that the structural holes are statistical significant in explaining the level of social resources (Seibert, Kraimer & Liden, 2001).

The third approach is the social resources theory (Lin, Ensel & Vaughn, 1981a; 1981b). Social resources focused on the nature of the resources embedded within a network. Any alter who possesses characteristics or controls resources useful for the ego's goals attainment can be considered a social resource. It simply means an individual is more likely to use a contact within his or her network regardless of tie strength who can provide the resource necessary to meet his or her goals (Hatala, 2009). Lin *et al*.'s research concluded that tie strength was negatively related to the alter's occupational prestige. In turn, the alter's occupational prestige was positively related to the job secured ego's prestige.

Modern theories of social capital (Seibert, Kramer & Liden, 2001) focus on the importance of bridging and networking functions. These authors compare and contrast three network-based theories, namely weak ties theory, the structural holes theory and the social resources theory. The premise behind the notion of social capital is rather simple and straightforward: investment in social relations with expected returns. This general definition is consistent with various renditions by scholars who have contributed to the discussion (Bourdieu, 1983/1986; Burt, 1992; Coleman, 1980, 1990; Erickson, 1995; Flap, 1994; Lin, 1982; Portes, 1998; Putnam, 1995). Individuals engage in interactions and networking in order to produce profits. Social relations are expected to reinforce identity and recognition. Being assured and recognised of one's worthiness as an individual and a member of a social group sharing similar interests and resources not only provides emotional support but also public acknowledgment of one's claim to certain resources (Lin, 1999).

2.2 Career Success, Information Power and SNS

Career success is conceptualized as both real/objective and perceived/subjective achievements in individuals' work lives (Judge, Cable, Boudreau & Bretz, 1995). Such achievements have been the subject of empirical inquiry, and can be classified into two broad categories. The first category includes objective career outcomes such as promotion and compensation (Dreher & Ash, 1990). The second category consists of subjective career outcomes. This includes more affective and less tangible signs of career success such as career satisfaction, career commitment, job satisfaction and turnover intentions (Koberg, Boss & Goodman, 1998; Noe, 1988).

From the standpoint of objective career outcomes, career success of employees has generally been defined in terms of performance and the two popular symbols of success ─ money and position (Feldman, 1989; Hall, 1976). Three classes of variables employed in many objective career progress studies are: (a) rate of advancement; (b) salary attainment; and (c) supervisory ratings of performance, success, and contributions (Bowen, 1986; Kanter, 1977; Kram, 1983; Levinson *et al*., 1978; Morris, 1969; Phillip-Jones, 1982; Speizer, 1981). Based upon established definitions of promotions (Whittely, Dougherty & Dreher, 1991), employee advancement or promotion is defined as including the criteria of significant increases in annual salary, significant increases in scope of responsibility, changes in job level or rank, and becoming eligible for bonuses, incentives, or stock plans.

For current study, objective career outcomes will be used. This is consistent with Maslow's hierarchy of needs theory which states employees will try to fulfil physiological needs first (Maslow, 1970). Herzberg (1968) two factor theory also shares similar argument which states hygiene factors that are considered inadequate by employees, can cause dissatisfaction with work. Hygiene factors include among others quality of inter-personal relations, wages, salaries and other financial remuneration. Subjective career outcomes such as career satisfaction and job satisfaction are motivational factors, they are of higher level needs. Furthermore, objective career success is observable career accomplishments and typically consists of highly tangible outcomes that can be reliably judged by others. Objective career outcomes are defined by the criteria of: (a) significant increases in annual salary; (b) significant increases in scope of responsibility; (c) changes in job level or rank; and (d) becoming eligible for bonuses, incentives, or stock plans. For this proposed study, subjective career outcomes are defined by the criteria of: (a) job success, (b) interpersonal success, (c) financial success, and (d) hierarchical success.

Several studies show that networking is related to both objective and subjective career success (Forret & Dougherty, 2004; Langford, 2000; Michael & Yukl, 1993; Orpen, 1996). In this study, networking refers to behaviours that are aimed at building, maintaining, and using informal relationships that possess the (potential) benefit to facilitate work related activities of individuals by voluntarily granting access to resources and maximizing common advantages (Wolff & Moser, 2006; see also Forret & Dougherty, 2004). The construct is defined on a behavioural level (e.g., Michael & Yukl, 1993; Wanberg *et al*., 2000; Witt, 2004). These behaviours include the use of SNS, as one of the popular media for networking nowadays.

According to Serbeit *et al*. (2001), there are two reasons to expect access to information and access to resources to each be related to objective career success. First, greater access to information and resources should enhance individual work performance. They argued that individuals able to use their network position to fill a broker or boundary spanner role within the organization add greater value to the organization. Second, information and resources are fundamental bases of social power (French and Raven, 1968). Greater access to information and resources will increase the individual's organizational reputation (Kilduff & Krackhardt, 1994; Tsui, 1984) and the individual will be perceived as more powerful or influential in the organization (Brass, 1984; Brass & Burkhardt, 1993). These perceptions should make the individual better able to secure valuable organizational rewards independent of their actual level of performance (Ferris & Judge, 1991; Luthans *et al*., 1988).

The information choices have increased every day. Thus, we should not only rely on the expertise of 'gatekeepers'. The information literacy skills have become necessary on our lives to evaluate authority, to confirm accuracy and credibility of sources of information. Information literacy means more than just finding facts. It means being able to verify those facts and evaluate information in a complex technological environment (Bush, 2008). Noorriati, Saadiah and Raja Suzana (2012) stated that social networking sites increase knowledge sharing among peers at workplace. Information is a resource that should be shared among individual in an organisation.

There is some evidence to suggest that network resources can substitute for receiving mentoring and, furthermore, to suggest that when both types of resources are present the benefits of network resources for career success are greater (i.e., over and above) than the benefits of mentoring received (Eby, 1997; Higgins & Kram, 2001). New communication technologies have dramatically increased the opportunities for development of relationship ties with other organizational members in a variety of roles and hierarchical levels (Higgins & Kram, 2001). This makes it easier for individuals to obtain access to resources (e.g., information, power) they need in order to advance their careers without the assistance or intervention of the traditional mentor.

Based on the discussion related to career success, social network sites and information power, above, four hypotheses are formulated. They are:

H1: There is a positive relationship between social networks sites used and career success.

H2: There is a positive relationship between social networks sites used and Information power.

H3: There is a positive relationship between information power and career success.
H4: Use of social networks sites influences information power which in turn influences career success.

2.3 Conceptual Framework

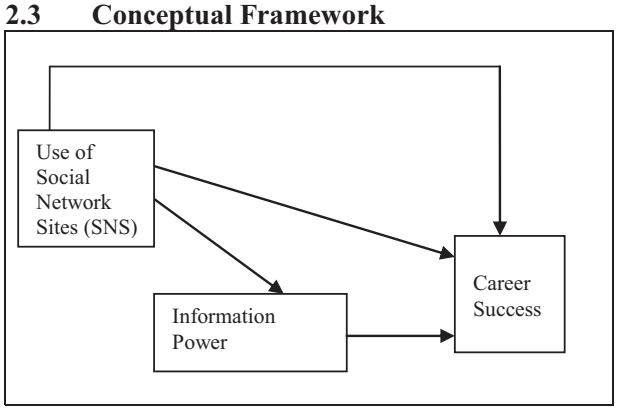


Figure 1: Conceptual framework between social networks sites, information power, and career success

3. METHODOLOGY

3.1 Measurement and Reliability Analysis

Information power consists of 10 items with Likert scaling, ranging from 1 = strongly disagree to 5 = strongly agree. Meanwhile, career success questions comprised of nine items 5-Likert scale with 1 = strongly disagree to 5 = strongly agree. There are seven items under the use of social network sites. The questionnaire surveyed how often they use several social network sites mentioned. The scales used for this part was 1 = never, 2 = less than once a month, 3 = once a month, 4 = 2-3 times a month, 5 = once a month, 6 = 2-3 times a week, and 7 = daily.

A pilot test was conducted before the actual data collection with 30 respondents as the sample. Table 1 states the standardised Cronbach Coefficient values for the pilot test and the actual reliability data. All items for information power and career success for pilot and actual tests were found reliable with Cronbach Coefficient of 0.847 and 0.914, respectively. All items for the use of social network sites were significant for pilot test, however, one item for the actual test was not reliable and the item was deleted. The Cronbach Coefficient had improved from 0.683 to 0.717.

Table 1: Pilot Test Reliability Statistics of Research Variables

Variables	Pilot (n = 30)	Actual (n = 308)
Use of social network sites	0.775	0.717
Information power	0.847	0.869
Career success	0.914	0.753

4. FINDINGS

4.1Respondents' Background

Respondents for this study consist of 308 employees from various organisations that came from different backgrounds. Table 2 showed the demographic background of the respondents. The respondents comprised of female and male respondents which represented 52.6 percent and 47.2 percent respectively out of the total respondents while 0.3 percent of the respondent did not state their gender. The respondents came from various level of education and qualifications with most of them were bachelor's degree holders (46.1 percent), followed by diploma holders (19.2 percent), master's degree holders (14.9 percent), professional degree (1.9 percent), doctorate degree (1.6 percent) and high school graduates (1.0 percent).

Furthermore, majority of the respondents were from age ranged of 21 to 30 years old with 48.7 percent, followed by respondent ranged 31 to 40 years old with 26.6 percent, 41 to 50 years old with 13 percent, while up to 20 years old and above 51 both were with 5.2 percent. 0.6 percent of the respondents did not mention their age. The respondents were from Malaysia with the ethnic group comprises of Malay as the majority (81.8 percent), Chinese (10.7 percent), Indian (4.2 percent), and others (2.9 percent) with 0.3 percent of the respondent did not provide for their ethnicity.

In term of working experience, 18.2 percent of the respondents have worked in the organisations for less than one year, 51.3 percent worked from one to five years, 14.6 percent worked between six to ten years and 15.3 percent worked for more than ten years. Next, executive being the majority respondents with 36 percent, followed by managers with 19.5 percent, individual contributor and supervisor with 17.5 percent and 17.2 percent respectively and 9.1 percent were directors. Out of the total respondents, 47.7 percent were from private sectors, 20.1 percent were from public sectors and 31.8 percent were consisting of government servants. Furthermore, 91.2 of the respondents were full time staffs of the organisations while 8.4 percent were part-time staff, and 0.3 did not indicate their employment status.

4.2 Descriptive Statistics

Table 3 and Table 4 show the descriptive statistics and correlation analysis of the use of social network sites, information power, and career success, respectively. The respondents demonstrated a low usage of social network sites (mean = 2.661), a moderate information power within the organisation (mean = 3.578), and a moderate career success (mean = 3.518).

The correlation analysis in Table 4 showed a very low relationship between the use of social network sites and information power with r = -0.033. The relationship between the use of social network sites with career success was negligible (r = -0.003). These two relationships were not statistically significant. However, the information power and career success have a moderate relationship and the relationship between them is significant (r = 0.401).
To examine whether the relationship between the use of social networking sites (independent variable) and career success (dependent variable) is mediated by information power, three steps regression were performed. In the first step, a simple regression was run to see whether the use of social network sites is a significant predictor of career success. The relationship between the use of social network sites and career success was further tested in the second regression analysis. Thirdly, a hierarchical regression was run to see whether information power has a mediating effect in between the use of social network sites and career success.

Table 5 shows that the use of social network sites variable is not a significant predictor of career success with β = -0.003, t = -0.046, and p = 0.963. When running a simple regression between the use of social network sites and information power, the results in Table 6 also depict that the relationship is not statistically significant (β = -0.033, t = -0.569, and p = 0.570). Hierarchical regression analysis in Table 7 presents that the use of social network sites is not a significant predictor of career success when information power was taken as a mediating variable (β = 0.038, t = -0.273, and p = 0.785). However, there was a significant positive relationship between information power and career success with β = 0.074, t = 7.325, and p = 0.000. The results do not support three of the hypotheses we developed for this study. Except for H3, which states there is a positive relationship between information power and career success.

5. DISCUSSION AND CONCLUSION

This study was conducted with the objective to examine to what extent the use of social network sites contributes to career success. It further examines the influence of information power on this relationship. The findings revealed that there is no significant relationship between the studied variables and there is no mediating effect of information power on this relationship.

The findings have important implications to the existing knowledge and practice. At present, empirical study on SNS and career success is very limited; no such study has been conducted at least in Asia. The present study gives a start-up for other researchers to further examine these variables in other context and group of population. Although networking has been proven by many researches as predictor to career success, based on the findings of this study,

employees have not tapped into the use of social network a media to network in organisation for the purpose of their career development. They use SNS more for personal use, to network with their friends, who has no contribution to their career development. Organisations, however has used SNS in employment recruitment activities (Osborn & LoFrisco, 2012).

Table 2: Respondents' Demographic Background

Variables	Frequency	Percent
Gender:		
Female	162	52.6
Male	145	47.1
Missing value(s)	1	0.3
Level of education:		
Lower than high school	3	1.0
High school	45	14.6
Diploma	59	19.2
Bachelor's Degree	142	46.1
Master's Degree	46	14.9
Doctorate Degree	5	1.6
Professional degree	6	1.9
Missing value(s)	2	0.6
Age:		
Up to 20 years old	16	5.2
21 to 30 years old	150	48.7
31 to 40 years old	82	26.6
41 to 50 years old	42	13.0
51 and above	16	5.2
Missing value(s)	2	0.6
Ethnic group:		
Malay	252	81.8
Chinese	33	10.7
Indian	13	4.2
Other(s)	9	2.9
Missing value(s)	1	0.3
Years the company:		
Less than 1 year	56	18.2
1 to 5 years	158	51.3
6 to 10 years	45	14.6
More than 10 years	47	15.3
Missing value(s)	2	0.6
Job function:		
Individual contributor	54	17.5
Supervisor	53	17.2
Executive	111	36
Manager	60	19.5
Director	28	9.1
Missing value(s)	2	0.6
Types of organisation		
Public	62	20.1
Private	147	47.7
Government	98	31.8
Missing value(s)	1	0.3
Employment status		
Full time	281	91.2
Part-time	26	8.4
Missing value(s)	1	0.3

Table 3: Mean Scores of the Variables

Variables	Mean	Std. Dev.
Use of social network sites	2.661	1.111
Information power	3.578	0.565
Career success	3.518	0.783

Table 4: Mean, Standard Deviation, and Correlation Analysis

Variables	Mean	Std. Dev.	1	2	
1. Use of social network sites	2.661	1.111	-		
2.Information power	3.578	0.565	-0.033	-	
3. Career success	3.518	0.783	-0.003	0.401**	

Notes: **. Correlation is significant at the 0.01 level (2-tailed)

Table 5: Simple Regression Analysis between the Use of SNS and Career Success

Variable	β	t	*p*-value
Use of social network sites	-0.003	-0.046	0.963

Notes: Dependent variable: Career success; Adjusted R^2 = -0.003.

Table 6: Simple Regression Analysis between the Use of SNS and Information Power

Variable	β	t	*p*-value
Use of social network sites	-0.033	-0.569	0.570

Notes: Dependent variable: Information power; Adjusted R^2 = -0.003.

Table 7: Hierarchical Regression Analysis of the Variables

Variables	β	t	*p*-value
Use of social network sites	0.041	-0.513	0.608
Use of social network sites	0.038	-0.273	0.785
Information power	0.074	7.325	0.000**

Notes: Dependent variable: Career success; Adjusted R^2 (Model 1) = -0.003; Adjusted R^2 (Model 2) = 0.153.

Employees in Malaysia should realize that SNS can do a lot more for them rather than just to get connected with their personal friends. SNS such as facebook and twitter has successfully helped world top politicians to be more interactive with citizens, and have helped them in their campaign. A form of social media (Barnes, 2006), SNS are web-enabled services that ''allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system'' (boyd & Ellison, 2007, p. 1). They feature prominent personal profiling, highlighting the connections between people and content (Cormode & Krishnamurthy, 2008), and allow people to visualize, interact with, and activate existing personal and professional networks, and to create connections with new ones unbounded by geographic distance (Essam, 2012). It is a great loss to the employees if they do not take advantage over such benefits. According to Neilson Online (2009), the time spent on social network and blogging sites accounted for 17 per cent of all time spent on the internet in August 2009 which has since tripled. The growth in user acceptance and business use of SNS suggests a change in the way these technologies are perceived and valued by both businesses and users, which makes them an interesting proposition for business/organisation use (Perez, 2009). In fact, co-worker sites have recently begun to appear as a viable social media platform. These sites are intended to enhance the workplace environment by encouraging staff to get to know more about peers, administrators, and others in the organisation. Therefore, for effective change and general acceptance of the role of SNS in organisation, senior managers need to develop a clear communication strategy, both vertically and horizontally, to promote the benefits and effects these tools have to offer.

Obviously, more researches are needed on SNS contribution to employees specifically on their career development and success. Since the use of SNS is more popular among the youths, and career success is not yet experienced by these youth, it is quite challenging to get evidence of how SNS help career success. This is the

limitation of the present study. But, it would interesting if future research can focus on successful managers and examine their use of SNS. The present study triggers more studies in this subject.

References

Barnes, S.B. (2006), ''A privacy paradox: social networking in the United States'', First Monday, Vol. 11 No. 9, available at: http://firstmonday.org/issues/issue11_9/barnes (accessed November 5, 2010).

Barros, C. P. And Alves, F. M. P. (2003), Human capital theory and social capital theory on sports management, International Advances in Economic Research 07/2003; 9(3):218-226. DOI:10.1007/BF02295445

boyd, d.m. and Ellison, N.B. (2007), ''Social network sites: definition, history, and scholarship'', Journal of Computer-Mediated Communication, Vol. 13 No. 1, available at: http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html (accessed October 9, 2008).Bush, G. (2008). Thinking around the corner: *the power of information literacy*, Phi Delta Kappan, 90(6), 446-447.

Bilton, N. (2010, March 09). Facebook will allow users to share location. *The New York Times.* Retrieved February 8, 2013. http://bits.blogs.nytimes.com/2010/03/09/facebook-will-allow-users-to-share-location/

Burt, R. S. (1997), The contingent value of social capital, *Administrative Science Quarterly,* 42, 2, 339-365.

Dreher, G. F. and Ash, R. A. (1990). A comparative study of mentoring among men and women in managerial, professional, and technical positions. *Journal of Applied Psychology,* 75,5, 539-546.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), article 1. Retrieved February 8, 2013. http://jcmc.indiana.edu/vol12/issue4/ellison.htm

Essam Mansour, (2012),"The role of social networking sites (SNSs) in the January 25th Revolution in Egypt", *Library Review*, 61, 2 pp. 128 – 159.

Giordani, P. (2006). Technology influences the profession. *National Association of Colleges and Employers Journal,* 67, 18

Granovetter, M. S. (1973), The strength of weak ties, *American Journal of Sociology*, 78, 5, 360-80.

Hatala, J. P. (2009), Assessing individual social capital capacity: the development and validation of a network accessibility scale, *Performance Improvement Quarterly*; 2009; 22, 1; ProQuest Education Journals, 53.

Higgins, M. and Kram, K. (2001), "Reconceptualizing mentoring at work: a developmental network perspective", *Academy of Management Review*, 26, 2, 264-88.

Judge, T. A., Cable, D. M., Boudreau, J. W., & Bretz, R. D., Jr. (1995). An empirical investigation of the predictors of executive career success. *Personnel Psychology, 48*, 485-519.

Lin, N.; Ensel, W. M.; Vaughn, J. C. Social Resources and Occupational Status Attainment, *Social Forces*, 59, 5, 1981a, 1163-81.

Lin, N., Ensel, W. M.; Vaughn, J. C Social Resources and Strength of Ties, American Sociological Review, 46, 4, 1981b, pp. 393-405. McPherson, J. M.; Popielarz, P. A.; Drobnic, S. Social Networks and Organizational Dynamics ,American Sociological Review, 57, 2, 1992, pp. 153-70.

National Association of Colleges and Employers. (2009). 2009 student survey: Research brief. Bethlehem, PA: Author.

Noorriati Din, Saadiah Yahya, and Raja Suzana Raja Kassim. (2012). Online social networking for quality of life. *Procedia Social and Behavioral Sciences, 35*, 713-718.

Osborn, D. S. and LoFrisco, B. M. (2012). How do career centers use social networking sites? *The Career Development Quarterly,* 60, 3, 263-272.

Raine, C. (2008). Uses and gratifications of facebook for political information. Unpublished M.S. dissertation, University of Kansas, United States. Retrieved February 8, 2013. Dissertions & Theses: Full Text (Publication No. AAT 1460175).

Ray, M. (2007). Needs, motives and behaviors in computer-mediated communication: An inductive exploration of social networking websites. Cited in Elder Jubelin, J..face(book)ing a crowd?: An exploration of audience, context, privacy, and self-presentation on facebook. Unpublished M.A. dissertations, York University, Canada. Retrieved February 8, 2013. Dissertations & Theses: Full Text (Publication No. AAT MR51624).

Seibert, Se. E. , Crant, J., and Kraimer, M. L. (1999). Proactive personality and career success. *Journal of Applied Psychology*, 84, 416-427.

Extending Trigger By Example Approach to Implement Reactive Agents in Active Databases

Kornelije Rabuzin
University of Zagreb, Faculty of organization and informatics Varazdin
Pavlinska 2
42000 Varazdin, Croatia
+385(0)42/390-847
kornelije.rabuzin@foi.hr

ABSTRACT
There is a connection between reactive agents and active databases that has been already explored earlier; so far it is known that both fields rely on reactivity and that reactive agents can be implemented in active databases. When talking about active databases triggers are unavoidable. Although one way to implement triggers is to use the Structured Query Language (CREATE TRIGGER statement in SQL), another (more interesting) one is based on the idea of Trigger By Example (TBE) approach. TBE is (on the other hand) based on Query By Example (QBE) that is another language used to work with databases (QBE is a graphically oriented language a bit easier to learn and use than SQL). In this paper a new approach (based on the already mentioned Query By Example and Trigger By Example approaches) to implement reactive agents in active databases is presented, the so-called Agent By Example (ABE) approach.

Categories and Subject Descriptors
I.2.11 [**Computing Methodologies**]: Distributed Artificial Intelligence – *multiagent systems.*

General Terms
Design, Experimentation.

Keywords
Reactive Agents, Active Databases, SQL, Query By Example, Trigger By Example, Agent By Example.

1. INTRODUCTION
Although (sometimes) misunderstood during the past decade, agents are nowadays accepted and used to solve different problems and complex tasks. Many books are written on the subject and many articles exploring different aspects of agency can be found (for example [4] or [12]).

Although many different types of agents emerged during the years, reactive agents are recognized as one important type. Reactive agents are capable to react to certain events and this capability (to react immediately) became popular during the years. The so called stimulus-response model is of great importance in that context and such systems, although it might seem a little strange (at first), can exhibit (very) complex behavior.

Databases have been with us for even longer (more than four decades). Although the relational data model is very important, many other types of databases emerged during the years as well as active databases. An active database is such a database that can react to certain events by performing some actions (actions are performed automatically). Active databases rely on active rules (active rules are explained later on) that are (mainly) implemented by means of triggers.

In order to work with (active) databases one has to be familiar with SQL that is a standard (and standardized) language that is used to work with databases. It is a well known fact that 90 % of all database management systems support SQL, but some other languages exist and can be used to work with databases as well (some of them are Query By Example, Tutorial D, QUEL, etc.). Although SQL should be simple (when it was introduced it was supposed to be used by end users), sometimes complex queries cause problems even for professionals. On the other hand Query By Example should make the process of posing queries much easier. The basic idea of Query By Example (QBE) language is quite simple; one sees a two dimensional table and has to fill it with some symbols in order to produce a query (one can design queries more or less graphically by entering some symbols in the visible table structure). Trigger By Example (TBE) approach is based on QBE and is used to implement triggers more easily (graphically) i.e. in the same way that QBE is used to pose queries.

Since reactivity is very important property immanent to (reactive) agents and active databases, the idea that active databases could be used to implement reactive agents is (however) not new, but only few papers have been written on the subject ([1], [3], [9]). In this paper the connection between the two mentioned fields is examined and a new approach for implementing reactive agents in active databases is proposed. Since triggers can be used to implement reactive agents, for the purpose of this paper Agent By Example (ABE) approach is proposed and developed (ABE relies on QBE and TBE).

The rest of the paper is organized as follows: after the introduction part reactive agent and active databases are described. In the third part query languages are described and Trigger By Example approach is presented. Further on, the connection between active databases and reactive agents is explored and Agent By Example approach is described. In the end the conclusion is given.

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.26

2. REACTIVE AGENTS AND ACTIVE DATABASES

One can say that the term agent has been a buzzword in the past. While some authors tried to define "what" an agent really was, other authors tried to identify agents by enumerating properties that agents should possess. Today authors agree that when we talk about agents we have in mind intelligent components that are trying to achieve certain objectives and that are capable to react (to certain events), communicate (with other agents), perceive (the environment) and act in someone's favor. Many different definitions can be found and different properties are immanent to agents. Although many different types of agents can be found as well, reactive agents are surely one of them.

When talking about multiagent systems one must be aware that there are some other important questions regarding multiagent systems that are not addressed in the paper (architectures, coordination, cooperation, negotiation, etc.). Within the scope of this paper it is assumed that readers are familiar with the subject. Details on multiagent systems are skipped and we take for granted the fact that reactive agents exist and that they are interesting in the context. Although the topic is quite complex and broad, more on agents can be found (for example) in [4] and [12].

Although some authors thought that reactive agents didn't deserve much attention, it was shown that many (purely) reactive agents could exhibit complex behavior. It is a well known fact that reactive agents don't possess memory and that they just react by performing some actions (the so called stimulus response model). Reactive agents operate in presence and they do or do not react to certain stimuli [4]. Reaction is used when there is no time for reasoning which is time consuming [5]. Reactivity is suitable for dynamic environments ensuring an immediate response to some recognized changes.

The behavior of one reactive agent can be described as follows [6]:

1. observe any input at time T,
2. (optionally) record any such input,
3. match conditions of condition-action rules with the inputs,
4. (optionally) verify any remaining conditions of the rules using information in the knowledge base, using for steps (3) and (4) a total of R units of time,
5. select an atomic action which can be executed at time T+R+2 from the conclusions of rules all of whose conditions are satisfied,
6. execute the selected action at time T+R+2, and (optionally) record the result,
7. cycle at time T+R+3.

Perhaps one of the best known reactive architectures is the so called subsumption architecture (author is R. Brooks). Brooks had certain assumptions about intelligence, real world, and deduction as such, and for his architecture the so called *situation -> action* rules are of great importance. Since there is an obvious similarity between *situation -> action* rules and active rules (that are used in active databases), the connection is described later on.

People are usually aware of triggers and use them quite often, but they don't know what active databases are. One can say that active databases can react to certain (types of) events by performing some actions. Active databases rely on the concept of active or ECA (Event - Condition - Action) rules; ON Event - IF Condition - THEN Action. The rule could be interpreted as follows; when something happens and some conditions are fulfilled, some actions are executed as reactions to recognized changes. Actions that are executed may be simple (sometimes almost trivial), as well as quite complex, but they would be executed only if the condition evaluation phase was successful. Events require intervention and can be divided into simple and complex events. Simple events can be further decomposed:

- database operations: INSERT, UPDATE and/or DELETE,
- time events (absolute, periodic and relative)
- method events,
- transaction events (ON BEGIN/ON COMMIT), and
- abstract events.

Several simple events can be combined and they can produce a complex event. For example, if we had two simple events (E1, E2), then E1∧E2 (E1 and E2) would represent a complex one. When talking about active databases one has to have in mind several important projects that used to deal with the subject (for example Starburst, Ariel, Ode, Sentinel, SAMOS, REACH, NAOS, TriGS, Chimera). Many things were discovered within those projects, but some of them were never widely used. During the years several special constructs were introduced that can be used to build complex events as well (negation, repeat, sequence, etc.) but are not discussed in the paper. More on complex types of events can be found (for example) in [2], [8] and [10].

3. ACTIVE DATABASES AND SQL

In order to work with (active) database management systems one must be familiar with SQL. SQL is a standardized language that consists of a certain number of statements that users use to work with their databases. It is already known that triggers are used to implement active rules. According to the SQL standard a CREATE TRIGGER statement should be used in order to implement a trigger; the CREATE TRIGGER statement syntax is given below (in a slightly reduced form):

CREATE TRIGGER <trigger name> <trigger action time> <trigger event>

ON <table name> [REFERENCING <transition table or variable list>] <triggered action>

<trigger action time> ::= BEFORE | AFTER

<trigger event> ::= INSERT | DELETE | UPDATE [OF <trigger column list>]

A quick view reveals that it has certain limitations; trigger events are limited just to three basic DML statements (INSERT, UPDATE and DELETE) although in the theory we distinguish many different types of events (discussed earlier).

An alternative solution (language) to work with databases is QBE (Figure 1 and Figure 2). Briefly, you see a two-dimensional table and you have to select (check) attributes that you want to include in the answer; you can filter the rows you get in the result (by using some constants) and you can join two (or more) tables (by using some variables). As one can see in Figure 1, one has to select the fields that are needed (Last Name and Job Title), one can enter certain criteria (Job Title = Manager), sort the result (ascending), etc.:

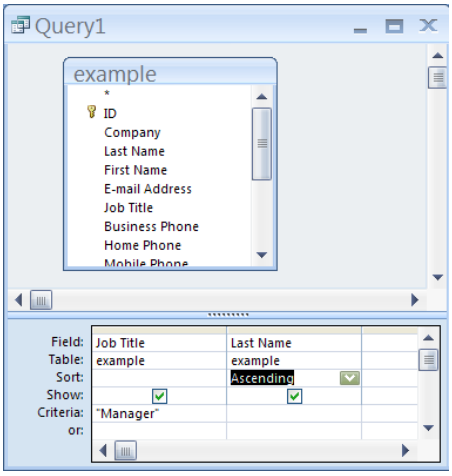


Figure 1. Query By Example in Microsoft Access

So basically, QBE allows us to build queries graphically and in that way simplifies the process of posing queries. Another example comes from [7]; P. means that the column should be included in the result i.e. "printed", while _d and _e are used to join tables (their values should be the same in both tables):

emp	Eno	Ename	DeptNo	Sal
	_e	Tom P.	_d	

dept	Dno	Dname	MgrNo
	_d		_e

Figure 2. Query By Example [7]

Sometimes it is not easy to write complex triggers and new ways for easier implementation seem to be interesting. One such approach is the so called Trigger By Example approach [7]. Authors accepted the idea of QBE language and decided to extend that idea in a way that triggers could be specified graphically, as well as queries. In that way the QBE idea could be used not only for posing queries (SELECT statement), but for creating triggers as well (Figure 3 shows the CREATE TRIGGER statement in SQL as well as the TBE specification that authors have proposed for trigger implementation):

```
CREATE TRIGGER TwiceSalaryRule AFTER UPDATE OF Sal ON emp
FOR EACH ROW
WHEN EXISTS (SELECT * FROM sal-change WHERE Eno = NEW.Eno
      AND Year = CURRENT_YEAR AND Cnt >= 2)
BEGIN ATOMIC
      UPDATE sal-change SET Cnt = Cnt + 1
          WHERE Eno = NEW.Eno AND Year = CURRENT_YEAR;
      INSERT INTO log VALUES(NEW.Eno, NEW.Sal);
END
```

E.emp	Eno	Ename	DeptNo	Sal
AFT.R.	_n			U._s

C.sal-change	Eno	Cnt	Year
	NEW(_n)	_c	CURRENT_YEAR

Figure 3. Trigger By Example [7]

4. AGENT BY EXAMPLE

We can say that active databases and reactive agents have been developed independently and that reactive agents and active databases rely on reactivity. According to [9] reactivity is a common denominator for both fields; the main distinction is that databases are used for knowledge organization and intelligent systems try to exhibit intelligent behavior.

One of the oldest articles written on the subject was "Active databases and agent systems – a comparison" [3]; authors compared the structure and functionality of active databases and reactive agents. It was shown that both fields relied on reactivity, that both fields were developed independently and that some similarities between them existed. Further on, it was stated that active databases were used in a reduced form (usually just for constraint checking and view maintenance) while agents were used for accomplishing different (complex) tasks.

As one can see up to now we know that triggers can be implemented by means of SQL, but that TBE could be used as well. In order to implement reactive agents in active databases (based on the previous discussion), the idea of Agent By Example approach is proposed and described.

In this part we describe the problem in a very precise and concise manner; the whole model is not described and the focus is on a specific problem that was solved. Briefly, the problem that had to be solved was efficient resource allocation in a multiagent system. Since many agents allocated many resources, resource allocation caused many conflicts and deadlocks and a new solution had to be found. Although conflicts can be solved in many different ways (centralized, decentralized, cooperative, non-cooperative, etc.), conflict avoidance is a possible approach as well [11]. The main goal was to ensure that there were enough resources available, that agents returned (de-allocated) resources in time and that conflicts were avoided (if possible).

A database that contained information on resources was built and the problem was solved with a reactive agent that had two main tasks;

- reactive agent could allocate resources by using several different algorithms (LRU, random, etc.) and
- reactive agent could ensure that allocated resources were successfully returned (de-allocated) in time (in the past resources hadn't been returned due to oblivious agents, technical (network) problems, human factor, etc. and a control mechanism for de-allocation had to be built).

In order to implement this reactive agent (especially this second feature) the presented "Agent By Example" approach was used. Based on the idea of QBE and TBE approaches and in order to return allocated resources an Agent By Example interface was implemented which enabled that certain time event (and belonging conditions) were defined on a (visible) table structure. This interface (Figure 4) allowed us to specify when the reactive agent should remind other agents to return resources; one could specify a time event (absolute or periodic) and a constraint that had to be checked in a certain point of time. The important part is that the TBE approach (idea) was extended in a way that other types of simple events were supported as well; in this context the most important events were time events.

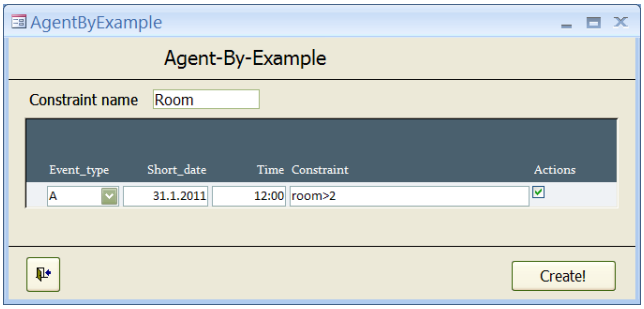


Figure 4. Agent By Example - interface

First thing that is important to specify is a name for a certain constraint as well as time and date when the constraint should be checked. For now only absolute and periodic time events are supported (one can select A or P in the event type combo box); if a periodic event was specified, it would be checked daily whether the constraint is satisfied or not and in the case of absolute event a specified constraint would be checked only once (in a certain point of time). Further on, in the visible table structure one could specify the condition that has to be satisfied (i.e. a constraint). Actions check box ensured that messages were sent to agents (and by default to administrator) to ensure that resources were returned in time.

From the technical point of view the solution was implemented in the following way. After the user clicked the "Create!" button certain parameters were written as a row in a specified table (PostgreSQL DBMS was used). After the row was added, a trigger started a Perl script that added relevant (date and time) parameters to a file that was read by the CRON system (this script is not presented in the paper due to size limitations). The CRON system then started a PHP file that sent messages accordingly.

Initially it was important to set certain constraints in order to keep "the model alive"; during the execution phase reactive agent added new constraints dynamically, based on allocated resources. A few algorithms were implemented for that purpose (least recently used, most recently used, random allocation, etc.). In order to allocate resources transactions were used and based on allocated resources certain time triggers were added into the system (by the reactive agent) that were used to de-allocate resources and (in that way) physical conflicts were avoided.

Further on, the reactive agent had another important role; to perform some task (i.e. a service) we needed a few different types of resources. As soon as some resource was de-allocated, the reactive agent determined which services could have been performed based on the list of non-allocated resources (in the database). In the same way, as soon as some resource was allocated, it was possible that some service became unavailable because a certain resource type was not available any more. This was used to speed up the system and to know in each point of time which services were available.

5. CONCLUSION

The problem that was originally solved was resource allocation in a multiagent system. For that purpose the Agent By Example approach is proposed. This approach enabled us to implement the reactive agent graphically by filling in the visible table structure. Basically, it was important to specify date and time parameters and conditions that had to be checked and based on specified parameters the behavior of reactive agent was determined. For that purpose the TBE approach was extended in order to support time events. In ABE one has to define a condition on a visible table structure and determine when to check specified conditions. The proposed Agent By Example approach enabled us to build and determine the behavior of the reactive agent graphically, based on the QBE and TBE ideas.

Very important part of this research is based on the so called Trigger By Example approach; certain modifications and improvements of the TBE approach resulted in the Agent By Example approach. Although this was an important step toward the final solution where reactive agent should determine (alone) when to send certain messages, the Agent By Example idea is interesting and can be upgraded and used in other fields as well.

6. REFERENCES

[1] Akker, J. and Siebes, A. 1997. Enriching active databases with agent technology. Lecture Notes In Computer Science, Proceedings of the First International Workshop on Cooperative Information Agents, 116-125.

[2] Andler, S. F. and Hansson, J. 1998. *Active, real time, and temporal database systems*. Berlin, Springer.

[3] Bailey, J., Georgeff, M., Kemp, D., Kinny, D., and Ramamohanarao, K. 1995. Active databases and agent systems – a comparison, Lecture notes in computer science 985, 342-356.

[4] Ferber, J. 2001. *Multiagenten-Systeme, Eine Einführung in die Verteilte Künstliche Intelligenz*. Addison-Wesley, USA.

[5] Hexmoor, H. 2003. Evolution of Agent Architectures, In: Truszkowski, W., Rouff, C., and Hinchey, M. (Eds.), WRAC 2002, LNAI 2564, 469-470, Springer-Verlag, Germany.

[6] Kowalski, R. and Sadri, F. 1997. An Agent Architecture that Unifies Rationality with Reactivity, www.doc.ic.ac.uk/~rak/papers/agents-97.pdf.gz, Accessed 20 August 2004.

[7] Lee, D., Mao, W., Chiu, H., and Chu, W. W. 2005. Designing Triggers with Trigger-By-Example, Knowledge and Information Systems, 7(1), 110-134.

[8] Paton, N. W. 1998. *Active rules in database systems*. New York, Springer.

[9] Rundensteiner, E. A. 1988. The role of AI in databases versus the role of database theory in AI: an opinion. Artificial Intelligence in Databases and Information Systems, Proceedings of the IFIP TC2/TC8/WG 2.6/WG 8.1 Working Conference, 233-252.

[10] Tan, C. W. and Goh, A. 1999. Composite event support in an active database. Computers & Industrial Engineering, 37(4), 731-744.

[11] Tessier, C., Chaudron, L., and Müller, H. 2001. *Conflicting Agents*. Boston, Kluwer Academic Publishers.

[12] Wooldridge, M. 2009. *An introduction to multiagent systems.* Wiley Publishing, UK.

Basic treatment principles for bladder cancer with Chinese herbal medicine: an application of text mining

Cheng Xiao
China-Japan Friendship Hospital
Yinghuayuan Dongjie Chaoyang District, Beijing, China
+86 10 84205442, 100029
xiaocheng2002812@163.com

Shenglong Jing
Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, China
+86 10 64014411, 100700
1538736048@qq.com

Miao Jiang
Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, China
+86 10 64014411, 100700
miao_jm@126.com

Cheng Lu
Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, China
+86 10 64014411, 100700
lv_cheng0816@163.com

Xiaojuan He
Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, China
+86 10 64014411, 100700
hxj19@yahoo.com.cn

Peng Xu
Jiangxi University of Traditional Chinese Medicine,
Nanchang, China
+86 791 7118826, 330004
xp0420@163.com

Dan Luo
Beijing University of Chinese Medicine,
Beisanhuang Donglu 11# Chaoyang District, Beijing, China
+86 10 64213841, 100030
330996648@qq.com

Guang Zheng
School of Information Science-Engineering, Lanzhou University,
Lanzhou, China
+86 931 8910864, 730000
forzhengguang@163.com

Aiping Lu*
Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, China
+86 10 64067611, 100700
lap64067611@163.com

ABSTRACT
Bladder cancer occurs in the majority of cases in males with a male/female sex ratio of 3:1. Of the three main histological variants of epithelial malignancies arising from the urothelium of the urinary bladder, transitional cell carcinoma (TCC) is the most prevalent in Japan, North America, and other developed countries, while squamous cell carcinoma and adenocarcinoma are diagnosed less frequently, and it is also the most expensive cancer to treat. Antineoplastic resistance is one of the most common problems in bladder cancer which can cause treatment failure and serious complications. Antineoplastic-resistance has been increasing in the past decades worldwide. Chinese herbal medicine (CHM) has been an integral part of Traditional Chinese Medicine (TCM) for thousands of years. Some Chinese herbs in the study were proved to have antineoplastic activity, which prompted their compound prescription use in the management of bladder cancer. Many herbal formulations have been developed and used in the treatment of Bladder cancer and were proved to be effective and safe. Yet the principles of treating bladder cancer with CHMs are hard to manage due to the complexity of TCM theory. In this study, a novel text mining method was development based on a comprehensive collection of literatures in order to explore the treatment principles more intuitively. Networks of TCM patterns and CHMs which are most frequently used in bladder cancer treatment are built-up and analyzed, two major principles are explored in treating bladder cancer from 14097 records of literature: Clearing the damp-heat with strengthening healthy qi. These findings might guide the clinicians in treatment of bladder cancer.

Categories and Subject Descriptors
Algorithms: Language Constructs and Features – *abstract data types, polymorphism, control structures.* In the process of data mining, we construct an data slicing algorithm called *discrete derivatives.*

General Terms
Algorithms, Management, Documentation, Design, Reliability, Experimentation, Human Factors, Standardization, Languages, Theory, Verification.[1]

Keywords
Bladder cancer, Chinese herbal medicine, Pattern, Traditional Chinese medicine, Text mining.

1. INTRODUCTION
Bladder cancer is the 10th most common cancer worldwide, with the highest rates reported in Europe, North America and Australia,

* Corresponding author: Hong Kong Baptist University School of Chinese Medicine, Kowloon, Hong Kong

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.27

and accounting for an estimated 261 000 new cases diagnosed and 115 000 deaths each year; by comparison, relatively low rates are found in the Far Eastern countries[1, 2]. It affects men more frequently than women[3]. Typical of solid tumours, bladder cancer incidence increases with age. Tumours of the bladder rarely occur before the age of 40 – 50, arising most commonly in the seventh decade of life[4, 5]. The median ages at diagnosis are 69 years for men and 71 for women[6].

