SKIMA 2010

PROCEEDINGS Of SKIMA 2010

The Fourth International Conference on
Software, Knowledge and Information Management and Applications
Towards Happiness and Sustainable Development

August 25-27, 2010

Paro, Bhutan

ISBN 978-974-672-556-9

Organized with the collaboration of :

eLink

KANTIPUR Engineering College

上海交通大學 Shanghai Jiao Tong University

UNIVERSITY OF BRADFORD MAKING KNOWLEDGE WORK

UNIVERSITÉ LUMIÈRE LYON 2

Università degli Studi del Sannio

มหาวิทยาลัยเชียงใหม่ พ.ศ. ๒๕๐๗

BUDAPESTI CORVINUS EGYETEM

Royal University of Bhutan

College of Science and Technology

STAFFORDSHIRE UNIVERSITY

SKIMA 2010

Proceedings of the Fourth International Conference on Software, Knowledge, Information Management and Applications

"*Towards Happiness and Sustainable Development*"

August 25 – 27, 2010

Paro, Bhutan

http://skima.info

Edited by

Ouyporn Tonmukayakul

Manawin Songkroh

Pradorn Sureephong

College of Arts, Media and Technology

Chiang Mai University

Thailand

Editors:

Ouyporn Tonmukayakul
Manawin Songkroh
Pradorn Sureephong

College of Arts, Media and Technology, Chiang Mai University

ISBN 978- 974-672-556-9.
©SKIMA Organizing Committee

This work is subject to copy right. All rights are reserved, whether the whole or part of the material is concerned. Permission to make digital or hard copies of portions of the work for personal or educational use is granted without fee with acknowledgement that the copies are not made or distributed for profit or commercial advantage and those copies bear the full citation.

Welcome Note

Dear Colleagues,

On behalf of the conference committees, I would like to welcome you to SKIMA 2010: the Fourth International Conference on Software, Knowledge and Information Management and Application.

This conference is held in Bhutan, the land of Thunder dragon, where many people long for a visit but few can make it. It is hosted by the only university in Bhutan; Royal University of Bhutan with a warmly support from all eLink project members.

This event demonstrates the continuous effort to uplift SKIMA to an international reference level in the arena of ICT and Knowledge Management. It receives a great honor from the Government of Bhutan to approve this conference and allow ICT and KM scholars from at least 10 nations to experience and share their knowledge in this wonderful country. Moreover, His Excellency Minister for Education and the Secretary of Ministry of Information and Communications kindly accept the conference invitation for the opening ceremony and giving a keynote speech so we can learn how Bhutan move "Towards Happiness and Sustainable Development."

Two other leading keynote speakers from the Royal University of Bhutan and Freie Universität Berlin will present on the second day.

With the help of the 53 reviewers from around the world and by the use of the previous conference standard measure, fifty eight papers are accepted to be presented in this conference. They are classified into 11 different areas. These presentations will provide extra enrichment to the conference.

I would like to express my deep appreciation to all sponsors, the conference boards and committees, especially the local organizing team, for their generous support and kind assistance to make this conference viable. Moreover, my gratitude goes to all contributors who provide valuable elements for the success of this conference.

OuypornTonmukayakul, Program Co-chair

ISBN 978- 974-672-556-9

SKIMA 2010 Organizing Committee

CONFERENCE ADVISORY BOARD

Abdelaziz Bouras, University of Lyon 2, France
Keshav Dahal, Bradford University, UK
Andras Gabor, Corvinus University of Budapest, Hungary
Alamgir Hossain, Bradford University, UK
Nayyer Masood, Mohammed Ali Jinnah University, Pakistan
Rameshwar Rijal, Kantipur Engineering College, Nepal
Hongnian Yu, Staffordshire University, UK
Suiran Yu, Shanghai Jiao Tong University, China
Pasquale Daponte, University of Sannio, Italy

GENERAL CO-CHAIRS

Matteo Savino, University of Sannio, Italy
Nopasit Chakpitak, Chiang Mai University, Thailand

PROGRAM CO-CHAIRS

Ouyporn Tonmukayakul, Chiang Mai University, Thailand
Chowdhury Mofizur Rahman, United International University, Bangladesh

PUBLICITY CO-CHAIRS

Pradorn Sureephong, Chiang Mai University, Thailand
Yacine Ouzrout, University of Lyon 2, France

LOCAL ORGINIZING CHAIR

Nidup Dorji, Royal University of Bhutan, Bhutan

STAFF

Manawin Songkloh, Chiang Mai University, Thailand
Walaiporn Mahamai, Chiang Mai University, Thailand
Visaruta Thongsupapol, Chiang Mai University, Thailand
Staff of University of Bhutan, Bhutan

ISBN 978- 974-672-556-9

ICP Members

Ahmed, Khawza Iftekhar Uddin
United International University, Bangladesh
Alam, M. Shafiul
University of Dhaka, Bangladesh
Anwar, Farhat
International Islamic University of Malaysia, Malaysia.
Asghar, Sohail
Mohammad Ali Jinnah University, Pakistan
Atkins, Anthony
Staffordshire University, UK
Benmoussa, Rachid
ENSA Marrakech, Morocco
Beqqali, Omar El
University of Fes, Morocco
Bista, Sanat Kumar
Kathmandu University, Nepal
Brun, Alessandro
Politecnico of Milano, Department of Management Engineering, Italy
Cang, Shuang
Bournemouth University, UK
Carrington, Alison
University of Aberdeen, UK
Ceccarelli, Michele
Bio.Gem, Genetic Research Institute Gaetano Salvatore, Italy
Chandarasupsang,Tirapot
Chiang Mai University, Thailand
Cowling, Peter
University of Bradford, UK
Delattre, Miguel
University of Lyon 2, France
Derrouiche, Ridha
Ecole supérieure de Commerce de Saint-Etienne, France
Dulal, Khada Nanda
Kantipur Engineering College, Nepal
Eynard, Benoit
Université de Technologie de Troyes, France
Guizzi, Guido
University of Naples, Federico II, Italy
Gupta, Omprakash
University of Texas in Dallas, USA
Harnpornchai, Napat
Chiang Mai University, Thailand
Henry, Sébastien
University of Lyon2, France
Hossain, Akhter Syed
University of Dhaka, Bangladesh
Huda, Mohammad Nurul
United International University, Bangladesh
Iqbal, Shaukat
Mohammad Ali Jinnah University, Pakistan
Jiang, Dan
Shanghai Jiao Tong University, China
Kim, Jay
Hanyang University, Korea
Lamouri, Samir
ENSSMA Paris, France
Lbath, Ahmed
University of Grenoble, France
Meksamoot, Komsak
Chiang Mai University, Thailand
Moalla, Néjib
University of Lyon 2, France
Mostefai,Sihem
University of Constantine, Algeria
Murino, Teresa
University of Naples Federico II,Italy
Nadeem,Aamer
Mohammad Ali Jinnah University, Pakistan
Neubert, Gilles
Ecole supérieure de Commerce de Saint-Etienne, France
Noriega, Len
Staffordshire University, UK
Nota, Giancarlo
University of Salerno, Federico II, Italy
Pernelle, Philippe
University of Lyon 2, France

ISBN 978- 974-672-556-9

Rachuri, Sudarsan
NIST Maryland, USA
Rahman, S. M. Monzurur
United International University, Bangladesh
Riane, Fouad
Fucam, Mons, BE
Sakib, Kazi Muheymin
University of Dhaka, Bangladesh
Santillo, Tina
Department of Production Engineering, University of Naples, Federico II, Italy
Sarowar, Hasan
United International University, Bangladesh
Sekhari, Aicha Seklouli
University of Lyon 2, France
Shaheed, Hasan
Queen Mary College, University of London, UK
Sharma, Nabin Chandra
Kantipur Engineering College, Nepal
Tassabehji, R.
Bradford University, UK
Terzi, Sergio
University of Bergamo, Italy
Yan, Xiu-Tian
University of Strathclyde, UK
Zhang, Yong
Shanghai Jiao Tong University, China
Zimeo, Eugenio
University of Sannio, Department of Engineering, Italy

ISBN 978- 974-672-556-9

Table of Contents

ISBN 978- 974-672-556-9

ISBN 978- 974-672-556-9

ISBN 978- 974-672-556-9

Session 7: Supply Chains Management and Applications

Session 8: Network Management and Collaboration

ISBN 978- 974-672-556-9

Session 9: Software Agent and Soft Computing Applications

Session 10: Education and Applications

ISBN 978- 974-672-556-9

Session 11: Identifications. Modeling and Robotics

ISBN 978- 974-672-556-9

Session 1: Information Systems

ISBN 978- 974-672-556-9

A Survey of Data Level Conflicts in Database Integration

Ghulam Ali Mirza, Nayyer Masood, Sohail Asghar
Department of Computer Science,
Muhammad Ali Jinnah University, Islamabad, Pakistan
alimirza@jinnah.edu.pk, nayyer@jinnah.edu.pk, sohail.asghar@jinnah.edu.pk

***Abstract*—Database integration plays a significant role in different application domains where data has to be accessed from multiple, distributed, autonomous and heterogeneous databases. Semantically similar data residing in different databases is prone to have schematic and data conflicts. For integrated access from multiple databases, these conflicts have to be resolved. Data conflicts come into play when database schemas have been merged after the resolution of schematic conflicts, like data value conflicts, data precision conflicts etc. Without identification and resolution of data conflicts, the data from source databases cannot be retrieved and correctly interpreted. The focus in this paper is explicitly on these conflicts; what they are, how they create problems in data access and how they have been treated in literature. The findings of this survey are; (1) a lot of work has been done on schematic conflicts but the data conflicts have been relatively ignored, (2) schematic and data conflicts are mixed with each other which undermines the importance of later, (3)the lack of attention in literature leaves many data conflicts unidentified. This survey is an attempt to highlight the importance of study about data conflicts that will ultimately help in performing more accurate database integration.**

***Index Terms*—Conflicts, Data Integration, Data Merging, Heterogeneities, data conflicts**

I. INTRODUCTION

Database integration is a process that is performed to achieve the unification of data from different database [1]. It is an important field of computer science that has significance in industry as well as in academia. Accessing heterogeneous data repositories is a requirement encountered in various applications [4]. Fig. 1 shows a heterogeneous data access setup in a multidatabase system framework that is mainly based on the database integration process. Database integration has two main steps, schema integration (SI) and data integration (DI). The SI process results an integrated schema also called the global schema. The user initiates a query on the global schema considering the result of his query will be obtained from the data underlying the (global) schema. The global schema, as a matter of fact, does not contain any data. The user query is transformed into local queries which are passed on to local/component databases where these queries are executed and results from local sites are sent back to the global layer where the DI is performed generating the merged data that is presented to the global user.

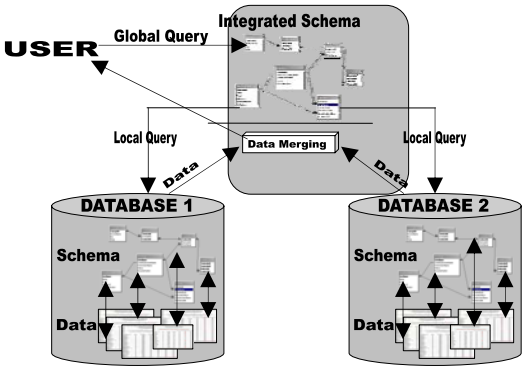


Figure 1. Heterogeneous data access

The major challenge in database integration is the handling of heterogeneities (also termed as conflicts) among the component databases [3]. These conflicts create problems in SI and DI phases of database integration. On one side, identification and merging of corresponding schema elements is hampered during SI and on the other side, the merging of data fetched from the local databases becomes problematic during DI process [16].

A lot of work has been done on the issue of heterogeneities among component databases. However, not too much work can be found focusing specifically on data level conflicts [2,4,8,9,10]. They are generally treated during the SI process along with the schematic conflicts. A deeper look into the issue of heterogeneities reveals that the schema level conflicts and data level conflicts need to be studied separately; reasons being

- They occur at different times/processes in database integration
- Not all data level conflicts are handled during the SI

As is evident from the second point, some of the conflicts are handled automatically/implicitly during the SI process. For example, missing attribute value, granularity of the information unit, heterogeneity where an element in one schema (like stdName)

matches more than one elements in other schema (like stFName, stLName). The schema merging approaches [2,9] concatenate the two elements before merging them with one element of the first schema. However, there are certain types of conflicts that cannot be managed during the SI process, for example, unit difference, like weight is modeled in kilogram in one database and in pounds in another database. There are many other such conflicts; some of them have already been identified in literature and some were undiscovered.

The objective of this study is to focus on this, relatively ignored, issue of database integration. The previous work has been analyzed with the specific perspective of data level conflicts, some new conflicts have been identified and finally a comprehensive taxonomy of conflicts has been presented. The work in this paper will highlight the importance of study of data level conflicts that will ultimately lead to proposing solutions for these conflicts.

The paper has been structured as follows: section II discusses the data level conflicts; what they are, what different types are and how they impact the integrated data access. Section III presents the literature survey on data level conflicts. Section IV contains a critical analysis of the work done so far on data level conflicts. Conclusion and future work is discussed in section V.

II. The Data Level Conflicts

Data Integration (DI) has been mentioned in the context of database integration in the previous section. This is one major application area in which DI is required but this is not the only one. Many applications have been mentioned in literature where DI is required for application specific purposes. Application domains mentioned in [21] focuses on highlighting the importance of schema matching but all these domains ultimately require the data integration. Like in schema integration, after merging the component schemas, data fetched from the local databases as a result of global query has to be merged that involves DI. In data warehouses, data extracted from a data source is to be merged with that already present in the data warehouse. It involves finding the matching elements between the two; once they have been identified the data from the new data source is to be merged with that of data warehouse that involves DI. E-commerce has been mentioned as another application domain in which trading messages are exchanged that could differ from each other in syntax and schema. In order to present messages in the required format of different trading partners, the messages have to be translated which partly involves schema matching. After translating a message the underlying data has to be presented in the required (by a particular partner) format. The required format of the data, like data type, range, allowable values etc., could be different from the source data. For this purpose source data has also be transformed into the required format what is, in a sense, is application of DI. The semantic query processing is a run-time scenario where user initiates an SQL query to an ad-hoc database. The differences between the concepts presented by the user in the query and the way they have been represented in the database schema could be different. Resolving these differences involves SI, however, resolving the differences between the format required by the user in the query and the way they have been stored in the database is the task of DI process. Another application of DI is in “a comparison shopping website that aggregates product offers from multiple independent online stores” [21]. In this case, the comparison website will pick the matching data items in the product catalogs of the independent sites, for example, productId in one may match with the prId and storeId (combined) in the other. After establishing the matching between the data items the underlying data will have to be merged to be presented together for the sake of comparison by the end user.

Previous paragraph mentions some of the applications of DI; many others can be picked from the literature. Next, we are going to discuss the data level conflicts, their examples and the way they have been treated in literature to have a clear understanding of what they mean and how they can cause problem in proper understanding of the merged data

- **Data Value Conflicts:** One situation covered under this category is encountered when corresponding elements use enumerated values but the bases on which these enumerations are defined are different. This leads to *enumeration conflict***.** For example, the value “Adult” in one database means a person with age greater than 16, while same value in the corresponding attribute in other database means age greater than 18. The values of these two attributes when shown together will be reflecting different/conflicting situation. Another type of data conflict that falls in this category is *data pattern conflict*, which means that data stored in corresponding elements is of different patterns. These patterns are defined as per the standards of the organizations. For example, the pattern to store phone number in one database could be as (DDD) DDD-DDDD; in which parenthesis denotes the country code whereas another database can store the telephone number in the DDD-DDD-DDDD format. This conflict can cause problem in querying as same information is being represented according to different patterns.
- **Data Representation Conflicts:** The same data

is represented using different formats in different databases. One such conflict is **format conflict**. A typical example of this conflict different formats of storing date, like "dd/mm/yy" or "mm/dd/yyyy". Difference in the format of storing time is another example of this conflict. Same value of time can be stored as "22:30" or "10:30 PM" or as "10:30:00". Another conflict in this category is use *of different signs*, i.e., different databases use different signs to mean same thing. For example, two attributes storing salaries of employees may contain value 20,000 and 20.000 both representing twenty thousand salary. The former is the English style and the later one is in German style. Then there is a *different data types* conflict where corresponding elements contain data of different data types. Like, ID can be defined using Numbers or String. Similarly, there is a conflict of *different lengths* where corresponding attributes have same data types but different lengths. For example, address in one database can be of type char[30], and in another it is of type char[40]. Merging them into an attribute of length 40 will add spaces to the one with length 30 and if merged into a length of 30 it would truncate 10 right most characters from the attribute with the length 40.

- **Data Unit Conflicts**: These conflicts occur when corresponding attributes store data with different units. Example of these conflicts can be attributes storing heights (in cm or in inches), weight (kilogram or lbs), currencies (Rupee or Euro), liquid measure (liters or gallons), time duration (minutes or hours) and many other such examples. Such attributes if presented without resolving the conflicts will not portray the proper picture of the real situation.
- **Data Precision Conflicts:** Numeric data in corresponding attributes could be of different *floating point precision*. One attribute might be storing fraction part of floating point number up to 2 places whereas in the other attribute precision can be up to 4 digits. This difference in precision if not resolved shows an inconsistent level of accuracy. Another conflict of this type could be due to *grouping conflict***.** This conflict occurs when corresponding attributes define different groups within the same set of values. For example, two attributes modeling the grades of students. One might be storing 5 grades [A, B, C, D, F] where A means 90+ marks and below 50 is F. Other attribute may have 11 grades [A, A-, B+ B, B-, C+, C, C-, D+, D, F]. When merging such data we have to resolve this discrepancy in grading to present the data in a meaningful way.
- **Granularity of the Data:** Data is represented at different levels of granularity in different databases. One example is *Attribute vs. Entity conflict*, when information in one table is presented as an attribute while in other it is represented as complete table. For example, address of employees might be stored as an attribute in one database whereas in another database it might be stored as a complete table in other database with a link to the table storing the other data of employees. Another conflict of this type is *one to many Attributes conflict* where same data is stored in one attribute in one database and in other database it is stored in multiple attributes. For example, name attribute storing the full name of Person in one database whereas another database stores first name and last name separately in two attributes. Same situation can happen with address, date or time also.
- **Attribute Constraint Conflicts**: Constraints or checks are defined on attributes as per the business rules of the organization. One type of constraint conflict is *default value conflict*, which covers two situations. One where default value for an attribute has been defined in one database whereas it has not been defined in other. Secondly, different default values have been defined for corresponding attributes. For example, the default value for qualification attribute in one database may be 'BS' and in another database it could be 'MS'. Another conflict is *Null/Not Null conflict*, which means that an attribute is allowed to have Null value whereas its corresponding attribute in anther database is not allowed to have Null value.

The data level conflicts mentioned above are the ones that have been discussed in literature. The list is not exhaustive, there are still many conflicts that have not yet been discussed in literature. Next section presents the literature review regarding the data conflicts.

III. LITERATURE REVIEW

The focus of the literature review is to discuss the data level conflicts that have so far been discussed in literature and the resolution approaches defined for these conflicts.

Park and Ram [11] presented their research about the concepts of having interoperability between distributed and heterogeneous information systems information systems. Many Semantic conflicts are proposed by them. They developed a framework that can measure the level of interoperability between two information systems. This framework can detect and resolve the conflicts. Their developed system resolves the schema and data level conflicts. Approach is tested on systems that can be integrated.

Apart from the schema level conflicts they presented few of the data level conflicts, these conflicts are;

- Data-value conflicts
- Data representation conflicts
- Data-unit conflicts
- Data precision conflicts

This work presented by the researchers is more on schema level conflicts. Although their systems resolves the conflicts of between heterogeneous databases but the conflicts they provided are of very basic level, they didn't presented conflicts of heterogeneous databases in detailed.

Ram and Park [17] also proposed a formal structure of a common ontology called Semantic Conflict Resolution Ontology (SCROL). This ontology basically describes the problem of lack to semantic richness and limited domain generality in the conventional approaches of integration of heterogeneous database systems. This SCROL provides a systematic way of automatically detecting and resolving semantic conflicts in heterogeneous databases. SCROL basically captures the following data level conflicts;

- Data value conflict
- Data representation
- Data unit conflict
- Data precision conflict
- Known data value reliability conflicts
  (Data present in different databases may be subject to data reliability i.e., measurement of error, measuring instruments, precision of measurements, topological properties, and treatment of time dimension)
- Spatial domain conflict
  (Specifications of geographic regions or objects are "differently" but "legally" defined by different people)

Ontological system is presented in this research; defining the mapping between the ontology is a complex task that why less concentrations on data level conflicts are given in this system, apart from the already defined classification of data level conflicts two more types are proposed that is; Known data value reliability conflicts and Spatial domain conflict.

Peristeraset al [6] presented the semantic interoperability issues in Pan-European Public Services (PEPS) and Pan-European E-Government Services (PEGS). They examined the interoperability issues that may arise when a citizen of one MemberState requests a public service from the public administration of another MemberState. They identified and classified the semantic conflicts according to the Governance Enterprise Architecture (GEA) object model concepts. GEA model is a semantic gateway for their system that can resolve the issues at a pan-European level. The basic theme of their research was to identify the schema and data level conflicts that may occur in their system. They identified few of the data level conflicts, these conflicts includes;

- Data value concepts
- Data representation conflicts
- Data unit conflicts
- Data precision conflicts
- Granularity of the information unit

This paper is basically presents their work about E-Government Service. Identification of data level conflicts is basic and generic. Expect the definition of the conflict, detailed are not provided.

According to Naumann and Häussler [10] when integrating the autonomous data sources quality of data is affected and the cause of the decreasing in quality is data conflicts that comes after the integration. For this problem they defined the resolution function for merging the conflicting data. They used the common queries through grouping, aggregation, partition and joining for merging the relational data sources. Their quires use the resolution function for data transformation and or for view integration of data. They also developed a framework that basically deals with schema mapping and data transformation. They handled three types of data level conflicts in their framework, that is;

- Null attribute value
- Missing attribute
- Conflicting attributes (provide by different data sources)

They presented the formal resolution function for null value attribute and other resolution functions are defined using the SQL functions.

When transforming the data sources many conflicts can be occur, presented conflicts are not detailed for transforming system.

Naiman and Ouksel [7] proposed a classification of semantic conflicts along with two dimensions, naming and abstraction, they added the third dimension that is, level of Heterogeneity, to assist in the schematic mapping between two databases. They claimed that classification provides a systematic representation of alternative semantic interpretations of conflicts when integrating the databases. They also claimed that their classification is sound and minimal. In their article they presented the problem of data representational conflict in the term of having different length of the attribute value; they also mentioned four of the data level conflicts;

- Data scaling conflicts
- Data precision conflicts
- Default value conflicts

- Attribute integrity constraint conflicts

Classification does not seem to be sound for the data level conflicts because they did not provide the details and did not defined any other conflicts beside from the standard conflicts.

Sheth and Kashyap [3] provided a semantic taxonomy of schema and data level conflicts they also discuss possible semantic similarities between two objects that have various types of schematic and data conflicts. Their taxonomy includes the following data level conflicts;

- Naming Conflicts
- Data Representation Conflicts
- Data Scaling Conflicts
- Data Precision Conflicts
- Default Value Conflicts
- Attribute Integrity Constraint Conflicts

Naming conflict that they explained in their research basically deals with the schema level conflicts, the example they provided is also more relevant to schematic conflict, rest of the conflicts are standard.

According to Dong and Naumann [8], conflicts can arise because of incomplete data, erroneous data and out-of-date data. They distinguish two kinds of data conflicts; uncertainty and contradiction. Uncertainty is conflict between a non null value and one or more null values that are all used to describe the same property of real-world entity and uncertainty is cued by missing information. So they defined the defined two types of conflicts;

- Duplicate Record Detection
- Null Value of Attribute

According to them the data fusion can be performed using SQL Join and Union function at tuple level. Data conflicts are not just the duplication and null value and to obtain the integration more steps are required than Union and Join operation.

According to Naumann, Bilke, Bleiholder, and Weis [13] data fusion is performed in tree steps schema, tuple and then value-level. They developed a tool, called HumMer, that matches the schema and detects the duplication of data and finally fusion of data is performed. This tool is guided by SQL queries, the main focus of this research is on duplication detection of record and they also provided the formal notation of duplicate detection. Representational conflict of data is solved using SQL queries

Conflicts defined by many authors are not discussed for their tool HumMer that fusion the data.

Bleiholder and Naumann [12] also worked on conflicts handling strategies. These strategies are cry with the wolves, roll the dice, meet in the middle and keep up to date; they have implemented these strategies in their system, description of these strategies defines that it would try to resolve the conflicts but types of conflicts are not discussed by the researchers.

In [2] [5] [9] [19] schema matching or schema merging approaches are discussed. These all techniques discuss the schema level conflicts and related issues but they do not cover data level conflicts and related issues.

IV. Critical Evaluationof Data Level Conflicts

Data level conflict are a must to exist in data access from multiple and autonomous databases. This is because every database is created with organization specific requirements and preferences that introduce conflicts when compared with the database of another organization. Identification and resolution of data conflicts is a must for proper integration and understanding of the merged data. We see little focus on data conflicts in the relevant literature because more attention is paid to schema level conflicts which are also a must to resolve for unified access of data from multiple resources. We have evaluated the work done on data level conflicts in three points that we have established on the basis of activities that are required for resolving the data conflict. First activity is the identification (mention) of the conflict; it tells us the particular conflict(s) that have been discussed in a particular paper. Second point is that how a conflict has been represented (defined) in the literature. Conflicts have been given either the textual definition or a formal one. Third point is resolution, which means whether an approach gives technique/strategy to resolve data conflict(s). This review has been presented in Table 1.

As can be seen from the above table, data level conflicts have not been given much attention in literature. The reason that we can analyze from our survey is that generally it is believed that data level conflicts will be resolved with the schema conflicts. This, as a matter of fact, is not true in all cases, that is, in certain cases data level conflicts will be resolved through schema level conflict resolution operations but not all. That means we need to study the data level conflicts in depth and propose the resolution operation for them.

TABLE 1: TREATMENTOF DATA LEVEL CONFLICTSIN LITERATURE

Ref No.	Identification	Representation	Resolution
[18]	Duplicate Record Detection, Null Value of Attribute	Textually	Using Join and Union, Tuple Level
[9]	Schema Integration Approach, No Date level conflicts discussed in Schema Integration	No	No
[14]	Discussed the Dynamic data integration problem for the information of different web sources. Identified Data Integration as Separate process	No	No
[6]	Data value concepts, Data representation conflicts, Data unit conflicts, Data precision conflicts, Granularity of the Information unit	Textually	No
[19]	Schema Mapping Level Issues Discussed, No identification of Data Level Conflicts in Schema Mapping Process	No	No
[15]	Authors proposed a new problem called *Veracity*, i.e. conformity to truth, which studies how to find true facts from a large amount of conflicting information on many subjects that is provided by various web sites.	No	No
[8,12]	Duplicate Record Detection, HumMer (Integration System)	Textually	Discussed Resolution Strategies, Textual Solves the Representational Conflicts for Tuple Level
[11]	Data value, Data representation, Data –unit, Data precision	Textually	No
[10]	Null Value Attribute Conflict	Formal Notation	Yes (Formally)
[7]	Data representation conflicts (data length), Data Scaling conflicts, Data precision conflicts, Default value conflicts Attribute integrity constraint conflicts	Textually	No
[3]	Data representation conflicts, Data precision conflicts, Default value conflicts, Attribute integrity constraint conflicts	Textually	No
[20]	Duplicate Record Detection	No	No

V. CONCLUSION AND FUTURE WORK

Data level conflicts are a must to resolve for unified access of data. We have analyzed the literature related to database integration with a focus on data level conflicts and it is our finding that very little attention has been given to data conflicts. Most of the work on heterogeneities has focused on schema conflicts as they are to be dealt during schema matching, a process prior to schema or data merging. In this study we have identified the data level conflicts that have been mentioned in the literature along with the treatment given to them. The treatment has been studied on the basis of three points that we have established are necessary for resolving the data conflicts. Next, we are in the process of identifying the data conflicts that have not yet been identified in the literature and build a complete taxonomy of such conflicts. We are also working on the formal description of the data conflicts and approaches to resolve them.

REFERENCES

[1] M. Lenzerini, "Data Integration: A Theoretical Perspective," in *PODS*. pp. 233-246, 2002.

[2] H. Do and E. Rahm, "COMA - A system for flexible combination of schema matching approaches," presented at the 28th Intl. Conference on Very Large Databases (VLDB), 2002.

[3] A. Sheth and Vipul Kashyap, "So Far Schematically yet So Near Semantically," Bellcore, 444 Hoes Lane, Piscataway, NJ 08854-418, 1992.

[4] M. Gagnon, "Ontology-based Integration of Data Sources," presented at the 10th International Conference, Quebec, QC, Canada, 2007.

[5] Z. Linková, "Ontology Based Schema Integration," in *Proceedings of SOFSEM*, pp. 71-80. Prague, 2007.

[6] L. Peristeras, "Semantic interoperability conflicts in pan-european public services,"in *Proc.of the 15th European Conference on Information Systems*, 2007.

[7] C.E. Naiman and A.M. Ouksel, "A classification of semantic conflicts in heterogeneous database systems," *Journal of Organizational Computing*, vol.5, no.2, pp. 167-193, 1995.

[8] X.Dong and F. Naumann, "Data fusion – resolving data conflicts for integration," *PVLDB*, vol. 2, no. 2, pp. 1654-1655, 2009.

[9] A. Radwan, L. Popa, I. R. Stanoi, and A. A. Younis, "Top-k generation of integrated schemas based on directed and weighted correspondences," in *Proc. of SIGMOD Conference*, pp.641-654, 2009.

[10] F. Naumann and M. Häussler, "Declarative data merging with conflict resolution," in *Proc. of the International Conference on Information Quality (IQ 2002)*.

[11] Jinsoo Park, and Sudha Ram, "Information systems interoperability: What lies beneath?"*ACM Transactions on Information Systems*, vol. 22, no. 4, pp. 595-632, October 2004.

[12] J. Bleiholder and F. Naumann. “Conflict handling strategies in an integrated information system,”presented at the International Workshop on Information Integration on the Web (IIWeb), Edinburgh, UK, 2006.

[13] F. Naumann, A. Bilke, J. Bleiholder, and M. Weis. “Data fusion in three steps: Resolving inconsistencies at schema-, tuple-, and value-level,” *IEEE Data Engineering Bulletin*, vol.29, no.2, pp.21–31, 2006.

[14] L. Berti-Equille, A. D. Sarma, X. Dong, A. Marian, and D. Srivastava, “Sailing the information ocean with awareness of currents: Discovery and application of source dependence,” in *Proc. of CIDR*, 2009.

[15] X. Yin, J. Han, and P. S. Yu, “Truth discovery with multiple conflicting information providers on the web,” in *Proc. of SIGKDD*, 2007.

[16] Y. Halevy, A. Rajaraman, and J. J. Ordille, “Data integration: The teenage years,” In *Proc. of VLDB*, pp. 9–16, 2006.

[17] S. Ram and J.Park, “Semantic conflict resolution ontology (SCROL): An ontology for detecting and resolving data- and schema-level semantic Conflicts,”*IEEE Trans. Knowl. Data Eng*., vol.16, no. 2, pp.189–202, 2004.

[18] J. Bleiholder and F. Naumann, “Data fusion,” *ACM Computing Surveys*, vol.41, no.1, pp.1–41, 2008.

[19] P. A. Bernstein and S. Melnik, “Model management 2.0: manipulating richer mappings,” In *Proc. of SIGMOD*, pp. 1–12, 2007.

[20] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, “Duplicate record detection: A survey,” *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol.19, no.1, pp.1–16, 2007.

[21] S. Melnik and H. Garcia-Molina, “Similarity flooding: A versatile graph matching algorithm and its application to schema matching,” presented at the 18th Intl. Conf. on Data Engineering (ICDE), San Jose CA, 2002

Classification Framework for Digital Preservation Platforms

An Evaluation Approach for Digital Preservation Platforms in Long Term Preservation Scope

Prem Timsina, Fei Teng, Néjib Moalla, Abdelaziz Bouras
Laboratory LIESP, University Lumîre Lyon 2
160 Boulevard de l'Universiť,
69676, Bron, France
Prem_timsina28@yahoo.com, Fei.Teng@univ-lyon2.fr, Nejib.Moalla@univ-lyon2.fr,
abdelaziz.bouras@univ-lyon2.fr

***Abstract*—Recent years, long term digital preservation has become a crucial mission to enterprises. Many techniques on long term digital preservation are proposed, and among all the techniques, enterprises have to identify which techniques or tools would be sufficient according to their own situations. Digital repository platform is the basics and the significant part of long term digital preservation. Thus in this paper, we propose a evaluation approach "Weighted Mean" for digital preservation platform. We use this approach to evaluate the most widely accepted three open source digital repository platforms (i.e. DSpace, Fedora repository and EPrints) and present their actual technological strength and weakness on different criteria concerning long term digital preservation.**

Index Terms- digital preservation platforms; evaluation; DSpace; Fedora; EPrints

I. INTRODUCTION

A research about digital information has conducted by IBM, Microsoft and Tessella Technology. They showed the volume of digital information held by companies will rise 25-fold over the next decade, from an average of 20 Tbytes to 500 Tbytes.[1] Thus, the archiving of electronic information has become a crucial mission of current world. Without suitable preservation and strategic planning process, digital information may be inaccessible over long term. In production industry, the product life cycle (e.g., aircraft 50 years) is much longer than the life cycle of computing hardware, storage media or used applications. Therefore, without sufficient Long Term Knowledge Retention (LTKR), companies will encounter with the technical and commercial problems after long time. Therefore, long term digital preservation has become the unavoidable circumstances for every competitive enterprise.