Gross hematuria (presence of blood in urine) is the most common symptom in BC which was noted in 63 to 88% of the cases [7-9]. Dysuria (painful voiding) has been reported as the second most common symptom. Urinary obstruction, abdominal pain, urinary tract infection and weigh loss have been reported occasionally[7, 10]. Rare cases of paraneoplastic syndromes such as ectopic ACTH secretion and hypercalcaemia were also reported [11]. There are two major biological pathways in human bladder tumor development, leading to two major subtypes of bladder cancer: superficial/non-muscle-invasive (NMIBC) and advanced/muscle-invasive (MIBC)[12].

An article by Bassi et al. [13]confirmed that delays in diagnosis and initiation of therapy have adverse effects on stage and survival. The median age at diagnosis is 65 years with as many as 10% of patients >85 years of age. Studies in this group are essential to define the optimal approach to care because a substantial number of people with advanced BC are elderly or have other poor-risk features (characterized by visceral metastasis, poor performance status (PS), and >5% weight loss).

Currently, cisplatin-based combination chemotherapy is considered to be the standard therapy for this disease. Regimens such as MVAC (methotrexate, vinblastine, adriamycin, cisplatin), CMV (cisplatin, methotrexate, vinblastine) and GC (gemcitabine, cisplatin) have been employed with relative risks (RRs) reported in up to 70% [14, 15]. Despite these high RRs, toxicity and survival outcomes remain suboptimal. For example, MVAC therapy results in a median survival time of approximately 13 months and is associated with severe toxicities including myelosuppression, nephrotoxicity, stomatitis, and emesis[16].In addition, treatment-related death rates have been up to 3%. In an attempt to minimize these toxicities, studies are under way to determine which patients would best benefit from therapy [17, 18].

Chinese herbal medicine (CHM) has been an integral part of Traditional Chinese Medicine (TCM) for thousands of years. Some Chinese herbs in the study were proved to have antineoplastic activity [19-21]. These properties had prompted their compound prescription maybe use in the management of BC.

However, due to the complexity of TCM theory, the treatment principles of BC are complicated and mysterious. In order to explore the treatment principles more intuitively, a novel text mining method was development based on a comprehensive collection of 14,097 records of literatures[22]. The study would provide an accessible way for understanding the treatment principles for BC with CHMs.

2. MATERIAL AND METHODS

2.1 Data Collection

The dataset were downloaded from SinoMed (http://sinomed.cintcm.ac.cn/index.jsp) with the query term of "Bladder cancer" on Feb. 23, 2012. This dataset conatins 14,097 records of literatures on clinical practices or theoretical research on bladder cancer. In this dataset, each record/paper is tagged with a unique ID. These records contain the title, keywords, and abstract of published papers[22].

2.2 Data Filtering

Please use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

2.2.1 TCM Patter

Pattern (also called as Syndrome, or Zheng) differentiation is regarded as the key role in the clinical practise of TCM traditional Chinese medicine[23]. Usually, pattern identification is the basis of the prescription of herb formulae, CHMs, or other TCM therapies. Thus it is natural and intuitive to filter out the pattern and then try to find the associate rules between pattern and CHMs. The top TCM patterns in BC are: *qi deficiency (Qi xu), blood heat (Xue re), blood stasis (Xue yu) and damp-heat stagnation (Shi re yun jie).*

2.2.2 Chinese herbal medicine

Based on the keyword list of CHMs (both legal names and other popular names are included for calculation), we filtered the CHMs in the plain text format, and then converted all popular names into legal names. All the CHMs were tagged with their unique paper ID. Based on the unique paper ID, we could construct the pairs of co-existed CHMs as they coexisted in literature. For example, in one paper, CHMs of Huangqi (*Radix Astragali seu Hedysari*), Renshen (*Radix Ginseng*), and Shengdihuang (*Radix Rehmanniae Recens*) are mentioned. Then, the pairs of co-existed CHMs of "Huangqi-Renshen", "Huangqi-Shengdihuang", and "Renshen-Shengdihuang" are constructed.

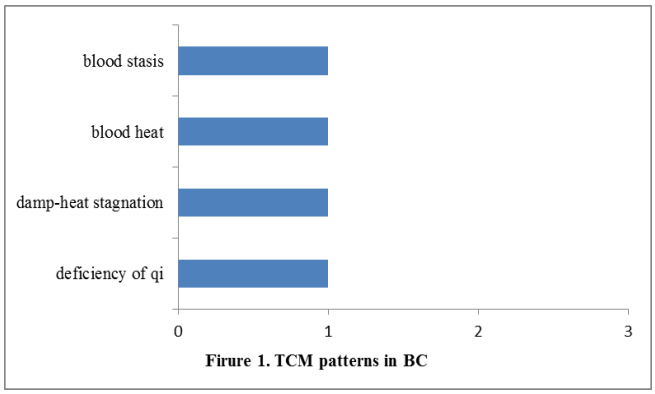
3. RESULTS

In this paper, focused on BC, we explored the principles of pattern differentiation and CHMs prescription and the association between the two aspects under the framework of TCM theory from 14,097 literatures. The network construction is based on the analysis of networks of pattern and CHM correlated with BC in literature. The connections among these networks are built-up under the professional knowledge of TCM.

3.1 Major TCM Patterns in BC

Pattern identification is regarded as the first step during TCM clinical practice procedure. After the pattern is approved, the treatment principle can be determined. For example, when the pattern of *blood stasis* is approved, then the treatment principle of *active blood and resolve stasis* is

determined. In our results, 4 TCM patterns are detected to be related with BC, and the TCM patterns in BC are presented in Fig. 1. Remarkably, the patterns of BC less amount due to lack of literatures.



Firure 1. TCM patterns in BC

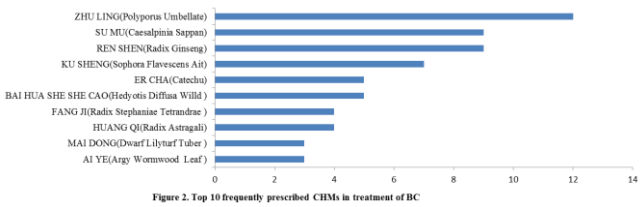


Figure 2. Top 10 frequently prescribed CHMs in treatment of BC

3.2 Most frequently prescribed CHMs in BC treatment

P- Altogether 64 CHMs are mined from the literature in treatment of BC. As herbal formulae are composed by the CHMs, the list of most frequently used CHMs can certainly provide the information of TCM treatment principles more effectively due to the stablity and uniqueness of each CHMs rather than formulae which can be renamed easily after slight regulation. The top 10 frequently prescribed CHMs are shown in Fig. 2. It is demostrated that most CHMs prescribed in BC management are those with functions of clearing away heat evil and removing carbuncle, only 3 CHMs stand for tonification, which can help strengthen the principal curative action of clearing damp-heat and invigorating qi.

3.3 Networks of the Pattern and CHMs in BC

The networks of patterns and CHMs in BC treatment can be constructed based on the co-existence frequency among patterns or CHMs, respectively. By checking these two networks, the correlation between TCM patterns and CHMs can be analyzed and explored. In order to achieve better visualization, the CHM network is simplified to preserve 13 CHMs which are the most frequently used in combination in treating BC. The networks of patterns and CHMs which with their correlation on BC is demonstrated in Fig. 3.The major correlation between TCM patterns identification and CHMs are demonstrated with arrows.

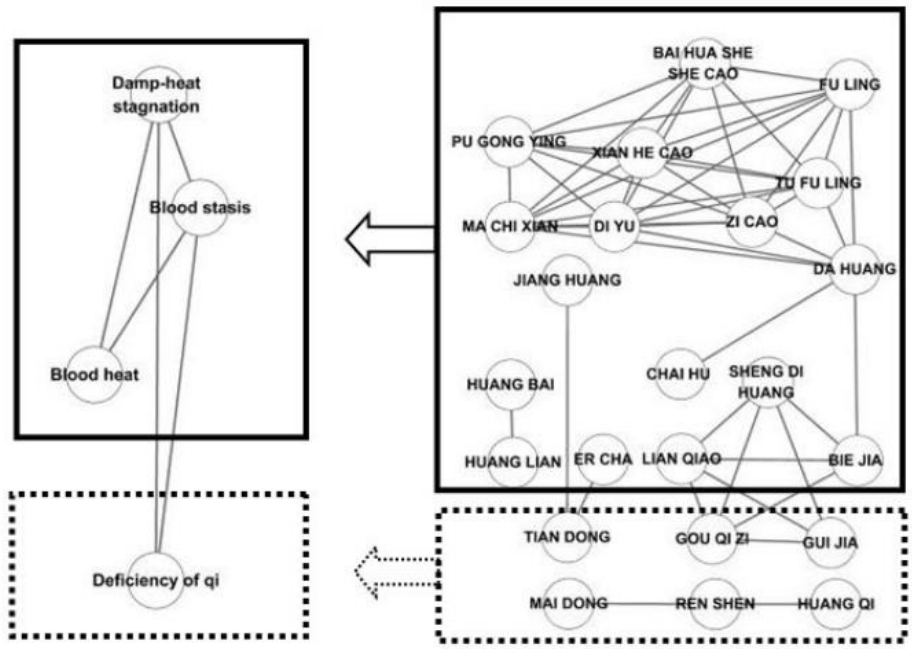


Fig. 3. The networks of TCM patterns and CHMs in treatment of BC. Network of patters is shown in the left part, network of CHMs in right part. Bigger shape represents higher frequencies. The lines between the shapes represent the co-existent correlations between the two patterns/CHMs. Arrows represent the correlation between TCM patterns and CHMs.

4. CONCLUSION AND DISCUSSION

Based on the analysis described in previous section, it is naturally come to the point that TCM treatment principles of a disease can be reasonably mined out and presented from dataset downloaded from SinoMed. Compared with the knowledge of BC in text book, the most knowledge is covered by the simple and succinct networks demonstrated in Fig.3 which can be summarized with following points and their internal connections.

4.1 TCM Networks of Patterns and CHMs can be Constructed and Analyzed

In this study, through mass calculation on dataset on BC, the main aspect of TCM networks were built-up. The pathogenesis related with BC includes Dampness-heat, blood heat, Deficiency of qi in spleen and kidney. To follow the matter of course, CHMs most frequently prescribed in BC treatment can be grouped in to 2 major classes, one group is responsible for clearing damp-heat, the other for help the principle action and reinforcing the healthy qi. These major principles might guide the clinicians in treatment of BC.

4.2 Internal Connections among Networks

Through directed text mining, the internal connections among TCM networks were also found. These internal connections can be grouped into two major hierarchical clusters. Each cluster is associated with one major kind of patterns. The major treatment principles of TCM treatment of BC can be explored by text mining method and summarized in a succinct figure.

4.3 TCM Network might be Useful in both TCM Clinical Practices and Scientific Researches

The network demonstrated in Fig. 3 can be taken as a high level of abstraction on the treatment of BC out of dataset contains 14,097 records. From the view point of clinicians, it can be taken as a kind of reference. From the view point of basic researchers, this result might be useful to illuminate some further studies in BC.

5. ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation of China (No. 30902003 and 81072982). 2012' Traditional Chinese medicine Professional project (No. 201207012).

6. REFERENCES

[1] Meliker JR, JO N: Arsenic in drinking water and bladder cancer:review of epidemiological evidence. Trace Metals and other Contaminants in the Environment 2007, 9:551-584..

[2] Larsson SC, Andersson SO, Johansson JE, A W: Diabetes mellitus, body size and bladder cancer risk in a prospective

[3] Kuper H, Boffetta P, HO A: Tobacco use and cancer causation: association by tumour type. J Internal Medicine 2002, 252:206-224.

[4] Zeegers MPA, Kellen E, Buntinx F, PA vdB: The association between smoking, beverage consumption, diet and bladder cancer: a systematic literature review. World J Urol 2004, 21:392-401.

[5] Shariat SF, Milowsky M, MJ D: Bladder Cancer in Elderly. Urol Oncopl 2009, 27:653-667.

[6] Volanis D, Kadiyska T, Galanis A, Delakas D, Logotheti S, V Z: Environmental factors and genetic susceptibility promote urinary bladder cancer. Toxicology Letters 2010, 193:131-137.

[7] Choong NW, Quevedo JF, JS K: Small cell carcinoma of the urinary bladder. The Mayo Clinic experience. Cancer 2005, 103(6):1172-1178.

[8] Cheng L, Pan CX, Yang XJ, Lopez-Beltran A, MacLennan GT, Lin H, Kuzel TM, Papavero V, Tretiakova M, Nigro K et al: Small cell carcinoma of the urinary bladder: A clinicopathologic analysis of 64 patients. Cancer 2004, 101(5):957-962.

[9] Siefker-Radtke AO, Dinney CP, Abrahams NA, Moran C, Shen Y, Pisters LL, Grossman HB, Swanson DA, RE M: Evidence supporting preoperative chemotherapy for small cell carcinoma of the bladder: a retrospective review of the M. D. Anderson cancer experience. J Urol 2004, 172(2):481-484.

[10] Abrahams NA, Moran C, Reyes AO, Siefker-Radtke A, AG A: Small cell carcinoma of the bladder: a contemporary clinicopathological study of 51 cases. Histopathology 2005, 46(1):57-63.

[11] N I: A rare bladder cancer--small cell carcinoma: review and update. Orphanet J Rare Dis 2011, 6:75.

[12] McConkey DJ, Lee S, Choi W, Tran M, Majewski T, Lee S, Siefker-Radtke A, Dinney C, B C: Molecular genetics of bladder cancer: Emerging mechanisms of tumor initiation and progression. Urol Oncol 2010, 28(4):429-440.

[13] Bassi P, Ferrante GD, Piazza N, Spinadin R, Carando R, Pappagallo G, F P: Prognostic factors of outcome after radical cystectomy for bladder cancer: a retrospective study of a homogeneous patient cohort. J Urol 1999, 161(5):1494-1497.

[14] Stein JP, Lieskovsky G, Cote R, Groshen S, Feng AC, Boyd S, Skinner E, Bochner B, Thangathurai D, Mikhail M et al: Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients. J Clin Oncol 2001, 19(3):666-675.

[15] Dalbagni G, Genega E, Hashibe M, Zhang ZF, Russo P, Herr H, V R: Cystectomy for bladder cancer: a contemporary series. J Urol 2001, 165(4):1111-1116.

[16] Hussain SA, ND J: The systemic treatment of advanced and metastatic bladder cancer. Lancet Oncol 2003, 4(8):489-497.

[17] Sternberg CN, Yagoda A, Scher HI, Watson RC, Geller N, Herr HW, Morse MJ, Sogani PC, Vaughan ED, Bander N ea: Methotrexate, vinblastine, doxorubicin, and cisplatin for advanced transitional cell carcinoma of the urothelium. Efficacy and patterns of response and relapse. Cancer 1989, 64(12):2448-2458.

[18] Racioppi M, D'Agostino D, Totaro A, Pinto F, Sacco E, D'Addessi A, Marangi F, Palermo G, PF. B: Value of current chemotherapy and surgery in advanced and metastatic bladder cancer. Urol Int 2012, 88(3):249-258.

[19] Chiu PH, Hsieh HY, SC. W: Prescriptions of traditional Chinese medicine are specific to cancer types and adjustable to temperature changes. PLoS One 2012, 7(2):e31648.

[20] Liu C, Li XW, Cui LM, Li LC, Chen LY, XW Z: Inhibition of tumor angiogenesis by TTF1 from extract of herbal medicine. World J Gastroenterol 2011, 17(44):4875-4882.

[21] Fang L, Wang Z, Kong WY, Feng JG, Ma SL, NM. L: Anti-tumor and apoptotic effects in vitro and in vivo of a traditional Chinese medicine prescription. Chin Med J (Engl) 2011, 124(21):3583-3587.

[22] Zheng G, Jiang M, He X, Zhao J, Guo H, Chen G, Zha Q, A L: Discrete derivative: a data slicing algorithm for exploration of sharing biological networks between rheumatoid arthritis and coronary heart disease. BioData Min 2011, 4:18.

[23] Jiang M, Zhang C, Zheng G, Guo H, Li L, Yang J, Lu C, Jia W, A L: Traditional chinese medicine zheng in the era of evidence-based medicine: a literature analysis. Evid Based Complement Alternat Med 2012, 2012:409568.

A New Graphical Method to Analyze the Tones of Chinese Monosyllable, based on the Modified AWT Time Domain Accumulative Energy

Zhiyong Deng
College of Music
Capital Normal University
No.105 Xisanhuan North Road
Haidian, Beijing, CHINA
+86 18610319780, 100048
dzy@cnu.edu.cn

Daiwei Wang
College of Music
Capital Normal University
No.105 Xisanhuan North Road
Haidian, Beijing, CHINA
+86 15201113815, 100048
davidyx@qq.com

ABSTRACT
As well known, there are four types of tones for the Chinese language. For the limitation to analyze these tones of Chinese monosyllable by the classical Fourier transform and spectrogram, by means of a auditory wavelet transform with the approximate auditory wavelet function based on db4, a new graphical method named the modified AWT time domain accumulative Energy (MECN) is introduced in this paper. Especial for the third tone of the typical Chinese monosyllables, the inverse 'S' shape of MECN has its unique character compared to the other three tones'.

Categories and Subject Descriptors
G.4 [**MATLAB**]

General Terms
Algorithms

Keywords
Auditory wavelet, Approximate auditory wavelet function, Time domain accumulative energy.

1. INTRODUCTION
As the developing of national economy and information technology, a more detailed research would be needed for the Chinese speech processing. Since the research by Yuanren Zhao from 1920s, many speech processing for Chinese language focused on spelling, words recognition, sound synthesis, speech information retrieval and voice coders are based on the characteristics of time-domain and the Fourier transform and spectrogram [1]. However, due to the Fourier transform and spectrogram of their own limitations, it is difficult to obtain the high-dimensional information for speech, such as tone or emotion characters [2].

Over the last several decades, numerous studies have shown that the wavelet transform with a high resolution of low-frequency and a low one of hi-frequency is suitable for analysis of time-varying and non-stationary transient signal for human hearing due to its similarity to the auditory filter. In 1990s, the concept of auditory wavelet transform (AWT) is put forward by means of using an auditory filter as the wavelet function. That means if a kind of auditory filter could be transit to a wavelet function, a better psychoacoustic model for speech processing would be easily acquired [3][4].

2. AWT WITH THE APPROXIMATE AUDITORY WAVELET FUNCTION BASED ON db4
Set $\varphi(t)\in L^2(R)$, when its Fourier transform $\overline{\Psi}(\omega)$ meets the allowable conditions of $\int_R \frac{\Psi(\omega)}{\omega}d\omega<\infty$, $\varphi(t)$ is called a wavelet or wavelet basis function, thus, when a signal $x(t)\in L^2(R)$, its continuous wavelet transform (CWT) is shown as followed:

$$CWT_x(a,\tau)=\left\langle x(t),\varphi_{a,\tau}(t)\right\rangle=\frac{1}{\sqrt{a}}\int_R x(t)\cdot\overline{\varphi(\frac{t-\tau}{a})}dt \quad (1)$$

a – Scaled factor (related to frequency domain)

τ – Shift factor (related to time domain)

Since an auditory filter can be represented to a wavelet basis function $\varphi(t)$, the CWT refers to an AWT.

Then take the scaled factor a as formula (2), when the sampling frequency of signals sets to 44.1kHz, the db4 wavelet basis function can be modified to a set of approximate auditory filters similar to the critical band-pass filters [5] shown in Figure 1, while its continuous wavelet family $\varphi_{a,\tau}(x)$ as formula (3). Then as Table 1 shown, the frequency corresponded to each scaled factor a can be within 31Hz to 31500Hz related to the range of auditory frequency from 20Hz to 20000Hz, while the other wavelet functions with incompatible ranges in Table 2.

$$a_i=2^{i/3} \quad (2)$$

$$\phi_{a,\tau}(x)=\frac{1}{\sqrt{a}}\overline{\phi(\frac{x-\tau}{a})} \quad (3)$$

Thus based on this approximate auditory wavelet family of db4, the AWT spectrum vs. its FFT spectrum and spectrogram for a typical Chinese monosyllable of '跟(gēn)' are shown in Figure 2.

Supported by a Youth Funding Project (No.11YJCZH026) of Humanities and Social Sciences Foundation from the Education Ministry of CHINA.

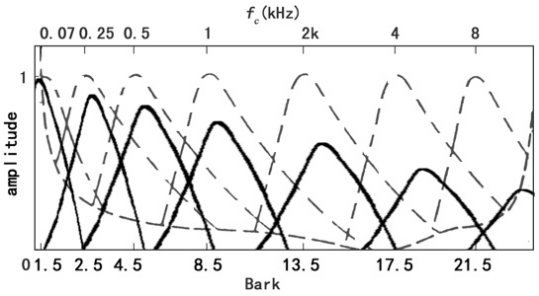


Figure 1. Frequency response of db4 wavelet functions (solid lines) vs. the auditory critical band-pass filters (dotted lines)

Table 1. The Central Frequency of db4 wavelet functions

i	a_i	f_c（Hz）	z（Bark）	Δf_G（Hz）
1	1	31500	–	–
2	2	15750	23.5	3500
3	2	15750	23.5	3500
4	3	10500	22.5	2500
5	3	10500	22.5	2500
6	4	7875	21.5	1800
7	5	6300	19.5	1100
8	6	4500	18.5	900
9	8	3938	17.5	700
10	10	3150	16.5	700
11	13	2423	14.5	380
12	16	1969	12.5	280
13	20	1575	11.5	280
14	25	1260	10.5	210
15	32	984	8.5	160
16	40	788	7.5	150
17	51	618	5.5	120
18	64	492	4.5	110
19	81	389	3.5	100
20	102	308	3.5	100
21	128	246	2.5	100
22	161	196	2.5	100
23	203	155	1.5	100
24	256	123	1.5	100
25	323	98	1.5	100
26	406	78	0.5	100
27	512	62	0.5	100
28	645	49	↓	↓
29	813	39	0	0
30	1024	31		

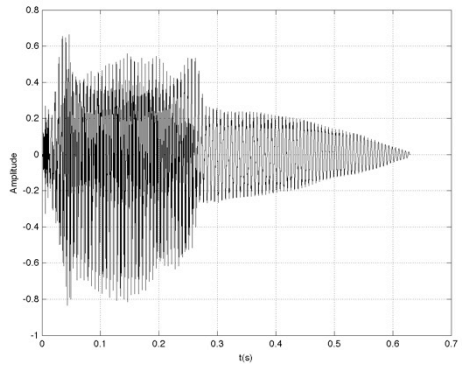
i – Wavelet sub-band, a_i – Scaled factor (related to frequency domain)

f_c – Central frequency, z – Auditory critical band

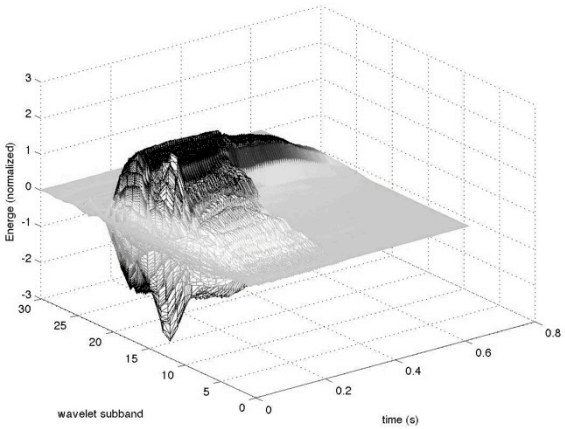
Δf_G – Auditory critical bandwidth

Table 2. Frequency response range of db4 vs. other common wavelet functions, sampling frequency of 44.1kHz

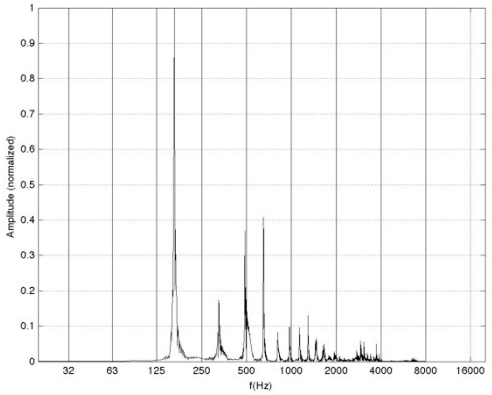
Wavelet function	Frequency response range (Hz)
db4	31~31500
Marr	11~11025
Haar	43~43928
Morlet	35~35831



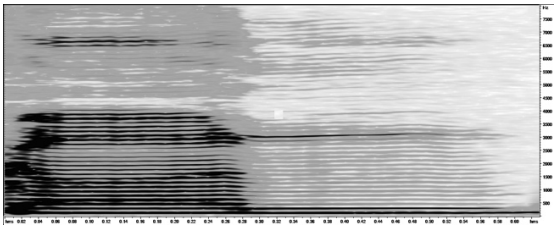
a. Signal waveform



b. AWT spectrum



c. FFT Spectrum



d. Spectrogram

Figure 2. Auditory Wavelet Transform for a typical Chinese monosyllable of '跟(gēn)'

3. MODIFIED AWT TIME DOMAIN ACCUMULATIVE ENERGY

3.1 CWT Accumulative Energy

The continuous wavelet transform (CWT) accumulative energy defined as formula (5) was suggested as a parameter to describe the distinctness of varied musical instruments tones by Bzoena Kostek in 1996 [6].

$$E_c(n) = \sum_{i=1}^{n} |c_i|^2 \tag{5}$$

n – sample number, c_i – CWT coefficient

3.2 MECN, a Modified AWT Time Domain Accumulative Energy

According to the linear additive for the CWT coefficients, the CWT accumulative energy as formula (5) can be redefined in the whole time duration as followed.

$$E_c(n) = \left[\sum_{j=n_0}^{n} \sum_{i=m_0}^{m} |c_{i,j}|^2 \right] \tag{6}$$

n – sample No. n, m – wavelet sub-band m

n_0 – the first sampling point of time domain in CWT

m_0 – the first wavelet sub-band of CWT, $c_{i,j}$ – CWT coefficient

Taken Δt as the whole duration of a transient signal, provided the initial moment of $E_c(0)=0$, t moment of $E_c(t)=1$, $E_c(n)$ can be normalized to [0,1], then the change of time-frequency domain is shown in a simple line of $E_c(n)$ temporal curve. If using the approximate auditory wavelet function as formula (2) and formula (3), this temporal curve can be named as a modified AWT time domain accumulative energy (MECN, see Figure 3).

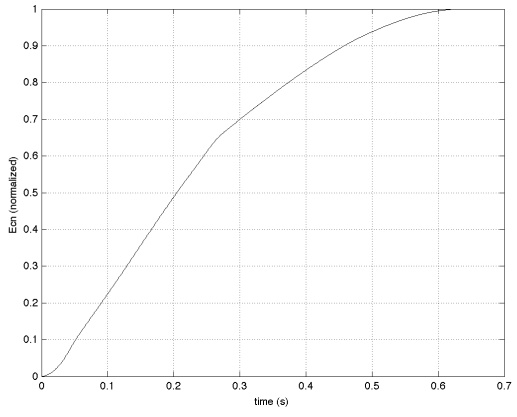
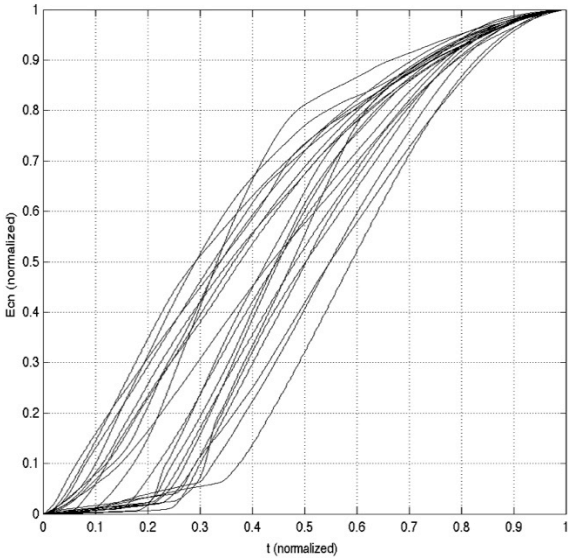


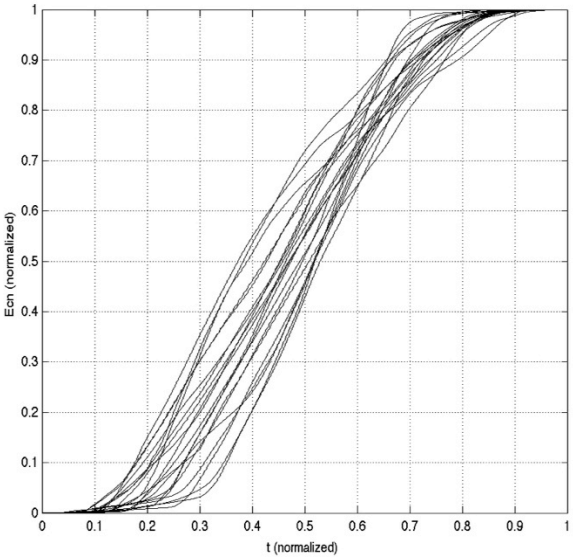
Figure 3. Modified AWT time domain accumulative energy (MECN) curve for a typical Chinese monosyllable of '跟(gēn)'

4. APPLICATIONS AND DISCUSSION

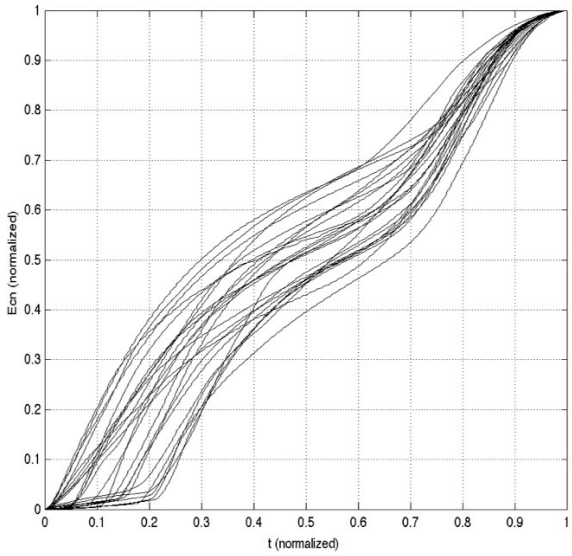
As well known, there are four types of tone for the Chinese language, included 'Yin Ping' (the first tone), 'Yang Ping' (the second tone), 'Shang Sheng' (the third tone) and 'Qu Sheng' (the fourth tone). In order to compare the MECN curves of all the signals with different lengths, according to the formula (6), normalized the whole duration of each signal to [0, 1], then, for some typical Chinese monosyllables, the MECN curves for each tone type are shown in Figure 4.



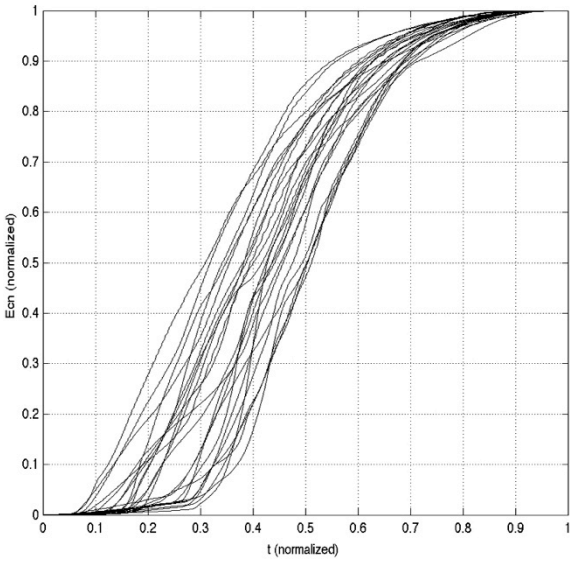
a. MECN curves for 21 typical Chinese monosyllables in the first tone ('Yin Ping')



b. MECN curves for 20 typical Chinese monosyllables in the second tone ('Yang Ping')



c. MECN curves for 25 typical Chinese monosyllables in the third tone (‘Shang Sheng’)



d. MECN curves for 24 typical Chinese monosyllables in the fourth tone (‘Qu Sheng’)

Figure 4. MECN curves for 90 typical Chinese monosyllables, classified by the tone type

If put all the curves in Figure 4 as Figure 5 together, it is obvious to find that, the curves for those Chinese monosyllables in the third tone have an unique inverse ‘S’ shape compared to the other three tones’. That means MECN curve tells a big difference for the Chinese monosyllables in the third tone from the others.

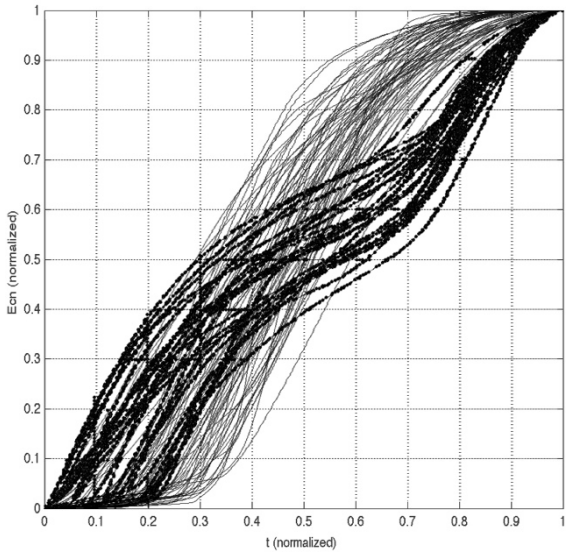


Figure 5. MECN curves of those in the third tone (dark dotted lines) vs. those in the other three tone types (solid lines)

Compared to the spectrum and spectrogram in Figure 2.c and Figure 2.d, the MECN curve as Figure 3 is simpler to a single line. And furthermore, this kind of curves can not only show the characters in the time domain and frequency domain at the same time, but the value for each point contains the accumulative energy both in the frequency domain and time domain, which is similar to the overall cognitive and the time persistence effect during the process of the real auditory perception for that kind of transient speech signals like Chinese monosyllables.

5. SUMMARY

According to what had been discussed above, the modified AWT time domain accumulative energy (MECN) curves provide a simple and graphical method to analyze the tones of Chinese monosyllables. This can be a cue to detect the relationship between the signal processing characters and the auditory perception. However, based on the limited word samples, the method and analysis are acquired in ideal laboratory conditions, many researches on factors effecting to the MECN curve, such as the speech accent, the sound loudness, the speaking rate, the hearing habits and even the ear training background, would be the needs for further study. And then, the wavelet function for AWT is also a critical feature for this graphical method.

6. REFERENCES

[1] Li Zhao. 2011. *Speech Signal Processing*. China Machine Press.

[2] L. R. Rabiner, R. W. Schafer. 2011. *Theory and Applications of Digital Speech Processing*. PHEI (China).

[3] M. N. Souza and L. P. Caloba. 1997. A comparison between fourier and biological auditory based time-frequency distributions. *Proceesdings of IEEE*, vol. 97, 807-810.

[4] Y.Salimpour, M.D.Abolhassani. 2006. Auditory wavelet transform based on auditory wavelet families. *Proceedings of the 28th IEEE EMBS Annual International Conference*, NY (Aug. 2006),1731-1734.

[5] R. PLomp, W.J.M Levelt. 1965. Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.*, vol. 26, 548-560.

[6] Bozena Kostek. 2005. Perception-Based Data Processing in Acoustics-Application to Music Information Retrieval and Psychophysiology of Hearing. Springer.

Using Business Intelligence to Analyze Google Circles

Kornelije Rabuzin
University of Zagreb, Faculty of organization and informatics Varazdin
Pavlinska 2
42000 Varazdin, Croatia
+385(0)42/390-847
kornelije.rabuzin@foi.hr

ABSTRACT
Google company often surprises us with their new services; one such service is "Google circles". The main concept of this service is the possibility to organize friends and (for example) co-workers in circles in order to share information more efficiently or just to stay in touch with one another. The main idea of this paper is to analyze how such circles are formed by using business intelligence capabilities; it is strongly believed (assumed) that people in a circle have (share) similar habits. This paper analyzes how business intelligence (i.e. analytical processing) can be used to analyze sales data based on the circle membership.

Categories and Subject Descriptors
H.2.7 [**Information Systems**]: Database Administration – *data warehouse and repository.*

General Terms
Design, Experimentation.

Keywords
Google Circles, Business Intelligence, Data Warehouses, SQL.

1. INTRODUCTION
An interesting new Google service i.e. Google circles is very important and analyzed in this paper. Google circles represent (in fact) a way to organize family, friends, co-workers, acquaintances, etc.; one circle can include people from work, other circle can include family members, etc. ([1]). By forming such circles you can easily share some interesting things or you can just stay connected etc. (Figure 1).

Business intelligence (BI) and data warehouses are used very much these days. It is a well known fact that more money is invested in data warehouses and business intelligence (i.e. data integration) then in databases (as such). Although databases represent an excellent mechanism that is used to store and manipulate data, the problem that has become obvious is that different applications have been built during the years and no one had a picture of how they should fit together. Although those applications contained relevant data, it would be almost impossible to integrate them (from several applications) and produce a meaningful report (scattered data represents a huge problem). Physical integration of data results in data warehousing that is accompanied by business intelligence i.e. advanced data analysis mechanisms that can be used to analyze data in the data warehouse ([3], [4], [5]).

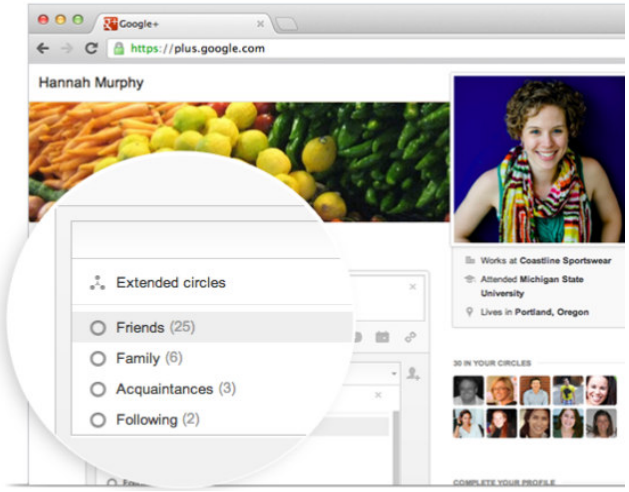


Figure 1. Google circles [1]

It is assumed that people are familiar with the basic concepts of data warehouses and business intelligence. Basically, in a data warehouse we distinguish two types of tables (dimension and fact) that usually form a star schema (or sometimes a snowflake schema). Because of that form of data organization many front-end tools can be used to intuitively analyze the data. This intuitive data organization is easy to understand and because of that even less educated (experienced) end-users can use the developed data warehouse to build reports. Combined with advanced mechanisms for data analysis (drill down, roll up, slice, dice, etc.), the data from the data warehouse can be turned into very useful piece of information.

However, although data warehousing is intuitive, data organized in this way can be exposed to data mining techniques as well. During the data warehousing Extract-Transform-Load (ETL) process the data is extracted from many sources, cleaned and stored in a unified format suitable for data mining as well.

Since Google circles represent certain structures, they can be used to analyze the data by means of BI tools and advanced analytical capabilities. The rest of the paper is organized as follows; the next part contains and describes the developed data warehouse (it is small but serves the purpose). After that the BI tool is used for analysis purposes taking into account already formed Google circles (some interesting reports are built). Finally the conclusion is presented.

Research Notes in Information Science (RNIS)
Volume13, May 2013
doi:10.4156/rnis.vol13.29

2. THE DATA WAREHOUSE

For the purpose of this paper the following scenario is used (it is artificial); let's assume that people (that one person has in his circles) shop in one store (they buy some products listed in the products table). Since people can be assigned to many different circles, the assumption is that it would be possible to use this "structural information" to analyze sales data i.e. circles could be used to analyze and predict customer's behaviour (for certain circles).

Let's assume that we analyze sales data in this artificial store; we have several dimension tables (date, time, product etc.) and a fact table containing the sales data. Regarding the data model it is obvious that one person can have (form) several circles and each circle belongs to exactly one person. Further on, one circle can contain many people and a person can be added to many different circles (owned by different people). People from those circles buy in our artificial store and their purchases are recorded in a transaction table.

The data warehouse was implemented in MS SQL Server 2008. As a BI tool the Business Objects XI was used; the model was first implemented in the Designer tool and reports were built using the Desktop Intelligence tool. The model that was implemented in the Designer tool is given below (Figure 2).

Although some dimension tables should (normally) have much larger number of attributes, this is not the case in this example because it is not necessary to add many attributes to understand the basic idea. Because of that the example seems to be rather simple, but it will serve the purpose.

More on Business Objects XI can be found in [2]. More on data warehousing can be found in [3], [4] and [5].

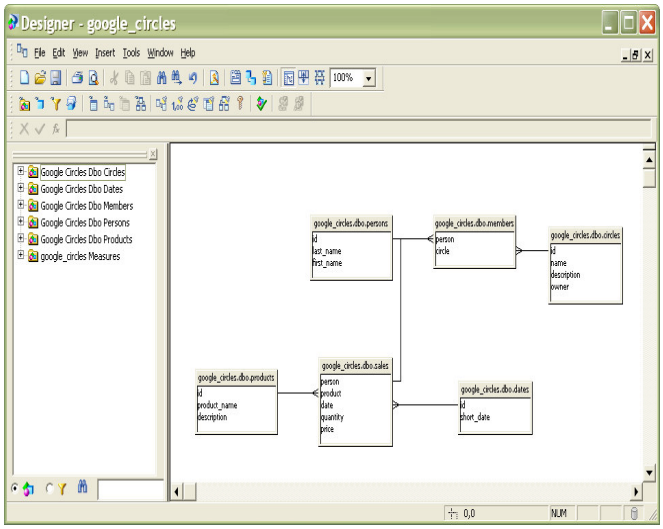


Figure 2. The data warehouse model

3. DESKTOP INTELLIGENCE TOOL

Once the data warehouse model is implemented i.e. tables are created and filled in with some test data, the Desktop Intelligence tool can be used for data analysis. The first report (table) shows how circles are formed:

Table 1. Circle structure

Circle_name	Circle_owner	Last Name	First Name
Family	Smith Jack	Smith	Joan
Family	Smith Jack	Smith	John
Family	Smith Jack	Smith	Joodie
Football	Smith Jack	Brown	Peter
Football	Smith Jack	Jones	Tim
Friends	Smith Joan	Jones	John
Friends	Smith Joan	Jones	Mary
Friends	Smith Joan	Smith	Joodie
Work	Jones Tim	Smith	Jack
Work	Smith Jack	Jones	Tim

We see that the first circle (Family) belongs to Smith Jack and that it contains three family members. The second circle (Football) belongs to the same person and contains two persons (Brown Peter and Jones Tim), etc. We can see that Smith Jack has formed three circles.