Besides simple bit preservation, LTKR also cares about logical preservation. In the bit preservation, it is necessary to preserve the files of digital data. On the logical level it is obligatory to preserve, in the long run, digital object persistence to technological change, and document usability [2]. In real case, when new technology arrives, the oldest will be discontinued. This makes the digital preservation in very complicated position, and probably unable to retrieve the information relying on obsolete technologies.

For sufficient long term digital preservation, we must be aware of file-format, underlying software and hardware, and the volatility of host organizations. Furthermore, we should be sure of authenticity and integrity of the stored content. To establish such a digital preservation system, we must get supports from the digital preservation tools, in other words, digital preservation platforms. The digital preservation platforms provide different functionalities and services. For the scope of LTKR, we must identify and clarify the functionalities that not only support digital preservation, but also support long term requirements. The purpose of this work is by introducing the long term digital preservation needs and requirements, to evaluate the capacity of open digital preservation platforms and their functionalities, in order to implement the LTKR system in certain platform(s).

II. RESEARCH PROBLEMS

The whole task of long term preservation is worthless if we can't address the needs of end users in the future. Although we try to identify all the additional information needed to be preserved along with the data bits, actually it is difficult to find out all the information that the future users want from the stored content. However, at least we could define the basic functional needs of future end users as follows:

- The end user should be able to access the preserved digital document.
- The content should be executable, i.e. the host machine should be able to render the document in its original environment.
- The end user should be able to interpret and understand the content of the digital document.

ISBN 978- 974-672-556-9

- All the above should be possible without the end user experiencing any errors or complexity [3].
- The preserved document should be authentic in the sense that it is the same document that was preserved and both its content and behavior have remained unchanged.
- Metadata should accompany the digital document instructing the future end user on how to execute the document, as well as explaining the document content, its intended behavior and a description of the software required to run it

Also the information should remains undead over the infinite period of time. The main barrier preventing achieving such goal is that the period of time may be much longer than the life time of any storage media, formats of the file and the hardware and software component. Therefore, none of the existing media, hardware or software can be trusted fully by the data producers. But any how the following minimal requirements should be fulfilled while establishing a preservation system.

- Media, software hardware, and file encoding formats must flow over the time and it should be replaced before it fails, or become obsolete.
- The system mustn't have even individual point of failure; it must be designed to persist every single failures of the system. It should also tolerate the simultaneous failures of many components of the system.
- The system should check the data consistency at the interval frequent enough to keep the errors below the acceptable levels [4].

III. Proposed Methodology

A. Techniques for Digital Preservations

The huge quantity of the preservation technique is proposed, but most of them can be interpreted into two headings [5]:

1) *Preservation of technological environments:*

Preserve technology and Emulation of technology. It involves on preserving the whole circumstances needed to retrieve and interpret the digital information. The first approach guidelines us to preserve all the related hardware and the software, and decoding techniques for retrieving the digital data. It is the simplest but most unfeasible technique because for the long term it is unrealistic to guarantee that all the hardware and software components are available and still usable. The second approach referred as *Emulation of Technology*, says us to develop the emulator of the outdated software and hardware which runs on the up to date system environments. But the problem to create the emulator is; we should know the details of the original environments used for retrieving the information.

2) *Persistence of technological obsolesces of data format:*

Migration and Encapsulation. The first approach, namely data migration tells us to periodically transfer the information from one software/hardware to the new up to date environment. It can also be done to transfer diverse type of format to standard format which is easy and economical for preservation. For e.g., all doc file can be transfer to the pdf file, as it is used universally. The second approach, referred as the encapsulation is based on idea that the preserved content should be self describing. The wrapper is build in which, together with the digital object, some side information, like the documentation about the original data format, and all necessary information to provide access to the object, are stored. But this strategy dramatically increases the record information and may not be suitable when original data format is not well documented.

B. Repository for Digital Preservation

The digital preservation repository serves the stewardship mission of the institution by providing a single shared solution for the preservation, management, and controlled dissemination of digital collections that support research, teaching, and learning. There may raises the question, how digital repositories differentiate from other data collection? Here we present minimal requirements of repositories which make it different from other data collection [6]:

- The repository architecture manages content as well as metadata, unlike data collection which keeps only data.
- The repositories offer the basic requirement of the searching and browsing.
- Content is created in the system either by content created owner or the third party system.
- The repositories must be sustainable and trusted i.e. it must guarantee the high level of protection from the unwanted attack and it must be persistence to disaster.
- Again it must be well-supported and well-managed.

Now, from the functionalities of the repository we can see that how repositories covers some requirements of the long term digital preservation.

C. Proposed evaluation approach

Here we present our key purpose of the evaluation of digital preservation platforms:

- To find out the design, architecture and the implementation details of short listed candidates.

- To understand the functionalities of the repositories according to the given criteria.
- To employ the members of the open source repository where needed.
- Finally, to find out the perfect match repository that can address our institution needs.

We could notice in the purpose of evaluation, the criteria which would lead the evaluation. Many needs can be characterized by the functional requirements. However, others, which relate to extensibility of the software, sustainability of the user/developer community and usability, are equally important. These needs are sometimes conflicting to each other resulting in difficulty to find out coherent criteria on evaluation of the system. Anyhow, here we have identified the following important criteria:

- Ease of system deployment and testing
- Versioning
- Scalability
- Interoperability
- Security
- Archiving and Database management
- Submission
- System Configuration
- Working with the code
- Archival and administrative concerns
- Globalization
- Searching and Browsing
- Community and support

This evaluation methodology of the repositories is primarily depended on the following methods;

- Deploying and testing the system.
- Examining of the release documentation of repositories system.
- Examining the papers from the user community i.e. from research works, reports and papers as well as wikis, institutions and projects websites.
- Analyzing the earlier comparison of the repositories.
- Analyzing the mailing list of the repositories community.

Each selected criteria is given the important rating when evaluation is done on the different repository. The headline criteria are also broken into sub-criteria for the transparency of the heading and each carry the important marking. Since each sub-criterion is marked using the range, these marking can be display as:

- 0: worst or/and feature doesn't exist.
- 0.25: carry poor support or /and it can be accomplished with significant labor.
- 0.5: reasonable but needs the adaptation to reach the desired condition or/and still need some effort for completeness.
- 0.75: good but still needs some minimal effort
- 1: as desired or /and need virtually no extra effort.

We intend to use this evaluation approach to classify different repository platforms, in order to select efficient digital preservation platform for implementation of LTKR system.

Among the different preservation platforms used around the globe, we have taken the systems as the short listed candidates which match the following criteria:

- Open source
- Massively used
- Strong community and support
- Development history and future forecast of the system

At last, we decide the short listed candidates are DSpace, EPrints and Fedora. In next section, we will introduce how the evaluation approach is performed, with detail description of our case study on the selected three digital preservation platforms.

IV. Detail Evaluationof Digital Preservation Platforms

A. Ease of System Deployment and Testing

The system is no-longer wanted if it is technically expensive and unrealistic. Here we analysis the technical complexity of the system.

TABLE I. Comparisionon System Deployment

	Platforms		
	DSpace	***Fedora***	***EPrint s***
Software requirement	0.75	0.5	0.75
Copy	0.75	0.75	0.75
Repository installation steps	0.5	0.25	0.5
As a total	0.67	0.5	0.62

1) DSpace

DSpace is the out of box system, once installation is finish, we are done. For the general purpose, we don't need to write any module or Plugins. It doesn't have hardware overhead, it is sufficient to have hosting machine capable of running apache server and the database. DSpace needs the following software [7]: Java 1.4 or higher, Apache Ant 1.5 or higher, Postgre7.3/oracle9i or higher, and Jakarta tomcat4.x or higher.

2) Fedora

It is complicated system; we should install Fedora front end application for the full functioning. Fez, which is front end of Fedora, needs common software components of Linux, Apache, Mysql and PHP (LAMP). It requires a pre-installation of "tidy" and GD PHP extensions. It also needs Image Magick, Graphiz and JHOVE software to enable it to

operate. Fedora needs the following software [8]: Apache, PostgreSQL / Oracle, Java, and additional front end application. The hardware requirement can be considered similar to DSpace.

3) EPrints

It can be also called out of box system. But we should install additional software such as Catdoc 0.94.2 for indexing the word file by the EPrints. The EPrints works well in UNIX system. But for Microsoft it is still in the phase of development. It is tested in the xp and vista. But no instance till date has been encountered in the window7. It requires the following software [9]: MySQL, Apaches, Active Perl, optional software for e.g. Ghost Script 8.6 is used to convert Ps and PDF format, Java.

B. Versioning

Allows the repository to keep older versions of metadata and files.

TABLE II. COMPARISIONON VERSIONING

	Platforms		
	DSpace	***Fedora***	***EPrint s***
Notice of similarity of submission	0	0	1
Keep Versions	0	1	1
Acquiring data and metadata for new version from old version	0	1	1
Log of versioning	0.5	1	1
As a total	0.125	0.75	1

1) DSpace

There is not any inbuilt feature of versioning in the DSpace. But user can manually keep the log by adding information on the metadata.

2) Fedora

We can declare any item in the system as versionable or non versionable. For the versionable item, Fedora shows the different versions of the same information packages in the timeline. As in the EPrints the Fedora system can't identify the similarity of the new item compare to the existing item inside the system.

3) *EPrints*

For ingesting the newer version of the item, the system makes exact copy of the older version. Then we change the metadata as well as the data. When we browse any item, the system provides links of all the versions of that specific file. Another important feature is that system can identify, if new ingesting item resemble with existing item in the repository.

C. Scalable

At the current context the volume of the digital archives may be the small. The balance will change over the time and archive can expect to receive huge quantities of digital materials in near future. The volume of the metadata will also sure to increase over the time.

Scale up: Capacity of the system to scale larger by adding more resources (processor, memory and other.)

Scale out: The system supports the caching, adding more instances and other related techniques.

Architecture: Ability of breaking the repository in different parts and host in geographically located different machine.

TABLE III. COMPARISION ON SCALABLE

	Platforms		
	DSpace	***Fedora***	***EPrints***
Scale up	0.5	0.75	0.5
Scale out	0.75	0.75	0.75
Architecture	0.5	1	0.75
As a total	0.58	0.84	0.67

1) DSpace

The Dspace @Cambridge says that when the amount of archived digital objects reaches 100,000 the browse will be slow [10]. It means that as the number of items gets larger, memory problem occurs. The reason of that is the memory will never be cleaned up when connecting the system. OARNZ (Open Access Repositories in New Zealand,) project has suggested that adding more memory to deal with this memory problem [11].

2) Fedora

The University of Karlsruhe has tested the Fedora with 14 million digital objects. They said that the ingest time was fairly constant over the duration [12]. But the retrieval of the digital object may have the problem according to dissemination method exposed. Fedora supports the additional operation on the digital object such as getting text from the digital object using OCR software. Fedora addresses this workload as it cans proxy the complex operation to different machine [11]. The Fedora community is planning to make scalable up to 30 million digital objects [13].

3) EPrints

ORANZ project has checked the EPrints up to 100,000 digital objects, and the system goes fine [11]. It writes every digital object as the static files which is the easiest and the fastest way to access from the web. The database layout is not normalized and it uses non-conventional way. It can index 100,000 digital objects in each run but it doesn't have incremental indexing method. Therefore, EPrints indexes fast even when increasing the cache.

D. Interoperability

It is desirable that our system can interoperate with the diverse systems inside the organization.

TABLE IV. COMPARISIONON INTEROPERABILITY

	Platforms		
	DSpace	***Fedora***	***EPrints***
OAI-PMH	1	1	1
SRW/SRU	1	0	0
SOAP	1	1	0
Bulk import and export	0.75	1	0.5
Integration with the other web pages	0.25	1	0.25
As a total	0.8	0.8	0.35

1) DSpace

DSpace supports the OAI-PMH, METS and Dublin core metadata. Again it supports the REST and SOAP web services. It also supports the SRU/SRW and LDAP authentication.

2) Fedora

It supports the SOAP and REST web services since it is web service rather than web application. It doesn't support the SRU/SRW services. It supports the FOXML and METS for the bulk import and export. It also supports the authentication by LDAP.

3) EPrints

EPrints supports the OAI-PMH and standard Dublin core metadata but yet it doesn't support any web service API such as SOAP and REST. Again it should be noted that the bulk import and export only supports the metadata. The LDAP authentication can be made possible.

E. Security

Security is of the topmost importance in building the confidence of potentials donors. Even the most flexible but poor security repository is undesirable.

TABLE V.COMPARISIONON SECURITY

	Platforms		
	DSpace	***Fedora***	***EPrints***
Data transmission	1	1	1
Server security	0.75	1	0.5
Roles and Authentication	1	0.5	0.5
As a total	0.92	0.83	0.75

1) DSpace

The security features of the DSpace are the finest as it use the SSL of the tomcat. It also has the good server side security and it uses the LDAP for the authentication. It has different roles which can be combined to produce the tight security.

2) Fedora

For the Fedora, it also uses the data encryption technique. It doesn't have inbuilt different roles. But different roles can be created by the XACML access policies. The authorization and access policy is given to the front end application like fez. The authentication is done by the LDAP access policy.

3) EPrints

The security of the EPrints can also be considered as reliable since, it also supports the data transfer encryption. But it doesn't have the good server side security as desired. And it have three kind of roles i.e., user, editor andadministrator.

F. Archiving and Database Management

The database used by the repositories must be finest and archiving must be ease.

TABLE VI. COMPARISIONON ARCHIVINGAND DATABASE MANAGEMENT

	Platforms		
	DSpace	***Fedora***	***EPrints***
Archival media and database	0.5	1	0.25
Storage hierarchy	0.75	0.5	0
Backup and disaster recovery of archived content	0.5	1	0.25
As a total	0.59	0.84	0.17

1) DSpace

It offers two means for storing bit streams. The first is in the file system on the server. The second is using SRB (Storage Resource Broker). [7] The SRB is the option which can be used for the replication of data or storage of repository inelsewhere. For the database DSpace uses the Oracle and the Postgres which are the advance and have huge query processing capacity. It inbuilt storage hierarchy of communities and collection. The method on organizing the backup for DSpace is available. Open source backup software is also present.

2) Fedora

It supports the huge no of the media since any digital content mustn't compulsory be inside the system. It uses the Mysql and Oracle as the database. But other database can also be used in the condition; the database needs to be JDBC-compliant and must support common SQL-92 syntax [13].

3) EPrints

It uses the Mysql traditional database which is easy but may not support the complex queries like the oracle. There is no any inbuilt technique for the backup. But any how it can be done by backing up the database and the repository.

G. Submission

The procedure of submission must be user friendly with appropriate authentication.

TABLE VII. COMPARISION ON SUBMISSION

	Platforms		
	DSpace	***Fedora***	***EPrints***
user interface	1	0.25	0.75
Authorization	1	0.25	0.5
Individuation of user interface	0.25	0	0
submission report	0.5	0.75	0.75
Workflow	0.75	0.75	0.5
As a total	0.7	0.5	0.5

1) DSpace

The user interface of the DSpace is finest. In the system we have different roles from the e-person to administrator. The DSpace provides the individualization of UI according to collection or community home page but can't accord to the item page. There is not facility of submission report but user can subscribe to any collection or community. DSpace have inbuilt workflows for the submission and also we can change the workflow by the user interface.

2) Fedora

The Fedora front end application must be installed for the UI, authorization and workflows. But it has great flexibility on those purposes as we can integrate with diverse type. Fedora has the log information where we can monitor the entire update made in particular digital object.

3) EPrints

The UI of EPrints is also good as all the task is nicely presented in web UI. It also has the facility of history where every change made in particular object is stored.

H. System Configuration

The system may not overlap with our need; so many configurations in the system are needed.

TABLE VIII. COMPARISION ON SYSTEM CONFIGURATION

	Platforms		
	DSpace	***Fedora***	***EPrints***
Configuration of UI	1	0.25	0.75
Configuring system policies	0.5	0	0.5
Configuring module of information	0.5	1	0.5
Configuring archival strategy	0.25	1	0.25
As a total	0.56	0.56	0.5

1) DSpace

The configuration of the interface is very easy in the DSpace. There we can change the home-page as well as the communities and the collection pages. Again the DSpace can be configured to support the multi-language and the system can be configured to the huge number of the workflows.

2) Fedora

For the easiness, the configuration on the Fedora is very complicated and difficult. Since it is not the complete repository, for the complete functionality instead of configuration of the old system, the new module should be installed. It can be configured and adapted into the wide range of the system but technically it is quite complex.

3) EPrints

The UI configuration of the EPrints is also easy. It home page can be configured as the desired. But the UI cannot be changed as like the DSpace which supports the individuality of the UI.

I. Working with the code

TABLE IX. COMPARISION ON WORKING WITH CODE

	Platforms		
	DSpace	***Fedora***	***EPrints***
Writing Plug ins or other packages	0.5	0.5	0.5
Alter the digital object type including metadata	0.75	1	0.5
Documentation and understanding of code	0.5	0.75	0.5
As a total	0.58	0.75	0.5

1) DSpace

DSpace has a user interface to add new metadata and namespaces. The database layout that stores the metadata supports adding new metadata. DSpace supports adding different workflows to a collection that holds different digital objects.

2) Fedora

Fedora has more code than other system so at a glance it may be confusing but its code consistency is good. Adding new content type is supported; content type is defined by new XSD document [13].

3) EPrints

The code has consistence structure and standard and changing to the content type looks fine by consulting the documentation. But the more complex type of content type requires the database table change, and due to non-normalized database schema, it differs from one to another which make difficult on adding new complex content type.

J. Archival and Administrative Concerns

TABLE X. COMPARISIONON ARCHIVALAND ADMINISTRATIVE CONCERNS

	Platforms		
	DSpace	***Fedora***	***EPrints***
Complex inter object relationship	0.25	0.75	0
Referenced metadata	0	1	0
Support content model	0.5	0.75	0.25
Realistic learning curve of system	0.75	0.5	0.75
Stability monitoring of data and metadata	0.75	0.75	0
As a total	0.45	0.75	0.2

1) DSpace

The inter-object relationship is bounded by the communities and collections making difficult to create user desired inter-object relationship. Addition of the own content model isn't supported in the DSpace. DSpace supports the checksum checking via a command line tool for the integrity monitoring of the data and the metadata [7].

2) Fedora

Through RDF (Resource Description Framework) metadata, we can create complex inter-object relationship of multiple levels. METS(Metadata Encoding and Transmission Standard) structural maps are also ingested to reflect the original order of an archival accession. To ensure that this is preserved for the archivist who will catalogue the archive. [13] It also allows the user to define their content model.

3) EPrints

Its learning curve is similar to DSpace. The new administrator can fine easy to habitat with system by studying the documentation.

K. Globalization

The repository system should be capable of computing the multi-language and different symbols for its wide acceptability.

TABLE XI. COMPARISION ON GLOBALIZATION

	Platforms		
	DSpace	***Fedora***	***EPrints***
Multi-language	1	0.75	0.75
UNICODE	1	1	1
As a total	1	0.87	0.87

1) DSpace

It supports the UNICODE preservation and the editing, so virtually it is capable of preservation of all type of language. Again the DSpace system can be installed in about languages, which is possible by downloading the language packages.

2) Fedora

It also supports the UNICODE. The front-end FEZ supports the different language in terms of the PHP templates.

3) EPrints

UNICODE is supported by the EPrints. Currently other than English the system can be configured to Bulgarian, French, German, Hungarian, Italian, Japanese, Russian, Spanish and Ukrainian [14].

L. Searching and Browsing

All the system has good facility of searching and browsing that provides ease of locating the desired object. Fedora has quite nice search facility as it has 4 types of search i.e. simple search, G-search (The Generic Search Service that makes Fedora's features useful to different search engines), PrOAI (The OAI provider service that is designed to take advantage of Fedora's features) and RDF based resource index.

TABLE XII. COMPARISIONON SEARCHING AND BROWSING

	Platforms		
	DSpace	***Fedora***	***EPrints***
Search engine	0.75	1	0.75
Browser	0.75	0.75	0.75
As a total	0.75	0.87	0.75

M. Community and Support

TABLE XIII. COMPARISION ON COMMUNITY AND SUPPORT

	Platforms		
	DSpace	***Fedora***	***EPrints***
Development community	1	1	0.75
User community	1	0.75	0.5
Supports for the user	1	0.75	0.5
As a total	0.88	0.81	0.56

1) DSpace

It has the largest and strongest community. The Wikipedia says that 240 institutions around the globe are currently using the DSpace. For the assistance it has the free mailing list as well as the source forge page. And it gets largest number of mail from its user compare to other system.

2) Fedora

Currently DSpace and Fedora is under the same community namely Dura space. Its documentation is pretty good quality. A wiki, mailing list, and bug tracker is provided for the public. Till May 2009, 165 known archive of Fedora was noted [13].

3) EPrints

The first version was release in 2000, which becomes first repository widely accepted. 269 known archives of EPrints have been discovered [15]. Wiki page, free mailing list and the paid organization for assistance of the EPrints is available.

N. Average Observation

We here provide only the foundation for evaluation of repository and hence to choice repository relevant to particular organization. Using above data, we can use mathematical tools such as Standard Distribution, weighted mean etc., for evaluation. Here we propose the weighted mean method to choice the repository that matches our needs. . For e.g. we have institution where we want to settle the repository. The administrator of institution gave the following weight on different criteria studying its local situation;

system deployment = 0.2,
versioning = 0,
scalability = 0,
interoperability = 0.05,
security = 0.05,
archiving and database management = 0.1,
submission = 0.2,

system configuration = 0.05,
working with code = 0,
archival and administrative concerns = 0.05,
globalization = 0.2,
searching and browsing = 0.05,
community and support = 0.05.

By the weighted mean formula, using the weight from system administrator and data from our repository evaluation, we found that

DSpace=0.75,
Fedora=0.68 and
EPrints=0.57.

In this situation we found DSpace is relevant for us.

TABLE XIV. AVERAGE OBSERVATION

	Platforms		
	DSpace	***Fedora***	***EPrints***
System Deployment	0.67	0.5	0.62
Versioning	0.125	0.75	1
Scalability	0.58	0.84	0.67
Interoperability	0.8	0.8	0.35
Security	0.92	0.83	0.75
Archiving and Database Management	0.59	0.84	0.17
Submission	0.7	0.5	0.5
System Configuration	0.57	0.57	0.5
Working with the code	0.58	0.75	0.5
Archival and Administrative Concerns	0.45	0.75	0.2
Globalization	1	0.87	0.87
Searching and Browsing	0.75	0.87	0.75
Community and Support	0.88	0.81	0.56
As a total	0.71	0.75	0.58

V. DISCUSSION

Those enlisted all systems are incredible, but the fact we perform comparison is to find out which platform is best suited for long term knowledge representation. We have performed most of the tests in the laboratory LIESP of University Lumiere Lyon 2. Issues such as scalability weren't tested on our own laboratory because of the large resources requirement, but collected from the previous comparison or tests done by different institutions. Thus, in this paper different feedback of previous comparisons is also added.

A. DSpace

It can be used within the medium scale organization or institution. The overall glance of DSpace is good but due to scalability issue it can't be recommended for the national system where the number of object may exceed million. Furthermore this scalability issue can't be solved with simple effort, but required major re-write of program. Again another weakness is the versioning but can be maintain by local writing the module.

The strength lies on the interoperability and the security of the system. Again it has the inbuilt roles and access policy. Strongly speaking DSpace is the out of box system which can be best suited for the medium and small institution.

B. Fedora

The key point is it can be adapted to any kind of local system. The system can be configured to institution relevant workflows and branding. The security and the interoperation schema can be imposed as desired. More than the simple downloading, it also supports the additional operation inside repository. Furthermore it has the huge scalability power. Finally speaking about the Fedora, it is best suited to the huge organization that have vague amount of data and complexity and can afford reasonable value to overcome this.

C. EPrints

It is best suited for the self-configuring institution which are wanting to built and host its own archiving system. There is not very tough security and interoperability schema but on overall rating it still holds good position. The power of EPrints is, it is simple and out of box system which need very small technical and cost overhead. Those institutions which can't afford continuous technical staff there is paid organization for the overall assistance. At last, EPrints is the most suited for small organization which needs overall common features of repository but don't desired for complex and high level inter-operation.

VI. CONCLUSION

Nowadays, the issue of long term preservation of digital data has accelerated rapidly and has been offered more discussion in the literature. However, there are different technologies i.e., repositories for the long term digital preservation, but none of them offer the exact solution of the specific institutions. Furthermore, different repositories have different strength in the given criteria and the requirement of specific organization for long term digital preservations may be different i.e. some organization may demand more security while other may demand more interoperability. Again repository may lack the burning feature which is most essential for the long term digital preservation for e.g. versioning in DSpace. The aim of this paper was to address these issues, that is, to presents the technical strength of short-listed repository under different criterion and assists the organization to choose nearest match repository by using "weighted mean" method for the long term preservation of digital data. Since we present the power of repository, the organization can also make rough estimation of technical and economic value they have to spend to install the repository. By

pointing the weakness of specific repository in certain criterion, we are also focused to keep the development community to break out the backwardness of their repository.

ACKNOWLEDGMENT

This research has been supported by the EU Erasmus Mundus Project - eLINK (east-west Link for Innovation, Networking and Knowledge exchange) under External Cooperation Window – AsiaRegional Call (EM ECW - ref. 149674-EM-1-2008-1-UK-ERAMUNDUS).

REFERENCES

[1] Companies must plan for digital preservation, British Library. Available:http://www.computerweekly.com/Articles/2010/03/22/240672/Companies-must-plan-for-digital-preservation-says-British.htm.

[2] P. Campisi, E. Maiorana, E. Massella, D. Teri, A. Neri, "Challenges to Long Term Digital Preservation, A Glimpse of the Italian Experiences," The paper published in IEEE.

[3] H. M. Gladney, Principles for Digital Preservation: Focusing on End User' Needs rather than those of Archiving Institutions. *Communications of the ACM*, vol. 49,no 2, 2006.

[4] D. S. H. Rosenthal, T. Robertson, T. Lipkis, V. Reich, S. Morabito, *Requirements for Digital Preservation System*. Stanford University Libraries.

[5] K.H. Lee, O. Slatter, R. Lu, X. Tang, and V. McCrary, "The state of the art and practice on digital preservation," *Journal of Research of the National Institute of Standards and Technology*, vol. 107, no. 1, pp. 93-106, 2002.

[6] R. Herry, "Digital Repositories Review", JISC Commissioned.

[7] *DSpace 1.5.2 Manual*. The DSpace Foundation. Available: http://www.DSpacedev2.org/1_5_2Documentation/

[8] Fedora documentation, Available: http://www.Fedora-commons.org/confluence/display/FCR30

[9] EPrints Manual, Available: http://wiki.EPrints.org/w/Manual

[10] Scalability issue of DSpace, Available:http://wiki.DSpace.org/confluence/display/DSPACE/ScalabilityIssues/

[11] Open Access Repositories in New Zealand. Available: http://eduforge.org/projects/oarinz/

[12] Scalability issue of Fedora. Available: http://Fedora.fiz-karlsruhe.de/docs/Wiki.jsp?page=Main

[13] Fedora open source Repository Software. Available: http://fedora-commons.org/

[14] EPrints wiki page. Available: http://en.wikipedia.org/wiki/EPrints

[15] EPrints Repository. Available: http://www.eprints.org/

[16] DSpace Open Source Software. Available: http://www.dspace.org/

[17] White paper for Long Term Archiving and Retrieval of Product Data within the Aerospace Industry (LOATAR). Technical Aspects of an approach foe application of project; W Group, "LOATR" version 1.0, August 30, 2002.

[18] Repository Comparison. Available: http://outgoing.typepad.com/outgoing/2006/03/repository_comp.html/

[19] Comparison Repository Software for Personal Digital Archives. Available: http://www.paradigm.ac.uk/workbook/repository/comparing-findings.html/

[20] Long Term Archiving-LOTAR. Available: http://www.paradigm.ac.uk/workbook/repository/comparing-findings.html/

[21] M. Jafari, P. Akhavan, A. Mortezaei, "A Review on Knowledge Management Discipline," *Journal of Knowledge Management Practice*, vol.10, no. 1, March 2009 W.D.

[22] CASPAR- Cultural Artistic and Scientific Knowledge for Preservation Access and Retrival. Available: http://www.casparpreserves.eu/

Management of IT Services – A Snapshot on Hungary

Zoltán Szabó, Péter Fehér
Department of Information Systems, Corvinus University of Budapest
Budapest, Hungary

***Abstract*—Although the role of information technology became more and more important in the daily operations of organisations, beside consulting a few analysis there is no detailed exploration of IT service management practices. The presented survey explored the evaluation of the role of information technology, IT budgets, IT strategy planning questions, and IT service management practices. Survey results show that Hungarian companies have to face the crisis of the current economic downturn, and they have to find the right solutions in IT to support business processes, despite of the strongly decreasing budgets. Fortunately in several aspects organisations have good basis, like above average IT development budgets, IT strategy foundations, or strong IT service management processes.**

Index Terms: service management; information technology management

I. INTRODUCTION

Crucial importance of information technology service management (ITSM) cannot be questioned in the last few decades. ITSM methodologies and standards, especially ITIL (IT Infrastructure Library) have recently become a very popular approach and a widely used methodology to improve this organizational activity. ITIL as a comprehensive approach to manage IT services is based on best-practices. In the international IT literature numerous practical ITIL guidelines, case studies, connectional papers have been published. Though these topics are popular within the IT consulting domain, little, if any research activity has been focused on this topic in Hungary. In the context of the economic crisis the question remains whether Hungarian organisations have understood these possibilities, and if they have so, to what extent they are able to realize ITIL-based innovations.

II. RESEARCH BACKGROUND

Based on the survey our intention with this article is to investigate the following issues:

- the role of IT in competitiveness and innovation,
- the actual activities in IT strategy planning,
- the characteristics of IT infrastructure and applications,
- the actual and planned state of IT service management, and
- the IT budget and the role of IT in the financial planning.

In the rest of the study, these issues are analyzed illustrated with figures and tables. We compare the findings to other available studies based on similar surveys (e.g. “Information Management” block of the “Competing the World” research project carried out in 1996, 1999 and 2004; and a recent survey of The IFUA Horváth and Partners [6]).

This paper presents the findings of a research on IT challenges, performed among the Hungarian IT service management applier companies in the first half of 2009, and with identical questions in the first half of 2010. Research data were collected through anonymous questionnaires, and complemented with case study findings. The questionnaires were filled voluntarily, mostly by mid-level IT executives and IT employees.

The target sample consists of 57 organisations (48% of active members of ITSMF Hungary) that have critical IT infrastructure, and planning or using IT service management solutions, therefore they are valid representatives of the Hungarian information management practitioners. 15% of the organisations are micro or small enterprise, 20% are medium-sized enterprise, and 65% of them are large-sized enterprises. Organisations both represent the public sector (35%) and the private sector (65%).

III. ANALYSING IT COSTS

The practice of information technology is mostly one of the first areas, where executives try to start cost cutting [11]: providing IT and business services demands high operational costs, but the return of IT services is hard to measure. IT departments continuously facing the extreme challenge of showing the value of their work, but failing this challenge means, that business customers perceive information technology as a rather expensive, but mostly useless area, and results the increased interest of cutting IT costs.

A. Size of IT budget

IT costs usually varies between the average 3-5% of the revenue, but the current values strongly

ISBN 978- 974-672-556-9

depends of the size of the company, the industry, and the country of analysis. A few examples for previous measures: in 2008, in the pharmaceutical industry, IT spending was 3.61% of the revenue, or in an average US company the *operational* IT costs were 1.5% of the revenue [3] [8]. Another international survey presents the average 2% of IT budget (cca. 125 million Euros) comparing to the revenue [10]. Hungarian surveys show a developing tendency in the size of IT budges, Drótos and Szabó [5] reported 1% of the revenue in for 1996 and 1999, Drótos [4] also reported 2.2% based on a 2004 survey, and Drótos and Szél [6] reported a rather high 4% based on 2007-2008 data.