To produce such a report it is enough to use the co called "drag & drop" principle in the Query panel (Figure 3):

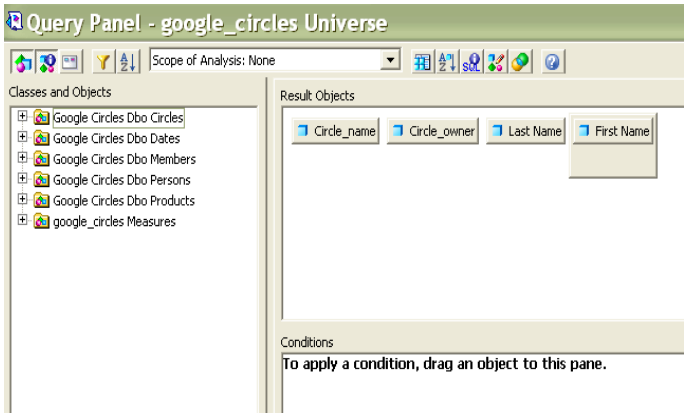


Figure 3. Query panel

For a start let's build a report to see who bought what (Table 2):

Table 2. Sales data

Last Name	First Name	Sum of Quantity	Sum of Price
Brown	Peter	2.00	250.00
Jones	Tim	3.00	350.00
Smith	Jack	3.00	350.00

We see that Brown Peter bought two things and that he spent 250$, and so on. Based on the previous report we can start posing following questions; how much money did the family members spend? How much money did the co-workers spend? Since people are placed in different circles, we can build a report that shows how much money the Smith's co-workers spent (Table 3; the circle name is "Work"):

Table 3. Work circle sales data

Circle_name	Circle_owner	Last Name	First Name	Sum of Quantity	Sum of Price
Work	Smith Jack	Jones	Tim	3.00	350.00

Now we can change the selection condition and check how much money the "Football circle" members spent (Table 4); this circle was formed by the same person i.e. Smith Jack:

Table 4. Football circle sales data

Circle_name	Circle_owner	Last Name	First Name	Sum of Quantity	Sum of Price
Football	Smith Jack	Brown	Peter	2.00	250.00
Football	Smith Jack	Jones	Tim	3.00	350.00
			Sum:	5.00	600.00

We see that the "Football" circle members spent 600$ and bought 5 items. Now we can change the view and see how much money did Smith's family members spend:

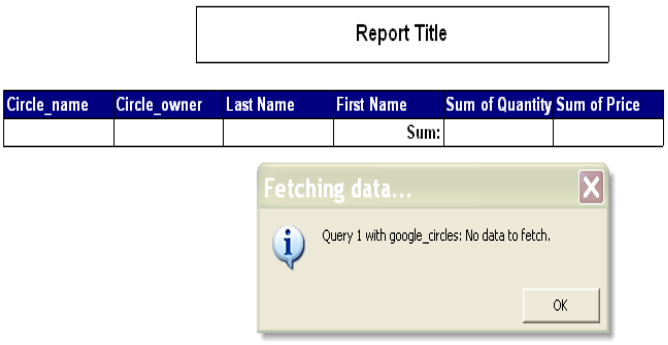


Figure 4. Family circle sales data

We can see that family members did not spend any money in that store. By using the slice and dice mode we can reorganize the view and show all data in one report; the report shows sales data for all circles formed by Smith Jack:

Table 5. Smith's circles sales data

	Football		Work	
Smith Jack	5.00	600.00	3.00	350.00

Family members i.e. the "Family" circle is not shown because of the way the tables are joined (outer join is not used); since they didn't spend any money, that circle is not shown on the report. However we can transform the query and use outer join syntax (this query was rearranged manually) to add the missing circle to the report (Table 6):

```
SELECT
  google_circles.dbo.circles.name,
  google_circles.dbo.circles.owner,
  sum(google_circles.dbo.sales.quantity),
  sum(google_circles.dbo.sales.price)
FROM
  google_circles.dbo.circles
LEFT JOIN google_circles.dbo.members ON (
google_circles.dbo.members.circle=google_circles.dbo.circles.id
)
  LEFT JOIN google_circles.dbo.persons ON (
google_circles.dbo.persons.id=google_circles.dbo.members.perso
n )
  LEFT JOIN google_circles.dbo.sales ON (
google_circles.dbo.persons.id=google_circles.dbo.sales.person )
WHERE google_circles.dbo.circles.owner = 'Smith Jack'
GROUP BY
  google_circles.dbo.circles.name,
  google_circles.dbo.circles.owner
```

Table 6. Smith's circles sales data (revised)

	Family		Football		Work	
Smith Jack			5.00	600.00	3.00	350.00

We can see that the "Football" circle members spent the most money and bought the most items. This could mean that guys exchange information before or after the game and this information sharing may be crucial to explain why they buy certain items.

Now let's assume that we are interested in circles in which Jones Tim can be found (Table 7):

Table 7. Circles that contain Jones Tim

Last Name	First Name	Circle_name	Circle_owner
Jones	Tim	Football	Smith Jack
Jones	Tim	Work	Smith Jack

We are now not interested in the circle owner (any more); we select one member (for example Jones Tim) and for that member we determine the amount of money that Jones spent as well as total amount of money spent in the circles where he is a member (Table 8):

Table 8. Member and circle sales data

Last Name	First Name	Circle_name	Circle_owner	Member quantity	Member amount	Circle quantity	Circle Amount
Jones	Tim	Football	Smith Jack	3.00	350.00	5.00	600.00
Jones	Tim	Work	Smith Jack	3.00	350.00	3.00	350.00

Further on, it would be interesting to add product categorization to some reports as well; in that way we could tell which circle shows interested for which product category and that would reveal even more interesting information by applying drill down / roll up mechanism.

Another interesting report deals with dates i.e. when purchases took place:

Table 9. Sales data by date

Short Date	Sum of Quantity	Sum of Price
2013-01-10	6.00	700.00
2013-01-17	2.00	250.00

One of the things immanent to BI tools is that same things can be displayed in different ways. The previous table can be rotated or turned into a graph in a matter of seconds:

Table 10. Rotated data view

Short Date	2013-01-10	2013-01-17
Sum of Quantity	6.00	2.00
Sum of Price	700.00	250.00

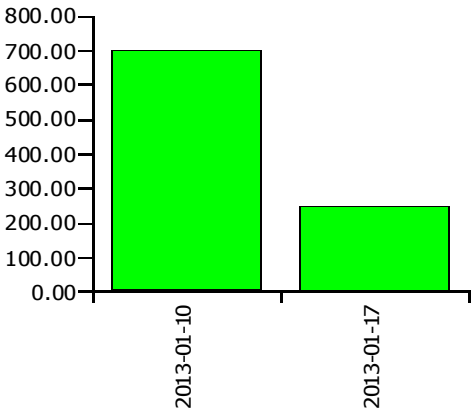


Figure 5. Chart view

Since we know that guys play football each Wednesday, it is quite interesting that so many products were sold on Thursday (i.e. the Football circle members usually shop the next day):

Table 11. Sales data by circle and date

Short Date	Circle_name	Sum of Quantity	Sum of Price
2013-01-10	Football	3.00	350.00
2013-01-17	Football	2.00	250.00

If we had a complete date dimension, then drill down and roll up capabilities would reveal us even more interesting things (for example that the previous statement is true but only in the beginning of a month).

4. CONCLUSION

The purpose of this paper was to show how Google circles and analytic capabilities could be used to analyze sales data based on the circle membership. It was assumed that people that could be found in one circle had similar habits. Although the example was artificial, on a number of reports we drawn some interesting conclusions based on the sales data and circle membership.

The example is artificial but the idea is applicable and interesting for analyzing the data. Data mining on circles could reveal even more information but this is to be done in the future papers.

5. REFERENCES

[1] Google + features, Google circles, http://www.google.com/intl/en/+/learnmore/features.html#2

[2] Howson, C. 2006. *BusinessObjects XI: The Complete Reference*, The McGraw-Hill Companies, USA.

[3] Inmon, H. W. 2002. *Building the Data Warehouse* – Third Edition. Wiley Computer Publishing, USA.

[4] Kimball, R. and Caserta, J. 2004. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Wiley Publishing, USA.

[5] Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., and Becker, B. 2008. *The Data Warehouse Lifecycle Toolkit* – Second Edition. Wiley Publishing, USA.

Performance of a Dynamic Population Solution agents in Water Flow-Like Algorithm

Ayman Ibraheem Srour
Faculty Information Science and Technology,
Universiti Kebangsaan Malaysia
aymansrour@gmail.com

Zulaiha Ali Othman
Faculty Information Science and Technology,
Universiti Kebangsaan Malaysia
zao@ftsm.ukm.my

Abdul Razak Hamdan
Faculty Information Science and Technology,
Universiti Kebangsaan Malaysia
arh@ftsm.ukm.my

ABSTRACT
A metaheuristic is a well-known optimization method for tackling a hard optimization problem. On the other hand, a population based metaheuristic has shown as promising performance in solution ability in many domains such as GA, ACS and Bees algorithm. However, its suffering from high computation time to reach solution especially for large data set, due to the nature of the algorithm is based on fixed number of the solution agent. Water flow-like algorithm (WFA) is inspired by a natural behavior of water flowing from higher level to lower one. The flows of water can split or merge according to scenery of the surface. WFA is self adaptive and dynamic in term of population size and parameter setting. Therefore, this paper presents the performance of WFA as dynamic population based solution using large data set in TSP. which are 1400 and 1665 number of cities). The result shows that WFA has reduced computation time up to 34% and 61% compare with ACS respectively. WFA also has quality of solution 4.1% and 7.4% from ACS respectively. This research concludes that WFA is suitable algorithm use for optimization problem solution especially for large data.

Categories and Subject Descriptors
I.2.8 **[Artificial Intelligence]**: Problem Solving, Control Methods and Search – heuristic methods.

General Terms
Algorithms

Keywords
Water flow-like algorithm; Dynamic solution agents, Metaheuristics; Combinatorial optimization

1. INTRODUCTION
Metaheuristics are general algorithmic frameworks that designed for solving a hard Combinatorial Optimization (CO). They are a viable choice for solving such NP-hard problems, as it recursively employs a computational method to optimize a problem by enhancing the candidate solutions for the objective function. A metaheuristic can optimize complex problems by searching through many candidate solutions with a few or no assumptions about the problem being solved and without any guarantee of finding the optimal solution. However, a wide range of metaheuristic algorithms have been developed over the past years that are varies in its nature, structure and complexity.

Two main categories of metaheuristics have been recorder in the literature, the single solution agent and population solution agents. An examples of the single solution agent include Simulated Annealing (SA) [1; 4] and Tabu Search (TS) Glover [3] which are designed based on a single solution search. On the other hand, the population based solution agents have designed based on multiple solutions search called ―population" and the objective would be to guide that search in state space to reach to the optimal solution. Example of population based metaheuristics include Genetic Algorithms (GA), Particles Swam Optimization (PSO), Ant Colony Optimization (ACO) [2].

In general, a single solution metaheuristics is usually search the solution space step by step. Although the search might be inefficient, a gradual improvement in an objective function can be obtained and used to effectively guide the solution search toward prominent areas. Therefore, single solution metaheuristics are suitable for combinatorial optimization problems with a smooth solution space; but not for difficult problems with several local optima. Population based solution metaheuristics, on the other hand, can search the solutions space more efficiently by using the set of solution agents; they suffer the quick convergence and redundant searches. However, computation resource is wasted in unavoidable redundant searches [10]. Yang and Wang [10] state another drawback of population based metaheuristics that come from the static number of the population size throughout the optimization process which make determining the proper population size is not easy task. However, it can be seen that both single and population methods are agile enough to perform an efficient and effective solution search. Therefore, to tackle this issues, Yang and Wang [10] proposed a new and novel metaheuristic algorithm so-called Water flow-like Algorithm (WFA). It uses a new concept of dynamic agent size. The authors assumption relies on two main issues that affect the efficiency of algorithm optimization. The first issue is the redundant search, which increases the computational cost to the algorithm during the optimization process. It causes a redundancy problem by combining the solution agents that share the same objective value. The second issue is the algorithm's adaptation ability and parameter tuning in the optimization process, i.e., setting up the optimal population size in a GA or an ACO variant based on a specific problem and problem space. In the conventional population based metaheuristic, the population sizes are usually fixed and cannot be tuned during the optimization process.

2. WATER FLOW-LIKE ALGORITHM
The water flow-like algorithm (WFA) [10] is categorized as a metaheuristic algorithm, inspired by the natural behavior of water flowing from higher to lower levels. The water flows can split or merge according to the surface scenery. The WFA has been successfully adapted and applied in different optimization problems, including the bin-packing [10], manufacturing cell fraction [9] and nurse scheduling problems [8].

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.30

The WFA offers a new method for computational optimization. The advantages of the WFA are that it is self-adaptive and dynamic in its population sizes and parameter settings. The solution agent size is not fixed, unlike in the ACS algorithm. The flow number is subject to increase or decrease during the optimization process. The population size changes are based on the problem diminution and solution quality found by the agents. However, Yang and Wang [10] describe and map the dynamic size of the solution agents based on the natural behavior of water flows as they split, move and merge.

The first WFA version was developed to solve the object grouping problem, called the bin-packing problem (BPP), which is a discrete optimisation problem and well known as an NP-Hard problem. The BPP with a tight capacity constraint has required several heuristic methods in terms of the derivation of feasible and optimal solutions. The traditional bin-packing problem minimises the number of bins used and is subject to a weight capacity constraint. However, the authors have used the BPP as a benchmark to measure the feasibility of the WFA for solving such optimisation problems. The proposed algorithm mainly relies on solution neighbour searches. The authors use a one-step move strategy to solution neighbour with a constant step moving for each flow. The performance of the WFA is compared with that of the GA, PSO and ACO. The experimental results show that the WFA outperformed the GA, PSO and ACO in both quality and execution time. Based on the experimental results, the authors conclude that WFA can solve difficult optimisation problems and suggest applying the WFA to solve sequencing problems such as the TSP.

In 2010, Wu et al. improved the WFA, known as the WFACF model, to solve the manufacturing cell fraction problem. The model utilises the similarity coefficient and machine assignment methods, as well as part assignments, to generate an initial feasible solution in the first stage and flow splitting and moving in the second stage to improve the solution using a neighbour search to obtain a near-optimal solution. The results show that the WFA outperformed the hybrid genetic algorithm (HGA) and the SA.

The WFA has also demonstrated good performance in multi-objective problems. [8] used the WFA to solve the nurse scheduling problem (NSP), which is a multi-objective optimisation problem. They compared the WFA with the differential evaluation algorithm (DE), and the results showed better solution quality when using the WFA.

The water flow-like algorithm uses an agent solution where the solution agents are mapped as water flows and the objective functions are mapped as terrains. Flow splitting occurs when rugged terrains are traversed. Conversely, water flows merge with each other when they join at the same point. The basic basic operation of the water flow-like algorithm can be seen in Figure 1.

3. WFA FOR TRAVELING SALESMAN PROBLEM

The travelling salesman problem (TSP) is a classic combinatorial optimisation problem that attracted many researchers from different fields, including operational research, mathematics and several scientific and engineering fields. The TSP solution has contributes to the real applications such as planning, scheduling, navigation, stock market, transportation and logistics problems. The TSP is well known as a non-deterministic polynomial (NP-hard) problem, and the determination of the exact solution is difficult [5]. The TSP searches for the shortest path among a set of cities with known distances between a pairs of cities, and it can be formulated as a complete graph with a set of vertices, which is a set of edges weighted by the distance between two vertices (cities). The problem is finding the most inexpensive Hamiltonian cycle associated with visiting each city exactly once and returning to the original city.

In this paper we use TSP problem as a case study to measure the performance and scalability of the WFA in a large size problem instances. This section presents the WFA algorithm applied to the TSP. Figure 1 show the basic WFA algorithm applying in TSP.

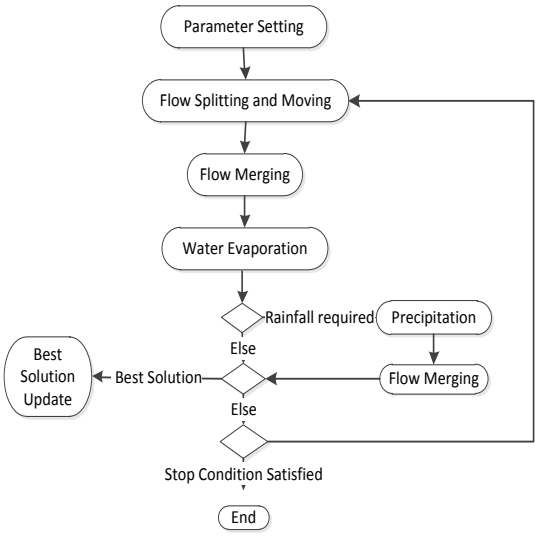


Figure 1: The basic WFA operations

Water flow movements (location changes) are influenced by the gravitational force and the energy conservation law. Iteration by iteration, water constantly moves to lower altitudes, correlating with improvements in the solution search. The WFA beings searching within the problem space using one solution agent (flow) with an initial momentum. Subsequently, the flow splits into multiple subflows when they encounter rugged terrains and if the flow momentum exceeds the splitting amount. A flow with more momentum generates more subflow streams than one with less momentum. A flow with limited momentum yields to the landform and maintains a single flow.

Many flows merge into one flow when they obtain the same objective values. The WFA reduces the number of solution agents when multiple agents move to the same location to avoid redundant searches.

Water flows are subject to water evaporation in the atmosphere. The evaporated water returns to the ground as rain. In the WFA, part of the water flow is manually removed to mimic water evaporation. A precipitation operation is implemented in the WFA to simulate natural rainfall and explore a wider area.

Based on the above WFA behaviours, the WFA contains four main operations: flow splitting and moving, flow merging, water evaporation and water precipitation. The following section describes further detail each operations:

3.1 Initial parameter settings

Before we start the searching of WFA, several parameters should be assigned as first. Maximum generation G, initial flow solution $flow_i$, initial mass W_0, initial velocity V_0, the base momentum T for splitting, the upper bound of subflows spited from single flow $\bar{n}$. The gravity g, and periodical precipitation generations t. The flows with larger kinetic energy have the potential to split into several sub-flows. And the number of sub-flows is decided from the kinetic energy. Hence, the value $M_0V_0=2T-3T$ was suggested by [10]. And they also suggested the limit of the number of sub-flows value, *2* or *3*, to avoid the rushing increase of the number of sub-flows. In addition, the periodical precipitation generation satisfies $t>G/20$.

A WFA starts with only one water flow (a single solution agent) whose location is assigned based on the initial solution value. The initial solution is generated using a greedy constructive method to generate a close approximate solution for the problem. The nearest neighbour (NN) heuristic is a famous heuristic that can generate a good near-optimal initial solution.

3.2 Flow Splitting and Moving

The WFA is uniquely characterised by dynamic agent sizes when searching the solution space, as mentioned above; the WFA begins with a single flow as the starting point from which to explore the solution space. Several flows branch from it based on the quality of the discovered solutions. Consequently, the flow splits based on its momentum; namely, a flow with a higher momentum generates more subflows than one with less momentum. Considering the gravitational force and the energy conservation law, the flow moves to a new location. The locations of the split subflows are derived from the neighbouring locations of the original flow.

Let N be the number of water flows in the current iteration. In the water flow operation, the number of subflows n_i that branched from $flow_i$, where $i \in 1,2,\ldots, N$, is determined by flow momentum, W_iV_i. A flow with zero momentum does not split and stays at the same location; its solution is considered a stagnant solution.

Conversely, a flow can split into subflows if its momentum value exceeds a base momentum T. The base momentum T is thus defined to compute the number of subflows into which a flow can split. If the value of W_iV_i is between zero and T, $0 < W_iV_i < T$, then $flow_i$ does not split and instead moves to a new location as a single stream solution. Avoiding the generation of extra subflows may cause unnecessary resource consumption. Yang and Wang [10] defined an upper limit $\bar{n}$ as the number of subflows that split from the original flow in each iteration. However, at any iteration, the number of subflows can be calculated based on equation (4):

$$n_i = \min\left\{\max\left\{1, int\left(\frac{W_iV_i}{T}\right)\right\}, \bar{n}\right\} \quad \ldots\ldots(4)$$

In WFA-TSP, assigning new location of subflows $flow_k$ is assigned by based on the neighbour structure, the new locations are obtained using a random insertion move of one city. If we have the location of all subflows x_{i1}, $x_{i2}, \ldots, x_{in}$, the second stage intensely searches for the neighbour solution of the new assigned locations until the best neighbour of the new location can be found using the 2-opt neighbourhood search procedure [6]. The main idea of the 2-opt procedure is to cut two edges of the tour, turn one of the two partial sequences and reconnect those edges on some way. Figure 2 illustrate the flow splitting and moving mechanism during the optimization process.

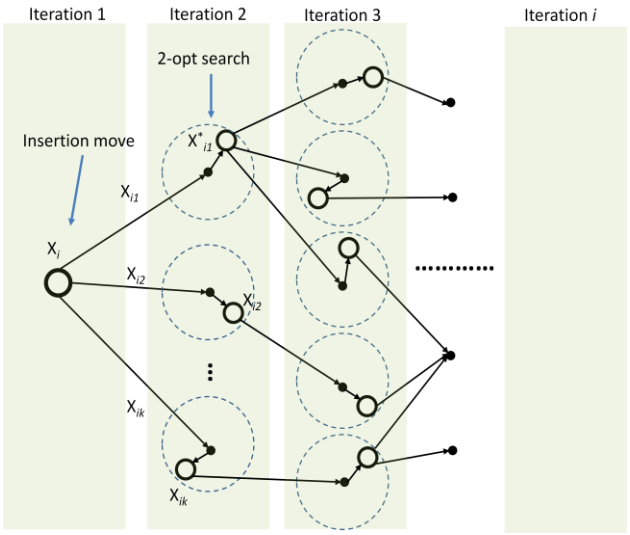


Figure 2: WFA Flows Splitting and Moving Mechanism

When $flow_i$ splits into subflow n_i, the mass of $flow_i$ is distributed to the subflows based on their ranks. The mass of subflow k, which splits from $flow_i$, is calculated using equation (5):

$$W_{ik} = \left(\frac{n_i + 1 - k}{\sum_{r=1}^{n_i} r}\right) W_i, \qquad k = 1,2,\ldots,n_i \quad \ldots\ldots(5)$$

The velocity of a subflow is calculated using the equation of energy conservation (6). μ_{ik} is the velocity of subflow k, which splits from $flow_i$.

$$\mu_{ik} = \begin{cases} \sqrt{V_i^2 + 2g\sigma_{ik}}\ , & V_i^2 + 2g\sigma_{ik} > 0 \\ 0. & \text{otherwise} \end{cases} \quad \dots\dots(6)$$

where g is the gravitational acceleration and σ is the improvement in the objective value of solution i to its neighbouring solution. σ represents the altitude decrease from solution i to solution k and can be obtained using equation (6). If $V_i^2 + 2g\sigma_{ik} < 0$, there is no improvement in solution i and it stagnates at the local optima with no splitting or moving. Such stagnant flow gradually evaporates into the atmosphere, returning to the ground as precipitation. At the end of the splitting and moving operation, the original flow is discarded because subflows have been generated. Information regarding the current number of subflows and solution sets is then recorded.

3.3 Flow Merging Operation

When two or more flows meet at the same location, they merge into a single flow with greater mass and momentum to reduce the number of solution agents at that location. The merging operation prevents redundant searching by solution agents with the same objective value.

Flow merging may also help stagnated flows escape from trapped locations. This operation regularly checks the current flows if they share the same location. Assuming that flows i and j share the same location, flow$_i$, mass W_i and velocity V_i are updated using equations (7) and (8), respectively.

$$W_i = W_i + W_j\ , \qquad \dots\dots(7)$$

$$V_i = \frac{W_i V_i + W_j V_j}{W_i + W_j}. \quad \dots\dots(8)$$

3.4 Water Evaporation Operation

As a neutral behaviour, water evaporates and returns to the ground as precipitation. The WFA uses the concept of water evaporation and precipitation in water flows after moving from one location to another to help flows escape from local optima and search within a greater solution space. In this operation, the flow mass is updated using equation (9):

$$W_i = \left(1 - \frac{1}{t}\right) W_i, \quad i = 1,2,\dots,N \quad \dots\dots(9)$$

W_i is the mass at the time when flow$_i$ was initially generated or when it merged with another flow. However, each water flow is subject to water evaporation into the atmosphere based on a fixed ratio $(1/t)$, as in equation (7). If a splitting or merging operation does not update a flow, the flow is removed after t iterations.

3.5 Water Precipitation Operation

Two types of precipitation have been applied to the WFA to mimic the natural lifecycle of water [10]. First, enforced precipitation is when all water flows are grounded with zero velocities. This indicates that there is no improvement in the solution search for all flows after t iterations. In this situation, all water flows must evaporate and become precipitation. All masses of poured flows should thus be updated without changing the current flow numbers, and an initial velocity V_0 should be assigned to them. However, the locations of all flows are changed stochastically from the original location. In WFA-TSP, we used a random swapping of two cities in the solution tor randomly reassign new locations to the flows. After relocating the position of flows, the initial mass W_0 is proportionally distributed to all flows according to their original mass using equation (10).

$$W_i' = \left(\frac{W_i}{\sum_{k=1}^{N} W_k}\right) W_0 \quad \dots\dots(10)$$

Regular precipitation is applied every t iterations to restore the evaporated water. After applying regular precipitation, the poured flows can join the current solution set. The accumulated mass of evaporated water is $W_0 - \sum_{k=1}^{N} W_k$, which is reassigned to ground flows, as in equation (11).

However, the revived flows may generate the same objective values and location assignments as those of other flows. The merging operation is thus performed after precipitation is applied to remove possible redundant solutions.

$$W_i' = \left(\frac{W_i}{\sum_{k=1}^{N} W_k}\right) W_0 - \sum_{k=1}^{N} W_k \quad \dots\dots(11)$$

4. EXPERIMENTS AND RESULTS

This performance of WFA is tested in two large TSP data sets (fl1400, d1655) obtained form TSPLIB[7]. The WFA was implemented using Java platform JDK 1.6, a Windows environment and a personal computer with an Intel core i5 (3.00 GHz CPU speed and 4 GB RAM). The WFA and ACS algorithms were implemented to compare them and measure their performance. Table 1 outlines the parameter settings for the experiments. The experiments measured the solution cost obtained from 10 runs of each dataset, with 10,000 cycles for each independent run. The number of iterations required to reach the best solution was also considered. The computation time was measured in seconds.

The results indicate the best, average and standard deviation (SD) of the solution cost for 10 independent runs. The distance between any two cities was calculated using the Euclidian distance and rounded off after the decimal point. The average computational time was also determined. The results were compared with ACS algorithms and run in the same computation environment

Table 1: Parameter Settings

Algorithm	Parameter	Value
WFA-TSP	Base momentum T	20
	Initial mass W_0	8
	Initial velocity V_0	5
	Subflow number limit $\bar{n}$	3
ACS	Number of Ants	10
	β	2
	ρ	0.1
	τ_0	$1/nC^{nn}$, where n is the number of cities and C^{nn} is the nearest neighbor value
	ε	0.1

Table 2 shows the experimental results obtained by WFA versus ACS in TSP. The Avg. Cots describe the average solution cost and the Best is the best solution cost, while the avg. Time is the average computation time for 10 independent runs. The best results are marked in bold. The table shows that WFA-TSP outperforms ACS algorithm in fl140 and d1655 data sets. The average WFA value of the best solution cost of the 10 runs for each data set was also lower than ACS algorithm Figure 3 shows That WFA-TSP performed faster then ACS with the improvement of the computation time for data set f1400 is reach up to 34%, while in d1655 is reach up to 61%. It can be seen in Figure 3. This means that WFA presented faster searching solutions space with faster convergence speed to the optima.

Table 2: A comparison of the experimental WFA results with the results of ACS

Dataset	Method	Avg. Cost	Best	Avg. Time
fl1400	**WFA**	**20445**	**20365**	**573.33**
	ACS	21326.09	21014	879.63
d1655	WFA	**64878.6**	**64313**	**746.33**
	ACS	70128.9	68602	1217.3

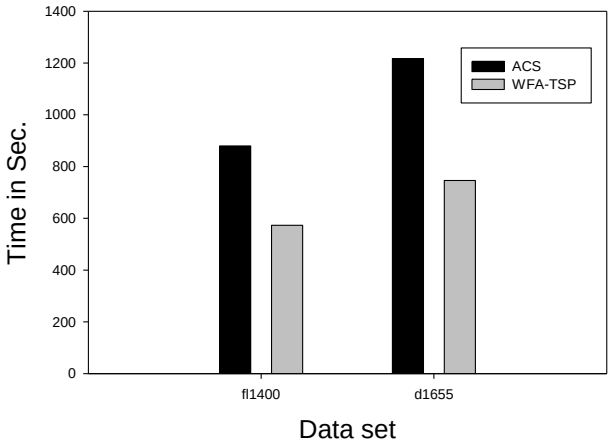


Figure 3: : The average computational time of WFA compared with ACS

Table 3 shows accuracy obtained by WFA versus ACS in TSP. The PD_{avg} describes the deviation of the average solution from the best known solution and the PD_{best} describes the deviation of the best solution from the best known solution, while SD is stand for standard deviation of the best solution cost of the WFA in 10 runs. The table shows that WFA obtained better result in both data sets. The result shows that the SD of WFA-TSP are lower than that of the ACS which indicates that WFA is more stable than ACS can always obtain solution in same range of quality. Furthermore, PD_{avg} and PD_{best} are significantly better than ACS, where the differences in the WFA solution cost deviation compared with the ACS algorithm deviation can be observed clearly in the table.

The result shows that ACS has improved 4.1% accuracy solution compare to ACS in data set fl1400, while in data set d1655 WFA improved 7.4% compare to ACS.

Table 3: The accuracy obtained by WFA versus ACS

Dataset	Method	$PD_{agv.}$	PD_{best}	SD
fl1400	WFA-	**1.5**	**1.1**	**52.28**
	ACS	5.9	4.4	253.24
d1655	WFA	**4.4**	**3.5**	**344.24**
	ACS	12.8	10.4	1058.69

Furthermore, in Figure 4, the performance of WFA can be seen clearly using log graph that showing the convergence speed toward the optimal of both algorithms. Figure 4 shows the convergence speed of WFA to reach the optimal solution is faster compare to ACS. In addition, the dynamic behaviour in population size of WFA can be seen clearly in Figure 5. The graph shows in 300 iterations when solving d1655 data set. The figure shows that the number of population size is changing during the optimization process that reflects the dynamic behaviour in population size

where the increasing of the flows number indicates that the algorithm find a promising solution area. Whereas, the decreasing in the flows number is mean illuminating the redundant search or solutions.

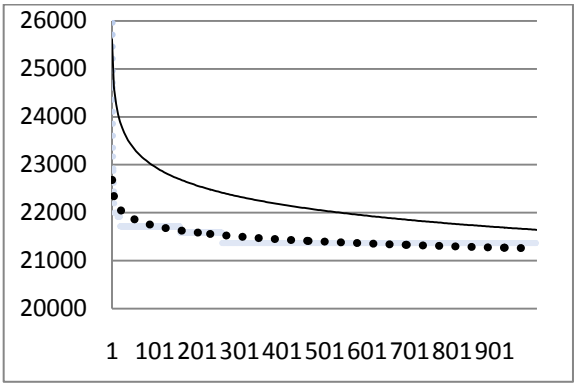


Figure 4: Log graph compare the convergence speed of WFA and ACS

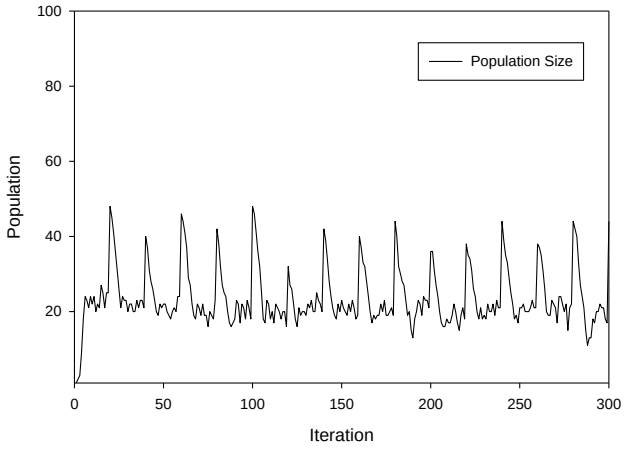


Figure 5:The dynamic behaviour of WFA-TSP in solving d1655 data set

5. CONCLUSION

This paper demonstrates the dynamic behaviour of population size in WFA has ability to speed up the process to reach solution in large size data sets of Travelling Sales Problem. Furthermore, the propose WFA-TSP also improved the accuracy of solution. The quality solutions and the computation time is much depending on the strong balancing between the exploration and exploitation of the solution search that adopted in WFA. Based on the result it can be concludes that WFA algorithm is suitable solution for large data set and the algorithm also has many potential for solution improvement by improving the flow moving mechanising using different neighbourhood structure such as 3 or k-opt neighbourhood structure.

6. REFERENCES

[1] ČERNÝ, V., 1985. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of optimization theory and applications 45*, 1, 41-51.

[2] DORIGO, M., MANIEZZO, V., and COLORNI, A., 1996. Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B 26*, 1, 29-41.

[3] GLOVER, F., 1986. Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research 13*, 5, 533-549.

[4] KIRKPATRICK, S., GELATT, C.D., and VECCHI, M.P., 1983. Optimization by simulated annealing. *science 220*, 4598, 671.

[5] LAWLER, E.L., 1985. The traveling salesman problem: a guided tour of combinatorial optimization. *WILEY-INTERSCIENCE SERIES IN DISCRETE MATHEMATICS*.

[6] LIN, S. and KERNIGHAN, B.W., 1973. An effective heuristic algorithm for the traveling-salesman problem. *Operations research*, 498-516.

[7] REINELT, G., 1991. TSPLIB—A traveling salesman problem library. *ORSA Journal on computing 3*, 4, 376-384.

[8] SHAHREZAEI, P.S., MOGHADDAM, R.T., AZARKISH, M., and SADEGHNEJAD-BARKOUSARAIE, A., 2011. Water Flow-Like and Differential Evolution Algorithms for a Nurse Scheduling Problem. *American Journal of Scientific Research*, 34, 12-32.

[9] WU, T.H., CHUNG, S.H., and CHANG, C.C., 2010. A water flow-like algorithm for manufacturing cell formation problems. *European Journal of Operational Research 205*, 2, 346-360.

[10] YANG, F.C. and WANG, Y.P., 2007. Water flow-like algorithm for object grouping problems. *Journal of the Chinese Institute of Industrial Engineers 24*, 6, 475-488.

800MHz~1700MHz Multi-Band Low Noise Amplifier with Interference Rejection Improvement

San-Fu Wang
Dept. of Electronic Eng,
Ming Chi University of Technology,
Taishan, New Taipei City, Taiwan,
R.O.C.
886-2-2908-9899-4867
sf_wang@mail.mcut.edu.tw

Jan-Ou Wu
Dept. of Electronic Eng,
De Lin Institute of Technology.
Tu-cheng, New Taipei City, Taiwan,
R.O.C.
janou@ms42.hinet.net

Yang-Hsin Fan
Dept. of Computer Science and
Information Eng.
National Taitung University.
Taitung, Taiwan, R.O.C.
yhfan@nttu.edu.tw

Jhen-Ji Wang
Dept. of Electronic Eng,
Yuan Ze University,
Taoyuan, Taiwan, R.O.C.
s1008508@mail.yzu.edu.tw

Chi-Chun Chen
Dept. of Cloud Service Technology
Center
Industrial Technology Research
Institute of Taiwan
chichun@itri.org.tw

ABSTRACT
In this paper, a differential multi-band CMOS low noise amplifier (LNA), operated in a wide range from 800MHz~1700MHz, with wide-band interference rejection, linearity improvement and the capacitive cross-coupling technology, is proposed. The proposed differential multi-band CMOS low noise amplifier with high linearity performance and good interference rejection performance. The post-simulation results of proposed LNA show that the gain is 13~17.5 dB, the noise figure (NF) is less than 3.4dB, the third-order intercept point (IIP3) is 9.98dBm. The LNA consumes 8.96mW under 1.8V supply voltage in TSMC 0.18-um RF CMOS process.

Categories and Subject Descriptors
B.7.1: [**Integrated Circuits**]: Types and Design Styles – *VLSI (very large scale integration)*

General Terms
Design, Experimentation.

Keywords
LNA, multi-band, linearity, interference rejection.

1. INTRODUCTION
In recent years, there has been a tendency towards of wireless communication with integrated of several functions on the chip, so the receiver which can support different wireless standards is required. However, the conventional narrow-band LNAs has poor performance on improving S11 offset for input matching components process variation and multi-standard support, because it is difficult to change the inductances and capacitances of input matching network, especially for the inductance of chip inductor which is not easy to be adjusted by digital or analog control [1][2][3]. Moreover, a wide-band LNA can satisfy different standards at the same time [4], but it also has the poor performances on interference rejection or image rejection [5]. For those reasons, the multi-band LNA is usually implemented with wide-band input matching, which has been discussed in the previous chapters. And the wide-band input matching technique must withstand the input power from different frequency. In other words, the broadband input power is integrated and amplified, and the circuit will be saturated easily. Therefore, the more broadband circuit can be saturated by other channel signals more easily. Therefore, the wide-band input matching technique has poor performance on interference rejection and input stage linearity. Especially in the state of the out-of-band channels constantly in use, the out-band interference rejection and input stage linearity will be more obvious [6]. Moreover, the narrow-band notch filter [7] only rejects single interference frequency, so it does not meet the wide-band interference rejection.

2. PROPOSED MULTI-BAND LNA
In this paper, a CMOS differential multi-band LNA, which employs common gate, tunable LC-tank load and capacitive cross-coupling techniques on its implementations, has been proposed. The wide-band input matching operations are more sensitive to out-of-band unwanted signals due to the use of out-of-band channels. These out-of-band interferences can severely degrade receiver's sensitivity and linearity. The proposed multi-band LNA has a high selectivity and sensitivity when it is constituted of several narrow-band operations at the single input frequency.

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.31

2.1 The theory of proposed LNA and tunable LC-tank load technique

The LNAs with wide-band operations are more sensitive to out-of-band unwanted signals (blockers) due to the transistor nonlinearity. These out-of-band blockers can severely degrade receiver's sensitivity [8]. Figure 1 shows the conventional multi-band LNA architectures utilize wide-band input matching technique to cover each operating frequency, and figure 2 shows the proposed multi-band LNA architectures utilize common gate LNA technique and tunable parallel LC-tank load technique to selecting input frequency. To compare figure 1 and figure 2, some out-of-band interference signal power is short to ground because the LC circuit is to provide low impedance at anti-resonance frequency. Moreover, some out-of-band interference signal power is eliminated by this path. In other words, the proposed LNA is operated in common gate LNA technique at LC-tank resonance frequency, and the circuit is provided a good performance of S11.

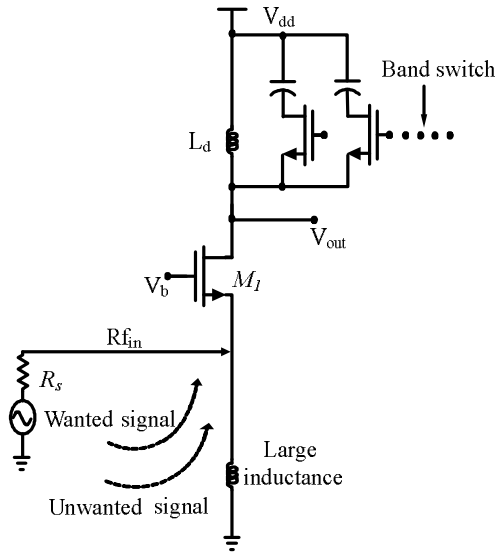


Figure 1. The principle of conventional common gate multi-band LNA.

Figure 3 shows the circuit of proposed differential multi-band LNA, and the proposed LNA can achieve multi-band function by tunable LC-tank resonance frequency. As shown in figure 3, the source terminal is used as an input terminal. Since the impedance can be written by the following formula

$$Z_{in} = \frac{1}{gm_{M1}} // Z_{eq} \quad (1)$$

Where gm_{M1} is the trans-conductance of device M_1, and the resonant circuit (L_2 and C_b) provides high impedance (Z_{eq}) at desired frequency (ω_r), and the impedance has been used as the load, which is given by

$$\left|Z_{eq}(s=j\omega)\right|^2 = \frac{(\omega_r{}^2 L_2{}^2 + R_{ind}{}^2)}{R_{ind}{}^2 C_b{}^2 \omega_r{}^2 + (1 - L_2 C_2 \omega_r{}^2)^2} \quad (2)$$

The equation 2 shows that the impedance does not go to infinity at any s=jω. We say the circuit has a finite Q (quality factor). The magnitude of Z_{eq} in equation 2 reaches a peak ($Z_{eq}=R_r$) in the vicinity of $\omega = \sqrt{L_2 C_b}$, but the actual resonance frequency has some dependency on R_{ind}. Where R_{ind} is the parasitic series resistance of the inductor L_2. And the R_r can be similar to

$$R_r = \omega_r{}^2 L_2{}^2 / R_{ind} \quad (3)$$

Based on those reason, the input impedance of the proposed differential multi-band CMOS low noise amplifier can be derived as

$$Z_{in} = \frac{1}{gm_{M1}} // \left(\frac{\omega_r{}^2 L_2{}^2}{R_{ind}}\right) \quad (4)$$

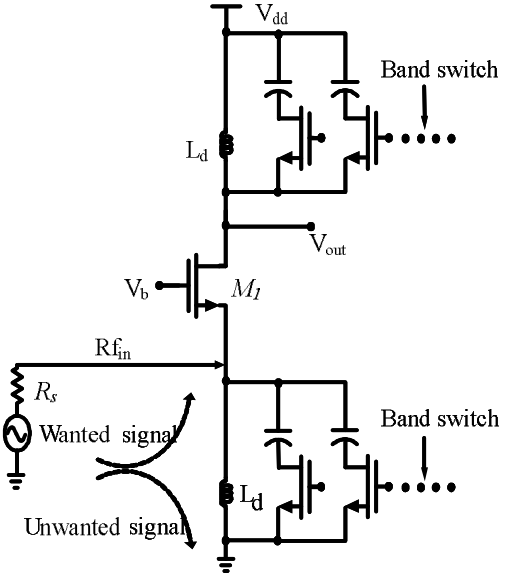


Figure. 2. The principle of proposed technique.

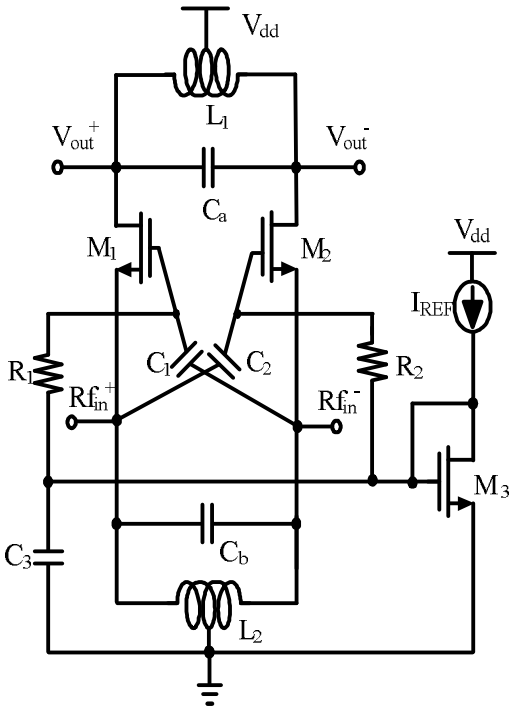


Figure. 3. The principle of proposed technique.

In other word, if the input signals frequency is far away from the resonance frequency, the Z_{in} will be close to zero. Therefore, the input signals power will be short to ground.

3. SIMULATION RESULTS

The multi-band LNA is simulated with Cadence's EDA-Spectre RF using TSMC 0.18-um RF CMOS process. The following figures give us the results respectively.

Figure 4 shows the simulated result of voltage gain parameter, which is operated in different standards by different capacitances (C_a and C_b shown in figure 3). The voltage gain of the LNA is more than 13dB, and the gain variable is about 4.5dB. It is operated in the range from 800MHz to 1700MHz.

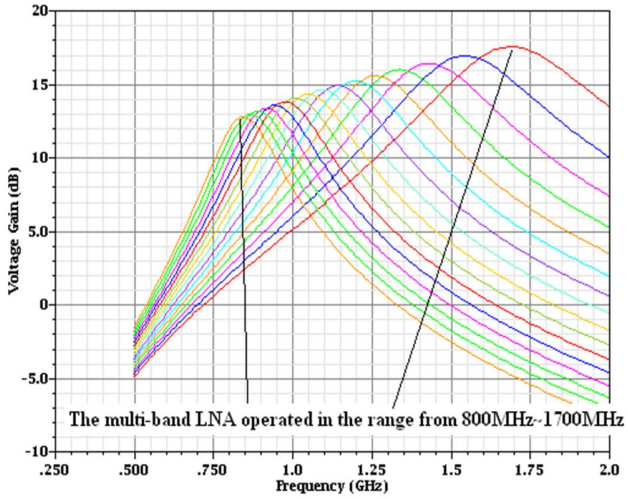


Figure. 4 The simulated voltage gain of proposed LNA.

Figure 5 shows the simulated result of S11 in Fig. 3 that S11 parameter is smaller than -18dB at the wanted switching mode, so the input impedance matching is close to 50ohm between 840MHz and 1700MHz. Figure 6 shows the simulated NF of the entire LNA, and the NF is below 3.42dB over the bandwidth.

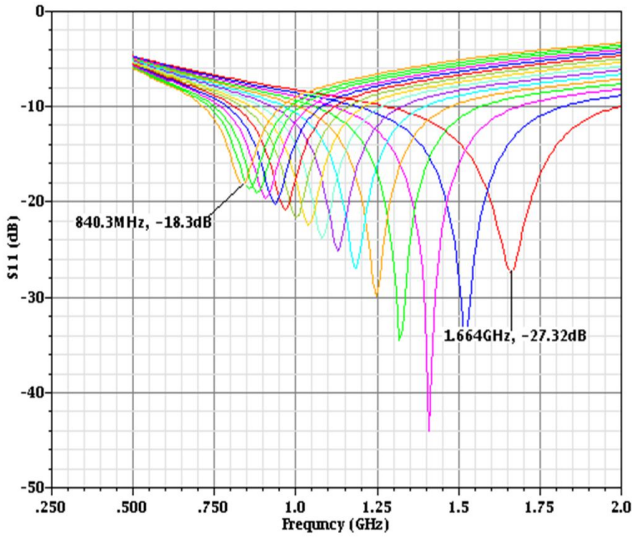


Figure. 5 The simulated input reflection coefficient of proposed LNA.

Figure 7 shows the simulated result of interference rejection improved of proposed LNA, and the result is compared with the conventional common-gate LNA. It manifests that the proposed LNA has the performance of wide-band interference rejection. Therefore, the out-band interference rejection and input stage linearity will be improved.

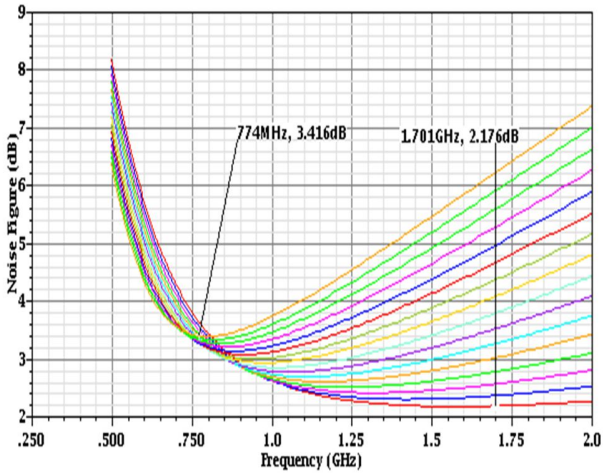


Figure. 6 The simulated noise figure of proposed LNA.

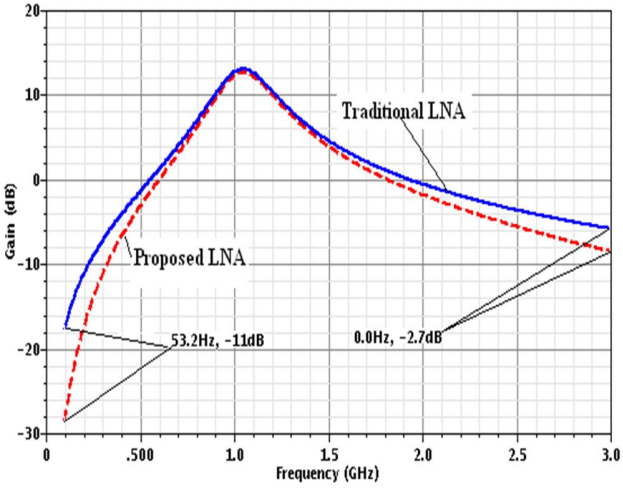


Figure. 7 The simulated interference rejection improved of proposed LNA.

4. CONCLUSIONS

This paper presents a new technique which can improve the multi-band input matching, wide-band interference rejection and input stage linearity by proposed technique. When more and more channels are used, the performance of proposed LNA will be highlighted. Moreover, it is manifested that the multi-band input reflection coefficient (S11), multi-band output voltage gain, and wide-band interference rejection have been improved.

5. ACKNOWLEDGMENTS

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC 101-2221-E-131-042, and the National Science Council and the Chip Implementation Center of Taiwan for project support.

6. REFERENCES

[1] L.C. Lee, A.K. A'ain, and K.A. Victor, "A 2.4-GHz CMOS tunable image-rejection low-noise amplifier with active

inductor, ”IEEE Asia Pacific Conf Circ Syst (2006), 1679–1682.

[2] M.K.Salama, A.M. Soliman, “Low-voltage low-power CMOS RF low noise amplifier”, AEU-International Journal of Electronics and Communications, Vol. 63, P. 478-482, Jun. 2009.

[3] Mohamed El-Nozahi, Ahmed A. Helmy, Edgar Sánchez-Sinencio, and Kamran Entesari, “An Inductor-Less Noise-Cancelling Broadband Low Noise Amplifier With Composite Transistor Pair in 90 nm CMOS Technology,” IEEE Journal of Solid-State Circuits, vol. 46, no. 5, pp. 1111–1122, May, 2011.

[4] Yuh-Shyan Hwang, San-Fu Wang, and Jiann-Jong Chen, “Design Method for CMOS Wide-band Low Noise Amplifier for Mobile TV Application” IEICE Electronics Express, vol.6, no. 24, pp. 1721–1725, December 2009.

[5] Ajay Balankutty and Peter R. Kinget, “An Ultra-Low Voltage, Low-Noise, High Linearity 900-MHz Receiver With Digitally Calibrated In-Band Feed-Forward Interferer Cancellation in 65-nm CMOS,” IEEE Journal of Solid-State Circuits, vol. 46, no. 10, pp. 2268–2283, October, 2011.

[6] Jonathan Borremans, Gunjan Mandal, Vito Giannini, Bjorn Debaillie, Mark Ingels, Tomohiro Sano, Bob Verbruggen, Jan Craninckx,, “A 40 nm CMOS 0.4–6 GHz Receiver Resilient to Out-of-Band Blockers,” IEEE Journal of Solid-State Circuits, vol. 46, no. 7, pp. 1659–1671, July, 2011.

[7] Hirad Samavati, Hamid R. Rategh, and Thomas H. Lee, "A 5-GHz CMOS wireless LAN receiver front end," IEEE Journal of Solid-State Circuits, vol. 35, no. 5, pp. 765-772, 2000.

[8] Hossein Hashemi and Ali Hajimiri, “Concurrent Multiband Low-Noise Amplifiers-Theory, Design, and Applications,” IEEE Transactions on Microwave Theory and Techniques, Vol. 50, No. 1, pp.288-301, January, 2002.

Artificial Chromosomes Structure for Extending the Diversity of Evolution in Sub-population Genetic Algorithms

Yen-Wen Wang
Department of Industrial Management, Chien Hsin University
229 Jiangsing Rd, Jhonglili, Taoyuan 32026, Taiwan, ROC
886-3-4581196
ywwang@uch.edu.tw

Chen-Hao Liu
Department of Information Management, Kainan University
1 Kainan Rd., Luzhu Shiang, Tao-Yuan 33857, Taiwan, ROC
886-3-4345705
chliu@mail.knu.edu.tw

ABSTRACT
Sub-population genetic algorithm is a population-based approach for heuristic search in multi-objective optimization problems. It has shown that this mechanic performs better than traditional genetic algorithms for some problem. In order to apply in the multi-objective problem, the novel structure we called artificial chromosome is developed and combined in the sub-population genetic algorithm in this research. This approach is applied to deal with multi-objective flowshop scheduling problems. Besides, the artificial chromosome with priority matrix will be introduced when the algorithm evolves to certain iteration for injecting to individual to search better combination of chromosomes, this mechanism can find a better Pareto solution set and avoid solutions converge toward local optima. Compares with other approaches, the experiments result show that this approach possess more balance convergence and average scatter of Pareto solutions simultaneously for solving multi-objective flowshop scheduling problems in test instances.

Categories and Subject Descriptors
I.1.3 [**Computing Methodologies**]: Evaluation Strategies

General Terms
Algorithms

Keywords
Sub-population genetic algorithm, flowshop scheduling problem, multi-objective problem

1. INTRODUCTION
In the operations research literature, Flowshop scheduling is one of the most well-known problems in the area of scheduling. Flowshops are useful tools in modeling manufacturing processes. A permutation Flowshop is a job processing facility which consists of several machines and jobs to be processed on the machines. In a permutation Flowshop all jobs follow the same machine or processing order and job processing is not interrupted once started. Our objective is to find a sequence for the jobs so that the makespan or the completion time is a minimum.

In this research, we take a close look at the evolutionary process for a permutation Flowshop scheduling problems and come out with the new idea of generating artificial chromosomes to further improve the solution quality of the genetic algorithm. To generate artificial chromosomes, it depends on the probability of each job at a certain position. The idea is originated from Chang et al.(2005) which propose a methodology to improve Genetic Algorithms (GAs) by mining gene structures within a set of elite chromosomes generated in previous generations. Instead of replacing the crossover operator and mutation operator due to efficiency concern, the proposed algorithm is embedded into simple GAs (SGA) and non-dominated sorting genetic algorithm-II (NSGA-II). The probability model acquired from the elite chromosomes will be integrated with the genetic operators in generating artificial chromosomes, i.e., off-springs which can be applied to enhance the efficiency of the proposed algorithm. Apart from our previous researches, Harik (1999), Rastegar (2006), Zhang (2005) have discussed and proved the genetic algorithm which is based on the probability models. For a complete review of the relative algorithms discussed above, please refer to Larranaga (2001), Lozano (2006), and Pelikan (2002). In most recent works of evolutionary algorithm with probability models, they all concentrate on solving continue problems rather than discrete problems. There are only few researches in applying evolutionary algorithm with probability models to resolve discrete problems.

The rest of the research is organized as follows: Section 2 introduces the flowshop scheduling problem. The algorithm of artificial chromosome embed in genetic algorithms will be presented in section 3. In Section 4, extensive experiments are conducted to test the performance of the proposed algorithm in two-objective Flowshop scheduling problems. Finally, the conclusion is discussed and future researches are also provided.

2. METHODOLOGY

2.1 Generating Artificial Chromosomes
The detailed AC3 steps are described in the following:

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.32

To convert gene information into Priority matrix: Before we collect gene information, selection procedure is performed to select a set of chromosomes. Because it is a two objective problem, thus, for each chromosome, we compute the fitness of two objectives and assign the chromosomes into two groups with better fitness. (as figure 1 shows)

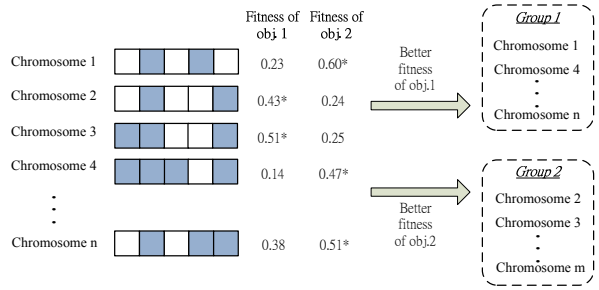


Figure 1. The group assignment of chromsome in the seletion procedure.

For each group after selection, all jobs allocation information is recorded into priority matrix in the following process. If job i is located in front of job j, the frequency in the priority matrix is added by 1. To demonstrate the working theory of the artificial chromosome generation procedure, a 5-job problem is illustrated. Suppose there are n sequences (chromosomes) in the group, then, we accumulate the gene information from these n chromosomes to form a priority matrix. As shown in the left-hand side of Figure 2. The procedure will repeat for the rest of the position. Finally, the priority matrix contains the gene information from better chromosomes is illustrated in the right-hand side of Figure 3.

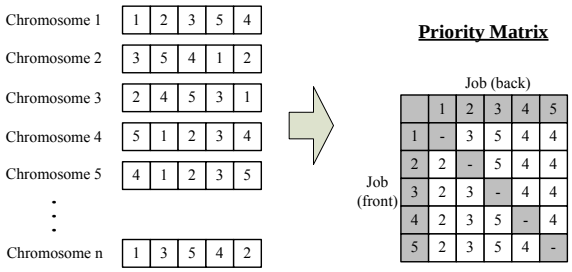


Figure 2. To collect gene information and converted into a priority matrix.

As soon as we collect gene information into priority matrix, we are going to assign jobs onto the positions of each artificial chromosome. The assignment sequence for every position is assigned randomly, which is able to diversify the artificial chromosomes. After we determine the assignment sequence, we select one job assigned to each position by roulette wheel selection method based on the probability of each job on this position. After we assign one job to a position, the job and position in the priority matrix are removed. Then, the procedure continues to select the next job until all jobs are assigned. (As shown in figure 3.)

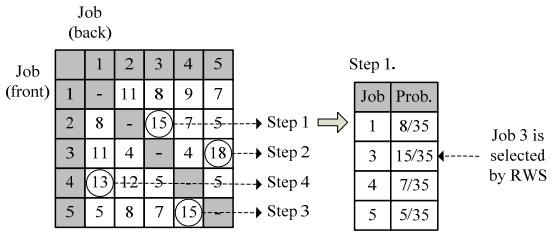


Figure 3. The updated priority matrix after assigning job 4 at position 3.

After embedding artificial chromosomes into the population, we use l + k strategy, which combines previous parent population and artificial chromosomes. Then, we select better chromosomes from the combined population. Consequently, better solutions are preserved to the next generation.

2.2 The procedure of AC3-SPGA

The procedure of the AC3-SPGA is explained as the following:

1. Initialize
2. Divide Population
3. Assign Weight To Each Objectives
4. while counter < Iteration do
5. for i = 1 to population number do
6. FindPareto
7. Fitness
8. Selection
9. Group assignment
10. Generate the Priority Matrix of each group
11. Generate the Artificial Chromosomes from each group
12. Elitism
13. Mutation
14. Replacement
15. counter + 1>
16. end while

Figure 4. The updated priority matrix after assigning job 4 at position 3.

Compared with SPGA, this approach is different in that it has the mechanism of creating AC , local search heuristic and the sorting information of chromosomes in each mutation is recorded for the use of creating AC and placing them in the mating pool for evolution.

3. EXPERIMENTAL TESTS AND CONCLUSION

The research uses the Flowshop scheduling case study by Ishibuchi (2003) in which four types were included in the biobjective flow-shop problems; they were 20, 40, 60, and 80 jobs in 20 machines. Two objectives are the total completion time (Cmax) and maximum tardiness (Tmax). The experimental results will be compared with those of MGISPGA(Chang 2005) and NSGA-II(Deb 2002) and AC2(Wang 2010). Table 1-4 shows the Pareto solutions of each algorithm. We could find that AC2 and AC3 can find large number of solutions than other two methods, it is because our AC model can record and keep the better chromosome combination while GAs evolving. Compare with all Pareto solutions from these four algorithms, the non-dominated solutions can be collected. We could find the AC3 algorithm always can find more Pareto solutions than other three

methods. It shows our AC3 not only can find a better s olution, but can avoid solutions converge toward local optima. Figure 5-8 shows the Pareto sets of eac h algorithm of four benchmark problems.

Table 1. The algorithm comparison of 20 jobs and 20 machines flowshop problem

Instance	20/20（Jobs/Machine）		
method	# Pareto solutions	# non-dominated solutions	% of non-dominated solutions
AC3	72	52	60.47
AC2	60	32	37.21
MGISPGA	18	2	2.32
NSGA-II	24	0	0

Table 2. The algorithm comparison of 40 jobs and 20 machines flowshop problem

Instance	40/20（Jobs/Machine）		
method	# Pareto solutions	# non-dominated solutions	% of non-dominated solutions
AC3	72	42	65.63
AC2	77	22	34.37
MGISPGA	12	0	0
NSGA-II	18	0	0

Table 3. The algorithm comparison of 60 jobs and 20 machines flowshop problem

Instance	60/20（Jobs/Machine）		
method	# Pareto solutions	# non-dominated solutions	% of non-dominated solutions
AC3	73	48	75.00
AC2	98	16	25.00
MGISPGA	17	0	0
NSGA-II	24	0	0

Table 4. The algorithm comparison of 80 jobs and 20 machines flowshop problem

Instance	80/20（Jobs/Machine）		
method	# Pareto solutions	# non-dominated solutions	% of non-dominated solutions
AC3	44	19	57.58
AC2	70	14	42.42
MGISPGA	7	0	0
NSGA-II	15	0	0

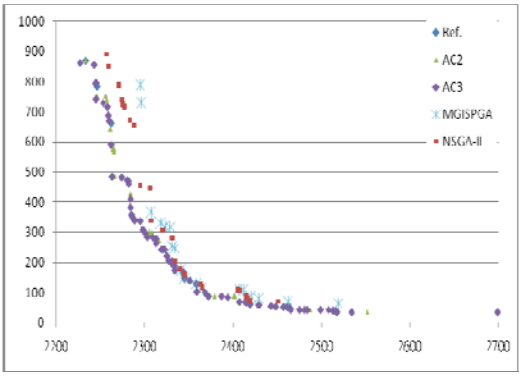


Figure 5. The plot of algorithms with reference Pareto set of algorithms for 20 jobs and 20 machines

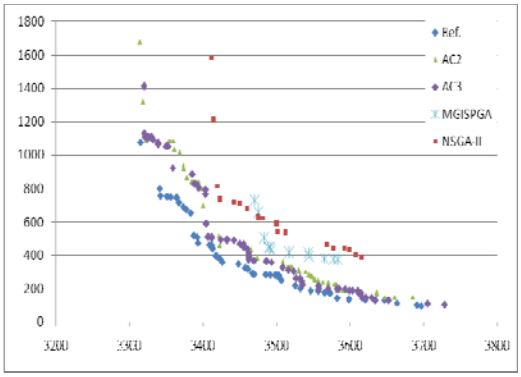


Figure 6. The plot of algorithms with reference Pareto set of algorithms for 40 jobs and 20 machines

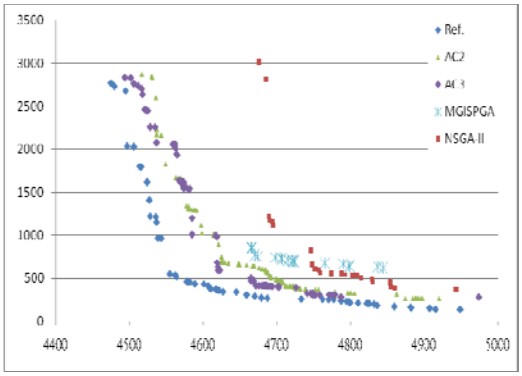


Figure 7. The plot of algorithms with reference Pareto set of algorithms for 60 jobs and 20 machines

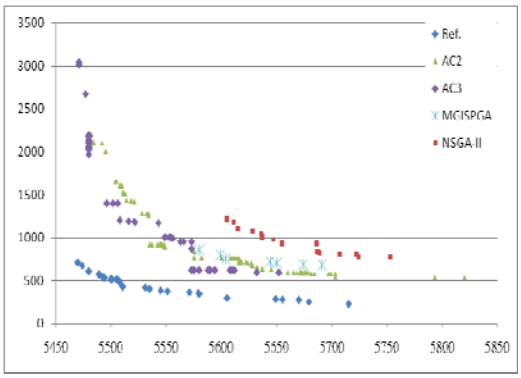


Figure 8. The plot of algorithms with reference Pareto set of algorithms for 80 jobs and 20 machines

To discuss the great effect on the Pareto solution seeking ability of AC2 and AC3. We can find th e convergence of AC3 compared with SPGA as shown in figure 9. The general GA-based algorithm(SPGA etc.) are developed to find the minimum average solution of each g eneration, but the AC structure emphasized the chromosome diversity of e ach generation. That is why the AC structure based algorithm can find more number of Pareto solutions.

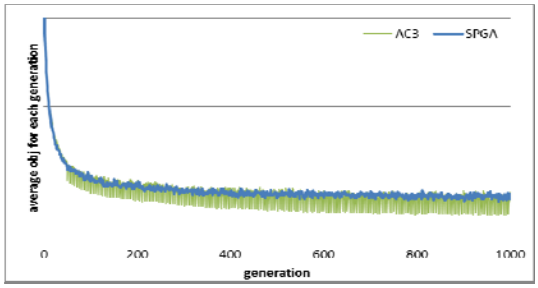


Figure 9. The convergence of AC3 and SPGA

4. REFERENCES

[1] Chang, P.C., Chen, S. H., and Lin, K. L. 2005. Two phase subpopulation genetic algorithm for parallel machine scheduling problem, *Expert Systems with Applications*, 29(3), 705–712.

[2] Chang, P.C., Chen, S. H., and Liu, C.H. 2007. Sub-population genetic algorithm with mining gene structures for multi-objective Flowshop scheduling problems, *Expert Systems with Applications*, 33, 762–771

[3] Chang, P.C., Wang, Y.W., and Liu, C.H. 2005. New Operators for Faster Convergence and Better Solution Quality in Modified Genetic Algorithm, *Lecture Notes in Computer Science*, 3611, 983 – 991.

[4] Deb, K., Amrit, P., Sameer, A., and Meyarivan, T. 2002. A fast and elitist multi-objective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, 6(2), 182 – 197.

[5] Harik, G. R., Lobo, F.G., and Goldberg, D.E. 1999. The compact genetic algorithm, *IEEE Transactions of Evolution Computing*, 3 (4), 287 – 297.

[6] Ishibuchi, H., Yoshida, T., and Murata, T. 2003. Balance between Genetic Search and Local Search in Memetic Algorithms for Multi-objective Permutation Flowshop Scheduling, *IEEE Trans on Evolutionary Computation*, 7(2), 204-223.

[7] Larranaga, P., and Lozano, J.A. 2001. Estimation of Distribution Algorithms: *A New Tool for Evolutionary Computation*, Kluwer, Norwell.

[8] Lozano, J.A., Larranaga, P., Inza, I., and Bengoetxea, E. 2006. *Towards a New Evolutionary Computation*, Springer.

[9] Pelikan, M., Goldberg, D.E., and Lobo, F.G. 2002. A survey of optimization by building and using probabilistic models, *Computational Optimization and Applications*, 21 (1), 5 – 20.

[10] Rastegar, R., and Hariri, A. 2006. A step forward in studying the compact genetic algorithm, *IEEE Transactions of Evolution Computing*, 14 (3), 277 – 289.

[11] Wang, Y. W., Liu, C. H., and Fan, C.Y. 2010. Develop a Sub-population Memetic Algorithm for Multi-objective Scheduling Problems, The 2nd International Conference on Computer and Automation Engineering, Feb 26-28, Singapore.

[12] Zhang, Q., Sun, J., and Tsang, E. 2005. An evolutionary algorithm with guided mutation for the maximum clique problem, *Evolutionary of Computing*, 9 (2), 192 – 200.

The Effect of Tracheal Stenosis on Airflow in the Trachea and Main Bronchi: A Numerical Modeling Analysis

Nasrul Hadi Johari
Faculty of Mechanical Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
+601-199111787
nhadi@ump.edu.my

Kahar Osman
Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
+60-196536201
kahar@utm.my

Mohammed Rafiq A. Kadir
Faculty of Biomedical Engineering & Health Sciences, Universiti Teknologi Malaysia, 81310 Skudai, Johor,
+60-137585553
rafiq@biomedical.utm.my

ABSTRACT
The location and size of tracheal stenosis are among the major factors that contribute significantly to the breathing difficulties. Hence, this study aims to establish correlation between the location and size of the stenosis to the possibility of breathing difficulties. This work used ideal trachea model based on realistic trachea model derived from Computed Tomography (CT) scan images. The stenosis was patched to the healthy trachea models at regular locations and sizes as proposed by medical practitioners. The changes in the flow behavior due to the different sizes and locations of the stenosis were then examined to determine the pattern of possible breathing difficulties. The present simulation confirms that the overall flow behavior could be significantly affected if the size of stenosis is more than 60% for location far from the bifurcation region, and at size of 50% for location close to the bifurcation region. The outcomes of this study may help the medical practitioners and researchers to understand how dramatic increase in pressure drops occurs inside the trachea and main bronchi with the presence of stenosis at different location and size.

Categories and Subject Descriptors
D.3.3 [**Computer applications**]: Life and Medical Sciences

General Terms
Algorithms

Keywords
Tracheal stenosis, Computational Fluid Dynamics (CFD) and flow behavior.

1. INTRODUCTION
The patients with tracheal stenosis often report a relatively sudden appearance of breathing impairment, which at the stage of admission to the clinic is observed when a loss of 75% or more of the airway lumen has occurred [1].

The flow characteristics and clinical symptoms caused by the stenosis do not usually occur until the trachea has become constricted to 30% of its original diameter [2]. Long-segment stenosis due to congenital cause is very critical, especially for infants such that the removal of stenosis will reduce the breathing problem [3]. Since most patients with minor stenosis referred for necessary treatment are typically asymptomatic, precise additional information is needed in order to perform proper diagnosis [1,4]. Researchers agree that location and size of tracheal stenosis are among major factors that contribute significantly to the possibility of breathing difficulties [5,6].

Within the last decade, Computational Fluid Dynamics (CFD) had become a popular choice among the researchers. Literature shows that the human airways consist of 25 generations (branches of bronchial tree), and the diameter of each generation will reduce to less than 1 mm in the last generation, alveolar [7]. Recently, Romula et al., (2011) [8], suggested the use of CFD simulation method to assess the tracheal stenosis using deformable shape models. Several researchers used CFD to simulate the breathing flow condition in the trachea and main bronchi with the presence of stenosis [9-13].

In this study, pressure drop of airflow and velocity distribution have been shown as the result analysis [10]. Brouns et al., (2007) [10] showed that the overall pressure drop at rest was only affected in case of severe constriction, approximately 70% of the normal diameter. The results also highlighted that the pre-critical stage can be detected using computed pressure drop. The effect of increasing stenosis was investigated by Jayaraju et al., (2006) [11] where they found that the pressure drop shows modest increment with the degree of narrowing up to 75% constriction.

It was expected that small size stenosis would have no effect on the airflow pattern. However, if the location of stenosis was nearer to the bifurcation, even a little stenosis was sufficient to modify the airflow pattern. This simulation study of tracheal stenosis is believed to be an alternative method to support the current clinical practice. Correct additional information is needed by medical practitioner to perform proper diagnosis [4]. The analysis of the airflow pattern can provide a possible potential risk indicator of stenosis severity to the patient [10].

2. METHODOLOGY
2.1 Geometric Model
In this study, we were used the model proposed by Schlesinger and Lippman [14] with some modification of stenosis shapes on the trachea wall. The asymmetric model was modified into several

models, which each model has different locations of stenosis. A details dimension of our model was summarized in Table 1.

Table 1. The model`s dimensions [14].

Region	Diameter [a] (cm)	Length (cm)	Angle [b] (degree)
Trachea	2.17	9.2	-
Right bronchi	1.7	4.2	15
Left bronchi	1.26	5.3	30

a,b These values represent the mean of transverse diameters at the midpoint of each branch and branching angle.

Previously, there were no adequate description on the relationships of stenosis size, location and type with patient breathing difficulties, until Freitag et al., (2007) [6] introduced the proposed classification scheme. The present model was employed several of their proposed classification, i.e. the basic stenosis types or shapes, and regular locations (Figure 1). The stenosis shape used in this work was a web-like shape (abrupt-transition), which was categorized the case of shrinking and scarring, and it was predominantly caused by post intubation stenosis, burn injuries and secondary healing after surgery [6]. The locations of stenosis were engaged based on their regularities among the patients [5,6].

For the first stenosis (T1), was at approximately 7 cm from the bifurcation region. The second (T2) and the third location (T3) were at around 5 cm and 2 cm respectively. Since all the models were generated from the same source, the bifurcation area was calculated from the inlet was maintained at 9.2 cm from the inlet. Two more web-like stenoses were added in between T1, T2 and T3T3 to increase the result's accuracy. For the size, it was reduced from 10% until 80% of normal trachea's diameter, with the gap of 10% each. Stenosis with the size ≥90% was considered to be the severe stenosis and patient should be referred to medical treatment immediately [15].

2.2 Boundary Conditions

For a human respiratory system, the inspiratory flow rates were in a huge rate depending on the intensity of physical activity. The Reynolds number, Re of air flow in the trachea may vary from 800 in light breathing (10 lit/min) to about 9300 in heavy breathing (100 lit/min) [16,17]. Thus, the flow in a trachea was considered to be transitional covering laminar to turbulence flow. Moreover, the presence of larynx and pharynx tend to stimulate disturbances and causes instability in air flow as the flow passed through the vocal folds [18].

For this study, the steady-state flow with different inhalation conditions based on previous researchers were simulated with a different range of Reynolds numbers; Re 1201 in resting, Re 3012 in moderate and Re 46600 for extreme exercise. The conditions imposed on the inlet and outlets of the models were summarized in Table 2.

The working fluid is air, with normal body temperature of 37 ^{o}C and was assumed to be isothermal and Newtonian. Since the flow rates have low Mach number, we used incompressible equation so that the constant density will be employed for the whole simulation [10,19]. Besides, no-slip boundary condition was invoked at the whole surface.

Table 2. The model parameter used for this study

Parameters	Resting [24]	Moderate [18]	Extreme[19]
Reynolds number, *Re*	1.201x10^3	3.012x10^3	4.66x10^4
Inlet diameter, *D* (m)	0.0217	0.0217	0.0217
Inlet velocity, *U* (m/s)	0.85	2.03	32.84
Outlet Pressure, *P* (Pa)	101325	101325	101325
Density, ρ (kg/m$^{3)}$	1.19		
Viscosity, μ (kg/ms^{-1})	1.82 x 10^{-5}		

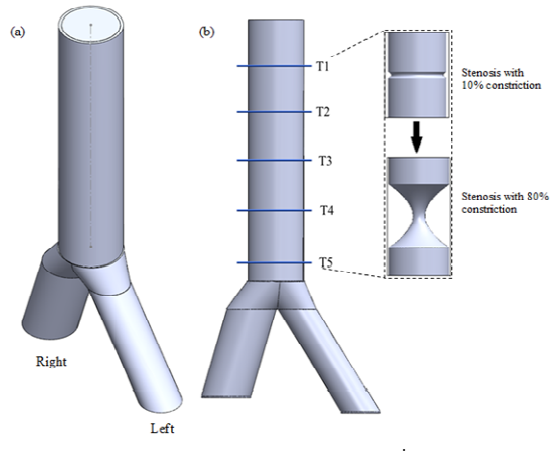


Figure 1.(a) The schematic model of normal trachea and main bronchi with the regular locations of stenosis (noted by T1- T5 respectively). (b) Trachea with stenosis model with the Inset is a zoom of stenosis shape that represents the radius of inflammation model inside trachea.

2.3 Flow Solver

A commercial finite-volume based software; Engineering Fluid Dynamics (EFD) was employed to solve the steady state governing equations below, Eqs. (1) and Eqs. (2) and the associated boundary conditions.

$$\nabla.\mathbf{u} = 0, \qquad (1)$$

$$\mathbf{u}.\nabla\mathbf{u} = -\rho + \frac{1}{\mathrm{Re}}\nabla^2\mathbf{u} \qquad (2)$$

The working fluid, air, was assumed to be a homogenous, Newtonian with constant pressure, p and viscosity, μ. In these equations, $x = (x, y, z)$ is the Cartesian position, $u = (u, v, w)$ is the velocity vector and p is the static pressure. All these variables are in dimensionless and they are defined with respect to the dimensional variables (designated by the superscript *) by the $X = x^*/D_{in}$, $u = u^*/U$ and $p = p^*/\rho U^2$. The Reynolds number is defined as $Re = \rho U D_{in}/\mu$ where μ is the air dynamic viscosity. The turbulence model *k-ε* was employed as part of

the consideration for occurrence of turbulent flow at low Reynolds numbers as a result of shear flow and it was proven capable to simulate the airflow [20,21]. The unstructured hexahedral mesh with local refinement contains of 300,000 cells and above. A preliminary grid independence study between 100,000 and 400,000 had indicated that the flow behavior inside the stenosis models were not altered (less than 1%) when the grid refinement is above 300,000 cells.

2.4 Validation

The experiment of flow inside axisymmetric constricted tube (stenotic model) by Ahmed and Giddens (1983) [22] was suitable to be compared with this study. The *k-ε* turbulent flow model was applied to the simplified stenosis model with local mesh refinement. Local axial velocity *Vx* at several locations were taken and later divided by the average velocity $V_{average}$. The dimensionless parameter is plotted in a function of position from the inlet of the stenosis region together with outcomes of experimental data and *k-ε* turbulent model solved by Luo et al., (2004) [20] is shown in Figure 2. Based on the plot, the numerical method used in this study is capable in delivering accurate pattern of axial velocity. Similar trend with experiment data was obtained. With finer meshing applied to the simplified stenosis model, it is found that the k-ε model has the potential to deliver flow pattern close to the experimental data, relatively better than what achieved by Luo et al., (2004) [20].

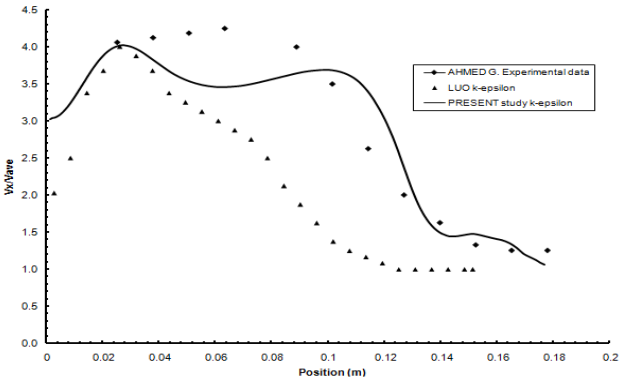


Figure 2. Comparison of axial velocity distributions obtained using k-epsilon (*k-ε*) model with experimental data [22], and *k-ε* model by Luo et al., (2004) [20]

3. RESULT

3.1 Location and size of stenosis affect the bronchi inlet flow rate

The different locations and sizes of the stenoses alter, not only the inlet pressure to the left and right bronchi, but the inlet flow rate as well. Figure 3 shows the changes of flow rates to the right and left bronchi during different locations of stenosis with 60% in sizes at resting condition. For this low inlet flow rate to the main airway, flow passing through the first two locations of the stenoses (Figure 3(b) and 3(c)) indicates the right bronchi receives more air as compared to that of the healthy model (Figure 3(a)).

However, as the location of the stenosis approaches the bifurcation (Figure 3(d)), the results show otherwise. The left bronchus seems to receive equal flow rate to the right bronchi approximately 50%. According to Horsfield et al., (1971) [23], the normal flow rates should be with 45% flow rates go to the left and the rest of 55% goes to the right. Similar findings also noted by another paper where they found 43% flow rates at the inlet of left bronchi and 57% to the right [24]. Our results show that the inlet flow rates of the main bronchi would be changed due to the circumstances. The existence of stenosis at severe sizes and distal locations to the bifurcation altered the normal breathing conditions and finally will affect lower lung generations. This is due to the faster air jet that leaves the stenosis and caught by the left bronchi. Furthermore, as the flow rates increases, the ratio difference of flow rates induced into right and left bronchi will be increased.

We can see in Figure 4 (a-c) that the upward sloping of pressure drops increases faster as the sizes of stenosis increases at all locations. The exponential graphs were divided into three, which represents different breathing conditions (Reynolds numbers). For Re 1201 and Re 3012, similar trends of graphs noted where the pressure drop gradually increased in numbers from 10% to 60% of stenosis sizes. However, the trend suddenly changes to be greater than before when the flow passes the stenosis higher than 60% in sizes.

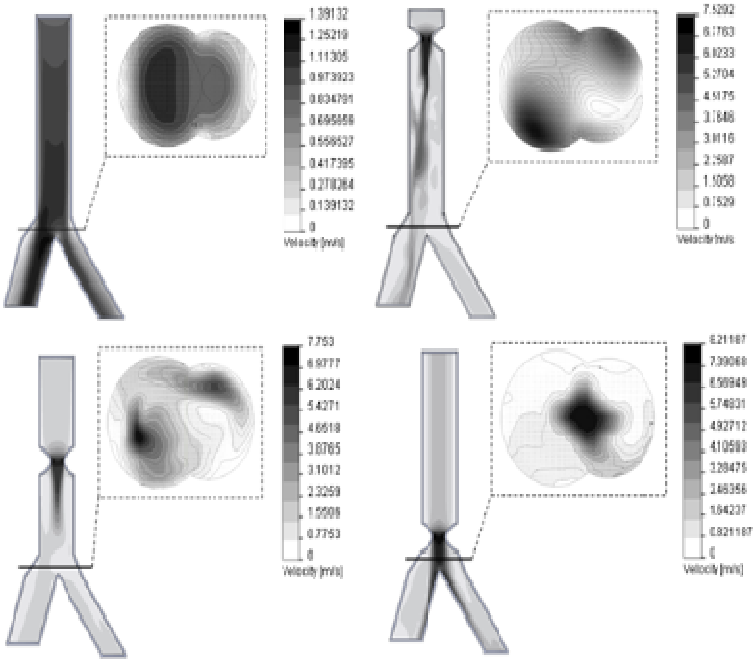
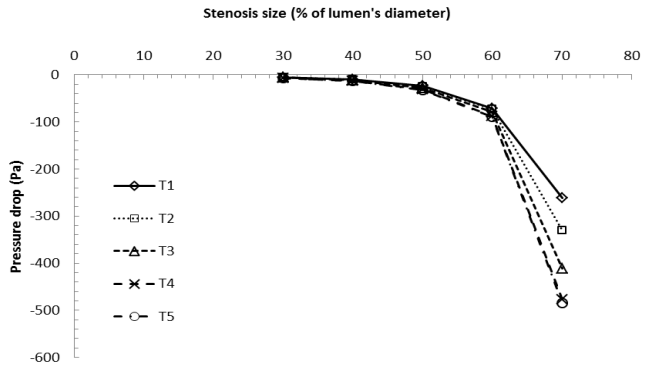


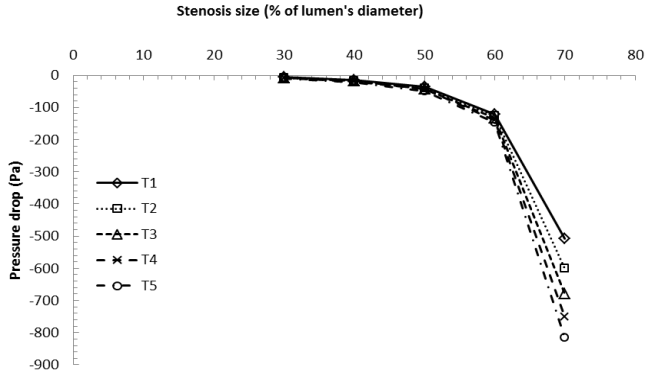
Figure 3. Velocity plots during resting condition (Re 1201) on the cross-sectional area for Normal model (a) and Trachea with 60% stenosis size at location 1, 2 and 3 (b, c and d respectively). Insets represent the horizontal cross-sectional area at bifurcation. Dark gray areas represent the regions where the higher velocity and flow rates intend to anywhere in the models.