Strong contrasts can be observed: in the case of innovative organisations, IT spending can reach, or even exceeds 10% of the revenues, while conservative companies spend only 0.5-1%. Such kind of innovative companies can be found in the financial sector (banks), where IT spending varies between the fabulous 8-20% of the revenues [7].

The conducted research among the Hungarian organisations shows a similar picture: IT costs vary between 0.1% to 20% of the revenue. The lowest spending can be identified in the public sector and in the constructions industry, while the biggest spenders belong to the financial industry (5-20% of the revenue).

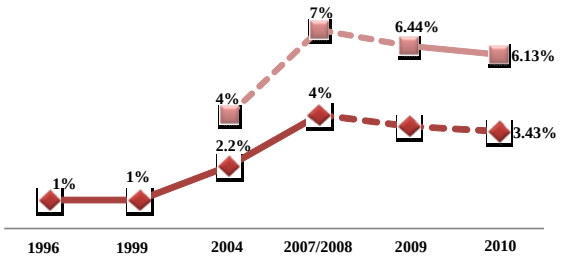


Figure 1. Share of IT budget comparing to annual revenue (Research findings)

By 2010 the IT budgets decreased again, by an average 4.85%.

Meanwhile the change of the budget size comparing to last year shows us, that IT departments have to face to serious budget problems, because of a significant decrease of financial background. The most significant decrease in IT budget can be identified in the case of large enterprises, and the average change of the surveyed organisations is around 11% decrease, below the average decrease of the Hungarian GDP. These data proves that the economic crisis reached not only the Hungarian companies, but also the IT departments.

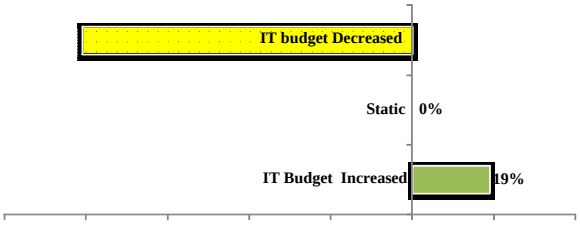


Figure 2.Is IT budget going up or down in real terms comparing to last year? (Research findings)

B. Determining IT Budget

Although researches concentrate on IT budgets, spending on IT can be two or three times more in an organisation beside the budget of the IT department, through business departments finance their own developments or investments (typically: equipments, unique applications and external services). Moreover, IT departments have limited decision right on their own spending, because business departments and decision makers can set the priorities, and determine the target of spending.

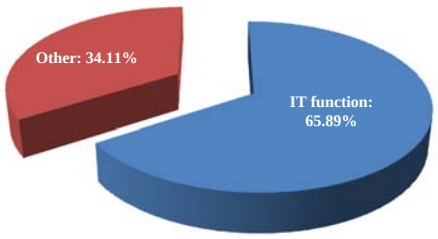


Figure 3. What portion of the IT budget is determined by the IT function? (Research findings)

The conducted research among the Hungarian organisations shows, that IT departments are the owner almost two-third of their budget, while a little more than a third of the IT spending is determined by other organisational units.

C. Components of IT Budget

For the cost cutting efforts the content of the IT budges raises the barrier of a relative high level of operational costs comparing to the development costs. As Cramm [2] states, “prior the credit crisis, around 75 cents of each dollar was spent to keeping existing systems up and operational.” Another survey in UK [3] shows that while in the small and medium level enterprises operational costs are lower (56% for SMEs), in the case of large enterprises 78% of the IT budges is used only to “keep the lights on”.

The conducted research among the Hungarian organisations shows that average level of operations costs in the IT budget is relatively low, only 55% of the IT budget. The rate is 52% for SMEs, and even in the case of the large enterprises this rate is only 58%. Beside of these raw statistics, if the results are analysed in more detailed, we can see, that

operational costs varies between 30%-80% of the total IT budget of the large enterprises. Comparing to the international results, the relative high share of development budget seems to be promising. A representative domestic survey in 2004 [4] showed the international 73%-27% rates for operational and development budgets, but a recent survey results shows similar results to our experiences with 51% operational and 45.2% development costs [6].

Analysing the 2010 year results, a decrease in development budgets was identified. Because of the economic crisis, development budgets were significantly decreased (and operation budgets were only moderately decreased), therefore its relative share in the whole is lower.

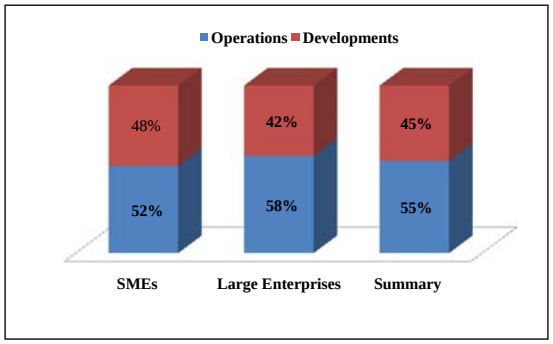


Figure 4. Operations vs. Development Budget (Research Findings)

IV. IT STRATEGY

IT strategy is a tool for converting organizational strategy and goals into operational understanding of IT goals and development projects. There is a dynamic relationship between IT strategy and organizational strategy: while the organizational strategy leads the way, sets priorities, and creates demands for information technology, IT trends and IT solutions can offer innovative business models, including new pricing methods, market segmentation, or new distribution channels. Our research aimed to analyze the existence and content of IT strategies, and the relationship with organizational strategies.

Information strategy planning (ISP) plays a crucial role in the utilization of the IT potential for improving competitiveness. A favorite question in similar studies is how much care is taken about this in Hungarian organizations. In 1996 27%, in 1999 22% of the surveyed companies formulated separate information strategy [4] [5] [6], Drótos et al. [4] reported 27.8%. In our survey the respondents are leading IT service providers, banks, service companies, for which the need to elaborate a formal IT strategy generally is very strong. The findings verify this preposition: vast majority, 83% of the respondents reported that they have prepared a separate IT strategy document. The average time horizon of the information strategy is half year shorter (2.9 years) than the results of the earlier surveys (3.5 years [5], but almost the same timeframe (2.75 years) reported in [4].

Concerning the content of information strategies we found that the information strategies covers most of the classic issues (application portfolio, standards, etc.), and a relatively new concept, IT Controlling is the second in the list.

These results, however, should be interpreted with caution, since the sample is significantly biased toward large organizations, for which the need to elaborate a formal IT strategy is almost inevitably necessary, and the methodological background is significantly more matured, than the "average" companies.

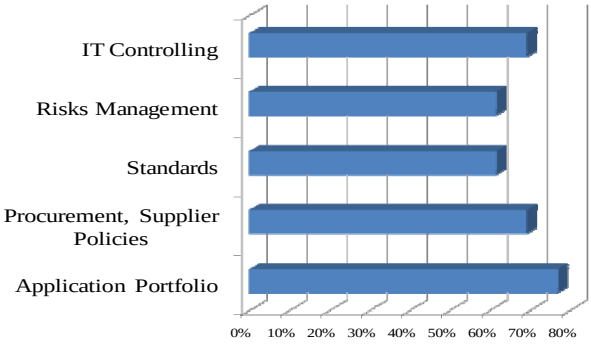


Figure 5. Some major issues in the information strategy (Research findings)

V. PERCEPTIONON THE ROLE OF INFORMATION TECHNOLOGY

Considering information technology as a factor of the organisational performance and competitiveness, the research analyses the innovative function of IT departments, and the uniqueness of IT solutions. It is as long-lasting challenge of information technology, how to align its activities with the business, and how can an IT department offer valuable, and innovative solutions for the business side. The challenge is that the language, perception and understanding are different in the various business, and even in IT units. The practice of judging information technology is rather different, starting with the perception of traditional cost centres, trusted suppliers to reaching the innovative partner position.

Respondents of our survey see the role of information technology more as a trusted supplier of innovative solutions, as accomplishing business innovations, rather than the source of innovative ideas. Although information technology has significant role in several business services, the perception of its role, and the acceptance of new

ideas is rather low.

These results mirror the findings of an international survey [10], in which although 70% of the organisations agreed on the critical role of information technology in business innovations, only around 24% of them accepted the IT function as the main originator of business innovations. Moreover, in a 6 elements list examining the initiator role of innovations, the IT function was ranked only fifth, below Sales, Marketing, Operations andResearch & Development, and onlyahead of Finance

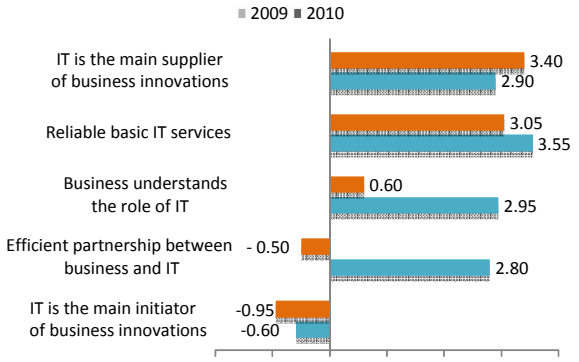


Figure 6. Role of IT (Research Findings: Average ranking transformed to -10 – 10 scales, where 10 = strongly agrees)

. The results for 2009 are a bit pessimistic in sense of the innovation performance of the IT function was ranked first in our survey, with only a relatively moderate average. Every answer tends to be around the middle area of the intervallum, there are only few outstanding companies (Fig. 6).

Although respondents have trust in their service quality (and it was expected in a sample of IT service management appliers), they do lack the efficient partnership with business areas, and continuously fighting the challenge make the role of information technology be understood by business partners. Considering the three main service provider categories of IT functions [9], comparing to the above mentioned international results, Hungarian companies act more in the role of utility player or trusted suppliers instead of business partners.

Surprisingly, there were significant changes by 2010. The crisis increased the relationship between business and IT, and business partners try to better understand the role of IT in these days. The crisis could create an opportunity for companies to better understand each other.

VI. IT INFRASTRUCTURE

As the responding companies are strongly dependent on IT, they probably have a massive IT infrastructure and relatively state-of-the-art applications. In the next session we briefly discuss the infrastructural findings, comparing them to the earlier surveys.

The average number of servers/organizations in the sample is 102, varies between 3 and 200, the average number of provided IT services is 109. The number of internal end-users is between 20 and 4000 (average: 1800), and some services are used by thousands of external users.

The sample shows some similar tendencies to the earlier surveys [4][5]. Accounting and finance applications are used in almost every responding organization. We have observed similar proportion of ERP applications (39% compared to 42,7% [4]. In contrast to this, inventory management, manufacturing management systems are not frequently used, that can be explained with the different sample (in our survey service oriented companies were overrepresented). Some of the classic functions, like HRM, controlling are also much less popular, than in [4], where these applications were reported in 90% of the respondents.

The integration of existing applications has been clear trend in the corporate IT practice in the last decade. We witness the same level of integration between the core systems, key functional applications in our survey than in [4].

Concerning the availability of cutting-edge applications, we have observed much higher penetration than in the Drótos et. al. 2006 survey [4]. They reported only 10% in the case of BI and knowledge-based applications, we found that 50-60% of the respondents use this categories. Similarly the workflow systems are more frequent in our sample (60% compared to their 30%). Another interesting finding is the high ratio of groupware applications.

VII. IT SERVICE MANAGEMENT

The processes and activities of IT departments should be based the IT strategy, but also heavily depends on the required services (number and content), complexity of IT environment, number of users and the size of the infrastructure, briefly discussed in the previous section. Our research now focuses on the analyses of the service management characteristics, and analyses the existence of the quasi standard ITIL v3 processes and functions.

According to the ITIL, that provides a framework of best practice guidance for IT Service Management, the key ITSM procedures can be grouped into 5 major categories:

- Service strategy (Service Strategy, Service Portfolio Management, Financial Management, Demand Management)
- Services Design (Service Catalogue

ISBN 978- 974-672-556-9

Management, Service Level Management, Capacity Management, Availability Management, Service Continuity Management, Information Security Management, Supplier Management)

- Service Transition (Knowledge Management, Change Management, Asset and Configuration Management, Transition Planning and Support, Release and Deployment Management, Service Validation and Testing, Evaluation)
- Service Operation (Incident Management, Problem Management, Event Management, Request Fulfillment Management, Access Management, Common Service Operations)
- Continual Service Improvement (Service Measurement, Service Reporting, 7-Step Improvement Process, Service Desk)

We asked the participating companies to submit data concerning their above listed key ITIL processes. We distinguished 3 potential status (implemented, in progress, planned), and an additional category (not expected).

Not surprisingly service desk, incident management, request fulfillment, security, BCP and availability management are amongst the most frequently reported existing processes. Interestingly configuration management and change management have been already implemented or under implementation in many organizations, showing that these companies recognized their central role in the service-oriented transition and operations.

Concerning the implemented and planned traditional core ITIL v2 service delivery and support processes, it is clear that they still constitute the majority of ITIL oriented activities and initiatives. Most ITIL v3 related initiatives are based on the existing ITIL platform and proposals (ITIL v3 emphasises the concepts of service life-cycle management, continual service improvement and the alignment of IT and business). Our findings are similar to Broussard's (2008) report, where 21.1% of survey respondents replied some implemented ITIL processes, and 21.8% of survey respondents replied that they are implementing ITIL processes now. Considering the current platform and future plans for ITIL, over one-half (56%) of respondents said they are implementing ITIL now or looking to implement ITIL within the next two years.

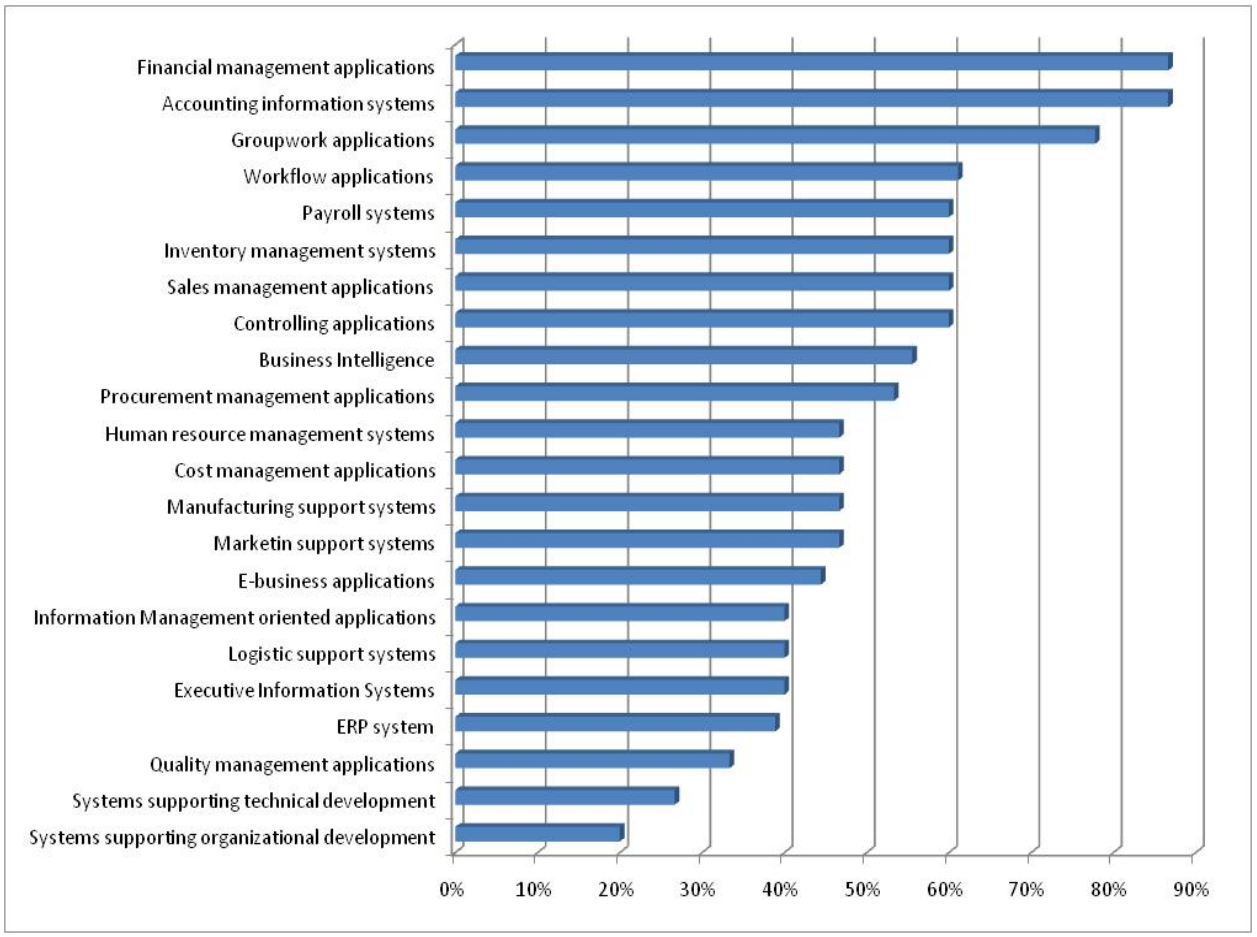


Figure 7: Penetration rate of the major application categories (Research findings)