These results are relatively in contrast with Brouns et al., (2007) [10] and Jayaraju et al., (2006) [11] where they reported that the steeply increment of pressure drop between inlet and outlet flow is beginning at 70-75% of stenosis size. This might be happened due to simplification of main airway that changed the orientation and magnitude of inlet air jet when compared to the more realistic geometry. During Re 46600, the pressure drop was seen to be significant increased beginning at the sizes of 30% of stenosis, which is earlier than others breathing conditions. As expected, the pressure drops noted was incredibly higher than others.

(a)



(b)



(c)

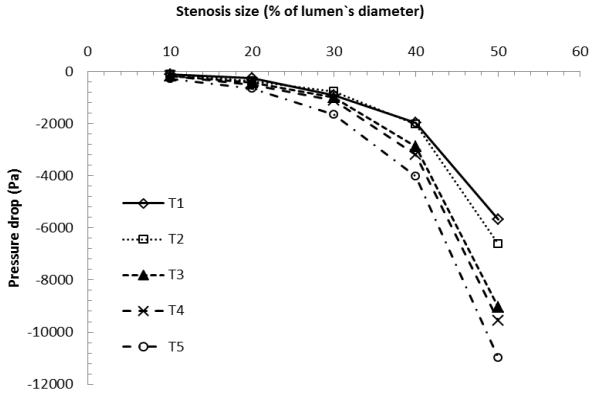


Figure 4. Pressure drops simulation during different breathing conditions (a) Re 1201, (b) Re 3012 and (c) Re 44600 as the function of various sizes and locations of stenosis.

4. CONCLUSION

The present simulations confirm that the overall flow behavior is only significantly affected if the size of stenosis is more than 60% and located further away from the bifurcation region, and at 50% or located close to the bifurcation region during resting and moderate exercise. However, the severity of stenosis can affect or causes the breathing difficulties early when the stenosis is 20% at any location while a person undergo extreme activity.

5. ACKNOWLEDGMENTS

The support of the Universiti Malaysia Pahang and Universiti Teknologi Malaysia are gratefully acknowledged.

6. REFERENCES

[1] Spittle, N., and McCluskey, A. Tracheal Stenosis after Intubation. PubMed Central, 321(7267): 1000–1002 (2000).

[2] Schuurmans MM, Bolliger CT Silicone airway stents. In: Interventional Pulmonary Medicine. Lung Biology in Health and Disease, edited by Beamis JF, Mathur PN, and Mehta AC. Dekker: New York, 2004, vol.189, p. 215–238 (2004).

[3] Yang. X.L., Y. Liu and H.Y. Luo. Respiratory Flow in Obstructed Airways, Journal of Biomechanics. 39: 2743–2751 (2006)..

[4] Hammer J., Acquired Upper Airway Obstruction. Paediatric Respiratory Review, 5:25–33 (2004)..

[5] McCaffrey TV. Classification of Laryngotracheal Stenosis, Laryngoscope, 1992; 102: 1335–1340 (1992)..

[6] Freitag L., M. Unger, A. Ernst, K. Kovits and C. Marquette. A proposed Classification System of Central Airway Stenosis. European Respiratory Journal. Vol. 30.no. 1, 7-12 (2007),.

[7] Gemci, T., V. Ponyavin, Y. Chen, H. Chen, and R. Collins. Computational Model of Airflow In Upper 17 Generations of Human Respiratory Tract. Journal of Biomechanics. 41: 2047-2054 (2008)..

[8] Romulo P., K. G. Tournoy and J. Sijbers. Assessment and Stenting of Tracheal Stenosis using Deformable Shape Models, Medical Image Analysis. 15, 250–266 (2011).

[9] Cebral, J., & Summers, R. Tracheal and Central Bronchial Aerodynamics using Virtual Bronchoscopy and Computational Fluid Dynamics. Medical Imaging, 8:1021-1033 (2004).

[10] Brouns, M., S.T.Jayaraju, C.Lacor, J.D.Mey, M.Noppen, W.Vincken. Tracheal Stenosis:a Flow Dynamics Study. J.Applied Physiology.102: 1178-1184 (2007).

[11] Jayaraju, S. T., Brouns, M., Lacor, C., Mey, J. D., & Verbanck, S. Effects of Tracheal Stenosis on Flow. European Conference on Computational Fluid Dynamics, Netherland (2006).

[12] Arpad Farkas and Imre Balashazy. Simulation of the Effect of Local Obstructions and Blockage on Airflow and Aerosol Deposition in Central Human Airways, Journal of Aerosol Science. 38, 865-884 (2007).

[13] Yang X.L., Y. Liu a, R.M.C. So a , J.M. Yang. The Effect of Inlet Velocity Profile on the Bifurcation COPD Airway Flow. Computers in Biology and Medicine 36: 181–194 (2006).

[14] Schlesinger RB, M. Lippmann. Particle Deposition in Casts of the Human Upper Tracheobroncial Tree". Am Ind Hyg Assoc J. 33:237-51 (1972).

[15] Cotton R.T. Pediatric laryngotracheal stenosis. J Pediatr Surg.19: 699–704 (1984).

[16] Ling W, Chung JN, Troutt TR, Crowe CT. Direct Numerical Simulation of a Three-dimensional Temporal Mixing Layer with Particle Dispersion. J Fluid Mech. 358:61–85 (1998).

[17] Pedley TJ. Pulmonary Fluid Dynamics. Ann Rev Fluid Mech. 9:229–74 (1977).

[18] Luo, H.Y, Y. Liu. Modeling the Bifurcating Flow in a CT-scanned Human Lung Airway. Journal of Biomechanics. 41:2681– 2688 (2008).

[19] Calay R.K., J. Kurujareon and A.E. Holdo. Numerical Simulation of Respiratory Flow Patterns within Human Lung. Respiratory Physiology & Neurobiology. 130:201-221 (2002).

[20] Luo X.Y, JS. Hinton, TT. Liew, KK. Tan, "LES modeling of flow in a simple airway model". J Med Eng Phys, 26:403 (2004).

[21] Allen G.M, BP. Shortfall, T. Gemci, TE. Corcoran and NA. Chigier, "Computational Simulations of Airflow in an In Vitro Model of the Pediatric Upper Airway". Journal of Biomechanical Engineering, 126, 604-613 (2004).

[22] Ahmed SA and DP. Giddens, "Velocity Measurement in Steady Flow Through Axis Symmetric Stenoses at Moderate Reynolds Number". J. Biomech., 16, 505-516 (1983).

[23] Horsfield, K., G. Dart, D.E. Olson, G.F. Filley, G. Cumming. Models of Human Branching Airways. J. of Applied Physiology. 31: 207-217 (1971).

[24] Zheng Li, C. Kleinstreur, Z. Zhang. Simulation of airflow fields and microparticle deposition in realistic human lung airway models. Part I: Airflow patterns. European Journal of Mechanics B/Fluids. 26: 632–649 (2007). Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (Nov. 1993), 795-825. DOI= http://doi.acm.org/10.1145/161468.16147.

Combining Web Mining and application of Structural Equation Modeling in E-commerce

Dinh Thuan Nguyen
University of Information Technology
Ho Chi Minh city, Vietnam
E-mail: thuannd@uit.edu.vn

Tuan Anh Nguyen
University of Information Technology
Ho Chi Minh city, Vietnam
E-mail: anhnt.0111@gmail.com

An Tung Phan
University of Information Technology
Ho Chi Minh city, Vietnam
E-mail: nlds.a279@gmail.com

ABSTRACT
The aim of this study is to develop a solution to collect product information from multiple websites, and then investigate the relationships of various factors that affect customer satisfaction on a product using Structural Equation Modeling (SEM), so that the system can suggest best products which are suitable for customer demands. These relationships were tested using a sample of 100 laptop shoppers from 4 popular electronic stores in HCM city. The result indicates that the satisfaction of customer with a product depends on some important factors such as product configuration, warranty time, post-sale service… which related to both product quality and service quality in general. All the steps involved in evaluating the coefficients of evaluation model are accomplished with the assistance of Statistical Package for the Social Sciences (SPSS), software used for statistical analysis. Implications for store managers and directions for future researches are also discussed.

Categories and Subject Descriptors
H.2.8 [**Information Systems**]: Database Applications - *Data mining,Statistical databases*

General Terms
Management

Keywords
Web mining, E-commerce, SEM

1. INTRODUCTION
With the rapid increase of available information online, especially with the growing popularity of electronic commerce, web data mining is being paid much attention, combination of web data mining and e-commerce has been a hot issue. The numbers of e-commerce websites with various brands and classes of products are growing in explotional rates make it difficult for customers to determine which products are suitable for them.

Combination of Web mining techniques and SEM provide us a solution for customer's problem. With web content mining, we can collect product information from many different ecommerce websites, and recommend products have highest satisfaction rating, which is calculated from customer survey data using SEM.

The role of causality in SEM research is widely perceived to be, on the one hand, of pivotal methodological importance and, on the other hand, confusing, enigmatic, and controversial. Structural equation modeling (SEM) is a research approach used in many academic disciplines, including information systems (Gefen, et al., 2000)[1] [7]. This method is often used to examine the perceptions of customers and has been used in e-commerce contexts to measure perceived value (Chen & Dubinsky, 2003)[5] and ratings of web-sites (Kwon, et al., 2002) [4][6]. The purpose of our paper, is not to build or test theory but to explore an alternative approach to survey data collection when creating SEM models that we can use for learning SEM technique and applying for electronic stores in Ho Chi Minh City.

2. METHODOLOGY
2.1. Web content mining
As the data on the web grows at explosive rates, it has led to several problems such as increased difficulty of finding relevant information, extracting potentially useful knowledge and learning about consumers or individual users. To solve the problem, many methods were proposed and developed, mainly including raw data acquisition, relevant data extraction and integration, wrapped data retrieval. Much work has been done on structured data process.

Raw data acquisition is implemented by web crawler. Data extraction is implemented by many approaches and this paper will focus on automatic wrapper generation. Data Integration research are not as many as data extraction.

Data extraction is a progress of extracting information from web pages. In this part, I will focus on structured data extraction. A program for extracting such data is usually called a wrapper.

Structured data are typically the data records retrieved from underlying database and displayed in the web pages following some templates. Sometime, the template is a table. Sometime, it is a form. Extracting such data records is useful because it enables us to obtain and integrate data from multiple sources (Web sites and pages) to provide value-added services such as: searching, comparative shopping, product consulting …

Basically, there are 3 approaches proposed:

+ Manual approach: by observing a web page and its source code, the human programmer finds some patterns and then writes program to extract the target data. There are several web data extraction libraries have been built to make it simplifier for programmers. This approach is simple, fast to develop, suitable for small number of websites but not scalable to large number of sites.

+ Wrapper induction: is the supervised learning approach, and is semi-automatic. In this approach, the users marks the target items in a few training pages, the system then learns extraction rules from these pages, and the rules are applied to extract target data from other similar formatted pages.

+ Automatic extraction: is the unsupervised approach, and is automatic. In this approach, a single or multiple pages are given, it automatically find patterns or grammars from them for data extraction. Since this approach eliminates the manual labeling effort, it can scale up data extraction to a huge number of sites and pages.

This paper will focus on the first approach because we only work with limited of sources. JSOUP, a Java library is used for extracted target data from websites.

2.1 Customer satisfaction and SEM

Structural equation modeling (SEM) is a statistical technique for testing and estimating causal relations using a combination of statistical data and qualitative causal assumptions.

Service quality is a focused evaluation that reflects the customer's perception of reliability, assurance, responsiveness, empathy and tangible. Satisfaction, on the other hand, is more inclusive: it is influenced by perceptions of service quality, product quality and price as well as situational factors and personal factors. Satisfaction is a broader concept, and can be measured by service quality.

To evaluate service quality, a service quality framework named SERVQUAL was developed in the mid 1980s by Zeithaml, Parasuraman & Berry with 10 aspects of service quality. By the early 1990s, the authors had refined the model with 5 aspects, and it is widely used in published and modified forms to measure customer expectations and perceptions of service quality.

With the background of SERVQUAL, we suggest an evaluation model with some adjustments for laptop market area, to measure customer satisfaction for a specific laptop product.

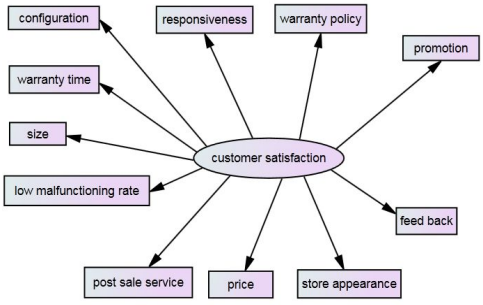


Figure 1. Theory evaluation model

H1: product configuration affects customer satisfaction on a product, satisfaction on configuration increases, satisfaction on that product increases and vice versa.

H2: warranty time affects customer satisfaction on a product, satisfaction on warranty time increases, satisfaction on that product increases and vice versa.

H3: product size affects customer satisfaction on a product, the size of a laptop decreases, satisfaction on that product increases and vice versa.

H4: malfunctioning rate affects customer satisfaction on a product, low malfunctioning rate increases satisfaction on that product and vice versa.

H5: responsiveness affects customer satisfaction on a product, satisfaction on responsiveness increases, satisfaction on that product increases and vice versa.

H6: warranty policy affects customer satisfaction on a product, strict warranty policy decreases satisfaction on that product and vice versa.

H7: promotion affects customer satisfaction on a product, interested promotion increases satisfaction on that product and vice versa.

H8: post-sale service affects customer satisfaction on a product, satisfaction on post-sale service increases, satisfaction on that product increases and vice versa.

H9: product price affects customer satisfaction on a product, satisfaction on price increases, satisfaction on that product increases and vice versa.x

H10: store appearance affects customer satisfaction on a product, satisfaction on store appearance increases, satisfaction on that product increases and vice versa.

H11: feedback customer perceives via media affects customer satisfaction on a product, good feedback increases customer satisfaction on a product and vice versa.

ode	Survey question	Answer						
onf	Does product configuration fulfill your demands?	1	2	3	4	5	6	7
rtt	Do you satisfy with product warranty time?	1	2	3	4	5	6	7
ize	How do you feel about product size?	1	2	3	4	5	6	7
alr	Does your product have low malfunctioning rate?	1	2	3	4	5	6	7
esp	Do company staffs willing to support and provide information about various kind of products?	1	2	3	4	5	6	7
rtp	How do you feel about company warranty policy?	1	2	3	4	5	6	7
rmt	Do company provide any promotion when you buy this product?	1	2	3	4	5	6	7

sts	Do the company willing to response to your concerns after buying product, or when you got problem with product?	1	2	3	4	5	6	7
ric e	Do you feel the product price is worthy?	1	2	3	4	5	6	7
bac k	How do you perceive the reputation of store through media (Internet, newspaper,…)?	1	2	3	4	5	6	7
tra pp	How do you feel about store appearance and product arrangement?	1	2	3	4	5	6	7
as	In general, do you satisfy with the product?	1	2	3	4	5	6	7

Data for constructing evaluation model are taken by conducting survey on customers who buy laptop product from 4 stores at HCM city. Below is the list of question corresponding to 11 hypothesizes in the model Multiple Linear Regressions is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of MLR is to model the relationship between the explanatory and response variables.

Population regression function (PRF):

$$Yi = B_1+B_2X_{2i}+B_3X_{3i}+\ldots+ B_kX_{ki}+ U_i$$

Where:

Y: independent variable and X_2, X_3, …, X_k is dependent variables.

B_1, B_2, B_3, …, B_k is coefficients of factors, used for building a prediction equation,

Ui stand for random errors.

We need to identify B_1, B_2, … B_k via n observations:

1st observation:

$$Y_1 = B_1 + B_2X_{21} + B_3X_{31} + \ldots + B_kX_{k1} + U_1$$

2nd observation:

$$Y_2 = B_1 + B_2X_{22} + B_3X_{32} + \ldots + B_kX_{k2} + U_2$$

nth observation:

$$Y_n = B_1 + B_2X_{2n} + B_3X_{3n} + \ldots + B_kX_{kn} + U_n$$

Let

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix} \qquad B = \begin{pmatrix} B_1 \\ B_2 \\ \ldots \\ B_n \end{pmatrix} \qquad U = \begin{pmatrix} U_1 \\ U_2 \\ \ldots \\ U_n \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & X_{21} & X_{31} & \ldots & X_{k1} \\ 1 & X_{22} & X_{32} & \ldots & X_{k2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{2n} & X_{3n} & \ldots & X_{kn} \end{pmatrix}$$

Then we have

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{21} & X_{31} & \ldots & X_{k1} \\ 1 & X_{22} & X_{32} & \ldots & X_{k2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{2n} & X_{3n} & \ldots & X_{kn} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \ldots \\ B_n \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \\ \ldots \\ U_n \end{pmatrix}$$

Now RF becomes: Y = X.B + U

3. CHIEVEMENTS

3.1 Web data collection

Considering the hot e-commerce, we choose laptop selling stores as the experimental object.

Collected domains: phongvu.vn, hoanlong.com.vn, tnc.com.vn, thegioididong.com in Ho Chi Minh city.

Search results are like below:

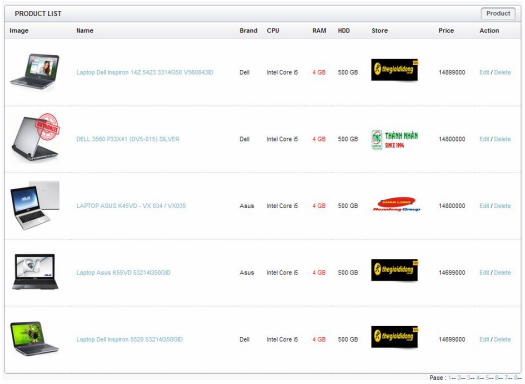


Figure 2. Search results

3.2 Constructing evaluation model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.895[a]	.802	.777	.567

a. Predictors: (Constant), strapp, mair, wrtt, resp, price, prmt, size, fback, psts, wrtp, conf

Figure 3. Model summary

R^2 = the coefficient of determination. This value determines how much of the variation in one variable is due to the other variable. In this research, R2 = 0.802, so 80.2% of the variation in the

customer satisfaction is determined by the predictor variables (and 19.8% of the variation is caused by undiscovered factors, or errors in survey data).

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	114.672	11	10.425	32.384	.000[a]
	Residual	28.328	88	.322		
	Total	143.000	99			

a. Predictors: (Constant), strapp, malr, wrtt, resp, price, prmt, size, fback, psts, wrtp, conf

b. Dependent Variable: sas1

Figure 4. ANOVA table

The ANOVA output reports whether the model results in statistically significant prediction. In this case, Significance = .000, so the outcome is statistically significant (i.e. p <.05), null hypothesis (all factor coefficients are zero) is rejected.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.961	.362		-2.653	.009
	conf	.390	.069	.384	5.629	.000
	wrtt	.159	.057	.173	2.796	.006
	size	-.094	.050	-.114	-1.898	.061
	malr	.155	.044	.198	3.518	.001
	resp	.208	.049	.238	4.197	.000
	wrtp	-.150	.063	-.161	-2.395	.019
	prmt	.097	.041	.140	2.387	.019
	psts	.115	.050	.145	2.312	.023
	price	.216	.056	.241	3.841	.000
	fback	.021	.055	.023	.375	.708
	strapp	.108	.050	.124	2.145	.035

a. Dependent Variable: sas1

Figure 5. Coefficients result

In this table, the column B provides the value for factor coefficients of the evaluation equation.

All of above factors have .sig value (or p-value) smaller than 0.1 except fback, it shows that these variables (except fback) are significantly different from 0 at the 0.1 alpha levels, or we can state that these factors (except fback) have affected on customer satisfaction.

In the case of fback, .sig value greater than 0.1 (0.708), it is not statistically significantly different from 0, in the other hand, we can state that the store credibility customer perceives via media (Internet, magazine…) does not have any influence on customer satisfaction, and need to be eliminated from evaluation model.

Analysis result after removing fback

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.895[a]	.802	.779	.565

a. Predictors: (Constant), strapp, malr, wrtt, resp, price, prmt, size, psts, wrtp, conf

Figure 6. Model summary without fback

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.945	.358		-2.640	.010
	conf	.389	.069	.383	5.648	.000
	wrtt	.160	.056	.174	2.832	.006
	size	-.094	.049	-.113	-1.898	.061
	malr	.154	.044	.197	3.518	.001
	resp	.209	.049	.241	4.278	.000
	wrtp	-.149	.062	-.160	-2.391	.019
	prmt	.099	.040	.143	2.470	.015
	psts	.117	.049	.147	2.358	.021
	price	.221	.054	.247	4.096	.000
	strapp	.115	.047	.132	2.439	.017

a. Dependent Variable: sas1

Figure 7. Coefficients result after eliminating fback

There are no significant differences, but the estimated values of factors conf – the coefficient for laptop configuration is 0.389, so for every unit increase in configuration, a 0.389 unit increase in customer satisfaction is predicted, and this is the factor that have the highest affection to general satisfaction in this research.

wrtt – the coefficient for warranty time is 0.160, for every unit increase in warranty time, 0.158 unit increase in customer satisfaction.

size – the coefficient for warranty time is - 0.094, for every unit increase in laptop size, 0.094 unit decrease in customer satisfaction.

malr – the coefficient for low malfunctioning rate is 0.154, for every unit increase in low malfunctioning rate, 0.154 unit increase in customer satisfaction.

resp - the coefficient for responsiveness is 0.209, for every unit increase in responsiveness, 0.209 unit increase in customer satisfaction.

wrtp - the coefficient for warranty policy is - 0.149, for every unit increase in warranty policy, 0.158 unit increase in customer satisfaction.

prmt - the coefficient for promotion is 0.099, for every unit increase in promotion, 0.099 unit increase in customer satisfaction.

psts - the coefficient for post-sale service is 0.117, for every unit increase in warranty time, 0.117 unit increase in customer satisfaction.

price - the coefficient for price is 0.221, for every unit increase in price, 0.221 unit increase in customer satisfaction.

strapp - the coefficient for store appearance is 0.115, for every unit increase in store appearance, 0.115 unit increase in customer satisfaction.

Customer satisfaction =

945 + 0.389*configuration + 0.160*warranty time 0.094*size + 0.154*low malfunctioning rate + 0.209*responsiveness – 0.158*warranty policy + 0.099*promotion + 0.117*post-sale service + 0.221*price + 0.115*store appearance

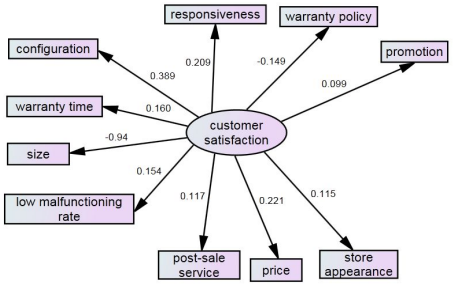


Figure 8. Result model

4. CONCLUSION

Customer satisfaction is the most valuable asset of all businesses. In highly competitive market nowadays, it is essential for businesses to effectively manage customer satisfaction. Therefore, businesses who want to success need to identify what factors affect customer satisfaction and have a suitable strategy to improve them. Via a research conducted on 4 popular electronic stores in Ho Chi Minh City, customer satisfaction on a laptop product has been seen to be depended on various factors such as product configuration, responsiveness from store, product price, post-sale service, … In all of them, product configuration fulfilling usage demands, product price and responsiveness of service provider seem to be most effective aspects that businesses need to focus on in order to increase customer satisfaction on their service.

Web mining with support from a HTML extracted framework like JSOUP allow us to collect products information from multiple sources available in the Internet to a single application. With extracted data and the result from customer satisfaction model research we can develop a system to consult the most satisfied products to customers, which reduce their effort to search for a good product, and in some ways support businesses in improving their services.

5. REFERENCES

[1] Kenny, D. A. and Milan, S. (2011*). Identification: A non-technical discussion of a technical issue*. In Handbook of Structural Equation Modeling (R. Hoyle, ed.). Guilford Press, New York, This volume.

[2] Williams, L. J. (2011). *Equivalent models: Concepts, problems, alternatives.* In Handbook of Structural Equation Modeling (R. Hoyle, ed.). Guilford Press, New York, This volume.

[3] Temme D, Kreis H, Hildebrandt L (2010). *A Comparison of Current PLS Path Modeling Software: Features, Ease-of-Use, and Performance."* In V Esposito Vinzi, WW Chin, J. Henseler, HF Wang (eds.), Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields, chapter 31. Springer-Verlag, Berlin.

[4] Lee, S.-Y., & Song, X.-Y. (2004). Maximum *likelihood analysis of a general latent ariable model with hierarchically mixed data.* Biometrics, 60, 624–636.

[5] Chen, Z., & Dubinsky, A. J. (2003). *A Conceptual Model of Perceived Customer Value in E-Commerce: A preliminary Investigation*. Psychology & Marketing, 20(4): 323-347.

[6] Kwon, O. B., Kim, C., & Lee, E. J. (2002). *Impact of website information design factors on consumer ratings of web-based auction sites.* Behaviour & Information Technology, 21(6): 387-402.

[7] Gefen, D., Straub, D. W., & Boudreau, M. (2000). *Structural Equation Modeling and Regression: Guidelines for Research Practice.* Communications of the Association for Information Systems, 4(7): 1-79.

Energy Efficient New Symmetric Key Algorithm (AP) for WSN

Archana Tayal
M.Tech Student
CSE and IT Department
ITM University, Gurgaon, India
+919466817931
tayal.archi@gmail.com

Prachi
Assistant Professor
CSE and IT Department
ITM University, Gurgaon, India
+919818992054
prachiah1985@gmail.com

ABSTRACT
Applications of wireless sensor network are increasing day by day. Data nodes in sensor network are easy to capture and confidential data of sensor nodes can be accessed by eavesdropper. Security has always been troublesome in the wireless communication. Cryptography algorithms are kernel of the WSN security. In this paper we present a new symmetric key algorithm (AP) based on shuffling, substitution and shifting to depict a security scheme for WSN which is energy efficient as well as difficult to crack. This paper features the time taken by algorithm for different key size and number of rounds along with the comparative analysis of proposed algorithm with AES on various parameters to prove its efficiency.

Categories and Subject Descriptors
C.2.0 **[Computer-Communication Networks]:** General – *Security and protection*; E.3 **[Data]:** Data Encryption – *code breaking;* H.3.4. **[Information Storage and Retrieval]:** Systems and Software – *Performance evaluation (efficiency and effectiveness).*

General Terms
Algorithms, Performance, Design, Experimentation, Security.

Keywords
Encryption, Security, Symmetric Key.

1. INTRODUCTION
In the last few years there has been a tremendous increase in the real life applications of wireless sensor network such as health care, environmental monitoring, structural monitoring, automobile services and data logging. Both civil and military applications are shifting towards WSN as sensor nodes are easy to deploy However, for these real life applications, security mechanisms are required to defend WSNs from malevolent attacks [1]. Battery life time of sensor nodes and energy of network are limited. It is a necessity to develop an energy efficient security mechanism which is hard to decrypt by third party and consumes low power.

Cryptography an art of hiding the text intelligence, is used to resolve crucial security issues. Cryptography is a way of secret writing where original meaningful data is transformed to a non-meaningful data which has no significance and non understandable. It uses encryption for conversion of original text to secret text known as cipher text and decryption to retrieve original text back from cipher text. This mechanism is based on key management. Based on the number of keys employed for encryption and decryption, there are mainly two types of algorithms as shown in figure 1 [2].

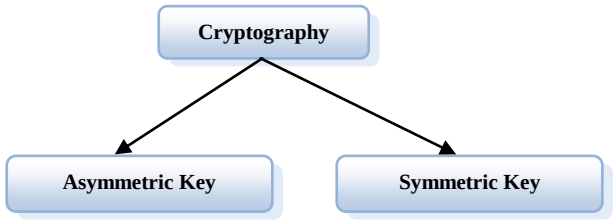


Figure 1. Classification of Cryptography.

1.1 Public Key Cryptography
Public key cryptography or asymmetric cryptography employs different keys for encryption and decryption. The key used for encryption is public key. Sender can encrypt the data using public key which is open for everyone, but only receiver has the private decryption key to obtain the original message.

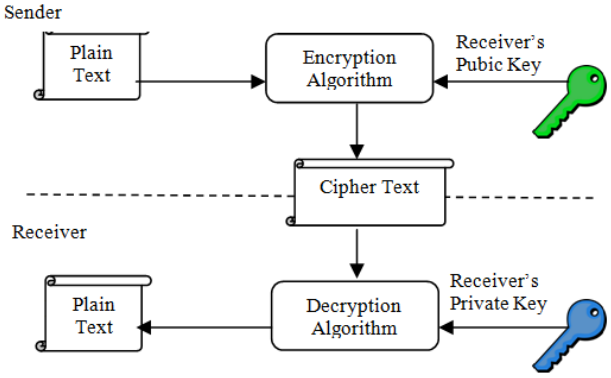


Figure 2. Public key cryptography.

Some asymmetric key ciphers as per [3] are :

- SSL handshake
- TinyPK
- RSA

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.35

1.2 Symmetric key cryptography

Symmetric key cryptography uses the single key for both encryption and decryption. It is shared by parties at both the ends. As represented in figure 2, the same key has been used by both sender and receiver.

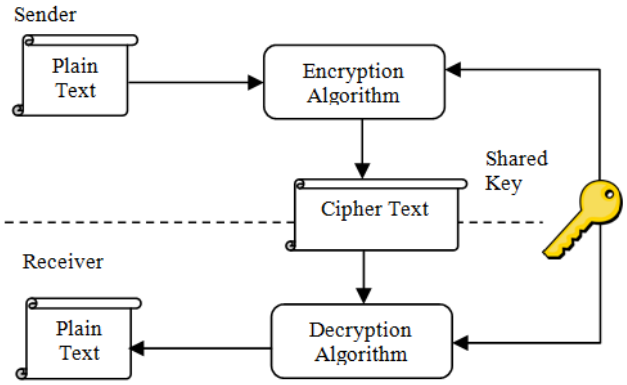


Figure 3. Symmetry key.

Some symmetric cryptography schemes as per [4] are:

- DES
- AES
- RC5

As stated by Shish Ahmad et. Al. in [5] symmetric key algorithms are preferred over the public key cryptography. Public key algorithms consume more energy and are less efficient than the symmetric key cryptography. The key issues for cryptographic algorithm designer are : cost, security and performance [6]. In order to resolve severe security issues, we present an algorithm on the basis of combination of existing techniques like P-box, vigenere cipher; circular shift left.

Many encryption/decryption techniques have been proposed by researchers and some most commonly used are discussed in Section 2. We have also briefly explained the threats and challenges to WSN in section 3. In section 4 new symmetric key energy efficient algorithm has been presented. Experimental results are shown in section 5. Section 6 concludes the paper with future work.

2. RELATED WORK

Many symmetric and asymmetric encryption ciphers have been proposed and developed by researchers. Literature demonstrates symmetric ciphers are more efficient as compared to asymmetric ciphers. Some common symmetric key ciphers are discussed below.

2.1 DES

DES is a 64 bit block cipher It encrypts 64 bits of data at a time using key of length 56 bits. 16 rounds of processing are applied to plain text after initial permutation and the last step is to apply final permutation to obtain a 64 bit cipher text. DES uses fiestel structure and its design is an inspiration for many block ciphers. Wuling Ren et. al. in [7] proposed an hybrid encryption mechanism based on DES and RSA. RSA is a public key encryption algorithm and used for the encryption of DES's key to be shared between shared and receiver. DES was prone to Brute Force Attack as there are only 2^{56} combinations of key are possible. It was quite easy to crack the DES and making it non-preferable approach for WSN security.

2.2 Triple DES

Triple DES or 3DES was proposed to enhance the security mechanism of DES. Key size was extended from 56 bits to 168 bits (56*3). Also it used 3 rounds of DES encryption. 3DES can be used in two ways either with three keys having 2^{168} possible combinations or with two keys having 2^{112} possible combinations. Cryptanalysis of 3DES shows that with so many combinations of key, brute force attack is practically impossible [8]. Large key size makes 3DES strongest encryption algorithm, but however it requires a large amount of time and memory and thus making it an expensive algorithm.

2.3 AES

As per [9] Advance Encryption Standard (AES) also referred as Rijndael cipher is a block cipher with block size of 128 bits and three different key sizes of 128, 192 and 256 bits. AES uses a non-fiestel structure for encryption decryption. Key size depends on the number of rounds which can be 10,12 or 14 for key size of 128,192 or 256 bits respectively. Round key used by AES is always of 128 bits. As explained in [10] basic Structure of AES is represented in figure 4. First transformation of round is *Substitution* where substitution of bytes is done via using S-box. After a state is substituted another transformation that is applied to state is *Shifting*. Number of shifts of state matrix depends on the row number that can be 0,1,2 or 3. *Mixcolumn* transformation is matrix multiplication where state column is multiplies by a constant matrix. Finally a round key word is added with each state matrix in *Addroundkey* transformation.

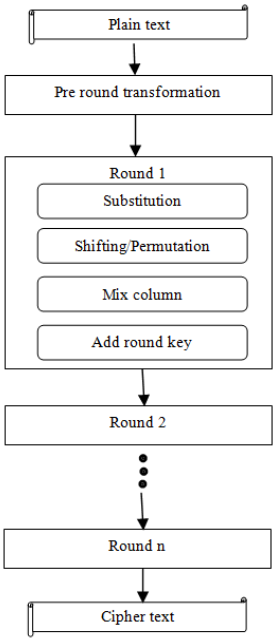


Figure 4. AES structure.

2.4 RC5

As per [11] RC5 is a potential cipher for data security in WBSN. It is a efficient and flexible algorithm where encryption attributes

like number of rounds, block size, key length can be adjusted according to the application. As discussed in [12] RC5 uses operations like modulo 2word-size addition, bit wise exclusive-or and a cyclic left rotation of word in each round. RC5 with short parameter value is susceptible to differential attack; whereas RC5 with large parameters value is time consuming.

2.5 Skip Jack

Skip Jack is also an symmetric key block cipher encryption algorithm. It require key of size 80 bits, data block of 64 bits and 32 rounds in an unbalanced Fiestel network [13]. F-Box stored in the system temporary memory RAM or program memory is used for functional calculations. Known attacks like brute-force cannot break Skip Jack but however the amount of time taken for F-box calculation makes it torpid. Also a large amount of RAM is required which further slowdowns the network.

3. CHALLENGES

Today wireless communication is facing many threats and challenges like confidentiality, integrity, masquedering, non-repudant, authenticity etc. Intruder attacks WSN in many ways in order to obtain the secret information. We have briefly reviewed them.

3.1 Confidentiality

A third party may access the data transferred over the internet and use it for his own benefits. Military data is highly secret. Any tampering with this confidential data can endanger the security of a nation. So it is inevitable that the transmitted data is highly secure.

3.2 Integrity

Data transmission over the internet is dynamic and ongoing process.. The changes are incorporated in the information from time to time but these should be made by authorized person only. In banking, suppose 'A' sends $100 to 'B'; but later on 'B' or someone else modifies the amount from $100 to $10000. This is an attack to data integrity.

3.3 Authenticity

Authorized persons are allowed to access the privileged information and kept it secret from the unauthorized access. In schools and colleges teachers mail question paper to each other. These question paper are accessible to teachers only. Students are unauthorized and therefore not allowed to access question papers.

3.4 Non-Repudant

A sender or receiver may deny later about the data transferred over the internet. It can be stated either the sender denies about the payment transaction made from his account or the shopping websites denying about the payment received for the bill.

3.5 Masquedering

Intruder may wear a mask of authorized entity befooling others and accessing the secret information to get his evil means done. This can be illustrated by a spy of an organization wearing a mask of an employee of another organization to know their future plans.

3.6 Freshness

Real time and continuous data transmission in required in WBSN. Thus it is desirable that the security algorithm must be efficient and requiring minimal time for sending fresh data to the hospital staff. Time is crucial parameter for WBSN security algorithm. Delay in data transmission may endanger the patient's life.

4. PROPOSED ALGORITHM

In this paper we are presenting a block cipher symmetric key algorithm. The set of operation known as 3S are shuffling, substitution and shift left those are to be applied to the plain text in each round. Rounds can vary from 2^0 to 2^{10}. Plain Text could be of any length. Though it is a block cipher but no padding is required in the proposed algorithm. Algorithm can be more elaborated from the figure 7.

AP_Encryption_Algorithm

1. Given Plain Text.
2. Randomly generate key k
3. Calculate key k2 and key 3 from the key k.
4. Repeat
   5. Divide the n bits of plain text P into r multiple blocks of key size k such that
      n = k * r + m
      where m is mod(n,k)
   6. Shuffle r blocks using key k
   7. Substitute the text (n bits) using key k2
   8. Shift the text in circular left shift with k3
9. Until all round done.

Figure 5. AP Encryption Algorithm.

AP_Decryption_ Algorithm

1. Given Cipher Text and key k
2. Calculate key k_inv , k2 & k3 from the key k.
3. Repeat
   4. Shift the text in circular right shift with k3
   5. Substitute the text (n bits) using key k2
   6. Divide the n bits of plain text P into r multiple blocks of key size k such that
      n = k * r + m
      where m is mod(n,k)
   7. Shuffle r blocks using inverse key k_inv
8. Until all round done.

Figure 6. AP Decryption Algorithm.

For explaining the algorithm, here we are considering the key of 16 bits and explaining how operations are processed in each round.

4.1 Shuffling

In our algorithm, initially we generated a permutation table using P-box of size 48 bits to shuffle the plain text. Thus, creating a key space of size 48 ! i.e. 1.2414e + 061 which is sufficiently large enough for the intruder to crack. To elaborate working of algorithm an example is considered with key of 16 bits.

If the plain text P1 is

"welcometounivers"

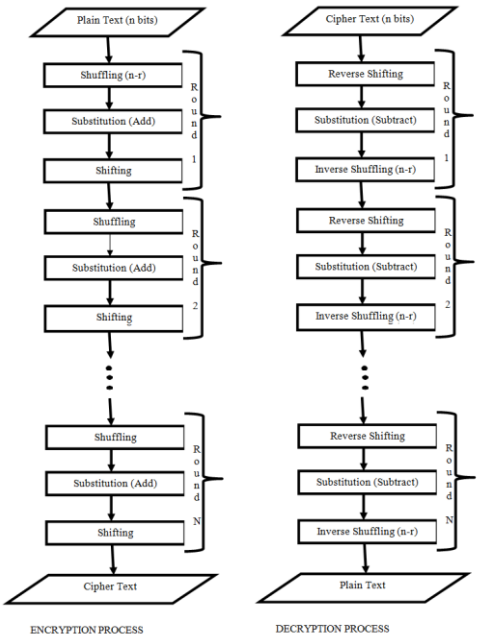


Figure 7. AP structure.

and P-box key k1 is

[4 8 5 3 11 14 6 13 10 16 2 15 12 7 1 9]

then the intermediary cipher text C1 will be

" ctolnemvuseriewo"

Thus cipher text has lost its actual meaning. Therefore shuffling the original text hides the meaning of message.

4.2 Substitution

Substitution is performed using Vigenere Cipher. Vigenere Cipher is a polyalphabetic substitution where text is envrypted using a series of additive cipher. An additive cipher is a traditional cipher where text is shifted ahead to a particular number. Text 'M' is substituted to 'U' for additive key = 8.

In our algorithm key to vigenere cipher will depend on the previous key k1. This key will be calculated by applying the operation.

$$k2 = \Sigma_{\,I = 1 \text{ to } 16}\, k1_i * 2^{(16-i)}$$

Applying the above operation we get

k2 = 993163

thus giving the intermediary cipher text C2

" lcrmthvextkurnzp "

4.3 Shift

Cipher text created in previous step is applied a circular shift by a certain number of times, and this number of shift depends upon the key k2 used by vigenere cipher.

Key k3 for circular left shift is the leftmost digit of key k2.

k3 = leftmost(k2)

In the given example k3 = 9 and generated final cipher text

C3 = "tkurnzplcrmthvex" .

This set of operation is applied a number of times depending upon the number of rounds and thus making it secure from various attacks like brute force, pattern attack, frequency test and impersonates this cipher with key of 48 bits and 100-150 rounds.

5. ANALYSIS

We implemented the proposed algorithm using MATLAB and calculated CPU time taken by algorithm to convert plain text to cipher text and vice-versa corresponding to different text size. Figure 8 show the CPU time for different values of number of rounds keeping the key size to 48 bits. Figure 9 shows the CPU time varying the key size keeping number of rounds to 100 rounds. In our setup we found that the CPU time increases with rounds whereas changing the key size don't affect the CPU time. Hence a key length of 48 bits provides 48! possible combinations which is much greater than the possible combinations for a 48 bit key of RC5 algorithm and therefore making it more secure.

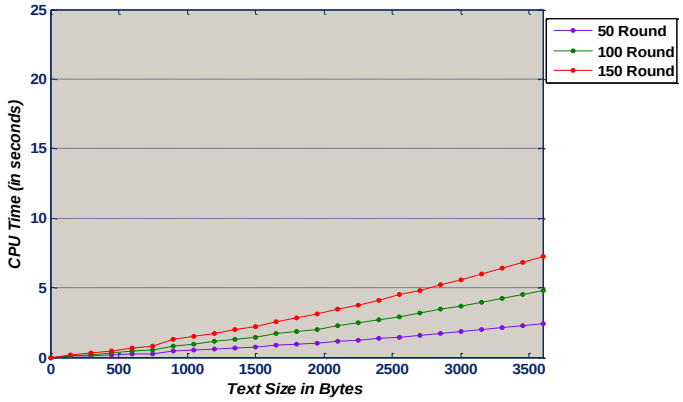


Figure 8. CPU Time for different number of rounds of AP.

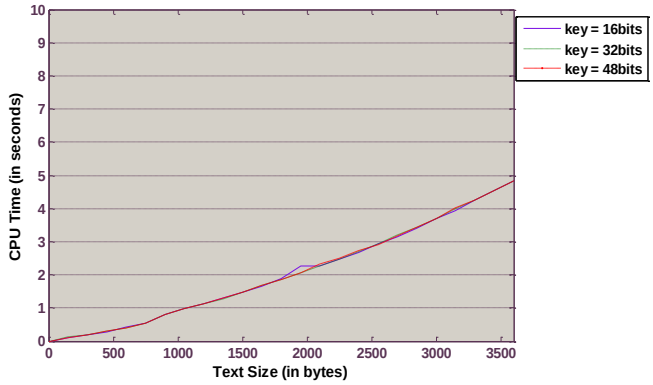


Figure 9. CPU Time for different key sizes of AP.

We also compared proposed algorithm AP (48 bits, 100 rounds) and AES (128 bits, 10 rounds). The comparison result as shown in figure 10 shows that time consumption of algorithm (AP) is much less than that of AES. Computation of energy consumption and throughput is done as per equations given in [14]. Voltage supply and time period is fixed for the chosen hardware, number of cycles is measured with the help of resource monitor. Average current drawn is taken as stated by Olaf Landsiedel et. Al. in [15]. Figure 11 and table 1 shows the comparative analysis of AP with

AES on the basis of various parameters like no of CPU cycles, key size, energy consumption and throughput.

Table 1. Comparison of AP and AES.

Algorithm / Parameter	AP	AES
No. of CPU cycles	5	25
CPU Time (in seconds)	7.27	1544.4
Key size (bits)	48	128
Energy Consumption (Joule)	27.027	135.135
Throughout	495	15.34

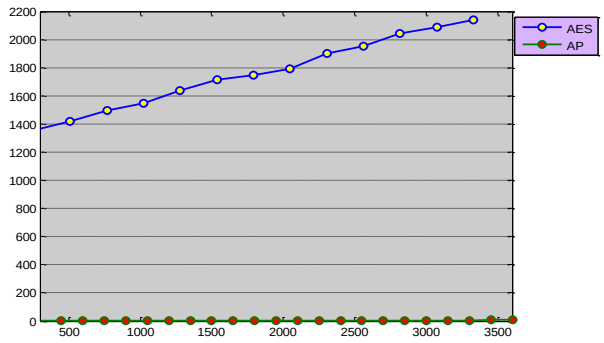


Figure 10. CPU Time comparison of AP with AES.

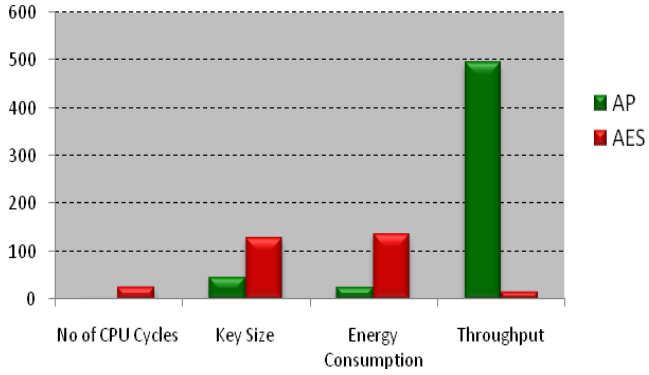


Figure 11. Comparison of AP and AES.

6. CONCLUSION

In this paper we briefly discussed about cryptography, various symmetric key algorithms, threats and challenges to WSN. We proposed an algorithm (AP) for security in WSN and shown that it is much faster, require less memory, energy efficient and have high throughput as compare to AES. AP can be used for securing WSN real life applications from malicious attacks in order to promote them to WSN. In future work, we will explore how we can deploy our proposed algorithm to safeguard various real life applications of WSN.

7. REFERENCES

[1] Zhou,Y., Fang, Y., Zhang,Y., 2008. Securing Wireless Sensor Networks:A Survey. In *Communications Surveys & Tutorials, IEEE, 10(3),*6-28.

[2] Chowdhury, M. J. M., Pal, T., 2009 A New Symmetric Key Encryption Algorithm based on 2-d Geometry. In *Proceedings of International Conference on Electronic Computer Technology* (Macau, February 20-22, 2009), IEEE, 541-544.

[3] Lokesh, J., Munivef, E., 2009. Design of Robust and Secure Encryption Scheme for WSN using PKI (LWT-PKI). In *Proceedings of the First international conference on Communication Systems and Network* (Bangalore, January 5-10, 2009), IEEE, 1-2.

[4] Hager, C. T. R., Midkiff, S. F., Park, J. M., Martin, T. L. 2005. Performance and Energy Efficiency of Block Ciphers in Personal Digital Assistants. In *Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications* (Kauai Island, Hawaii, March 8-12, 2005), 127-136.

[5] Ahmad, S., Beg, M. R., Abbas, Q. 2010. Energy Efficient Sensor Network Security Using Stream Cipher Mode of Operation. In *Proceeding of International Conference on Computer and Communication Technology (ICCCT)* (Allahabad, Uttar Pradesh, September, 17-19, 2010), 348-354.

[6] Karuppiah, A. B., Rajaram, S. 2012. Energy Efficient Encryption Algorithm for Wireless Sensor Network. In *Proceeding of International Journal of Engineering Research & Technology, ESRSA, 1(3)*, 1-7.

[7] Ren, W., Miao, Z., 2010. A Hybrid Encryption Algorithm Based on DES and RSA in Bluetooth Communication. In *Proceedings of the Second International Conference on Modeling, Simulation and Visualization Methods* (Sanya, May 15-16, 2010), 221=225.

[8] Agrawal, M., Mishra, P. 2012. A Comparative Survey on Symmetric Key Encryption Techniques. In *International Journal onComputer Science and Engineering, Inderscience 4(5),* 877-882.

[9] Forouzan, B. A. *Cryptography & Network Security.* Tata McGraw-Hill Publishing Company Limited, New Delhi, 2007.

[10] Shivkumar, S., Umamaheswari, G. 2011. Performance Comparison of Advanced Encryption Standard (AES) and AES key dependent S-box - Simulation using MATLAB. In *Proceeding of International Conference on Process Automation, Control and Computing* (Coimbatore, July 20-22, 2011), 1-6.

[11] Gawali, D. H., Wadhai, V. M. 2012. Rc5 algorithm: potential cipher solution for security in wireless body sensor networks (WBSN). In *International Journal Of Advanced Smart Sensor Network Systems, AIRCC, 2(3)*, 1-7.

[12] Rivest, R. L. 1994. The RC5 Encryption Algorithm. In *Proceedings of the Second International Workshop on Fast Software Encryption (FSE),*86-96.

[13] Koo, W. K., Lee, H., Kim, Y. H., Lee, D. H. 2008. Implementation and Analysis of New Lightweight Cryptographic Algorithm Suitable for Wireless Sensor Networks. In *Proceeding of International Conference on Information Security and Assurance* (Busan, April 24-26, 2008), 73-76.

[14] Salama, D., Kader, H. A., Hadhoud, M. 2011. Studying the Effects of Most Common Encryption Algorithms. In *International Arab Journal of e-Technology, 2(1),* 1-10.

[15] Landsiedel, O., Wehrle, K., Gotz, S. 2005. Accurate Prediction of Power Consumption in Sensor Networks. In *Proceedings of the 2nd IEEE workshop on Embedded Networked Sensors* (May 30-31, 2005), 37-4.

A Generalized Approach On Design And Control Methods Synthesis Of Delta Robot

Trinh Duc Cuong
Department of Mechatronics,
University of Technical Education
Ho Chi Minh City, Viet Nam
+84-903.839.238
cuongtdc@hotmail.com

Tuong Phuoc Tho
Department of Mechatronics,
University of Technical Education
Ho Chi Minh City, Viet Nam
+84-909.160.264
tuongphuoctho@gmail.com

Nguyen Truong Thinh
Department of Mechatronics,
University of Technical Education
Ho Chi Minh City, Viet Nam
+84-903.675.673
thinhnt@hcmute.edu.vn

ABSTRACT
This paper will describe the kinematics and dynamics of parallel robot named Delta with 3 degree of freedom (d.o.f). The use of dynamics coupled with kinematics for the control of parallel robot has been gaining increasing popularity in recent years. Relationship between generalized and articular velocities is established, hence jacobian and inverse jacobian analyses are determines. The inverse formulas are generally shown simply and the direct formulas are also described. Besides, this paper deal with the direct and inverse dynamics to determine the relations between the generalized accelerations, velocities, coordinates of the end-effector and the articular forces based on simulation and control. Parallel robots have become the important machines to manufacturing. They are used for various purposes in industry and life. The dynamic model of parallel robot with 3 dof is presented, and an adaptive control strategy for this robot is described. The robustness of the control system with respect to the nonlinear dynamic behavior and parameter uncertainties is investigated by computer simulation. Experiments were implemented to evaluate the responding of controlling system based on dynamics and kinematics controlling method for tracking desired trajectories. The results show that the use of the suitable control system based on dynamics model can provide the high performance of the robot.

Categories and Subject Descriptors
B.1.2 **[Control Structures And Microprogramming]** Control Structure Performance Analysis and Design Aids -Automatic synthesis, Formal models, Simulation.

General Terms
Performance, Design, Experimentation, Verification.

Keywords
Delta platform, Design, Dynamics, Delta Robot, Parallel robot,..

1. INTRODUCTION
Parallel robots are closed-loop mechanisms presenting very good performances in terms of accuracy, regidity and abality to manipulate large loads. Many applications in the field of production automation, such as assembly and material handling, require machines capable of very high speeds and accelerations. The parallel robots are able to work on some tasks with a much better performance. However, there are still several unanswered questions and few papers published studying robots with parallel architectures. This paper introduces a three d.o.f parallel manipulator dedicated to pick-and-place: Delta Parallel Robot. First a kinematics model of a Delta parallel robot is obtained using a generic geometrical formulation then the model is used for a workspace analysis. Delta robot has many advantages like operating required accurary, rigidity and manipulation of large loads. A Parallel Robot is a mechanism that has links that form closed kinematics chains. Because of this, Parallel mechanisms have many advantages compared to serial mechanisms, such as speed and accuracy. Generally, a parallel robot is made up of a mobile platform (end-effector) with *n* d.o.f, and a fixed base, linked together by at least two independent kinematics chains. Normally, each kinematics chain has a series of links connected by joints. Manipulators with *3* degrees prove extremely interesting for pick-and-place operations. Several prototypes have been suggested. The most famous robot with *3* d.o.f is Delta. All the kinematic chains of this robot are *3* rotary actuators allowing to obtain 3 dof in translation. This paper introduces a 3-dof parallel manipulator architecture Delta dedicated with kinematics and dynamics analyses to pick-and-place and developed to perform high speed and acceleration. In this article we have discussed the inverse and direct kinematics solution as well as dynamics for the Delta parallel robot. With this manipulator it is often difficult to determine the kinematics and dynamics analyses. Thus, this paper includes five seperated sections. The main properties of parallel robot is described in section II as well as focusing on kinematics and dynamics analyses, respectively. Experiment and discussions is established in Section IV. Finally, in Section V is shown the conclusion.

2. KINEMATIC AND JACOBIAN ANALYSES
In this section, the description and kinematics of the parallel robot – 3 dof are shown in **Fig.1**. Generally, parallel robot is a

Research Notes in Information Science (RNIS)
Volume13, May 2013
doi:10.4156/rnis.vol13.36

closed loop manipulator is more difficult to calculate the kinematics. The moving plate always stays parallel to the base platform and its orientation around the axis perpendicular to the base plate is constantly zero. Thus, the parallelogram type joints (forearm) can be substituted by simple rods without changing the robot kinematic behaviour. The revolute joints (between the base plate and the upper arms and between the forearms and the travelling plate) are identically placed on a circle. Thus, the travelling plate can be replaced by a point P which the three forearms are connected to.

The modelling of Delta robot has the assumptions like as: φ_1, φ_2, φ_3 are the rotate angle of 3 link, d_A is the distance from the center of the base (origin) to the spin axis of the transmission, F_1; F_2; F_3 are the center of the spindle attached to the transmission, r_A is the distance from the center stand on compared to the projection axis of the arm to stand on. And L_1, L_2 are the length of *2* link as describe in **Fig. 2**. Because, the inverse kinematics of Delta parallel robot is more easier than Direct Kinematics (Forward Kinematics), so firstly the inverse kinematics is shown. The inverse kinematics of a parallel manipulator determines the θ_i angle of each actuated revolute joint given the (x,y,z) position of the travel plate in base-frame.

$$\theta_1 = \arctan\left(\frac{z_{J_1}}{y_{F_1} - y_{J_1}}\right) \tag{1}$$

Such algebraic simplicity follows from good choice of reference frame: joint F_1J_1 moving in YZ plane only, so we can completely omit X coordinate. To take this advantage for the remaining angles θ_2 and θ_3, we should use the symmetry of delta robot. First, let's rotate coordinate system in XY plane around Z-axis through angle of 120^o counterclockwise.

We've got a new reference frame X'Y'Z', and it this frame we can find angle θ_2, θ_3 using the same algorithm that we used to find θ_1.

$$\begin{cases} x_0' = x.\cos\left(\pm 120^o\right) + y.\sin\left(\pm 120^o\right) \\ y_0' = -x.\sin\left(\pm 120^o\right) + y.\cos\left(\pm 120^o\right) \\ z_0' = z_0 \end{cases} \tag{2}$$

Now the three joint angles θ_1, θ_2 and θ_3 are given, and we need to find the coordinates (x_0, y_0, z_0) of end effector point E_0.

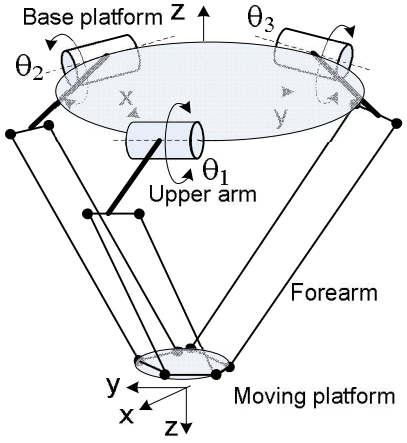


Fig.1. Modelling of Delta parallel robot.

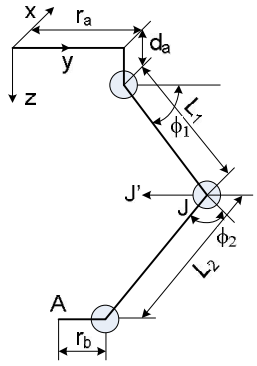


Fig.2. Shows model simplification of the Delta parallel robot.

Use of the vector translation of y-axis displacement, we have:

$$\overrightarrow{OJ_1'} = \overrightarrow{OF_1} + \overrightarrow{F_1J_1} + \overrightarrow{J_1J_1'} \tag{3}$$

With a length of the vector, the distance from the original quadrant to the swivel point of the transmission are:

$$OF_1 = OF_2 = OF_3 = \sqrt{r_A^2 + d_A^2} \tag{4}$$

Distance from center of the three spheres intersect at the center base

$$J_2J_2' = J_2J_2' = J_3J_3' = r_B \tag{5}$$

Radius of the sphere is L_2, so:

$$F_1J_1 = L_2\cos\varphi_1 \tag{6}$$

$$F_2J_2 = L_2\cos\varphi_2 \tag{7}$$

$$F_3J_3 = L_2\cos\varphi_3 \tag{8}$$

We have:

$$r = \sqrt{r_A^2 + d_A^2} - r_B \tag{9}$$

$$OJ_1' = OF_1 + F_1J_1 - J_1J_1' \tag{10}$$

And (x, y, z) is the coordinates of sphere centers J_1', J_2', J_3'.

So the coordinate of J_1' is:

$$\begin{bmatrix} 0 & r + L_2\cos\varphi_1 & L_2\sin\varphi_1 + d_A \end{bmatrix}^T = \begin{bmatrix} x_1 & y_1 & z_1 \end{bmatrix} \tag{11}$$

Similarly we have the coordinates of J_2' and J_3' as follows:

$$J_2' = \left(x_2; y_2; z_2\right) = \left((r + L_2\cos\varphi_2)\cos 30^0; (r + L_2\cos\varphi_2)\sin 30^0; L_2\sin\varphi_2 + d_A\right) \tag{12}$$

$$J_3' = \left(x_3; y_3; z_3\right) = \left(-(r + L_2\cos\varphi_3)\cos 30^0; (r + L_2\cos\varphi_3)\sin 30^0; L_2\sin\varphi_3 + d_A\right) \tag{13}$$

So the intersection of 3 sphere here:

$$\begin{cases}(x-x_1)^2+(y-y_1)^2+(z-z_1)^2=L_1^2\\(x-x_2)^2+(y-y_2)^2+(z-z_2)^2=L_1^2\\(x-x_3)^2+(y-y_3)^2+(z-z_3)^2=L_1^2\end{cases} \quad (14)$$

And, we have solutions like as:

$$x_0=\frac{a_1z_0+b_1}{d} \quad (13)$$

$$y_0=\frac{a_2z_0+b_2}{d} \quad (14)$$

$$z_0=\frac{-b\pm\sqrt{\Delta}}{2a} \quad (15)$$

With help from computer this equation system can be solved. There will be two solutions that describe the two intersection points of the three spheres. Then the solution that is within the robots working area must be chosen. With the base frame {R} in this case it will lead to the solution with negative z coordinate.

3. DYNAMIC ANALYSIS OF DELTA ROBOT

One important step in design process of a robot is to understand the behaviour of the device as it moves around its workspace or doing a specific task. This behaviour is determined through the study of the dynamics of the mechanism, where the forces acting on the elements and torques required by the actuators can be determined. Consequently, each component must be optimized in dimensions and material to be used in the manufacturing processes. In section, the dynamics of Delta parallel robot is described based on Lagrangian formulation, which is based on calculus variations, states that a dynamic system can be express in terms of its kinetic and potential energy leading in an easy way the solution to the problem. In addition, it is considered a good option to be used for real-time control for parallel manipulators [4]. The Lagrange equations can be derived.

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right)-\frac{\partial L}{\partial q_j}=Q_j+\sum_{i=1}^{k}\lambda\frac{\partial g_i}{\partial q_k} \quad (16)$$

Where L is the Lagrange function, where L = T - V, T is the total kinetic energy of the body, V is the total potential energy of the body, q is the k_{th} generalized coordinate, Q is a generalized external force, λ_i is the Lagrange multiplier and g_i is the constrain equation. By employing the formula above it is possible to determine the external forces of a body. However, friction forces are not constraints even though they play an important role in the dynamics analysis so they can be treated separately.

The Lagrange multipliers are derived as.

$$2\sum_{i=1}^{3}\lambda_i(p_x+h\cos\phi_i-r\cos\phi_i-a\cos\phi_i\cos\theta_{1i})=(m_p+3m_b)\ddot{p}_x \quad (17)$$

$$2\sum_{i=1}^{3}\lambda_i(p_y+h\sin\phi_i-r\sin\phi_i-a\sin\phi_i\cos\theta_{1i})=(m_p+3m_b)\ddot{p}_y \quad (18)$$

$$2\sum_{i=1}^{3}\lambda_i(p_z-a\sin\theta_{1i})=(m_p+3m_b)\ddot{p}_z+(m_p+3m_b)g_c \quad (19)$$

When the Lagrangian multiplies are found the actuator torque can be determined as.

$$\tau_1=\left(\frac{1}{3}m_aa^2+m_ba^2\right)\theta_{11}+\left(\frac{1}{2}m_a+m_b\right)g_ca\cos\theta_{11} \quad (20)$$
$$-2a\lambda_1\left[\left(p_x\cos\phi_1+p_y\sin\phi_1+h-r\right)\sin\theta_{11}-p_z\cos\theta_{11}\right]$$

$$\tau_2=\left(\frac{1}{3}m_aa^2+m_ba^2\right)\theta_{12}+\left(\frac{1}{2}m_a+m_b\right)g_ca\cos\theta_{12} \quad (21)$$
$$-2a\lambda_2\left[\left(p_x\cos\phi_2+p_y\sin\phi_2+h-r\right)\sin\theta_{12}-p_z\cos\theta_{12}\right]$$

$$\tau_3=\left(\frac{1}{3}m_aa^2+m_ba^2\right)\theta_{13}+\left(\frac{1}{2}m_a+m_b\right)g_ca\cos\theta_{13} \quad (22)$$
$$-2a\lambda_3\left[\left(p_x\cos\phi_3+p_y\sin\phi_3+h-r\right)\sin\theta_{13}-p_z\cos\theta_{13}\right]$$

The analytical inverse dynamics solutions for Delta parallel robot can be obtained from **Eqs**.(20-22)

4. EXPERIMENTS AND DISCUSSIONS

To valid the analyses of kinematics and dynamics in previous section, an experimental setup was built to perform the control of Delta parallel robot (**Fig.3**). The specifications of Delta parallel robot is shown in **Table 1**.

Table 1. Specifications of Delta parallel robot

Parameters		**Value**
Upper robot arm m_a [kg]		1.1
Parallelogram m_b [kg]		0.9
Moving platform m_p [kg]		0.2
Radius of the fixed base a [mm]		150
Radius of the moving platform b [mm]		100
Upper arm length l_1 [mm]		250
Parallelogram length l_2 [mm]		480
No. of AC Servo motor		3
Motor power [W]		200
Encoder resolution [ppr]		1000
Maximum load capacity [kg]		5
Maximum moving platform velocity [m/s]		5.0
Position repeatibility [mm]		0.2
Workspace	Diameter [mm]	500
	Height [mm]	200

This experimental implementation is built on the PC and Delta robot. The software for Delta parallel robot is implemented in Matlab using the kinematics and dynamics analyses from above solutions to control the moving platform. The proposed analyses are applied the Delta parallel robot for material cutting and drawing. The program is used to control the moving platform with predefined trajectory. We will apply the kinematics, Jacobi and dynamics to control suitable trajectory of parallel robot based on positions, velocities. In the section, some experimental results by kinematics - dynamics control are addressed. To demonstrate the capability controller, several responses were taken into account with several various trajectories. In these experiments, a pen attached to moving platform of Delta parallel robot is regulated following the several predefined paths including curves of circle, butterfly, flower, heart.

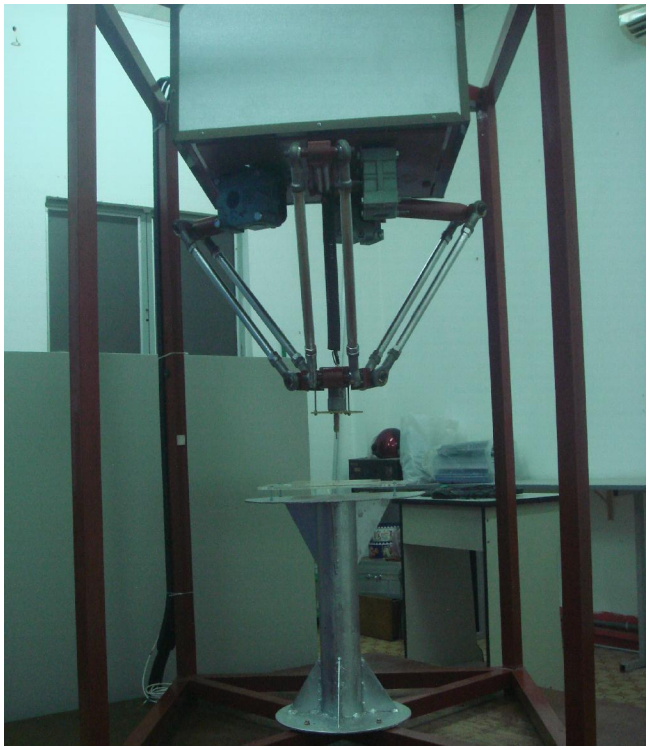


Fig.3. Delta parallel robot for experiments.

The first experimental results for controlling the moving platform with contour of flower are illustrated in **Fig.4**. A curve has the shape of a petalled flower and the polar equation of the rose is follows.

$$r = a\sin(n\theta) \tag{23}$$

The drawing on paper or cutting on acrylic reveals that the analysis results are almost near the desired ones shown in **Fig.4(a)**. Compared desired contour, we can see that the very small differences between the desired and experimental values may be attributed to the following reasons: first, there is error of mechanical transmission and calculation of kinematics and dynamics of Delta robot. The improvements will bring better results for generating trajectories. And responding of three AC servo motors with time is shown in **Fig.4(b)** .

Next, other responses for reference commands for butterfly contour are presented to evaluate the performance of the controller based kinematics and dynamics. The equation of butterfly curve is follows.

$$r = e^{\sin\theta} - 2\cos(4\theta) + \sin^5\left[\frac{(2\theta-\pi)}{24}\right] \tag{24}$$

Fig.5 shows output of responding trajectory and input responses for contour of butterfly. The control results for butterfly are good enough to track the perfect shape while moving path of pen has a little bit error.

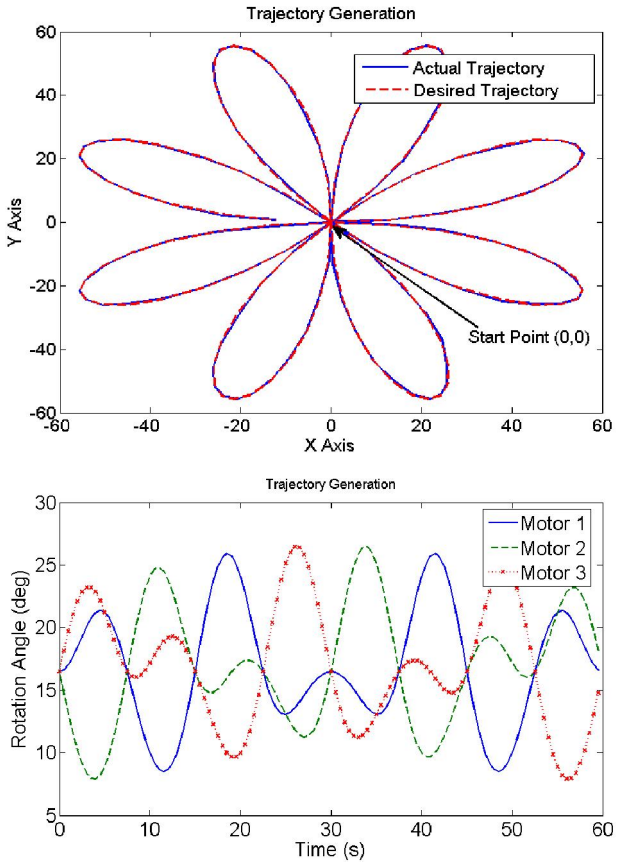


Fig.4. Trajectory of moving platform (a) and responding of 3 motors(b) with flower curve path.

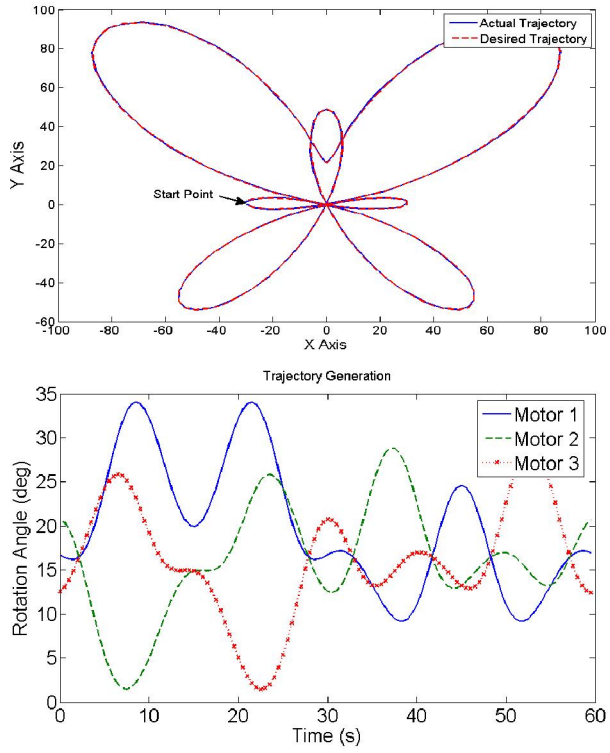


Fig.5. Trajectory of moving platform (a) and responding of 3 motors (b) with butterfly curve path.

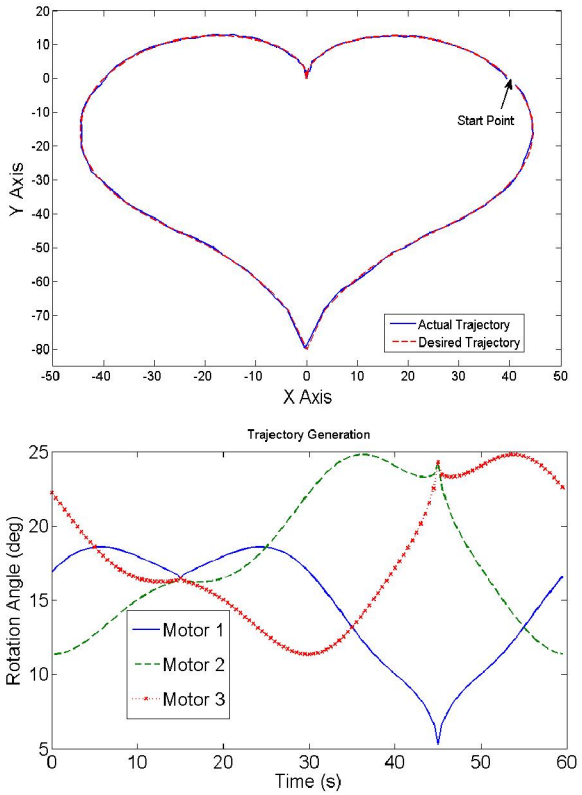


Fig.6. *Trajectory of moving platform (a) and responding of 3 motors(b) with heart curve path.*

Besides, we also generate the trajectory of heart curve with pole equation like as:

$$r = 2 - 2\sin\varphi + \frac{\sin\varphi\sqrt{|\cos\varphi|}}{\sin\varphi + \frac{7}{5}} \tag{25}$$

Fig.6(a) shows the actual time response signals and the command signals of parallel to a heart profile, and the time history of the controlled position output.

The movement of moving platform followed the commanded signals quite well for long time. Present results show that the analyses of kinematics and dynamics can be successfully applied to the dynamic tracking of various contour profiles.

5. CONCLUSION

This paper is mainly concerned with kinematic and dynamic analyses as well as the application of solutions of kinematics and dynamics to modeling and control of parallel manipulators. A practical implementation is completed to evaluate the results of an designed controller for Delta manipulator control system. It can be said that, excepted results has been achieved for these cases. The inverse and forward kinematics and velocity equations have been derived. The results presented in the paper will be valuable for both the design and development of Delta parallel robot for various applications. With the aid of computer, these equations with the design of this robot base on dynamic modeling and dynamic control in order to improve the behavior of the robot while reaching high acceleration. By fitting grippers or other tools to this small platform the delta robot can handle all sorts of items. Their design enables them to move both rapidly and accurately, and they are deployed for tasks varying from high-speed packaging to the assembly of miniature products.

6. ACKNOWLEDGMENTS

This study was financially supported Ho Chi Minh city University of Technical Education, Viet Nam (HCMUTE).

7. REFERENCES

[1] André Olsson, Modeling and control of a Delta-3 robot, 2009.

[2] Jon Martínez García, Inverse-Forward Kinematics of a Delta Robot, 2010.

[3] Manuel Napole and Cardona Gutierrez, *Kinematics Analysis of a Delta Parallel Robot*, 2011.

[4] S.M.Ha, P.V.B. Ngoc and H.S.Kim, "Dynamics Analysis of a Delta-type Parallel Robot," 2011 11th International Conference on Control, Automation and System, 2012.

[5] S.M.Ha, P.V.B.Ngoc and H.S.Kim, "Dynamics Analysis of a Delta-type Parallel Robot", 2011 11th International Conference on Control, Automation and Systems, pp.855-857, 2011.

Low Power Dissipation Model Analysis for Embedded Systems

Yang-Hsin Fan
Dept. of Computer Science and Information Engineering
National Taitung University
684. Sec. 1, Chunghua Rd., Taitung City, 95002, Taiwan
+886-89-350410
yhfan@nttu.edu.tw

Jan-Ou Wu
Dept. of Electronic Engineering
De Lin Institute of Technology
1. Ln. 380, Qingyun Rd., Tucheng Dist, New Taipei City 23654, Taiwan
+886-2-22733567#112
janou@ms42.hinet.net

San-Fu Wang
Dept. of Electronic Engineering
Ming Chi University of Technology
84 Gungjuan Rd., Taishan Dist. New Taipei City 24301, Taiwan
+886-2-29089899#4867
sf_wang@mail.mcut.edu.tw

ABSTRACT
Embedded systems serve diverse functionalities to meet the requirement for computer, communication equipment, consumer electronics and car (4C) products. It results the need of embedded systems to exponential growth. While hundreds of thousand embedded systems run on every corner of daily life, the consumed power consumption is extremely huge. As a result, embedded systems consumed low power consumption become a significantly issue. This study presents model analysis approach to obtain low power consumption for embedded systems. First, embedded systems are divided into a set of tasks that are implemented with hardware circuits and software applications. Second, various models with same tasks combination and height of tree of embedded system are analyzed for power dissipation. Next, dynamic and static power consumption for each task is measured for further calculating to gain an embedded system with low power dissipation. Four, we present a schema via formula for various models to fast assessing the consumed power consumption for embedded systems. Finally, the effectiveness of the proposed approach is demonstrated by assessing an adaptive pulse code modulation (ADPCM) system.

Categories and Subject Descriptors
C.3 [**Computer Systems Organization**]: Special-Purpose and Application-Based Systems – *real-time and embedded systems, microprocessor/microcomputer applications.*

General Terms
Theory, Experimentation.

Keywords
Low power consumption, Power saving, Embedded system.

1. INTRODUCTION
Recently, low power consumption of embedded systems become significant issue owing that the energy of earth is gradually consumed. The worst affected products include computer, communications equipment, consumer electronics and car, *etc.* In the energy shortage era, the Intel Company predicts that 15 billion embedded products will surf to internet in year 2015. Once those products simultaneously serve, energy dissipation must be rapidly raised even used up if power saving or energy efficient improvement is not achieved.

Executing task inside embedded systems consume power consumption. The states of task categories power dissipation into dynamic and static power consumption. Dynamic power consumption happens while task is performed. On the other hand, task consumes static power dissipation when its states being idle. For embedded system insides *n* tasks, executing task 1 results that consume dynamic power consumption and the other tasks arise static power consumption. In consideration task 2 runs, it consumes dynamic power consumption and task 1, task 3, task 4 to task *n* occur static power consumption. To iterate the process for every task execution, the power consumption of embedded system can be assessed. From these evaluating designs, the lowest power consumption of embedded system can be determined.

2. PRELIMINARY WORK
Smarter, smaller and portable characteristics make embedded systems to serve the functions becoming diversity. The products reside embedded systems that spread over computer, communication, consumer and car (4C). However, the more embedded systems serve, the more power dissipation consumed. For minimizing the CPU power consumption for real time embedded system, Vîlcu [1] first studies task execution in the power consumption of processor(s). Then, he finds the affection of optimal configuration processor(s) for energy consumption. Finally, he defines globally optimal scheduling which gains minimal energy consumption for homogeneous multiprocessor system. Silva-Filho and Lima [2] state memory hierarchy consumes power up to 50% in microprocessor system. Consequently, they propose an automated architecture exploration mechanism to NIOS II processor and memory hierarch with parameter variation. Experimental results show the reduction of energy consumption near to 27%. In 2008, Zeng *et al*. [3] present generalized *dynamic energy performance scaling* (DEPS) framework to hard real-time embedded systems for exploring application-specific energy-saving potential issue. Three energy performance tradeoff technologies called DHRC, DVFS and DPM are integrated into DEPS. Experiment results of simulation show the static DEPS improves 13.6% and 13.7% in DVFS and DHRC, respectively. Also, dynamic DEPS improves 5.7% than static DEPS.

Qiu *et al*. [4] discuss the execution time of tasks with conditional instructions or operations problem. They adopt probabilistic random variable approach to model execution time of tasks. Then,

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.37

they propose practical algorithm VACP to minimize energy consumption for uniprocessor embedded systems. Gao *et al*. [5] present *energy-efficient architecture for embedded software* (EAES) and dynamic energy-saving method to solve energy-saving problem. The former uses a processor with dynamic voltage scaling capability, FPGA modules and extends directed acyclic graph to embedded system. The latter adopts pre-assignment to achieve dynamic runtime scheduling and minimizing energy consumption.

Real-time power information is a valuable data for software designer for battery-powered embedded systems. Genser *et al*. [6] propose power profiling approach to collect real-time power information at early design stages. Moreover, they present an emulation-based power profiling approach to achieve real-time power analysis for embedded systems. Because of the power information is collected at early design stages, the development efficiency and time-to-market is improved. In 2008, Elewi *et al*. [7] first discuss the real time scheduling of dependent tasks problem and then present enhanced *multi-speed* (MS) algorithm for energy saving. With energy consumption problem of battery-powered embedded systems, Casares *et al*. [8] aim embedded smart camera to analyze the power consumption and performance. Not only graph of energy consumption but also instruction of collections is presented. They conclude the important of lightweight algorithm, the time of transfer data and transferred data type.

3. LOW POWER MODEL ANALYSIS

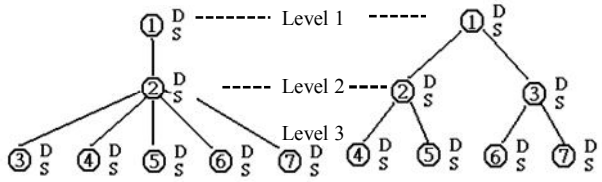
Modern embedded system executes sequentially a set of tasks to provide multimedia, social network and diverse applications. Performing these tasks consume a lot of power energy that depend on the deploying architecture of embedded system. In order to design embedded system with low power dissipation or power saving, minimizing power dissipation for every task is one intuitive approach.

Each task consume individually the power dissipation that affect the design becomes low power dissipation products or power saving equipment. Consequently, the candidate of tasks combination for embedded system becomes a significant issue. According to the hardware-software codesign theory, the candidate of tasks is either hardware circuits or software applications. As a result, the degree of power consumption that depends on the power dissipation with designed tasks with implementation via hardware circuits or software applications.

Tree topology is generally used to analysis the power dissipation for embedded systems. It consists of node, arc, control and data flow, height and level. Figure 1 shows two tree topologies called *M* and *N* of embedded system with 7 tasks. The embedded system executes task from node 1. Then, it departs for next nodes that depend on the arcs after run a period of time. The arc connect sink and destination node. It is used to exhibit the control and data flow. Another terminology in tree topology called height, *H*, of tree that is used to exhibit the architecture for embedded systems. The other terminology named level, *L*, is used to indicate the control or data flow in specific time for embedded systems. We apply tree topology to model embedded system and then further to analysis their power consumption.

From the appearance of view, Figure 1(a) and (b) are two kinds of design of topology for embedded system. Both of them are consist of 7 nodes, some arcs, control and data flow and levels. In particular, it is as same as *H* in two tree topology. Besides, level 1 to 3 happen the control or data flow in time t_1, t_2 and t_3.

Embedded systems consume dynamic or static power consumption that depends on the status for task. Dynamic power consumption *D* occurs while task is executed. On the contrary, the



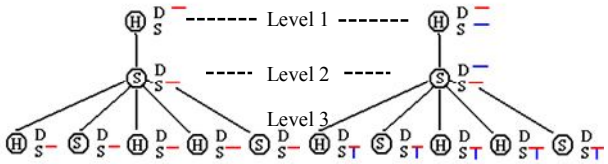
(a) *M* topology (b) *N* topology

Figure 1. Two tree topologies of embedded system with 7 tasks.

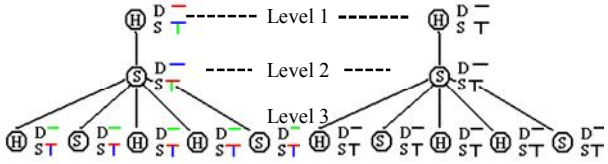
idle status of task consumes static power consumption *S*. For example, task 1 of Figure 1(a) in time t_1 is working for a particular function resulting it dissipates *D*. At the same time, tasks 2 to 7 dissipate *S* since they are idle in t_1. Similarly, task 2 consumes *D* in time t_2 and task 1 and tasks 3 to 7 dissipate static power consumption. On time t_3, task 3 to 7 consumes *D* and task 1 and 2 dissipate *S*. Based on the stated rules, any topology of embedded systems can be analysis for their total power consumption for assessment.

Figure 2 demonstrates *M* topology with 7 tasks, height and level as 3. Labels *D* and *S* beside tasks indicate dynamic and static power consumption. We analyze power consumption via the order of level in the following. Figure 2(a) shows the power consumption model analysis for level 1. Label *D* besides task *H* in level 1 is marked and the other tasks on level 2 and 3 are marked in symbol *S*. Such results imply dynamic and static power consumption happen simultaneously upon whole embedded system. Figure 2(b) displays the analysis for power consumption in level 2. Label *D* next to task *S* is marked and the other tasks are marked with symbol *S*. Similarly, the power dissipation in level 3 is discussed in Figure 2(c). Five tasks named *H*, *S*, *H*, *H* and *S* in level 3 next to label *D* are marked and tasks located on level 1 and 2 are marked with symbol *S*. Figure 2(d) exhibits the total power consumption for *M* topology.

Another model referred to topology shown in Figure 3 is introduced in the following. It characteristics include 7 tasks, the height and level as 3, the number of tasks with *H* and *S* is as same as Figure 2. In other words, both topology *M* and *N* have four *H* and three *S*. The power consumption model analysis adopts stated steps and approach. First, we analyse the power consumption for Figure 3(a) for level l. We label *D* next to task *H* in level 1 and the

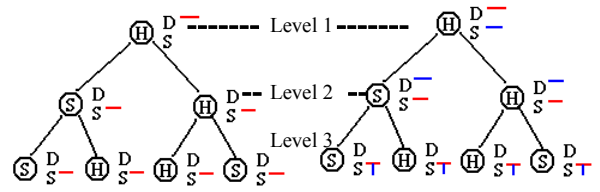


(a) Level 1 power consumption (b) Level 2 power consumption

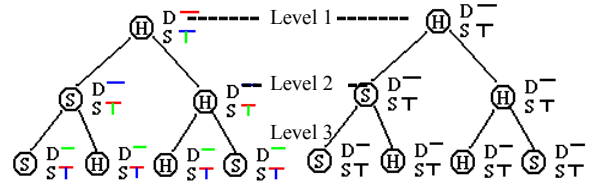


(c) Level 3 power consumption (d) Total power consumption

Figure 2. Power consumption model analysis for M topology.