ISBN 978- 974-672-556-9

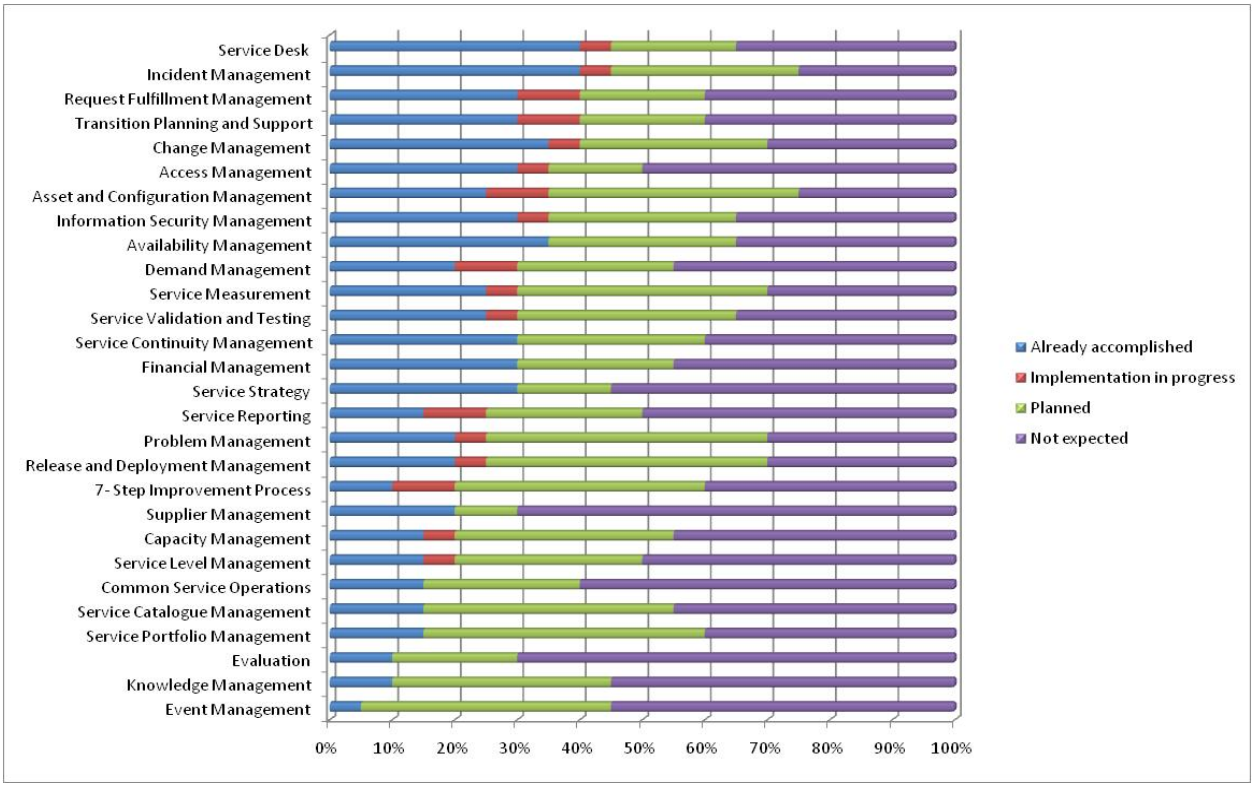


Figure 8. Approaches to the major processes of ITIL in the sample (Research findings)

The survey shows intensive development plans towards the mature level of ITIL oriented service management. Interesting to note, that service evaluation and supplier management are not so popular, although they are really crucial issues of a sustainable service quality. Another interesting figure that Common Service Operations, that are probably really common in all organizations, were not recognized as implemented processes. This shows that the transition to process-oriented IT operations is a tough project.

We should emphasize, that in the sample the pioneers of ITIL are over-represented, and the proportions of the implemented ITIL processes are higher, than in the real Hungarian situation. On the other hand the trends that can be drawn from the plans and initiatives are clear: the final destination of ITSM in the next decade is the implementation of ITIL based processes.

VIII. Conclusion and Further Research

Our research was aimed at surveying some IT related aspects of domestic companies that manage their IT infrastructure above average consciousness. During the time of the survey companies have to face the current crisis of the economic downturn that has a serious impact on their IT budgets. IT budgets are significant parts of organizational budgets, and this fact pressurizes IT departments to find cost efficient and innovative solutions to support the expected business services.

Finding answers to this challenge, organizations may choose to way to develop both their supporting IT applications and their IT service management practice, based on a matured IT strategy.

The survey results comparing to other international and domestic survey results shows that Hungarian companies are close to the international trends in the field of information technology, although there are significant differences in budget composition or in the IT application portfolio. As the biggest problem issue, CIOs have to face the strongly decreasing IT budgets of their organizations, well below the decrease of the GDP. To balance this pressure, Hungarian developments budgets are still larger than average development budgets in the international surveys.

This study was presented on a quick survey, in order to explore the most important and significant phenomena in the current practice of IT department. Although there are several interesting results, this research has its limitations in sample size, composition, and distortion towards IT savvy organizations.

Results of this survey are good bases for further, more concentrated researchers. Further researches are planned to explore more organizational questions, customer behavior and perception, described not only in numbers but with detailed case studies. Moreover, future researches are planned to fully explored IT dependent sectors like the financial organizations (banking and insurance), and telecommunication.

Future researches will of course reflect the current study results on IT budget, costs, applications, and IT service management processes.

REFERENCES

[1] F.W. Broussard, "IT Service Management Needs and Adoption Trends: An Analysis of a Global Survey of IT Executives." IDC White Paper. 2008. Available: http://whitepapers.techrepublic.com.com/abstract.aspx?docid=395611 (accessed June 28, 2009)

[2] S. Cramm, "Smaller IT Budget? Pursue Value-Driven Development," *Having IT your way – HBR Voices*, November 20, 2008. Available: http://blogs.harvardbusiness.org/hbr/cramm/2008/11/valuedriven-development.html. (accessed June 23, 2009)

[3] *DBVIT: Driving Business Value from IT*. The Batwick Group, December 2007, London

[4] D. György, G. Károly, M. Péter, and V. György, *Az információmenedzsment fejlettsége és a versenyképesség. Versenyképesség Kutatások Műhelytanulmány-sorozat, 28. sz. műhelytanulmány*, Corvinus University Budapest, 2006. Available: http://web.uni-corvinus.hu/vallgazd/kutatas/versenykepes2005/28%20mht.pdf (accessed June 28, 2009)

[5] D. György and S. Zoltán, "Corporate Information Technology in Hungary at the Millennium: Myth and Reality," in *Chikán Attila - Czakó Erzsébet - Zoltayné Paprika Zita: National Competitiveness in the Global Economy*, Akadémiai Kiadó, 2002, pp. 248-263

[6] D. György and S. Zoltán, " Mennyit költünk az informatikára?" in A. Mártonffy, Ed., *IT Business*, 27. January 2009, p. 24.

[7] J. Girnt , "Rendszerváltó projektet vezényel az év főinformatikusa, megjelent: Napi Online," 2008.09.09, NAPI online - Napi Gazdaság. Available: http://www.napi.hu/cikkek/?nID=380553. (accessed March 19, 2009)

[8] *ITSPE: IT Spending, Staffing & Technology Trends, Computer Economics.* Irvine, California, 2008

[9] P. Roehrig, " New Market Pressures Will Drive Next-Generation IT Services Outsourcing," *Forrester Research Report*, Oct. 9, 2008

[10] O. Sevillia and E. O – Monnoyer, "The Role of the IT Function in Business Innovation, Innovator vs. Operator," *Global CIO Survey*. Cap Gemini, 2008

[11] S. Spang, "Five trends that will shape business technology in 2009," McKinsey *Quarterly*, Feb. 2009. Available: http://www.mckinseyquarterly.com/Five_trends_that_will_shape_business_technology_in_2009_2296 (accessed March 5, 2009.)

Impact of the Economic Crisis on Using Information Technology

Péter Fehér
Department of Information Systems, Corvinus University of Budapest
Budapest, Hungary

***Abstract*— The economic crisis has rapid impact on organizational, and also on IT budgets. Neither companies, nor IT departments were prepared for such changes, and first reactions followed a conservative cost cutting approach. Most IT departments had to suffer a significant decrease in their budget, and the slow recovery will last for years. This paper explores the challenges, how Hungarian organisation can deal with the crisis, and how they choose their answers. A 2 years long survey among Hungarian IT departments showed, that the organisations try to find the balance in these challenging times, and as promising economic indicators arise, the status of IT departments is also progressing**.

Index Terms: economic crisis, information technology use

I. INTRODUCTION

StevieSacks, an executive of a leader IT service management supplier prepared a study in 2004 about IT cost reduction (years before the current economic crisis). The study started as the following: "In an economic climate of slow and cautious growth, IT is under intense pressure to manage costs and ensure that spending is well aligned with business drivers."[21] If these thoughts were true in 2004, in the area of slow and cautious growth, how should be information technology managed during downturn?

An economic downturn always creates the pressure for organisations to rethink their operations. There are already several studies and analysis on managing critical situations (e.g. [18] [19] [24]). Conclusions of these studies suggest traditional solutions that are useful in any economic climate: cost reduction, conscious debt management, analysis of customer behaviour, reducing organisational complexity, increase the visibility of organisational processes. Bőgel [2] summarises these advices as the following: "...take the issues seriously, you have not take before. Do what you have not done."

Although there are several possible solutions, organisations started to concentrate on cost cutting issues, stop investments, developments and hires. The question is, whether these actions will increase the efficiency of the organisations, and what will happen, when the crisis is over?

Through stopping developments inside, and postponing strategy executions, organisations are not able to enhance their productivity, and while the world is changing, these organisations have no answer to the present and future challenges. Of course organisations should decide wisely where to put the money in these times, but they have to consider both short-term and long-term investments.

The pressure of the crisis is also a good opportunity to reconsider the current practice of the organisation: Are we doing everything perfectly? In which points can we enhance our practice? Where are the bottlenecks? Are our resources used efficiently and effectively?

II. THE IMPACT OFTHE ECONOMIC CRISISON ICT

Czech Republic and Hungary are good examples of successful economies in the past. It was hard to foresee the impact of the current economic situation on the developing economies of Central Europe, which is closing up to the EU average in every indicator. The Crisis presented, that these countries are not more fragile than other EU countries, although it still presents significant problems in several areas (Fig.1).

Indicator	European Union	Central Europe (including Czech Republic, Hungary and Poland)	India	China
Real estate market	⇩⇩⇩	⇩⇩	⇩⇩⇩	○
Consumer spending	⇩⇩⇩	⇩⇩	⇩	⇩
Inflation rate	○	⇧	⇩	⇩
Business environment	⇩⇩	⇩⇩⇩	⇩⇩	⇩
Trade momentum	⇩⇩⇩	⇩⇩⇩	⇩⇩	⇩⇩
Consumer confidence	⇩⇩⇩	⇩⇩⇩	⇩⇩	⇩⇩⇩
Business confidence	⇩⇩⇩	⇩⇩⇩	⇩⇩	⇩⇩⇩
Cost of capital	⇩⇩	⇩	⇧⇧	⇧
Foreign-exchange discontinuities	⇩	⇩⇩⇩	⇩	⇩

Severely worse ⇩⇩⇩ *Significantly worse* ⇩⇩ *Moderately worse* ⇩
Near normal ○ *Pre-Crisis or better* ⇧ *Significantly better* ⇧⇧

Figure 1.Impact of the economic crisis on EU, Central European Countries, India and China, based on selected indicators. Changes comparing to the pre-crisis period are indicated. [27]

Information Technology is expected to have an innovative role and significant contribution to the economic growth [13]. For example in Hungary the growth of the ICT sector is continuously above the average GDP growth of the economy, and contributes almost around a third of the economic growth, although the growth has a continuous

ISBN 978- 974-672-556-9

slowing trend.

Therefore it is also expected, that ICT solutions should contribute to the rebuild of the economies, through the following factors [28]:

- Contribution to productivity and growth through technological progress,
- Improvement of productivity of other factors of production, and
- Spillover effects on the rest of the economy.

Information technology is both a source and enabler of innovative solutions. Despite of these expectations, general reality is that although the demand for basic ICT services is still stabile while the demand for advanced applications is more uncertain [1].

The question is how ICT solutions can contribute to the competitiveness and productivity of organizations? There are no general ICT solutions for every business problem, moreover different organizations, and different organizational units have unique requirements to be fulfilled. Results are expected in the following areas [16]:

- Cost reduction, via automation and standardisation of back- and front-office activities, centralised solutions and changes sourcing methods.
- Enhanced time-to market ration or quality, via productivity enhancement, and new solutions for customer analysis, marketing and alternative distribution channels.
- Enabling business innovation, via the use of ICT solutions. ICT has an important role to offer novel solutions and possibilities, and find ways to support new business ideas.

Fulfillment of these expectations is limited by the relationship through the perception of the role of ICT by business partners. Generally information technology is perceived as critical, but IT function (e.g. departments) are not: in a survey of the innovative role of ICT solutions 70% of the respondents agreed that ICT has a critical role in business innovation, but less than a 50% of the respondents agreed that the ICT function itself is a driver of business innovation. Moreover, only 24% identified the IT function as a main initiator of business innovation [22].

There is also a difference in interpreting the role of information technology: while the most of the business decision makers believe, that the role of IT is to enable revenue growth, less IT executives agrees with it [22].

ICT providers and users have to deal with the above mentioned challenges, and decide on the above mentioned dilemmas in an environment that pressure low risk taking and punishes every mistake. In the following sections the paper explores the current practice and answers to the economic crisis in two central European countries.

III. CUTTING IT COSTS

Cutting IT costs are usually one of the first areas that seem to be promising for business areas. In the following, the main cost factors of an IT department are identified, especially concentrating on the cost cutting and optimisation possibilities, both in the areas of operations and development. The main focus of this research is to analyse cost reduction possibilities, considering both short-term and long-term impacts.

A. Operation costs (short term impacts)

Operation costs related to providing IT services and business as usual, concentrating to the management of existing services, and the background infrastructure. The main cost categories are: human resources, licence fees, public utilities (like electricity), renting fees, back-ground materials (like a printer toner), security tools, etc. In the following the main cost cutting possibilities are analysed.

• Human resources: Although operating an IT infrastructure demands high cost level itself, human resources take serious part of the budget. The first idea for reducing costs is the optimisation of human resources, meaning considering layoffs, or stopping new hires. If these steps are based on a conscious analysis, and performed together with the optimisation of operational processes, the results can be promising. But without a well-grounded plan, layoffs can result declining work efficiency, overloaded workforce, or increasing number of mistakes.

• Process management: In order to provide visible, controllable IT operations, process management is the first step for any further enhancement. Exploring and modelling the processes themselves is a good basis for performance development, because executives can see a clear structure of activities and required resources. Recorded processes create a clear responsibility and accountability structure; therefore it is a good basis for human resource optimisation. For IT process management international best practices, like ITIL, MOF or eSCM are a good basis to start with.

• Conscious supplier management: Organisations tend to follow a low-risk, conservative supplier policy, meaning maintaining long-term partnership with existing, good per-forming suppliers. Over the time, as a supplier builds strongly in daily operations, this long term relationship results a high level of dependency on existing partners. In these situations the supplier has power to raise prices. In a more

conscious supplier management approach, suppliers should be selected along with the goals of the information strategy, considering long term risks and benefits.

• Sourcing models: For reducing operation costs, business decision makers usually consider the option of changing the sourcing model for IT services, reorganising the IT department into an internal or external centralised service centre, or outsourcing the whole function. Although these options can result a significant cost reduction, based on the mass-efficiency, these approaches cannot provide a universal solution. Partly or fully outsourcing services, or centralising function are mostly efficient in case of services that are easy to standardise, and flexibility is not the first priority. Moreover, the results of these models in cost will appear only in medium term, although these cost cuttings have long term impact. The challenge of these decisions is that outsourcing of an unorganised internal operation cannot be changed in a few days, but it can cause additional costs in short term. To prevent this trap, prior changing the sourcing model process and service optimisation is necessary.