(a) Level 1 power consumption (b) Level 2 power consumption



(c) Level 3 power consumption (d) Total power consumption

Figure 3. Power consumption model analysis for N topology.

other tasks on level 2 and level 3 are marks *S*. Figure 3(b) presents the power dissipation for level 2. Symbol *D* next two tasks named *S* and *H* in level 2 are marked and the other tasks in level 1 and level 3 are marked with label *S*. Figure 3(c) aims level 3 to discuss. Four tasks, *S*, *H*, *H* and *S*, are marked with *D* in level 3 and tasks locate in level 1 and 2 are marks with *S*. After every level of embedded system is analyzed, the total power consumption for topology *N* is shown in Figure 3(d).

According to Figure 2(d) and Figure 3(d), we conclude topology *M* dissipates the same power consumption as topology *N*. *i.e.*, both topology consume total power consumption as one dynamic power consumption, *D*, and two static power consumption, *S* for each node. Such conclusion can be applied to other topology of embedded systems. As a result, we summary the sum of power consumption for embedded system in *Equation* (1). It consists of the levels of tree, *L*, a set of static power consumption $P_{s,t1}$, $P_{s,t2}$,…, $P_{s,tn}$ and dynamic power consumption $P_{d,t1}$, $P_{d,t2}$,…, $P_{d,tn}$ for each node.

$$P = (L-1) \times \left(P_{s,t_1} + P_{s,t_2} + \cdots + P_{s,t_n}\right) + \left(P_{d,t_1} + P_{d,t_2} + \cdots + P_{d,t_n}\right) \quad (1)$$

where

L is the height of tree,

P_s is static power consumption,

P_d is dynamic power consumption,

t_1, t_2,…, t_n is a set of tasks

4. EXPERIMENTAL RESULTS

This study is evaluated by *adaptive pulse code modulation* (ADPCM) system which is comprised of encoder and decoder modules. The encoder module consists of four tasks namely T_a, T_b, T_c and T_d. On the other hand, two tasks called T_e and T_f are designed for decoder module. Each task is implemented separately to hardware and software form by using Verilog and C language. The measured data of each task includes dynamic and static power consumption with implementation of hardware circuits and software applications. Table 1 shows the measured data for T_a to T_f.

Table 1. Measured data of tasks for ADPCM system.

Application		Power Consumption			
ADPCM system	Tasks	H/W (mW)		S/W (mW)	
		Dynamic	Static	Dynamic	Static
Encoder	T_a	14.58488	2.76628	984.1802	984.18
	T_b	2.91698	0.55326	984.1809	984.18
	T_c	13.61256	2.58186	984.1817	984.18
	T_d	10.69558	2.02860	984.1816	984.18
Decoder	T_e	23.22750	5.40143	984.1819	984.18
	T_f	11.00250	2.55857	984.1814	984.18

The experiments are set to five scenarios which depend on the number of tasks with implementation by hardware circuits and software applications. These five scenarios are different to the number of hardware tasks and software tasks. The first scenario is that embedded system implement with five hardware tasks (*i.e.* 11111) and one software task (*i.e.* 0) and level as 6. Figure 4 illustrates the power dissipation of ADPCM with six designs with five hardware tasks and one software task. There are six kinds of designs that are 011111, 101111, 110111, 111011, 111101 and 111110. Experimental results display the implementation with 111101 which gains low power consumption as 6.01mW. On the contrary, the design of 101111 consumes the most power consumption as 6.055mW.

The second scenario is that embedded system implement with four hardware tasks (*i.e.* 1111) and two software tasks (*i.e.* 00) and level as 6. Figure 5 displays these designs with four hardware tasks, two software tasks and level as 6. It has 15 types of combinations while develop an embedded system. The lowest power dissipation is 11.887mW which are developed via a set of tasks with 011101. On the other hand, the most power consumption of design as 101011 that consume 11.939mW.

The third scenario is that embedded system implement with three hardware tasks (*i.e.* 111) and three software tasks (*i.e.* 000) and level as 6. Figure 6 demonstrates these embedded systems that comprises of three hardware and software tasks and level as 6. There are totally 20 types to complete the designs. Experimental results indicate the lowest power consumption as 17.7656mW which tasks are made of 010101. In contrast, the most power dissipation as 101010 design that consumes 17.8204mW.

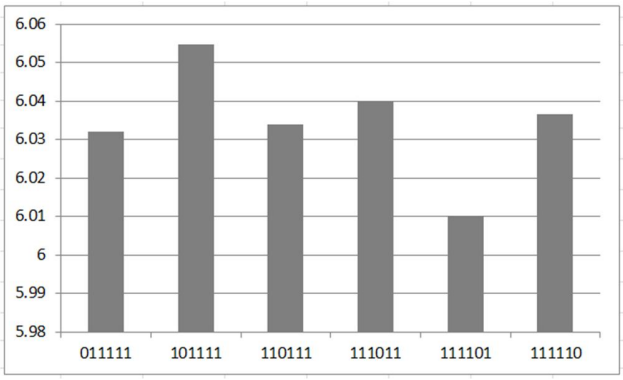


Figure 4. Power dissipation of ADPCM embedded system with five hardware tasks and one software task.

The fourth scenario is that embedded system implement with two hardware tasks (*i.e.* 11) and four software tasks (*i.e.* 0000) and level as 6. Figure 7 illustrates the 15 results that consist of two hardware circuits and four software applications. Among 15 designs, the embedded system with 010100 performance gains low power consumption as 23.6468mW. On the contrary, the design of 100010 consumes the most power consumption as 23.699mW.

Finally, the fifth scenario is that embedded system implement with one hardware task (*i.e.* 1) and five software tasks (*i.e.* 00000) and level as 6. Figure 8 exhibits the outcomes which comprise of one hardware circuit and five software applications. The lowest power consumption is made up of 010000 design. It consumes power dissipation as 29.5311mW. In contrast, the design with 000010 consumes the power dissipation as 29.5756mW.

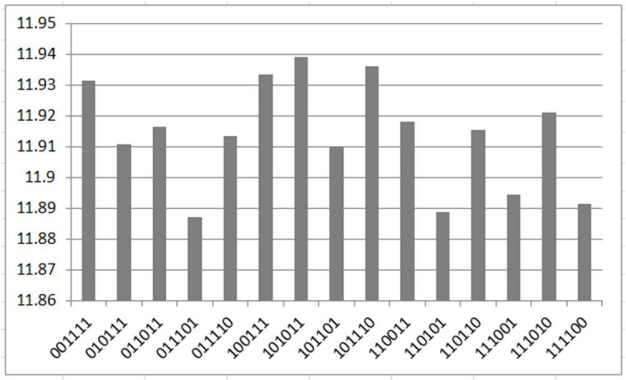


Figure 5. Power dissipation of ADPCM embedded system with four hardware tasks and two software tasks.

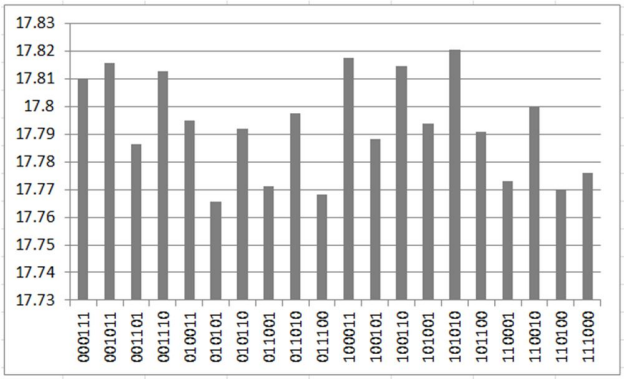


Figure 6. Power dissipation of ADPCM embedded system with three hardware and software tasks.

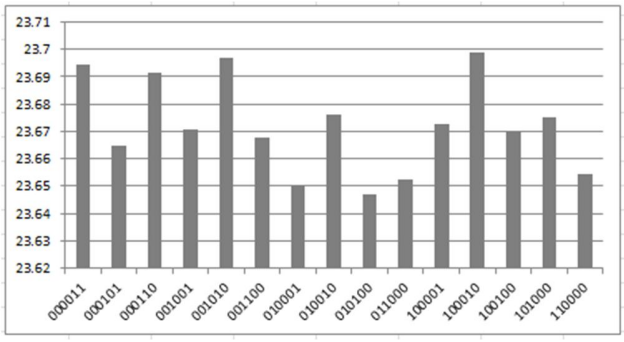


Figure 7. Power dissipation of ADPCM embedded system with two hardware tasks and four software tasks.

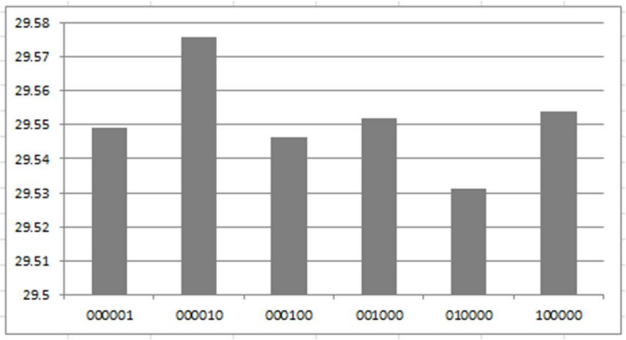


Figure 8. Power dissipation of ADPCM embedded system with one hardware task and five software tasks.

5. CONCLUSIONS

This study aims for power consumption of embedded systems to achieve the low power dissipation design. We categorize tasks into hardware or software form with corresponding to implementation by hardware circuits or software applications. Each task consumes dynamic and static power consumption that is taken into account. Then, we present various models with same tasks combination and height of tree of embedded system to gain low power consumption. Based on the proposed approach of low power model analysis, embedded system with minimizing the power consumption can be determined among diverse designs. Experimental results of ADPCM prove the effectiveness for gaining low power consumption of embedded systems that apply low power model analysis approach.

6. ACKNOWLEDGMENTS

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC 101-2221-E-143-002.

7. REFERENCES

[1] D. Vilcu. 2008. Real time scheduling and CPU power consumption in embedded systems. In *Proceeding of the IEEE International Conference on Automation, Quality and Testing, Robotics* (Cluj-Napoca, Romania, May 22 – 25, 2008). IEEE, 261-266. DOI= 10.1109/AQTR.2008.4588746.

[2] Silva-Filho, A.G., and Lima, S.M.L. 2008. Energy consumption reduction mechanism by tuning cache configuration usign NIOS II processor. In *Proceeding of the IEEE International SOC Conference* (Newport Beach, CA, USA, September 17 – 20, 2008). IEEE, 291-294. DOI= 10.1109/SOCC.2008.4641530.

[3] G. Zeng, H. Tomiyama, H. Takada, and T. Ishihara. 2008. A generalized framework for system-wide energy savings in hard real-time embedded systems. In *Proceeding of the IEEE/IFIP International Conference on Embedded and Ubiquitous Computing* (Shanghai, China, December 17 – 20, 2008). IEEE, 206-213. DOI= 10.1109/EUC.2008.101.

[4] M. Qiu, J. Wu, F. Hu, S. Liu, and L. Wang. 2009. Voltage assignment for soft real-time embedded systems with continuous probability distribution. In *Proceeding of the IEEE International Conference on Embedded and Real-Time Computing Systems and Applications* (Beijing, China, August 24 - 26, 2009). IEEE, 413-418. DOI= 10.1109/RTCSA.2009.50.

[5] Z. Gao, G. Dai, P. Liu, and P. Zhang. 2009. Energy-efficient architecture for embedded software with hard real-time

requirements in partial reconfigurable systems. In *Proceeding of the IEEE International Conference on Embedded Computing, Scalable Computing, and Communications* (Dalian, China, September 25 - 27, 2009). IEEE, 387-392. DOI= 10.1109/EmbeddedCom-ScalCom.2009.76.

[6] A. Genser, C. Bachmann, J. Haid, C. Steger, and R. Weiss. 2009. An emulation-based real-time power profiling unit for embedded software. In *Proceeding of the IEEE International Symposium on Systems, Architectures, Modeling, and Simulation* (Samos, Greece, July 20 – 23, 2009). IEEE, 67-73. DOI= 10.1109/ICSAMOS.2009.5289259.

[7] A.M. Elewi, M. Awadalla, and M.I. Eladawy. 2008. Energy-efficient multi-speed algorithm for scheduling dependent real-time tasks. In *Proceeding of the IEEE International Conference on Computer Engineering & Systems* (Cairo, Egypt, November 25 – 27, 2008). IEEE, 237-242. DOI= 10.1109/ICCES.2008.4772942.

[8] M. Casares, A. Pinto, Y. Wang, and S. Velipasalar. 2009. Power consumption and performance analysis of object tracking and event detection with wireless embedded smart cameras. In *Proceeding of the International Conference on Signal Processing and Communication Systems* (Omaha, Nebraska, September 28 – 30, 2009). IEEE, 1-8. DOI= 10.1109/ICSPCS.2009.5306326.

High Definition Network Camera Design Based on TMS320DM368

Miao Zang[1,2]
[2]Beijing University of Post & Telecom.
No.10 Xitucheng Road, Haidian District, Beijing, China
88803130, 010. 86
Zangm@ncut.edu.cn

Yongmei Zhang[1]
[1]North China University of Technology
No.5 Jinyuanzhuang Road, Shijingshan District, Beijing, China
88802212, 010. 86
zhang_yong_mei@sohu.com

Yuhua Wang[1]
88803130, 010. 86
wangyh@ncut.edu.cn

ABSTRACT
Existing network video surveillance sy stems generally use the standard definition analog camera, the definition and the information quantity are limited, and further intelligence analy sis is hardly impossible. This paper provides a design and implementation scheme of high definition digital network camera based on TI TMS320DM368 video processing chip. It can output the H.264 HD video stream with the m aximum definition of 1080P@30fps, and can transmit the compressed stream to the far end by IP network. This sche me is com patible with m any encoding formats, and it can support both ty pes of CMOS and CCD image sensors, which make th e system to be flexible and practical.

Categories and Subject Descriptors
C.3 [**Special Purpose and Application Based System**]: Real time and embedded system –*arm, dsp.*

General Terms
Design, Experimentation, Verification.

Keywords
Network Camera, High Definition, TMS320DM368

1. INTRODUCTION
With the development of co mputer network and digital multimedia technology, the requirem ent for the digital video information communication is raising rapidly. The network video applications such as video su rveillance, video conference and video telephone are used widely. Video camera, as the front end of these video application systems, is play ing a very important role in the whole s ystem. The existing network video sy stems generally use the front-end SD (standard definition) analog camera, the output analog video signals of which are encoded by digital video sy stem, and then the back-end video server stores and distributes the digital video signals. The structure of the whole system is complex and is not easy to maintain. In addition, the definition and the information quantity of the analog SD video signals are limited[1]. For example, in the PAL standard of SD video, D1 resolution is only 704*576, this can not satisfy the application in the big scene (suc h as campus and square) or the scene with high requirement for details (such as lift entrance and door entrance).

Combining the front-end camera and the encoding function in a single device can s implify the exis ting video system effectively. And this scheme can solve the problem that the HD (high definition) analog signal has hi gh requirement for transmission media. This scheme is the prototy pe of the HD network camera. This paper provides a design and implementation scheme of a HD network camera. The system can process the HD im age with the maximum resolution of 1080P. It is practical and flexible since the whole system cost also can be controlled.

2. WHOLE SYSTEM DESIGN
A complete network camera can be divided into four parts: video optical imaging unit, control processor, network transmission unit and storage unit. The whole sy stem block diagram [2] is shown in figure 1. In which, the encodi ng module is implemented by the control processor.

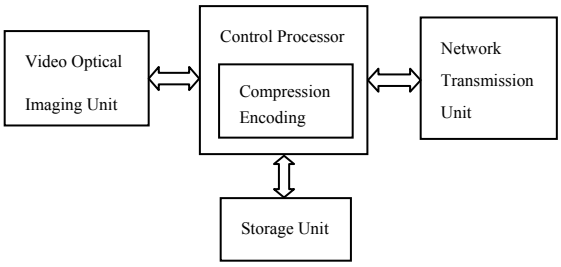


Figure 1. Network camera block diagrams.

As seen from figure 1, the optical im aging unit trans forms the visible lights of different bands into electrical signals with different levels by the wave length and light intensity, the electric signals are digitized to form the raw video data transm itted to the control processor; the video front-end of the control processor transforms the raw video data into video YUV data, and the encoding module compresses the YUV data into streams of H.264 or other compressed format, and then the s tream can be transmitted by network or be stored locally according to the system control. Users can s ee the HD video us ing the W EB

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.38

browser to configure the camera parameters, or can store, decode and play the received video file by the client application.

3. SYSTEM HARDWARE DESIGN

By investigation, the network camera solution based on TI TMS320DM36X series chips has distinct superiorities on system cost, image quality, and technique support, and so on. Therefore, this system chooses the TI network camera solution based on TMS320DM36X. The system hardware design block diagram [3] is shown in figure 2.

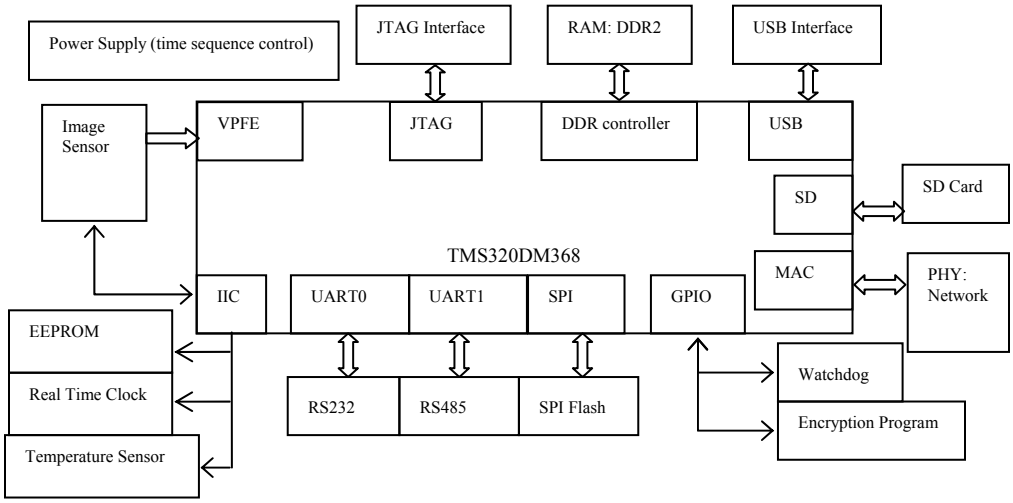


Figure 2. Network Camera Hardware Design Block Diagram based on TMS320DM368

As seen from figure 2, TMS320DM368 is the system control processor, it transmits the acquired raw video signals from Image Sensor to the VPFE (Video Preprocess Front End), and VPFE implements the preprocessing task, DM368 can compress the preprocessed video signals and transmit compressed stream to the far surveillance center through the MAC network interface.

3.1 Control processor

Control processor is the core of the whole system. It is mainly used to implement the function of image encoding compression and control processing. This system uses TI video processing chip TMS320DM368 as the control processor.

The DM368 is capable of achieving HD video processing at 1080p 30fps H.264, using the ARM926EJ-S core running at 432 MHz. It supports production-qualified H.264BP/MP/HP, MPEG-4, MPEG-2, MJPEG and VC1/WMV9 codecs providing customers with the flexibility to select the right video codec for their application. These codecs run on independent coprocessors (HDVICP and MJCP), supporting multi-channel, multi-stream and multi-format design. In addition, DM368 provides a seamless interface to most additional external devices required for video applications. The image sensor interface is flexible enough to support CCD, CMOS, and various other interfaces such as BT.656, BT1120.[3]

3.2 Video optical imaging unit

The main function of the video optical imaging unit is to transform the light signal into the electrical signal. Image sensor receives light irradiation during sensitization, transforming the different luminous flux into corresponding electrical signals. The electrical signals form the image matrix, which is transmitted to the processor.

The image sensor is the core of the video optical imaging unit.

Table 1. Image Sensor Parameters

Parameters	KAI-02150 CCD	MT9P031 MOS
Optical format	2/3 inch	1/2.5 inch
Number of effective pixels	1960[H]*1120[V]	2592[H]*1944[V]
Shutter type	Global shutter	Electronic rolling shutter (ERS)
ADC resolution	Decides on the back-end AD	12 bit, on-chip
Pixel size	5.5μm[H]*5.5μm[V]	2.2μm[H]*2.2μm[V]
Aspect ratio	16:9	4:3
Pixel dynamic range	64dB	70.1dB
Maximum frame rate	Quad Output (Full resolution)：64 fps Dual Output：33 fps Single Output：17 fps	Full resolution：14 fps VGA：53fps

The existing image sensors generally are both types of CCD and CMOS. In this paper, the performance parameters of both types of image sensors are analyzed and compared using Kodak KAI-02150 [4] CCD chip and Aptina MT9P031 [5] CMOS chip as examples. The comparisons are shown in table 1.

As seen from table 1, both type of image sensors support 1920*1080 HD images, and have large dynamic ranges. By

comparison, the CCD sensor has a bigger size of camera lens, and the pixel size is bigger, thus the acquired images are clearer. CCD sensor supports multi-channel output, and has a higher frame rate with a full resolution. Because of the differences of the imaging principle and production craft, CCD image sensor has a higher sensitivity, which is suitable for applications at night; and CMOS sensor has a higher integration, a smaller size, and a higher cost performance. In order to be convenient for users to choose, this system design scheme supports both types of image sensors.

As to implementation, this system uses the Aptina MT9P031 CMOS digital image sensor. It is a 1/2.5-inch CMOS active pixel digital image sensor with an active imaging pixel array of 2592*1944. It incorporates sophisticated camera functions on-chip such as windowing, column and row skip mode, and snapshot mode. It is programmable through a simple two-wire serial interface.[5]

The connection diagram between MT9P031 and DM638 is shown in figure 3. In order to receive video data from CMOS image sensor, the video port of DM368 must be configured to be raw data mode. The data acquisition rate is controlled by the pin PIXCLK of CMOS sensor. Thus, the system can control the frame rate of the video signals output from the image sensor by controlling the clock signal. DM368 can configure the MT9P031 registers by I2C bus (SCL and SDA), controlling the working mode of the image sensor. The MT9P031 image data is read out in a progressive scan. Pixels are output in a Bayer pattern format consisting of four "colors"—GreenR, GreenB, Red, and Blue (Gr, Gb, R, B)—representing three filter colors. [5] The ISIF(Image Sensor Interface) of DM368 is responsible for accepting raw image/video data from a sensor. It support for conventional Bayer pattern. [3] This ensures the seamless connection between MT9P031 and DM638. FV is the frame sync signal and LV is the line sync signal.

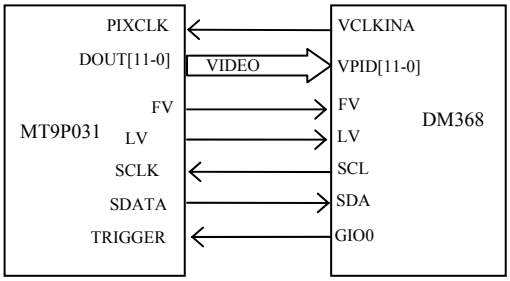


Figure 3. Connection between MT9P031 and DM368

3.3 Network transmission unit

Network transmission unit is the signal transceiver of the whole system. It consists of EMAC (Ethernet Media Access Control) module and MDIO (Management Data Input/Output) module. This system implements the network transmission using DAVICOM chip DM9161B [6]. It is a physical layer, single-chip, and low power transceiver for 100BASE-TX and 10BASE-T operations. On the media side, Through the Media Independent Interface (MII), the DM9161B connects to the Medium Access Control (MAC) layer, ensuring a high inter operability.

The interface circuit communicating to DM368 is shown in figure 4. The EMAC controls the flow of packet data from DM368 to the PHY. The MDIO module controls PHY configuration and status monitoring by two-line interface. Both the EM AC and the MDIO modules interface to DM368 through an EMAC control module that allows efficient data transmission and reception. The control module is also used to multiplex and control interrupts. [3]

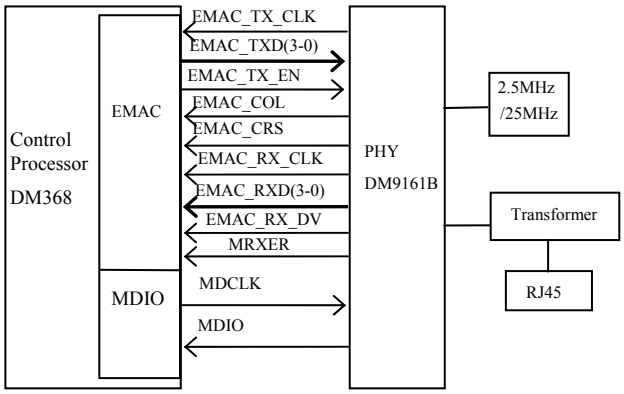


Figure 4. Connection between DM368 and DM9161B.

3.4 Other external devices

The functions of other external devices are introduced as follows:

Power supply: In order to ensure device reliability , TMS320DM368 requires restricted power supply power-on and power-off sequences. For example, the following steps should be followed for the restricted power-on method: a) Power on the Main core (1.2V); b) Power on the Main I/O (1.8V); c) Power on the Main/Analog I/O (3.3V).

JTAG: simulator debugging interface.

RAM: implemented by the DM368 DDR controller, with 2Gbit DDR2 RAM by two MT47H128M8HQ chips.

External storage interface: USB interface, SD card.

Program storage: Implemented by connecting to SPI Flash using SPI interface.

I2C device: EEPROM store the system parameters, and the real time clock can ensure the system acquires the right time after power off. The temperature sensor is used to check the device working environment.

UART: Two UARTs are used respectively to implement the RS232 and RS485 interfaces.

GPIO: implement the watching dog, and ensure the system to restart under the abnormal cases, and the program encryption chip is used to protect the intellectual property.

4. SYSTEM SOFTWARE DESIGN

4.1 System Software Environment

This system uses Linux as the operating system of high definition network camera. Linux system can be used widely in embedded system because it is open, safe, pollable and flexible.The embedded system cross developing environment is as shown in Figure 5.

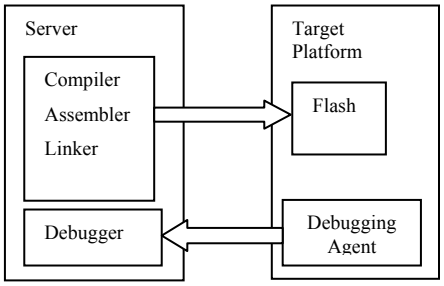


Figure5. Embedded system cross developing environmnet

The cross developing environment of embedded software consists of server and target platform. Application is edited and cross compiled on server, and then downloaded to the target platform to run. Debugging the application is mainly implemented by debugging tool such as the serial port and network of target platform. The cros s compiling tool us ed in this system is mvl_5_0_0_demo_lsp_setuplinux_02_10_00_14.bin。

4.2 BootLoader

BootLoader is a section of program run before the operating system kernal start. It is implemented by hardware in TMS320DM368. After power on, TMS320DM368 runs the Bootloader， loading UserBootloader, that is, UBOOT. This system uses BOOT 1.3.4. It is configed according to the target hardware configuration, and th en compiled and burned into SPI Flash by simulator. Again afte r power on again, UBOOT starts from SPI, starting Kernel and file system from SPI Flash or by NFS.

4.3 Embedded Linux System

This system uses embedded Li nux system Redhat Enterprise Linux 6.0 as the sy stem software platform. Embedded Linux system mainly includes Linux kernal file, virtual disc file and Linux file system.

4.4 UBOOT Add New Devices

Add Sensor driver based on the third party dvsdk_dm368_setuplinux_2_10_01_18.bin. This process includes configuration of Sensor by I2C and outside trigger to Sensor, such as exposure starting, auto exposure setting, etc.

4.5 User Application

This system application incl udes embedded Web server and stream media server. The main functions of embedded Web server are monitoring client service request, and provide corresponding service, such as user authorization management, IP network parameters modification, video parameters configuration, and RS323/485 control and so on. This system uses the open boa Web server as the embedded Web server, controling the web server by cross compiling boa source program and then modify ing the configuration file boa.conf. Stream media server transports H.264 video stream by RTP protocol. RTP uses UDP to transport data. When the IP network cam era received video m onitoring and control commands from client, it builds the UDP connection, encodes and encapsulates the H.264 video, and then send data package. When it received the video monitoring stop command, it stops encoding and closes the UDP connection. The controling instructions are trans ported by TCP, and RTP data packages are transported by UDP. This sy stem use the open stream media library Live555, which supports the standard RTP/RTCP, RTSP protocol.

5. SYSTEM TESTING

In the IP network environment, the video image transmission is stable, the im age quality is clear, and the power consum ption is lower. Therefore, the s ystem can satisfy the design requirements. The maximum resolution of the output video image is 1080p@30fps. The actual resolution of the output image is relevant to that of the image sensor. Because of the small size and convenience to install of the sy stem, it can be used widely in all kinds of the surveillance fields.

6. SUMMARY

This paper m akes the res earches on the hardware design and implementation of HD netw ork camera based on the TMS320DM368. The system innovations lie in: (1) Using digital network camera to replace the analog camera, this scheme not only reduces the whole complexity of system installation, but also implements the HD video output, which provides the possibility of video intelligent analy sis. (2) The design schem e based on TMS320DM368 supports both CMOS and CCD image sensors, which provides the flexibility to user's choice. The further research on this project is to im plement the HD network camera based on CCD image camera.

7. REFERENCES

[1] Shengyue, X. 2010. Network Camera Technology and Application Trend, *China Security and Protection*. 3(Mar 2010), 43-44.

[2] Xuelian Y., Qian C., GuoHua G, Design of Embedded Wireless Network Camera Based on MX27, *Microcomputer Information*. 25, 10(Oct. 2009), 65-67.

[3] Texas Instruments Incorporated, TMS320DM368 digital media system-on-chip, http://www.ti.com.cn/cn/lit/ds/symlink/tms320dm368.pdf, 2012.

[4] Kodak Company, KAI-02150 data sheet, http://www.kodak.com/ek/uploadedFiles/Content/Small_Business/Images_Sensor_Solutions/Products/KAI-02150LongSpec.pdf, 2011.

[5] Aptina Image Corporation, MT9P031 data sheet, www.aptina.com/products/image_sensors/mt9p031i12stc, 2011.

[6] Davicom, DM9161 data sheet. http://www.davicom.com/DM9161-datasheet.html, 2002.

Using Grey Decision to FPGA Multi-Tasking Scheduling Reconfigurable Systems

Jan-Ou Wu
Department of Electronic Eng.
De Lin Institute of Technology
Tu-Cheng, Taipei, Taiwan, ROC
janou@ms42.hinet.net

Nian-You Lin
Graduate Institute of Computer and Communication
Nation Taipei University of Technology
Taipei, Taiwan, ROC
T100418081@ntut.edu.tw

Yang-Hsin Fan
Department of Computer Science and Information Eng.
National Taitung University
Taitung, Taiwan, ROC
yhfan@nttu.edu.tw

San-Fu Wang
Department of Electronic Eng.
Ming Chi University of Technology
Taipei, Taiwan, ROC
sf_wang@mail.mcut.edu.tw

ABSTRACT
Recently, the FPGA hardware tasks scheduling have become the focal study. Most of the existing papers researched on single-objective optimal operation of reconfigurable assembly line. Although it can get the optimal result, it can spend the high cost in other aspects. Therefore, the single-objective is not an optimal scheme. In this propose, we embed the TGFF to the task graphs parameter format and produce the benchmark of task graph. We approach the Grey Decision Analysis Placement (GDAP) algorithm for multi-tasking scheduling to whole system in optimal performance from different work project.

Categories and Subject Descriptors
H.2.4 [**Database Management**]: Systems – *object-oriented databases*, *transaction processing, rule-based databases*, *query processing*.

General Terms
Design, Theory, Experimentation, Performance.

Keywords
FPGA, Grey Decision, Scheduling, TGFF, GDAP.

1. INTRODUCTION
FPGA have become an attractive the ASIC because of its short time to market, and low manufacturing cost. An FPGA chip consists of programmable logic blocks, programmable interconnections, and programmable I/O pads, which are placed on a two-dimensional array in the FPGA chip.
When FPGA turn off the power or need to change the function it must be reconfiguration and add the hardware function. Such as, The FPGA resource utilization is limited. In order to solve the problem mentioned above, Technology of reconfigurable system can offer FPGA to satisfy different design demand. Use reconfigurable system technology can plan FPGA into some of reconfigurable module and realize existing hardware resources to multifunctional system application, furthermore when reconfigurable system processing, it would not affect the running circuit in other module. Under the condition that just change parts of hardware resource could decrease the allocation time and increase utilization and system flexibility of FPGA, promote the overall performance, to achieve saving hardware resource and cost. Every task is an independent function, there's no priority and dependency in Hauck [1]. The time of reconfiguration hardware task occupies 10% of overall hardware executing time in FPGA, thus it can be seen that a good scheduling algorithm will influence the refuse frequency and all work executing time when FPGA reconfiguration directly in Kalra and Roman [2]. Proposes a new method, when hardware task been executed, it doesn't remove the task from the surface of RPU in Bassiri and Shahhoseini [3]. Use directed acyclic graph to express the dependent architecture of reconfigurable hardware task, and then to optimize static circuit in Belaid et al. [4]. When it has been executing with priority and this phenomenon will cause the delay on processing in Clemente et al. [5].
From reference mentioned above, scheduling in FPGA mainly determine the system configuration and the order of implementation of each work, A good scheduling could reduce the complexity of configuration and routing, so it is important that scheduling in design flow of FPGA. In this paper, we propose Grey Decision of Multi-Tasking of multi-goal to finish algorithm of multi-goal work scheduling, and then to better overall system efficacy. We associate the grey decision analysis [8] technology to construct FPGA multi-tasking scheduling re-configurable systems.
The remainder of this paper is organized as follows. Section 2 discusses the proposed grey decision grade. Section 3 presents our proposed methodology. Experimental results are reported in section 4. Finally, a conclusion and discussion of future research directions are given in section 5.

2. GREY DECISION GRADE
Since Deng [6] first investigated grey theory, the theory has been utilized for various applications [7-9]. The theory consists of five major parts - grey decision analysis, grey predication, grey decision making, grey programming and grey control. Grey decision theory is an effective method for solving uncertainty problems using discrete data and incomplete information. Grey decision making with grey cell or grey functions [11], Includes

situation decision making, hierarchy decision making and grey programming, and used in engineering [12], [13].

Systemic decision making was developed in the field of engineering. It involves events, strategies and targets. This work employs this method to determine the pairs of LUT clustering size N and input number k that optimize the performance of FPGA architecture. The general procedures of grey decision making theory are as follows.

First, define a set of strategies A = { a_i } (i = 1,2,…,n) and a set of events B = {b_j}(j = 1,2,…,m), and define the situation as S = A×B = {$S_{i,j}$ (a_i,b_j) ｜ $a_i \in$A, $b_j \in$B}. A strategy refers to any result of a decision-making process. An event and a strategy constitute a particular situation pair.

Second, define the target sequence as {1,2,…p}. This target refers to the indicators that are used in the evaluation of strategies. Let $u_{i,j}^{(p)}$ be the sample effect value of situation $S_{i,j}$ on target p. Let $r_{i,j}^{(p)}$ be the measure of the effect of situation $S_{i,j}$ on the target p. Effect measure $r_{i,j}^{(p)}$ and sample effect value $u_{i,j}^{(p)}$constitute a mapping pair, γ:$u_{i,j}^{(p)} \rightarrow r_{i,j}^{(p)} \in$ [0.1].The value of the sample effect will be transformed into a value between 0 and 1.

Third, define the effect measure. Three effect measures [14] are the upper effect measurement for the maximum target, the lower effect measurement for the minimum target and the medium effect measurement for the nominal effect weighting of effect sample.

Finally, define the decision-making matrix. A decision-making cell consists of a situation $S_{i,j}$ and its effect measure $r_{i,j}$. For a given strategies ai, event b_1, b_2,...,b_m yield situations $S_{i,1}$, $S_{i,2}$,...,$S_{i,m}$. For a given same event b_j, strategies a_1,a_2,...,a_n yield situations $S_{1,j}$, $S_{2,j}$,…,$S_{n,j}$. The S_i column in a decision-making matrix is S_i = [$r_{i,1}$/$S_{i,1}$, $r_{i,2}$/$S_{i,2}$, ..., $r_{i,m}$/$S_{i,m}$] and the decision-making line S_j is S_j = [$r_{1,j}$/$S_{1,j}$,$r_{2,j}$/$S_{2,j}$,.....,$r_{n,j}$/$S_{n,j}$]. Let M = $(r_{i,j}/S_{i,j})_{n\times m}$ be the decision-making matrix. $M^{(p)}$ is the decision-making matrix for target p, which is defined as follows.

$$M^{(\Sigma)} = \begin{bmatrix} \frac{r_{1,1}^{(\Sigma)}}{S_{1,1}} & \frac{r_{1,2}^{(\Sigma)}}{S_{1,2}} & \cdots & \frac{r_{1,m}^{(\Sigma)}}{S_{1,m}} \\ \frac{r_{2,1}^{(\Sigma)}}{S_{2,1}} & \frac{r_{2,2}^{(\Sigma)}}{S_{2,2}} & \cdots & \frac{r_{2,m}^{(\Sigma)}}{S_{2,m}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{r_{n,1}^{(\Sigma)}}{S_{n,1}} & \frac{r_{n,2}^{(\Sigma)}}{S_{n,2}} & \cdots & \frac{r_{n,m}^{(\Sigma)}}{S_{n,m}} \end{bmatrix}$$

Where $r_{i,j}^{(\Sigma)} = \sum_{i=1}^{n}\sum_{j,p=1}^{m} w_p r_{i,j}^{(p)}$, $\sum_{p=1}^{m} w_p = 1$

In the above formula, w_p is the weight of the target. If $a_i^*$is the best strategy, then

$$r_{i^*,j}^{(\Sigma)} = \max\{r_{i,j}^{(\Sigma)}\}$$

3. PROPOSED METHOD

This research adopts Multi-objective Task Scheduling of Reconfigurable Grey decision, it can follow user's different demand and according to every property of work to decide the order of task scheduling when FPGA loading works. Reconfigurable logic resource is made up by a two-dimensional array of CLB, it offer FPGA's use state and store the execution state of task currently.

Placement module is made up by scheduler, placer and loader in task scheduling. Scheduler is used to randomly decide the executing task presently. All of task will be stored in module library until scheduler randomly decides the task and the corresponding will be configuration into FPGA. Placer is used to manage space and find the best space to the loading task. And there is no priority and dependency problem, ever task is an independent function.

At the view of all executing time that time is a very important feature, so configuration FPGA mode must consider the executing time, and assume there is enough space besides it doesn't consider the routing problem. That is to say the important feature of task modes reconfigure that important features are area size, configuration, allocation and shape.

- *Size*: Every area size of task is equal to the amount of CLB.
- *Overhead*: If there's any task in FPGA, the configuration time will also be affected and cause time overhead when allocating.
- *Locatability*: Assume that every task can be allocated into FPGA, so it hasn't to consider I/O feature.
- *Shape*: The initial shape of each task is rectangle. But when task implemented in FPGA, not each task in rectangular shape. Therefore, in order to simulate this condition, we assume that the shape of these tasks could be changed.

So we establish reconfigurable scheduling will take Configuration Area (CA) size, Configuration Time (CT), Running Time (RT) important feature into condition.

Dick[10] proposed random Task Graphs generator named Task Graph for Free (TGFF) in 1998, and open its source code. This paper will use TGFF generator generate some tasks to process test of Grey decision and with Configuration Area (CA) size, Configuration Time (CT), Running Time (RT) important feature mentioned above to experiment.

4. EXPERIMENTAL RESULT

The proposed Grey Decision Analysis Placement (GDAP) algorithm was utilized TGFF construction in C# Builder and run on a desktop PC with the CPU of 3.4 GHz Intel Core(TM) i-7-2600. Re-configurable multi-objective task scheduling of grey decision scheduling work design discusses the Small configuration size, Short running time, short configuration time and multi-objective.

The following example is correspond strategy with Grey Decision System and correspond method which take area size, Configuration time and executing time into condition, use TGFF [10] to generate 10 set of task, as Task1、Task 2、…、Task10, defined following: Target 1 is Configuration Area (CA), Target 2 is Configuration Time (CT), Target 3 is Running Time (RT).

4.1 Quantitative result of the three effects

According to corresponding result with area, configuration time and running time was utilized TGFF construction then to do ten tasks qualitative analysis. In the task, configuration time is gain from behavior plus the part of column, about the CA, CT and RT effects property are shown in Table 1.

Table 1. Quantitative result of the three effects target

TGFF to generate 10 tasks	Quantitative result of the three effects target		
Task	CA	CT	RT

Task 1 (r_{1-1})	6	2	43
Task 2 (r_{1-2})	10	2	24
Task 3 (r_{1-3})	36	6	33
Task 4 (r_{1-4})	45	9	60
Task 5 (r_{1-5})	39	4	72
Task 6 (r_{1-6})	55	11	142
Task 7 (r_{1-7})	68	5	37
Task 8 (r_{1-8})	78	39	114
Task 9 (r_{1-9})	57	4	191
Task10 (r_{1-10})	69	23	86

4.2 Three smaller and better deviation generate result of target

According to the generation of effect measure, area, configuration time and running time are used to measure the data degree of deviation the smaller deviation and the better answer. It mean that the smaller the configuration time. According to lower effect measurement formula, three smallest and the better result can be obtained as shown in Table 2.

Table 2. Three smallest and better result of target

TGFF to generate 10 tasks	Three smallest and better result of target		
Task	**CA**	**CT**	**RT**
Task 1 (r_{1-1})	1.000	1.000	0.558
Task 2 (r_{1-2})	0.600	1.000	1.000
Task 3 (r_{1-3})	0.167	0.333	0.727
Task 4 (r_{1-4})	0.133	0.222	0.400
Task 5 (r_{1-5})	0.154	0.500	0.333
Task 6 (r_{1-6})	0.109	0.182	0.169
Task 7 (r_{1-7})	0.088	0.400	0.649
Task 8 (r_{1-8})	0.077	0.051	0.211
Task 9 (r_{1-9})	0.105	0.500	0.126
Task10 (r_{1-10})	0.087	0.087	0.279

4.3 Three Target State Results

As the effect measure above, if use 0.9, 0.05 and 0.05 to represent for weight value of three target state, CA, CT and RT complex result of the CA size of smallest target. If execute scheduling with the smallest CA size, Task 1 is shown the best executed. The sequence scheduling with the smallest configuration area size is shown in figure 1.

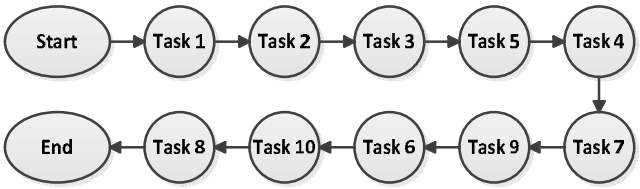


Figure 1.The sequence scheduling with the smallest CA size.

If use 0.05, 0.9 and 0.05 to represent for weight value of three target state, CA, CT and RT complex result of the CT size of smallest target. If execute scheduling with the shortest CT, then Task 2 is shown the best executed. The sequence scheduling with the smallest configuration time is shown in figure 2.

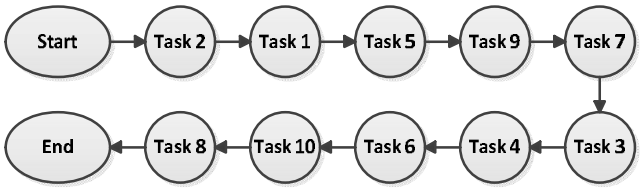


Figure 2.The sequence scheduling with the smallest CT size.

If use 0.05, 0.05 and 0.9 to represent for weight value of three target state, CA, CT and RT complex result of the RT of smallest target. If execute scheduling with the shortest RT, then Task 2 is shown the best executed. The sequence scheduling with the smallest running time is shown in figure 3.

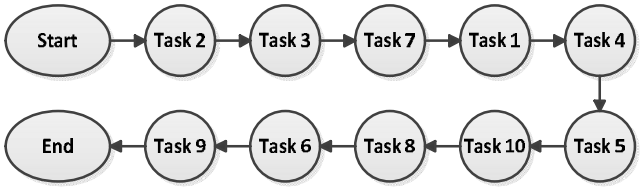


Figure 3.The sequence scheduling with the smallest RT size.

From the complex result of Table 3, Grey decision design of Grey System can according to CA, CT and RT, different demand to obtain various kinds of the best multi-objective task scheduling, and FPGA set the best task scheduling with the correspond effect target strategy decision of Grey decision finally.

Table 3.Three target state results

TGFF to generate 10 tasks	Three Target state results					
Task	**CA**		**CT**		**RT**	
	Time	***Seq.***	***Time***	***Seq.***	***Time***	***Seq.***
Task 1 (r_{1-1})	0.978	1	0.978	2	0.602	4
Task 2 (r_{1-2})	0.640	2	0.980	1	0.980	1
Task 3 (r_{1-3})	0.203	3	0.345	6	0.680	2
Task 4 (r_{1-4})	0.151	5	0.227	7	0.378	5
Task 5 (r_{1-5})	0.180	4	0.474	3	0.333	6
Task 6 (r_{1-6})	0.116	8	0.178	8	0.167	9
Task 7 (r_{1-7})	0.132	6	0.397	5	0.608	3
Task 8 (r_{1-8})	0.082	10	0.061	10	0.196	8
Task 9 (r_{1-9})	0.126	7	0.462	4	0.143	10
Task10 (r_{1-10})	0.097	9	0.097	9	0.260	7

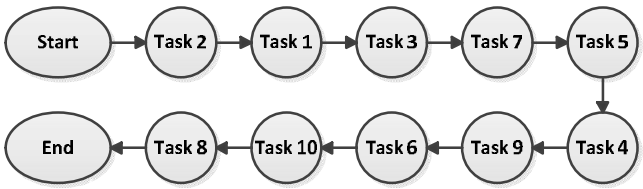


Figure 4.The sequence scheduling with the average result.

According to correspond Table 3 result with CA, CT and RT then to do qualitative analysis with average weight value of three target state, CA, CT and RT complex Average Execution Time of multi-Decision shown in Table 4. If execute scheduling with the average result, then Task 2 is the best executed. The sequence scheduling with the average result is shown in figure 4.

Table 4.Three target state average results

TGFF to generate 10 set of task	Three Target State Results	
Task	Average	
	Time	*Seq.*
Task 1 (r_{1-1})	0.852	2
Task 2 (r_{1-2})	0.866	1
Task 3 (r_{1-3})	0.409	3
Task 4 (r_{1-4})	0.252	6
Task 5 (r_{1-5})	0.329	5
Task 6 (r_{1-6})	0.153	8
Task 7 (r_{1-7})	0.379	4
Task 8 (r_{1-8})	0.113	10
Task 9 (r_{1-9})	0.243	7
Task10 (r_{1-10})	0.151	9

5. CONCLUSION

This study investigated a Scheduler problem in FPGA placement. We propose a method of Grey Decision Analysis to improve and discuss each task in FPGA and use C# language to process data test. TGFF will generate Benchmark, and then input to grey decision system according to different target, finally, gain the best reconfigurable multi-objective scheduling meanwhile to analyze and compare with other method.

6. ACKNOWLEDGMENTS

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC 101-2221-E-237-005-.

7. REFERENCES

[1] S. Hauck, “Configuration prefetch for single context reconfigurable coprocessors,” ACM Symposium on FPGA, pp.65-74, 1988.

[2] Rahul Kalra and Roman Lysecky, “Configuration Locking and Schedulability Estimation for Reduced Reconfiguration Overheads of Reconfigurable Systems,” IEEE Transcations on Very Large Scale Integration(VLSI) System, vol. 18, no.4, pp.671-674, April, 2010.

[3] Maisam Mansub Bassiri and Hadi Shahriar Shahhoseini, “A New Approach in On-line Task Scheduling for econfigurable Computing Systems,” ASAP pp.321-324, 2010.

[4] Ikbel Belaid, Fabrice Muller and Maher Benjemaa, “Optimal Static Scheduling of Real-Time Dependent Tasks on Reconfigurable Hardware Devices,” Communications, Computing and Control Conference, pp.1-6, Narch, 2011.

[5] Juan Antonio Clemente, Carlos Gonzalez, Javier Resano and Daniel Mozos, “A Hardware Task-Graph Scheduler for Reconfigurable Multi-tasking Systems,” International Conference on Reconfigurable Computing and FPGAs, pp. 79-84, Dec. 2008.

[6] J. L. Deng, “Introduction to grey system theory,” The Journal of Grey System, vol. 1, pp.1-24, 1989.

[7] Jan-Ou Wu, Chyun-Shin Cheng, and Chia-Chun Tsai, “The application of Grey multi-objective decision making in data structure of VLSI layout,” Journal of National Taipei University of Technology, vol. 34-1, pp.99-107, 2001.

[8] Chyun-Shin Cheng and Shu-Yueh Lee, “The application of Grey multi-objective decision making in performance evaluation of DAC and ADC,” Journal of National Taipei University of Technology, vol. 35-1, pp.149-156, 2002.

[9] Jan-Ou Wu, Chyun-Shin Cheng, and Chia-Chun Tsai, “Application of Grey relational analysis to minimal clock skew routing in SoC,” The Journal of Grey System, vol. 16, no. 3, pp.221-234, 2004.

[10] R. P. Dick, D. L. Rhodes and W. Wolf, “TGFF: task graphs for free,” in Proceedings of the Sixth International Workshop Hardware/Software Codesign, Washington, USA, Mar. 15-18, pp.97-101, 1998.

[11] Liu Sifeng, Guo Tianbang and Dang Yaoguo, “Grey system theory and application,” Beijing: Science Press, pp.174-194, 1999.

[12] Zhou Ziyang, Liu Sifeng, and Wan Jun, “Grey decision-making method in exit decision of venture capital,” Journal of Nanjing University of Aeronautics & Astronautics, vol. 38, no. 3, pp.393-396, Jun. 2006.

[13] Tang Yingying, and Zhang Xingzhou, “Application of grey-decision method in partner-selection of joint venture,” Construction Management Modernization, Total, 82, no. 3, pp.1-4, 2005.

[14] Luo Youxin, “Grey system theory and approach to mechanical engineering,” Changsha: National University of Defense Technology Press, pp.152-174, 2001.

Data Warehouses and Business Intelligence Used for Exam Analysis

Kornelije Rabuzin
University of Zagreb, Faculty of organization and informatics Varazdin
Pavlinska 2
42000 Varazdin, Croatia
+385(0)42/390-847
kornelije.rabuzin@foi.hr

ABSTRACT
Faculty of organization and informatics in Croatia represents an interesting choice for vast number of students. In order to enroll students had to pass the entrance exam that consisted of several groups of questions covering different courses. Based on students' high school knowledge (and grades) as well as entrance test results (and some additional activities) we used to determine the list of students that were eligible to enroll. Later on (during their study) those students passed some exams as well. This paper describes the solution that was built in order to analyze the success of students from different schools and regions. For that purpose a data warehouse was built that integrated data from several different sources and front-end business intelligence tool was used to build reports that were used to clarify the whole situation.

Categories and Subject Descriptors
H.2.7 [**Information Systems**]: Database Administration – *data warehouse and repository.*

General Terms
Design, Experimentation.

Keywords
Data warehouse, Business intelligence, Exams.

1. INTRODUCTION
In the past 15 years (or so) data warehouses have been used extensively to integrate data from different applications in order to enable the data analysis and in order to build reports that contain interesting piece of information. While in the past people used to spend 90% of time preparing the data (they needed) for analysis and only 10% of time analyzing the data, data warehouses have made it possible for number to switch their places. It is clear today that spending 90% of time to build a report is not acceptable (any more) and by using data warehouses producing reports is a matter of seconds.

The answer to the question how this is possible lies in the fact that data warehouses are special type of databases organized in a star (or snowflake) schema that is easy to understand and that can be used to build reports, even for un-experienced users. But to enable this report creation the daunting task called ETL (Extract Transform Load) has to be performed (in the background); ETL usually enhances the data quality and resolves the inconsistencies and other things that make the data (in its original form almost) useless. Once the data is in the data warehouse, OLAP (i.e. front-end) tools can be used to build reports and to analyze the data by using advanced capabilities (drill down, pivot, slice, dice, etc.).

Throughout the history it was a very common scenario that many different applications were built and used within business systems. Once the managers realized what IT can do, without any special planning many applications were built and deployed, but technologies used to build those applications were heterogeneous and incompatible. Although each application was useful in a sense that it supported some business processes, combining the data from heterogeneous systems represented a huge problem. Because of that it was almost impossible to build reports that would contain the data from different systems. But today we know that one has to place the data in some context to reveal its true value; for that purpose you have to analyze the sales data and usually compare them to financial data or global trends data to realize what is really going on. Data warehouses are here to do exactly that; data from different applications (systems) are moved to a single repository called a data warehouse. Once the inconsistencies are resolved and certain degree of redundancy is added into the system, the data warehouse represents a real gold mine that can be used to reveal many interesting things by building different reports and analyzing the data in different ways. At the same time the data organized in such a way can be used for data mining purposes as well, but we will not discuss this any further in the paper.

SQL is a dominant and standardized language used to work with databases and data warehouses (although some data warehouses can use MDX as well, but we will skip that). Although the language is standardized, during the years it has become quite complex. Further on, when the language was introduced the idea was that it was supposed to be simple so that end-users could build queries on their own; today we know that professionals can have problems with some queries as well and it is not reasonable to expect that end-users pose complex queries that are usually needed to answer some question. Further on, the logical data model is usually complex (containing many tables) and end-users are usually unaware of how to join them together (in different ways) in a single query. Since data can be found in many heterogeneous applications and SQL is complex, user friendly

Research Notes in Information Science (RNIS)
Volume13, May 2013
doi:10.4156/rnis.vol13.40

interface is a "must" in order to build successful reports. Business intelligence tool fit nicely as we will see later on.

This paper focuses on (entrance) exams at the Faculty of organization and informatics, University of Zagreb. The problem that was obvious was that in one point of time we have had five different applications that supported the exams (in one way or another). However, it was almost impossible to build certain (advanced) reports that would show us which students were the best i.e. which schools (regions) produce the best candidates, etc. Because of that a data warehouse has been built and certain reports have helped us to come to some very interesting conclusions.

The rest of the paper is organized as follows; the problem scenario is explained in the following section and then the data warehouse model is explained (briefly). Later on some reports (that were created) are presented and some remarks are given. In the end the conclusion is presented.

2. EXAMS

In order to enroll, students had to pass the entrance exam. Entrance exam contained questions from several different areas (courses) like math, informatics, foreign language, etc. (the list of courses sometimes changed from year to year). The entrance exam was used to check whether students know all the things that should know and based on the entrance exam results some students (i.e. candidates) were eliminated i.e. the best candidates were enrolled.

A very interesting point here is that students (or candidates) that came from certain schools (regions) usually achieved better results meaning that some schools were better than some other schools. Further on, students could get some additional points for certain activities as well (professional athlete, third foreign language, etc.) and it was obvious that certain schools produced candidates with higher number of additional points. Because of that it was crucial to understand what was really going on.

The entrance exam took place two times a year; in the summer (usually in July) and in the autumn (usually in September). After the entrance exam was over we had to build a few reports showing the number of points, average high school grades, etc. Although some of those reports were not complex, they were usually built within a few days (one report required several different queries to be combined and presented on a single report).

If we tried to compare the results to the ones from the previous years, this was not so easy to accomplish. Further on, once the students were enrolled they passed certain exams. Comparing their high school grades with entrance and passed exams results was almost impossible. Five different applications contained relevant data and there was no way to present them in just one report:

- one application was used to store the data on entrance exams;
- one application was used to store the data on high school grades and
- three applications were used to store the data on exams that students had passed.

For the passed exams three different applications were used:

- one old DOS application that was not used but still contained some relevant information on student exams (history data is relevant for data warehouses);
- the successor to that old DOS application was an application that was used for the same purposes for about 4-5 years, and was developed by our local team;
- and the third application (still) used for the same purposes is the one developed and supported on the university level.

In order to analyze the data a small data warehouse was built that integrated data from different sources; the data warehouse model is explained in the following section.

3. THE DATA WAREHOUSE MODEL

As we already know data warehouses are databases that are arranged according to some other design principles; unlike databases where redundancy causes anomalies, in data warehouses redundancy is desirable because it reduces the number of joins and because it makes the model understandable.

The star schema in the data warehouse contains a fact table and dimension tables that are (usually) organized as one can see in the Figure 1:

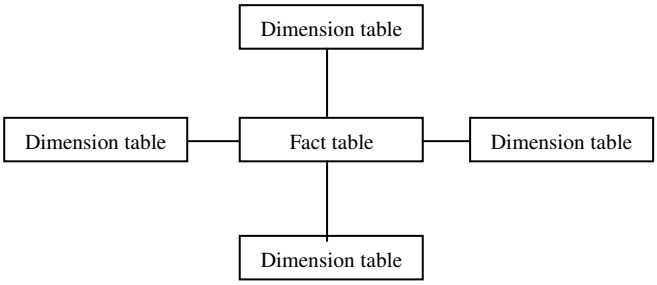


Figure 1. Star schema

Such a model is easy to understand, even for end-users. The number of tables to be joined is reduced and end-users can create reports on their own because they understand such a schema and distinguish attributes and measures (usually drag & drop principle is used to create reports). Of course, the whole situation is not trivial because many data warehousing projects fail, but we assume the reader is familiar (to certain extent) with the technology. There are some other design issues that we will not discuss in the paper (keys, SCD, aggregate tables, etc.).

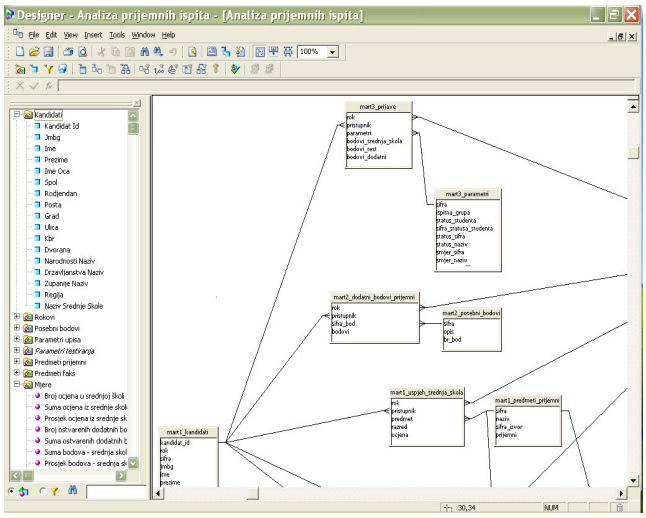


Figure 2. Data warehouse model

Since the data warehouse was rather small (one table had 200 000 rows and other tables much less), it was implemented in MS Access 2007 with four different fact tables:

one was used to store data on the high school grades;
one was used to store data on the entrance exams;
one was used to store data on the passed exams, and
one was used to store data on additional points.

We had several dimension tables as well (candidates, courses, additional point, etc) as can be seen in Figure 2 (Business Objects XI was used as a BI tool; the Designer tool was used to build the universe).

One can see (although in Croatian language) many measures and dimension tables with their attributes; all of them can be used to analyze the data and produce reports. This is done in the following section.

More on Business Objects can be found in [1]. More on data warehouses can be found in [2], [3], [4], [5], [6] and [7].

4. REPORTS

We have built a number of reports (what took us hours before was built in a matter of seconds) and some results were astonishing. It is important to have in mind that this warehouse was built within a few months and it has enabled us to build reports that nobody even tried to create before because this was just not feasible (and it was expensive as well). The reason why this was not feasible was that several incompatible technologies were used and nobody tried to extract data from those different systems. With this data warehouse reports that couldn't be built before have been built within minutes. The first report shows the number of candidates compared by years:

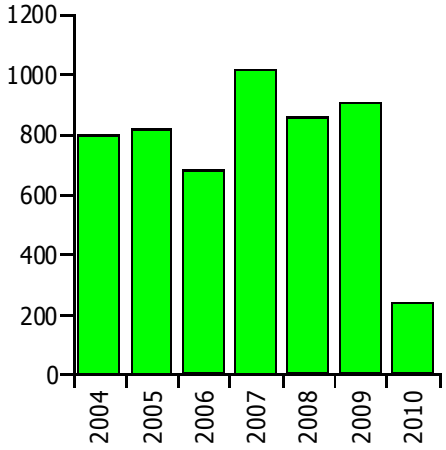


Figure 3. Number of candidates per year

This report is not complex but shows how the number of candidates changed during the years (the year 2010 is to be observed cautiously because data were not complete). This report is interesting when regions and counties are added; one can easily determine the regions and counties that should be targeted more aggressively to attract more students, especially when we take into account the regions and schools from which the best candidates come.

Another interesting report shows the average number of points achieved per several courses (Hrvatski jezik means Croatian language, Informatika means Informatics and Matematika stands for Math) and month of the entrance exam (7 or 9 i.e. July or September):

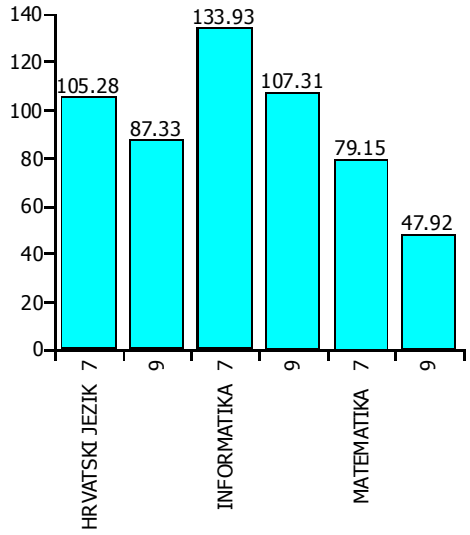


Figure 4. Average number of points per course and month

It is obvious that in July students achieve better results than in September (the maximum number of points is 200) as well as that Math causes the most problems.

Another interesting report is a slight modification of the previous report i.e. it shows the same data only per regions and course (green represents Croatian language, blue represents Math and yellow represents Informatics while the X axis represents the region names in Croatian). Since regions form a hierarchy (region – county – school), the report can be drilled down to county and school, but this will not be shown because it is then obvious which school produces the best candidates (but when this was done the results were surprising):

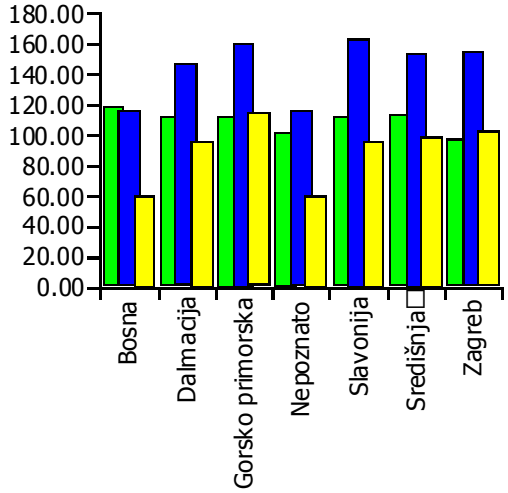


Figure 5. Average number of points per region and course

Another report is quite interesting because it shows the number of points achieved through some additional activities (table rows represent 1st – 3rd place in competitions, second foreign language, already passed entrance exam, professional athlete, second high school and third foreign language and columns represent years); each activity brings 20 points. We see that the most influential activity that brings many additional points is the second foreign language. Some cells are empty meaning that (for that specific year) we didn't have candidates with such type of additional activity:

Additional points

	2004/2005	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011
13. MJESTO NA TAKMIČENJU	20.00	140.00	80.00	80.00	100.00	120.00	20.00
DRUGI STRANI JEZIK KROZ 4 GODINE	5,140.00	6,060.00	5,400.00	8,080.00	6,920.00	6,300.00	1,320.00
POLOŽEN RAZREDBENI ISPIT U AKADEMSKOJ					80.00	20.00	40.00
SPORTAŠ 1. I 2. KATEGORIJE					120.00	40.00	40.00
ZAVRŠENA JOŠ JEDNA SREDNJA ŠKOLA				80.00	100.00	160.00	120.00
ZNANJE TREĆEG STRANOG JEZIKA	40.00	220.00			680.00	100.00	140.00

Figure 6. Additional points

Another very interesting report shows the high school grades, the entrance exam results and the passed exam results per region. The story repeats again and one can drill down and see the results on the school (county) level, but this is (again) not shown in the paper for obvious reasons:

Region	High school (max. 400)	Entrance exam (max. 600)	Passed exams (max. 5.00)
Zagreb	209.62	297.72	2.04
Nepoznato	222.78	262.24	2.07
Dalmacija	223.21	305.38	2.01
Središnja	227.05	306.60	2.08
Gorsko primorska	228.96	341.19	2.28
Slavonija	231.94	324.25	1.91
Bosna	258.57	268.21	2.42

Figure 7. Points per regions

It is important to have in mind that it was almost impossible to build such a report (earlier) because one had to look at several different applications (and their databases) to do such a thing, but this was too complex due to heterogeneous and incompatible technologies.

Here are just some reports but many others were built and many other reports could be built because of the vast number of attributes and measures in the data warehouse. The reports presented in the paper just show some of the possibilities and they are significant because they contain interesting piece of information, but some reports were not shown in the paper because they would reveal sensitive information.

One can easily say that this data warehouse represents a valuable source of information. Reports in the paper contain important information and could be helpful in many ways i.e. to improve the "input", to guide marketing activities, etc.

5. CONCLUSION

We can say that the results were (and still are) interesting. Many trends can be spotted and some assumptions that we had had were confirmed.

An interesting thing however is that people that were supposed to be interested in such a solution didn't show much interest at all. Since they are all familiar with IT, it is hard to explain their behavior.

In our future paper data mining techniques will used since data warehouses represent cleaned collections of data suitable for data mining techniques.

6. REFERENCES

[1] Howson, C. 2006. *BusinessObjects XI: The Complete Reference*, The McGraw-Hill Companies, USA.

[2] Inmon, H. W. 2002. *Building the Data Warehouse – Third Edition*. Wiley Computer Publishing, USA.

[3] Kimball, R. and Caserta, J. 2004. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Wiley Publishing, USA.

[4] Kimball, R. and Ross, M. 2002. *The data warehouse toolkit: the complete guide to dimensional modelling*, Wiley Computer Publishing, USA.

[5] Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., and Becker, B. 2008. *The Data Warehouse Lifecycle Toolkit – Second Edition*. Wiley Publishing, USA.

[6] Ponniah, P. 2001. *Data Warehousing Fundamentals*, John Wiley & Sons, USA.

[7] Silvers, F. 2008. *Building and Maintaining a Data Warehouse*, CRC Press, USA.

Cluster-Based Energy Aware Routing (CBEAR) - A Self-Healing and Self-Organizing Routing Protocol for Firefighter Communication Network

Mahin K. Atiq
Department of Information and Communication Engineering, Sejong University, Seoul, Republic of Korea
mahinatiq@yahoo.com

Kamran Manzoor
Department of Computer Science and Engineering, Aalto University, Finland
kamran.manzoor@aalto.fi

Hyung Seok Kim*
Department of Information and Communication Engineering, Sejong University, Seoul, Republic of Korea
hyungkim@sejong.ac.kr

ABSTRACT
The life of firefighters and other humans depend on reliable and efficient communication between firefighters and their commandant. In this paper, a concept of cluster based hierarchy along with energy aware routing has been established to form a self-organizing and a self-healing network for firefighter communications. The proposed scheme has been incorporated into the existing Ad-hoc On-demand Distance Vector (AODV) protocol. The cluster based approach; along with energy aware routing helps reduce control overheads and repair broken links, thus providing reliable communication to the firefighters.

Categories and Subject Descriptors
D.3.2 **[Programming Languages]:** General

General Terms
Algorithm, Design, Performance

Keywords
Firefighter, Self healing network, Cluster based network

1. INTRODUCTION
Recently, advanced fire rescue techniques have required sensing of environmental conditions and firefighter vitals, along with regular command messages between the firefighters and the commandant who oversees the entire fire rescue operation. The firefighter network usually consists of body-mounted sensors or sensors deployed along the path [1]. The sensors continuously sense environment and firefighter vitals and send updates to the base station. The firefighters continuously move inside the fire rescue scene to find the humans to rescue, while they extinguish the fire. Firefighters cannot rely on any previously deployed communication infrastructure because that might have been destroyed by the fire [2]. Firefighters need the support of a network that is easily established without any need for infrastructure to be installed [2, 3]. Wireless mobile ad hoc network [MANET] is the best solution for such an emergency situation [2]. Keeping in view the energy and memory constraints of firefighter's body-mounted sensors and lossy links of networks, proper routing protocols for mobile ad hoc networks should be developed. A robust, energy aware, self-healing and self-organizing routing scheme is needed that can make communication possible even in adverse environmental conditions.

This paper presents Cluster-based Energy Aware Routing (CBEAR) protocol for firefighter communication network. The movement of firefighters in the form of teams at an emergency scenario motivates, CBEAR to adopt a hierarchical cluster-based network data dissemination model, which reduces the number of transmissions at longer distance. Also this clustering approach helps to achieve self-organization as stated in [3]. To maximize the network lifetime, the cluster head (CH) responsibility is rotated among the members of a cluster to distribute the energy consumption evenly among the cluster members as in dynamic LEACH algorithm [4]. Furthermore, the CHs make routing decisions based on the number of hops to the destination, the number of neighboring CHs and their energy level. The CH tries to select the route with the maximum number of neighboring CHs, the minimum number of hops to destination as well as the CHs with maximum residual energy. This would help to maximize the reliability and minimize latency, by avoiding packet drop due to node failure. Regarding the self-healing nature of the protocol, the node at which the link is broken recovers the link, by sending the path request message to the neighboring nodes and selecting an optimal path. This helps to heal the link quickly without the intervention of the source CH.

The remaining of paper is organized as follows: Section 2 discusses the challenges for routing in firefighter network. Section 3 presents the existing routing protocols for firefighter networks and their limitations. Section 4 proposes the frame work for routing in firefighter networks. Section 5 concludes the paper and also presents some future recommendations.

2. ROUTING CHALLENGES IN FIREFIGHTER NETWORKS
The firefighters need to communicate in an environment where

* Corresponding author (hyungkim@sejong.ac.kr)

Research Notes in Information Science (RNIS)
Volume13,May 2013
doi:10.4156/rnis.vol13.41

there is fire, vapor and smoke [5]. The firefighter network consists of several mobile body-equipped nodes inside a building where localization is not possible [1, 2, 6] and the nodes need to send and receive data on regular basis to the commandant or base station (BS). The fire rescue operation depends on the reliable communication between the commandant and the firefighters.
The firefighters always work in teams at a fire scenario. Every team has a team leader and the basic mission is to rescue while staying together. No one should be left behind alone [7]. Fig. 1 shows the typical firefighter network model. The rescue teams are highly structured and organized, no one just moves around randomly. Every team has a team leader that manages the team solely and decides when to move and when to stop. These decisions are based on tactical reasons [8].

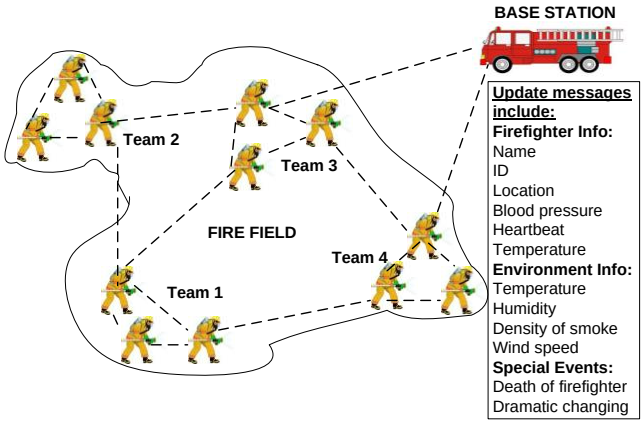


Figure 1.Typical Firefighter Network Model

Very stringent performance requirements in terms of latency, self-healing and reliability are there for the fire fighter communication network. The firefighter network should not only provide reliable communication but also provide sensor data to the firefighters and their commandant [2]. The firefighters cannot rely on any infrastructure that might be present at the fire place. Thus the ad hoc network is the only solution for the firefighters to communicate. Moreover, the firefighter network should be able to work in adverse environmental conditions, enduring a high temperature of up to 150 degrees Celsius, visibility down to 30 cm due to smoke, high environmental noise, walls and infrastructure covered with water [1].

3. ROUTING PROTOCOLS FOR FIREFIGHTER NETWORKS

In the literature, routing protocols for MANETs have been classified into three categories, i.e., proactive, reactive and geographical routing protocols. Flooding is incorporated as a basic routing mechanism in different routing protocols. Each routing protocol has its own pros and cons depending on the application scenario. A single routing protocol cannot outperform all others in every possible scenario [9].
Proactive routing protocols are better suited for fixed topology networks and have worst performance in MANETs where the network topology changes continuously [9]. Because the firefighters move continuously inside a fired building [7], the network topology changes frequently. As a result, proactive routing protocols cannot be used directly for the firefighter scenario. In the scenario of topology diversity, the reactive routing protocols incur low overhead. On the other hand, finding a route on request causes high route discovery latency and suboptimal route problems [9]. Geographic routing protocols are best suited for a continuously changing network. However, these protocols require updated location of each node at regular intervals, which results in high delay overhead due to the use of an exclusive location service [5]. Hybrid routing protocols successfully combine the benefits of both reactive and proactive routing protocols, but still a lot of work is need to be done in this field [9]. Rather than employing these MANET routing protocols, a new self-healing and energy efficient routing protocol is needed for the firefighter communication network.

The firefighter network needs to be established quickly, resistant to link breakage and efficient to local link repair. Studies on self-healing networks have been going on. A self-healing network should be able to reinstate its normal state as it was working before the failure [10]. The core idea of the self-healing network is to provide reliable communication across multiple nodes in the presence of continuous node and link failures [11]. In [12], a self-healing routing using broadcast communication and prioritized transmission with slotted time is presented, where the receiving node has to decide whether to forward, accept or discard a packet depending on distance from the destination. The routing scheme in [12] is not energy efficient and does not support node mobility. Another self-healing geographical routing protocol is proposed in [13]. An implicit location service is incorporated in the routing protocol to adapt the geographic paths according to the dynamic network topology. On-demand Geographical Path Routing (OGPR) in [13] achieves higher packet delivery rate and lower control overhead than AODV and dynamic source routing (DSR) protocols. However, OGPR in [13] has a greater delay and cannot be used efficiently for time-sensitive firefighter communication network. Therefore, in this paper, we propose a new self-healing routing protocol for firefighter communication network.

4. PROPOSED ROUTING FRAMEWORK FOR FIREFIGHTER NETWORKS

4.1 Network Model

We consider a network model in which the firefighters work in teams on a fire scene. Each team is led by a team leader. This approach motivates to use a hierarchical cluster based data dissemination structure for the implementation of self-healing routing algorithm. The clusters are organized using the clustering algorithm as presented in [14]. Another reason for using the hierarchical network model is that it is most suitable for the firefighter network because of the firefighter network's rigorous energy constraints.

Fig.2 shows the hierarchical clustering model used for the firefighter communication network. In this network model, the BS acts as the sink for all update messages. Each CH aggregates the data from its cluster members and forwards to the CH closer to the BS. Each team represents a cluster with a CH. The network operates as follows:

1) The CH collects data from the cluster members and forwards it to the CH closer to the BS.
2) Because the CH functions are more sophisticated and require more energy, the CH responsibility is rotated

among cluster members to distribute the energy consumption evenly among all the cluster members.

3) Each node can only transmit data to its own CH, hence conserves energy.
4) The cluster members are in close vicinity and sense the same data that is aggregated at the CH.
5) Only the CH needs to know how to forward the data to the next level CH or BS, so this reduces the complexity of the routing protocol.

It is also assumed that at any time instance, each CH is surrounded by at least three neighboring CHs, to make the communication possible with the BS.

4.2 Proposed Routing Protocol-CBEAR

The proposed routing protocol is a modified version of AODV but self-healing. The routing protocol incorporates on-demand routing along with residual energy level of the neighboring CHs. The routing table has two extra entries 1) number of neighboring one-hop distance CHs and 2) their residual energy levels. The energy level is characterized as high, low and in-danger. The energy level is scaled from 1 to 10, with 10 being the lowest energy and 1 being the highest energy. Whenever a new path is required, the neighboring nodes are queried and reply with the number of CHs attached to them and their power status. The routing tables are updated with the Hello messages as in the case of AODV but with two more entries of neighboring CHs and their corresponding energy levels. Within each cluster, the data is aggregated at the CH and then forwarded using CBEAR to the base station. The route request, the route error and maintenance phases of AODV need to be modified to incorporate the neighboring CHs and their energy status into the routing table. For route recovery at CH, finding neighboring CHs and their residual energy, invalidating broken routes, sending valid route updates and listing of new routes are involved. Fig. 3 explains how the route recovery works. The path selection variable for a CH is usually set according to the following equation:

Path selection variable = No. of hops to BS
+ No. of neighboring CHs
+ Energy status **(1)**

The CH with the least path selection variable is chosen as the next hop for delivering the packet to the base station. Table 1 shows the calculation of path selection variable for the nodes in Fig.3.

The previous route between source and destination is shown by black solid lines. The nodes with high energy are shown in black color and nodes with low energy are shown in grey color. The link breaks as soon as the node **Z** fails due to energy failure. The node **B** sends the path request to its neighboring CHs **M, C** and **K**. CHs **M, C** and **K** reply with their routing tables. The path selection variable is calculated using (1). CH **C** is chosen as the next hop because it has the least path selection variable as listed in Table 1. Following this technique, the route is repaired to **[source-A-B-C-D-E-F-destination]** which is good in terms of energy and consists of only high energy nodes. The repaired path is shown using dotted lines in Fig. 3.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed the concept of cluster-based energy aware routing (CBEAR) for the firefighter communication network. The clustering approach is selected for firefighter networks to achieve self-organization and conserve energy. The routing protocol avoids link breakage by constructing the path with highest energy CHs and repairs broken links without the intervention of the source. Furthermore, the protocol tries to select the path with minimum number of hops to the destination in order to decrease latency. The proposed approach can be a solution to problems like, link breakage due to node failure, high energy consumption, delay and control overheads faced when using simple mobile ad hoc network routing protocols for the firefighter communication network.

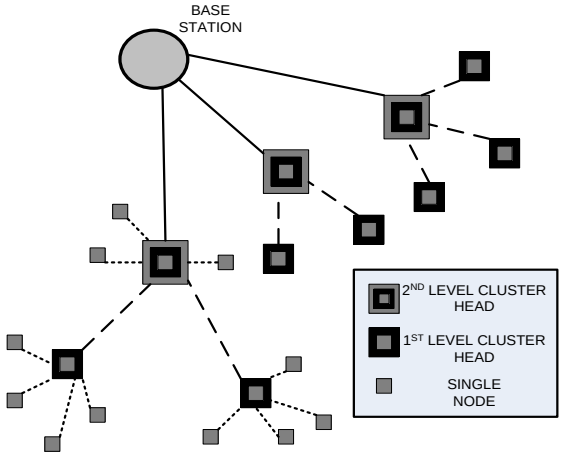


Figure 2. Hierarchical Network Model

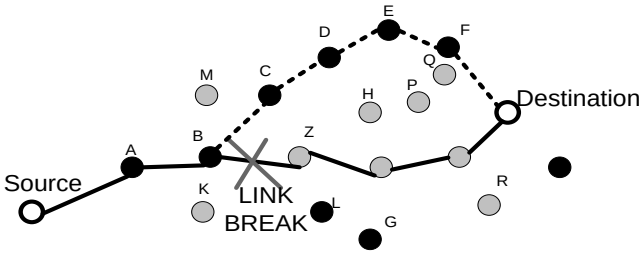


Figure 3. Link Repair

Table 1. Path Selection Variable Estimation

CH	No. of neighbor CHs	No. of hops to BS	Residual energy level	Path selection variable
A	3	5	1	**9**
B	4	4	2	**10**
C	3	4	1	**8**
D	4	2	3	**9**
E	3	2	1	**6**
F	4	1	2	**7**
G	3	3	3	**9**
H	5	2	8	**15**
K	4	4	8	**16**
L	4	3	3	**10**
M	3	4	7	**14**
P	5	1	9	**15**
Q	4	1	8	**13**
R	3	1	9	**13**
Z	4	3	9	**16**

Next we intend to implement the concept of Cluster Based Energy Aware Routing (CBEAR) presented in this paper and analyze the results

The link breakage due to mobility of the firefighters in the building is the next problem that needs to be solved. The presented idea can be extended to include solutions for link breakage due to mobility.

6. ACKNOWLEGMENT

This research was supported by Special Disaster Emergency R&D Program from National Emergency Management Agency (2012-NEMA10-002-01010001-2012), MKE (The Ministry of Knowledge Economy) under the CITRC (Convergence Information Technology Research Center) Support Program (NIPA-2012-H0401-12-1003), supervised by the NIPA (National IT Industry Promotion Agency) and Seoul R&BD Program (SS110012C0214831), MKE IT R&D programs (grant no. 10035610).

7. REFERENCES

[1] Thomas Hilebrandt and Marcel Kyas Heiko Will, "Wireless sensor networks in emergency scenarios: The FeuerWhere Deployment," in *SESP* , 2012.

2 . i - ie . i nithi . of nn n . r g, "Wireless and Ad Hoc Communications Supporting the Firefighter", 15th IST Mobile Summit, Myconos, Greece, June 4-8, 2006.

[3] F. heo eyre F brice V ois " se f-organization structure for hybri networks " "Ad-hoc Networks (technologies and protoco s)" Springer, vol.6, pp. 393-407, 2008 .

[4] o S Yoo Y. "An energy balancing LEACH algorithm for wireless sensor networks " In: Proc. 7th international conference on information technology: new generations (ITNG), Las Vegas, Nevada, USA; April 2010.

[5] Silvia Giorano, "Mobile Ad Hoc networks," in Handbook of Wireless Networks and Mobile Computing, New York: John Wiley & sons, Ch. 15, 2002

[6] nn M. et . "LifeNet: n -hoc Sensor Network and Wearable System to Provide Firefighters with Navigation Support" Ubi o p: De os Exten e bstr cts s tri 2007

[7] U. Witkowski, et . " -hoc network communication infrastructure for multi-robot syste s in is ster scen rios" In Proceedings of the International Workshop on Robotics for Risky Interventions and Surveillance of the Environment, 2008

[8] N. Aschenbruck, "Human mobility in MANET disaster area simulation - a realistic approach," in 29th Annual IEEE International Conference on Local Computer Networks, Germany, pp. 668 – 675, 2004

[9] Prasant Mohapatra n Srik nth " hoc Networks (techno ogies n protoco s)" Springer 2005

[10] J. Laprie, B. Randell and C. Landwehr A. Avizienis, "Basic concepts and taxonomy of dependable and secure computing," IEEE Transactions on Dependable and Secure Computing 1, 2004.

[11] H.W. Chong , M. Chan and K.F. Man S. Kwong, "The use of multiple objective genetic algorithm in self-healing network," Elsevier Science B.V., 2002.

[12] Ashwani Kush and Sunil Taneja, "Self-Healing and Optimizing Adhoc Routing ," International Journal of Engineering Innovation & Research, vol. 1, no. 1, 2012.

[13] Venkata C. Giruka and Mukesh Singhal, "A self healing On-demand Geographic Path Routing Protocol for mobile ad-hoc networks," Elsevier, Ad-Hoc networks, vol. 5, pp. 1113-1128, 2007

[14 W eize n h n r k s n B krish n. "Energy-efficient ro t ing protoco s for wire ess icrosensor networks" Proc 33rd Hawaii International Conferences on System Sciences (HICSS'00), 2000