• Reducing losses, managing risks: Operational risks can cause serious losses for every organisation, and IT services are strongly vulnerable. Indentifying and assessing risks is a good basis for risk management efforts, to prevent future losses. These efforts should consider the costs of risk management against expected losses, and approaches should target the main risk categories.

• Operation optimisation: Organisations maintain several IT services that are rarely used, and do pay for licenses and supports that are never used. In order to reduce costs the assessment of these services should be performed, analysing the rate and intensity of usage, business impact of a service, and real business requirements. Services, and the required infrastructure can be evaluated based on these assessments, and existing capacities can be used for more important services. Moreover, based on the newest technologies, like virtual servers, or centralised services, additional costs can be spared.

B. Development costs and innovation (long term impacts)

Comparing to operations, developments concentrate on creating new projects for business advantage. Developments are organised and managed in projects, with dedicated resources, time limits and clear expectations. Unfortunately the success rate and effectiveness of IT projects is rather challenging in every industry, and with more conscious project management efforts, significant cost reduction can be achieved. During downturn, the risk of stopping or reducing the number of development projects becomes high that results static IT and business capabilities. Main areas for a more conscious management of innovation are the following:

• Specify business demands: Changing business requirements during a project, especially in the second part of projects can result in significant and increasing level of additional costs. Business requirements should be explored and recorded in the early phases of a project, to avoid unnecessary changes. Unfortunately, decision makers try to reduce projects cost on reducing the resources and time for these specification efforts. A good requirement specification can reach 25-35% of project costs and 30-50% of project time. As development reaches newer and newer stages, any change in the requirements results in exponentially increasing costs, and can result a multiple level of project costs and time comparing to the original baseline.

• Conscious project management: There is a strong inverse correlation between project success and project complexity (project length, developed functions, impacted units, etc.). Reducing project risk (and costs) is to reduce the complexity of the projects through dividing project goals into independent segments, and managing more, but smaller projects. Another tool is to prepare regular milestones to check project health. With these simple efforts, projects become more controllable, and problems can be identified earlier, giving a chance to intervene.

• Unnecessary developments: A regular problem is the development of already existing solutions in organisations. The problem is not counting the existing capacities and capabilities. Why should a billing system be changed, if the existing one is working perfectly? A telecom company wanted to acquire a service management tool in a development project. The company selected a well-known, international supplier, and during the preparatory phase this supplier suggested to check the existing software the telecom company already has, because a few years ago they bought already licences. The supplier manager did not know this fact, and after checking the existing licences that can be also used for service management, the telecom company had to spend less on this procurement.

• Project business case: Although IT developments are usually hard to justify, and in order to secure development budget both business and IT departments have to provide clear calculations about the business costs and expected benefits of developments. Although calculations cannot be perfect, efforts on creating a business case provides a better understanding of business expectations and IT

possibilities, and helps to filter the not-viable solutions. This approach is a proactive risk (and cost) reduction solution.

• Open source solutions: Several companies exclude the use of open-source solutions that could cost only a fraction of other tools. Using open-source solutions does not mean free solutions, but considering the TCO (total cost of ownership) calculations, and the expected functionality, it can be a good alternative of traditional programmes.

IV. Research Background

This paper presents the findings of a research on IT challenges, performed among the Hungarian IT service management applying companies in the first half of 2009 and an identical questions in the first half of 2010, concentrating on downturn issues. Research data were collected through anonymous questionnaires, and complemented with case study findings. The questionnaires were filled voluntarily, mostly by mid-level IT executives and IT employees.

The target sample consists of organisations that have critical IT infrastructure, and the do plan or currently use IT service management solutions, therefore they are valid representatives of the Hungarian information management practitioners. 15% of the organisations are micro or small enterprise, 20% are medium-sized enterprise, and 65% of them are large-sized enterprises. Organisations both represent the public sector (35%) and the private sector (65%). 57 companies filled the questionnaire. While the answer rate is relatively high, because of the objective sample size, it consists several limitations

V. Roleof ICT Solutions Duringthe Crisis in Hungary

The presented results related to the current situation in Hungary were collected via anonymous questionnaires, and complemented with case study findings. The surveyed sample is consists of IT intensive organisations, which has a high level of consciousness of managing IT solutions. Organisations currently use or plan to use IT service management solutions, and were selected mostly from the members and participants of the Hungarian IT Service Management Forum. These organisations are a good representation of Hungarian IT practitioners. The questionnaire was filled by 68.9% of the ITSMF members, mostly by ICT managers.

Organisations both represent the public sector (35%) and the private sector (65%). The research was conducted in 2009, concentrating on the innovative role of ICT, and the impact on applying ICT solutions. 15% of the organisations are micro or small enterprise, 20% are medium-sized enterprise, and 65% of them are large-sized enterprises.

A. Role of ICT solutions

Similarly to the international trends, the perceived, expected and real role of information technology is observed differently. Respondents mostly believe that IT can provide the basic IT services in an adequate level, and also believe that IT is able to support effectively the business innovation. Taking into account, that respondents are mostly IT people, this self-confidence seems to be very natural, although the average of the given points (3.6 on the scale of 1-5) from this view could be even higher.

The survey indicates significant problems in the relationship of IT and business areas. From an IT point of view, this relationship was strongly questioned: there is a problem, that business areas cannot (or will not) understand the role and offerings of IT. This can be one of the sources, why innovative solutions from IT departments are not accepted and supported, and why the efficient partnership between IT and business in lacked (Fig. 5).

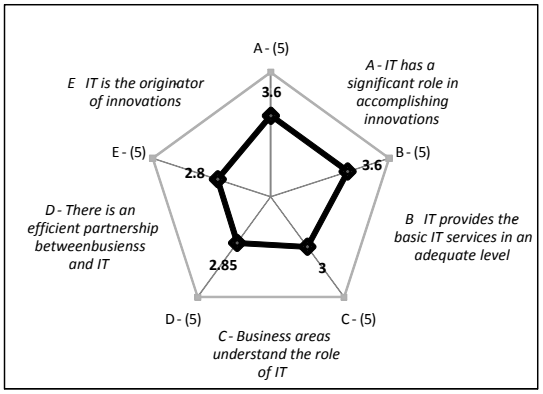


Figure5. Role of IT (Research Findings: Average ranking 1 to 5, where 5 = strongly agrees)

IT functions of the Hungarian organizations can be therefore categorized in the role of utility players, or trusted suppliers at most, and – except few examples – are not able to reach the business partner rank, if we follow the Forrester's categorization of the IT service providers [20]. It also means that researched organizations belong to the first or late majority user category.

Analyzing these results are in accordance with the international experiences: in an international survey, in which the initiator role of innovations were analysed, IT function was ranked only as fifth, below Sales, Marketing, Operations and Research & Development, and only ahead of Finance.

Interestingly as the crisis evolved the situation did not worsened, but got better. The research presented that business partners try to understand more the role of IT, and build a stronger relationship between

business and IT areas. A TOP5 bank CIO reported this changing situation, as a real advantage of the crisis. Because IT is a strong cost factor, business leaders have to explore and understand this area that requires deep discussions with IT leaders. Similar experiences were found in other big financial institutions.

B. Changes in the IT budget

IT budgets of Hungarian organizations (and also internationally) varies between 3-5% of the revenue, but organizational specific values strongly depend on the industry, and the size of the company. Hungarian survey showed and increasing trend in the size of IT budgets (starting from 1% in 1996, and reaching 4% in 2008) [5] [6] [7]. The average budget size of the researched organizations was almost 6.5% of the revenue that is a high share comparing to other Hungarian and International surveys, but acceptable result in the case of IT intensive organizations.

Of course there are differences between organizations: the lowest rates (0.1% - 0.5%) were indentified in the public sector and in the constructions industry, and the highest rates (5-20%, even above 20% of the revenue) belong to the financial sector. The question is how the budget was changed, because of the crisis?

Survey results indicate that Hungarian organizations have to face serious problems of a significantly decreasing budget: only 20% of the survey respondents reported increase in their budget, while other companies reported decrease. In the case of about 40% of the organizations reported more than 20% decrease in their budgets, and the average decrease of organizations were around 11%, strongly above the average decrease of the Hungarian GDP.

Hungarian organizations considered the year of 2009 as a critical year, but significant changes were not expected for 2010. The decrease of the IT budgets was pressured IT departments and IT leaders the rethink the current practice. Organizations reacted as cutbacks in the HR budget (stop hiring new employees, decrease salaries and benefits, even layoffs), and stopping developments, especially postponing procurements, and stopping consultancy and development projects.

Although IT executives believe that developments cannot be totally stopped, this budges became very limited. Development budgets of the Hungarian organizations were relatively high with around 45% share of the total IT budget, comparing to the international average of around 25% [3]. Therefore there is space for cost reduction, but there is a risk of losing important capabilities during these years.

Although development budgets are strongly decreasing, R&D budgets are still open for organizations, and used as additional resources even for ICT topics (it does not include direct development, but can cover experimenting and prototyping). This trend is also valid in an international environment: around 75% of the organizations having loss reported increase in their R&D budgets [10]. To motivate these really innovative projects, the Hungarian government motivates these efforts, although most of the organizations are still experimenting these new possibilities.

Considering the pressure of the crisis, the decreasing IT budgets, and the low considerations of the innovative role of IT, these R&D projects create a real opportunity for IT departments to show and develop their innovative capabilities.

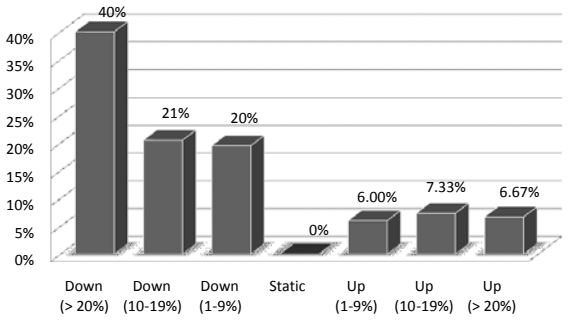


Figure 6. Changes of the IT budget of Hungarian organizations because of the crisis (years 2008 to 2009)

C. Organisation clusters

Based on the perceived role, and opinions about information technology, three clusters, three groups of companies were identified:

- *Trusted suppliers of basic IT services:* The business side understands the role of information technology, and the relationship between the two sides is close and good. Inthese companies the share of IT development budget is low (~20%), even comparing to the international average.
- *IT innovators*: Information technology has an important role in delivering and even initiating business innovations. In these cases the share if IT development budget is high.
- *Average companies*: These companies do not show any advanced performance, neither in innovations, nor in delivering basic IT services.

The identified groups are similar to Forrester's company categories, although there are differences (trusted suppliers, partner players, solid utility) [26].

During the research the change of IT budget was analysed in these groups. Companies, that are able to

deliver basic IT services reliably could increase radically (>20%) their IT budget, although their innovative capabilities are rather low. The emphasis is on reliable operations that have a high importance even during crisis. These companies are able to increase their IT development budget.

IT innovator organizations had to suffer radical cuts in their IT budgets (~ -20%). Although these organizations have an important role in business innovations, their high development budget is a main target of cost-cutting efforts, to support cash-flow management.

Average companies are average in every aspect: these organizations have to face with a decreased IT budget, but close to the average level (~-12%).

Although during normal economic periods the innovation ability has higher importance, during this economic period innovation is more a risk than an asset.

TABLE I. ORGANISATION CLUSTERS (RESEARCH FINDINGS)

	IT Innovator	Avarage company	Trusted supplier
IT Development budget	High	High	Far below avarage
Innovative role of IT	High	Avarage	Low
Evaluation as supplier of basic IT solution	Above avarage	Avarage	High
Partnership between business and IT	Above avarage	Low	High
Change of IT budget	Decrease above average (~ -20%)	Average decrease (~ -12%)	High increase (>20%)

VI. RESEARCH FINDINGS ON COST REDUCTION

The practice of information technology is mostly one the first areas, where executives try to start cost cutting [23]: providing IT and business services demands high operational costs, but the return of IT services is hard to measure. IT departments continuously face the extreme challenge of showing the value of their work, but failing this challenge means, that business customers perceive information technology as a rather expensive, but mostly useless area, and results increased interest of cutting IT costs.

Based on our research findings the crisis already reached the IT departments. Every re-searched organisation reported several measures in the area of IT to reduce costs. Other surveys [25] already indicated preliminary cost reduction activities: The CIO Executive Board polled its members, 59% of the surveyed CIOs already started negotiations to change parameters of the existing supplier contracts, and 61% of them started to rethink the actual IT budget.

A. Short term cost reduction

Our survey shows, that first measures concentrate on stopping as much thing, as possible, to reconsider the status and possibilities of the IT Department. IT departments identified the huge cost group of human resources: the most popular measure was stopping to hire any new employee to the IT departments, indicated by over 68% of the respondents, but layoffs started only in less than quarter of the companies, and only less than 18% of the organisations started to reduce the salaries of IT employees. Although salaries and other allowances create huge cost for organisations, companies plan to wait, instead of intervening in this area. If further short term plans are analysed, measures on human costs are the least wanted actions, comparing to other activities.

The second category of cost cutting concentrates on development projects, large enterprises already realised the importance of clear business cases to justify development projects, but the crisis catalysed these efforts also in other companies. Almost 65% of the surveyed organisations already prepared new rules for project decisions, and with the short term plans almost every company (94.12%) applies more strict policies.

The crisis has strong impact on a TOP5 Hungarian Bank. Because of the organization-wide cost cutting expectation (driven by the multinational parent-bank), several investments and developments had to be stopped. It has serious impact on the IT department, because a deep transformation project was on the way, that expected to have impact on the performance and long term operation costs of the whole IT department. The IT department managed to accomplish several short term cost cutting actions successfully, but whole, and part budgets were decreased. The bank has chosen the surviving strategy: beside ongoing developments, every new innovation was stopped. These measures have short term impact on costs, but in medium term the expected solutions will be missed, and will have negative impact on incomes. Bank employees hope that after these times of rigorous budgets, with newer solutions they will be able to close the gap that arises in the following months and years.

Following the international trends, CIOs concentrate on postponing procurements (especially hardware and software), stopping or limiting already running developments. Almost half of the respondents reported postponed procurements, but together with short term plans almost 95% of the companies is using this tool. Almost the third of the

organisations stopped or limited already the development projects but with the short term plans almost 90% of the respondents use this approach.

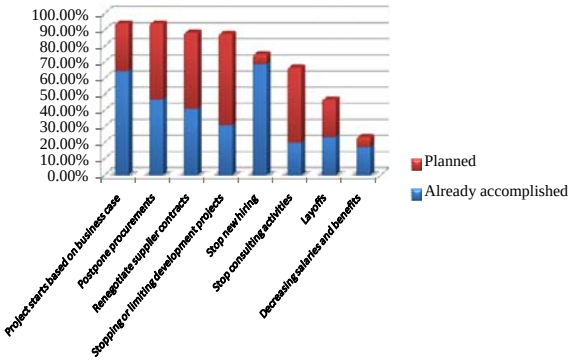


Figure 2. Planned and already accomplished short term crisis measures (Research Findings)

Renegotiating contracts and future plans with suppliers and consultants requires the cooperation of both parties. Suppliers can try to resist to customer demands, but risking their long term partnerships, and the renewing of their contracts, therefore they are leaning to-wards finding a compromise. A little more than 40% of the companies already started the negotiation, which shows that Hungarian companies are behind the international trends (comparing to the international 58% at the end of 2008). Considering the planned actions, almost 90% of the companies will start negotiations with suppliers in this year, which makes the third popular measure in this survey.

Consultants are in a little better position: Only 20% of the companies suspended the consulting projects, that mirrors the results of another international survey [22], in which consultants were evaluated as the most value adding players among external IT suppliers, while contractor agencies and hardware providers were the least recognised among CIOs. Consultants cannot be happy, because with the short term plans, altogether 55% of the companies try to cut costs on them.

B. Long term plans

In longer terms, organisations consider several other options that require a more conscious preparation, and have no immediate impact on costs. Following the pre-crisis trends operation optimisation efforts are the most popular solutions: organisations continuously look for new technologies and solutions to exploit their existing infrastructure, or to achieve higher service levels with lower costs. A little over 72% of the respondents plan to concentrate on operation optimisation.

Process management, especially BPR was one of the magic concepts in the 1990s [3] [9]. The approach passed the peak of inflated expectations and finally reached its maturity to be a productivity tool. Unfortunately process management creates a less flexible, but more controllable environment. In IT departments employees are not good in following rules and policies, but the complexity of IT departments demand these efforts, and more than 60% of the companies plan to investigate this opportunity, especially in the IT department.

The development of IT Service Management practices is in the mind of more and more Hungarian IT employee, especially if it is a basic expectation of employment. Unfortunately in practice only a few key processes are used daily, especially in the field of operations. Although the international best practices (ITIL, MOF, eSCM, etc.) cover the whole life-cycle of IT activities covering IT strategy, planning, developments, deployment, operations and improvement Hungarian companies mostly concentrate on operations (service desk, incident and problem management, demand management, etc.), and developments are slow. Although these challenges, half of the companies plan to expand or start their IT service management practice, expecting a more conscious and cost effective IT operation.

The crisis increased the interest on alternative sourcing models. The interest shown in this survey will not result change in the sourcing models necessarily, but the interest of almost 45% of the respondents show, that in the current economic situation companies do look for alternative solutions, especially to share the risk of operations.

Project management issues were already mentioned among the short term cost cutting efforts, but as third of the respondents see, more conscious project management efforts have also long-term impact on cost and effectiveness. Efforts include creating centralised project offices, project management teams, methodologies, trainings and support tools.

At the end of the list 16.7% of the respondents plan to consider open-source investments, partly because the promising total cost of ownership (TCO), and partly because of the reliability of special applications that became industry standards.

Partly surprising, that plan for IT risk management is the least important issue, but we should take into consideration, that IT risk management, especially IT security issues are already part of the daily practice in almost every organisation at least in a basic level, and in more than 40% of the respondents in a conscious level, and companies do not expect further cost reductions because of these actions.

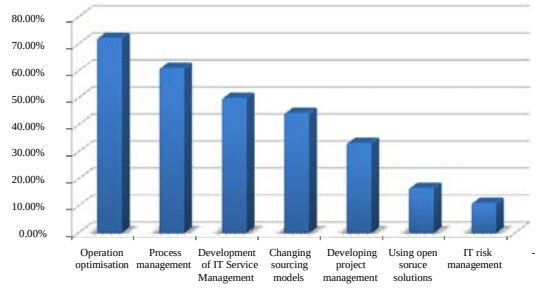


Figure 3. Planned actions to reduce cost in long term (Research findings)

VII. SLOW RECOVERYFROMTHE CRISISIN HUNGARY

As the tendencies show, 2010 is expected to be the end of the crisis, although there are differences between world regions. The dynamic countries of Asia can continue their raid economic growth, while Europe is less optimistic: although the economic growth can be started again in North America and Asia, it has only little impact on European economies, because European countries have to deal with the unbalanced national budgets, that has strong impact on private organisations: increased or industry specific taxes (e.g. for the financial sector), stagnating customer demand postpone the arrival of a more suitable economic environment [17].

The situation seems to be promising in Hungary: Although companies still report decrease in their annual revenue, it value is better comparing to last year. The positive economic changes have impact on IT budget. By 2009, 80% of the organizations reported decrease, by 2010 this value decreased to 58%. The average change of IT budget is still negative: - 4.85%

Organisations that were able to increase their annual revenue had more chance to increase their IT budget (r =0.546). While in the first year of the crisis budget cuts had impact on IT development budgets, by 2010 budget increases appeared in this category (probably partly compensating last year).

The research identified the following factors, or characteristics that have impact on securing additional budget for delivering IT solutions (based on Pearson correlation matrix, with minimum of 0.05 level of significance):

- There is an efficient partnership between IT and business (0.693)
- Business and IT strategy are strongly integrated (0.546)
- IT should deliver reliable basic IT solutions (0.461)
- IT has an important role in differentiating in competition (0.475)
- Business areas should understand the role of IT (0.433)
- IT has an important role in delivering business innovation (0.416)

Although several factors were identified, neither factor is determining IT budget alone.

TABLE II. CORRELATION MATRIX FOR ANALYSING IT BUDGET (RESEARCH FINDINGS)

		K1-R. Size of IT Budget, comparing to annual revenue	K2-R-OP. Share of IT Operation budget	K2-R. DEV. Share of IT development budget	K3a-R. Change of IT budget (2009 → 2010)
cK3a-R.Change of IT Budget (2009→2010)	Pearson Correlation Sig.(2-Tailed)	.474* .042	-.464* .030	.464* .030	.546** .003
K1-R. Size of IT Budget, comparing to annual revenue	Pearson Correlation Sig.(2-Tailed)	1	-.269 .265	.269 .265	.221 .349
K2-R-OP. Share of IT Operation budget	Pearson Correlation Sig.(2-Tailed)		1	-1.000** .000	-.596** .003
K2-R. DEV. Share of IT development budget	Pearson Correlation Sig.(2-Tailed)			1	.596** .003

* Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

VIII. CONCLUSION

Although the identified long term cost cutting plans can be independent of the current crisis, the economic downturn pressured organisations to rethink their current practices, and to speed up the promising developments. IT departments decided on several short term solutions, navigating themselves into a stasis, and waiting for better times. Only a few organisations were able to see the big picture and try to harmonise business and IT goals.

Beside short term cost cutting, another TOP5 Hungarian Bank started a long lasting cost reduction project. In the first part of the project financial data is analysed in order to identify possible overspendings, comparing to international and local benchmarks. In the second step (partly parallel with the financial analysis) international best practices are collected for cost cutting, and not only for IT departments, but also for the business units, but concentrating on IT solutions. In the final step the results, demands and possibilities are synthesized and clear development priorities are set. In this case the business and IT side should work strongly together to identify the role of information technology in business processes that is a key for long term cost reduction, while maintaining user satisfaction.

Environments, where operational costs have a high share in the IT budgets, short term solutions have only limited impacts. In order to achieve enduring results, long term solutions (process management, optimisation, IT service management)

efforts should be applied.

Beside the popular cost cutting efforts, the value creation capability of IT departments should be also taken into account. If costs are reduced along with decreasing performance, there is no benefit on company level. In order to understand these benefits, the business areas (and so the IT department) should understand the value chain and the cost structure of IT supported business services.

The forthcoming months and years provide a great challenge for IT departments and CIOs: operational efficiency and showing business value was never so important as it is now. IT departments should achieve results in a climate where cost cutting is a priority, and the business side leans to forget the importance of IT services and innovation.

REFERENCES

[1] P. Biggs and S. Rollet, *Confronting the Crisis – Its Impact on the ICT Industry*. International Telecommunication Union, 2009.

[2] G. Bőgel. (2009, Feb. 24). "Vezetés válság idején," in *IT lábjegyzet*.Available:http://www0.kfki.com/blog_8E210F31B2AC45668D7C83CB0FFFFB33.php?blogId=1269%234c1e7e3# (accessed March 15, 2009.)

[3] S.Cramm, (2008, Nov. 20). "Smaller IT Budget? Pursue Value-Driven Development," in*Having IT your way – HBR Voices*.Available:http://blogs.harvardbusiness.org/hbr/cramm/2008/11/valuedriven-development.html (accessed June 23, 2009)

[4] T.H. Davenport and J.E. Short, "The new industrial engineering: Information technology and business process redesign," *Sloan Management Review*, pp. 11-27. Summer, 1990.

[5] G. Drótos, K. Gast, P. Móricz, and G. Vas, "Az információmenedzsment fejlettsége és a versenyképesség," in*Versenyképesség Kutatások*,Working Paper No. 28.

[6] G. Drótos and Z. Szabó, "Corporate Information Technology in Hungary at the Millennium: Myth and Reality," in *Chikán Attila - Czakó Erzsébet - Zoltayné Paprika Zita: National Competitiveness in the Global Economy*, *Akadémiai Kiadó*, pp. 248-263, 2002

[7] G. Drótos and Z. Szél, "Mennyit költünk az informatikára ? " in A. Mártonffy, Ed., *IT Business*, vol. 27, p. 24, Jan. 24, 2009.

[8] A. Gábor, "Tudja-e Ön, amit tudnia kell? Áttekintés a tudástársadalomról és a tudásgazdaságról, " *Társadalom és Gazdaság*, 24/2, pp. 189-226 , 2002.

[9] M. Hammer, "Reengineering work: Don't automate, obliterate," *Harvard Business Review*, pp. 104-112, July-August 1990.

[10] B. Jaruzelski and K. Dehoff, "Profits down, spending steady: The global innovation 1000," *Strategy and Business*, Issue 57, Winter2009.

[11] J. M. Kaplan, R.P. Roberts, and J. Sikes, "managing it in a downturn," *McKinsey Quarterly*, pp. 66-71, Fall 2008.

[12] K. Ala-Mutka, P. Gáspár, K. Gábor, V. Vehovar, and M. Suurna, " Status and developments of eLearning in the EU10 member states: Cases of Estonia, Hungary and Slovenia," *European Journal of Education*, to be published, 2010.

[13] B. Kovács and R. Vas, "Az informatika gazdasági, társadalmi hatásai, jogi és etikai vonatkozásai," in A. Gábor, szerk, *Üzleti informatika*. Aula Kiadó, Budapest, 2007, pp. 285-307.

[14] A. Kő and G. Klimkó, "Towards a framework of information technology tools for supporting knowledge management," in E. Noszkay, Ed., *The Capital of Intelligence - the Intelligence of Capital*. Foundation for Information Society, Budapest, 2009, pp. 65-85,

[15] KSH-EITO, *Az infokommunikációs technológiák (IKT) szektor iparpolitikai akcióterve* - Nemzeti Fejlesztési és Gazdasági Minisztérium, 2009.

[16] S. Marwaha and P. Willmott, " Managing IT for scale, speed and innovation," *McKinsey Quarterly*,pp14-21. Fall, 2006.

[17] McKinsey, "Economic conditions snapshot," McKinsey *Global Survey results*, June 2010

[18] P.F. Nunes,C. Firstbrook, and J.M. Ellis, "New choices for new challenges,"*Outlook*, no.1, pp. 1-9, January 2009.

[19] D. Rhodes and D. Stelter, "Seize advantage in a downturn," *Harvard Business Review*,pp. 50-58,February 2009.

[20] P. Roehrig, *New Market Pressures Will Drive Next-Generation IT Services Outsourcing*. Forrester Research Report. October 9, 2008.

[21] S. Sacks, "Ten steps to modifying user behavior and reducing IT costs,"*Computerworld*, October 25, 2004.

[22] O. Sevillia, P. Nannetti, and E. Monnoyer, "the role of the it function in business innovation, innovator vs operator," in *Global CIO Survey 2008*. Cap Gemini, 2008.

[23] S. Spang, "Five trends that will shape business technology in 2009," *McKinsey Quarterly*, February 2009. Available: http://www.mckinseyquarterly.com/Five_trends_that_will_shape_business_technology_in_2009_2296 (accessed: March 5, 2009.)

[24] P.J. Williamson and M. Zeng, "Value for money strategies for recession times," *Harvard Business Review*, pp. 66-74, March 2009.

[25] B. Worthen, "Tech budgets become a victim of the credit crunch," *Wall Street Journal - Business Technology*, October 1, 2008.

[26] T. Dewett and G. R. Jones, "The role of information technology in the organization: a review, model and assessment," *Journal of Management*, vol.27, no 3, 2001.

[27] L. Bryan and B. Hoffman, "How managers should approach a fragile economy," *Mckinsey Quarterly*, September 2009.

[28] *i2010 High Level Group: Summary of: The impact of the economic downturn on ICT*, Issue no. 9, European Commission Information Society and Media, Feb. 24, 2009,

A Workflow-based Risk Management Approach

Giancarlo Nota, Maria P. Di Gregorio and Rossella Aiello

***Abstract*- The management of operational risks is receiving increasing attention as a fundamental part of monitoring, controlling and decision support systems. However, organizations are reluctant to support risk management programs probably because of the high cost of human resources necessary for acquisition, manipulation and analysis of risk data. The paper presents an approach to continuous risk management which exploits the workflow management technology. The capability to measure executable process instances provided by a Workflow Management System (WfMS) is assumed as a major premise for the design of a workflow based risk management system. The benefits deriving from this approach are manifold: first, the responsibility of data collection necessary to risk evaluation at workflow execution time is assigned to the WfMS that can automatically record execution data. Second the workflow performance measurement capability can be exploited as a fundamental mechanism that can feed a framework of key risk indicators. Third, a proactive police can be planned to risk treatment at workflow execution time when there is still time to react to adverse events. In the following, the model of a workflow oriented risk management system, the fundamental performance measures of workflow and some example of workflow based key risk indicators are discussed.**

Index Terms*- *risk management; workflow management; workflow management system

I. INTRODUCTION

Every enterprise can be affected by risks with potential impact on their single organizational parts or on their organizations as a whole. The need to identify, assess, and manage risks has motivated organizations to develop integrated frameworks to improve Enterprise Risk Management (ERM) [1]. The literature about risk proposes various techniques to identify and classify risks in different fields of knowledge or descriptions of various innovative approaches for managing risks. For example, in [2] two approaches for managing risks are compared: tactical risk management and systemic risk management. Tactical risk is traditional, bottom-up analysis defined as a measure of the likelihood that an individual potential event will lead to a loss coupled with the magnitude of loss. In contrast to the bottom-up analyses employed in tactical risk management, systemic risk management approach starts at the top with the identification of a program's key objectives. Once the key objectives are known, the next step is to identify a set of critical factors, called drivers that influence whether or not the key objectives will be achieved.

In order to minimize the impact of risks, Enterprise Risk Management framework typically includes four major areas corresponding to the achievement of enterprise objectives:

- *Strategic*: high-level goals, aligned with and supporting its mission
- *Operations*: effective and efficient use of its resource
- *Reporting*: reliability of reporting
- *Compliance*: compliance with applicable laws and regulations.

Many organizations are reluctant to support risk management programs, probably because of the high cost of human resources necessary for acquisition, manipulation and analysis of risk data. However, the management of operational risks is being given increasing attention as a fundamental part of monitoring, controlling and decision support systems because of the opportunity that Workflow Management Systems (WfMS) provide in terms of automatic collection of business process execution data.

The lack of models and systems in the field of real time management of operational risks encourages new research activity. In this paper we propose a model that integrates WfMS and Risk Management System (RMS) functionalities in order to represent operational risk management. The process oriented approach to continuous risk management, based on a top level model for the representation of qualitative and quantitative risks, is able to reduce effort and cost necessary to implement a risk management program. The capability to continuously measure executable process instances provided by a WfMS is assumed as a major premise for the design of a workflow based risk management system. We will show how the typical WfMS capabilities, in terms of process enactment and performance evaluation, can be represented within an augmented model that integrates WfMS capabilities and continuous risk management aiming at the monitoring and control of operational risks. The benefits deriving from this approach are manifold: a) the cost reduction for the risk management systems due to the automatic process execution data recoding provided by the WfMS; b) the definition and management of qualitative and quantitative risks within the unifying framework of process management; c) the definition of a proactive policy for the treatment of operational risks.

ISBN 978- 974-672-556-9

II. WORKFLOW APPROACHTO ERM

This section introduces the rationale and the building blocks of an approach for the design and implementation of a workflow oriented risk management system. When the management decides to follow a risk management program, one of the hurdles hindering the success of such initiative is that many roles, e.g. business administration or IT, perceive different views of risks [3]. This separation is mainly due to different goals pursued by different roles [4], [5]. From one hand, management roles adopt, more or less consciously, a system thinking approach [6] to the understanding of organizational structures, processes, policies, events, etc. This approach allows, once business processes have been designed and implemented, to monitor them at a high abstraction level; watching at 'the big picture' and transcending organizational boundaries, the manager focuses himself on business goals and on risks that could threat their achievement. On the other hand, operational roles have a completely different view of risks. For example, IT personnel are usually concerned about how data and information can be stored/retrieved and how to provide access to ICT services over the organization's 'digital nervous system'. In this case, the perception of risks mainly concerns the availability and performances of communication/database systems, application programs, access policies, etc.

As pointed out by Leymann and Roller in [7], workflow technology helps to bridge the gap between these different views of business processes because: a) management roles typically look at the process models and at their execution instances eventually asking for execution data to evaluate the process performance, b) operational roles implement process activities and perform them with the support of a workflow management system.

The model shown in Fig. 1 represents an integrated system aiming at the management of operational risks in a context where processes are enacted with the support of a workflow management system.

The process management subsystem comprises the usual tools for process definition, process instance creation and execution as well as maintenance services. One of the most appealing features of workflow management systems is the measurement capability offered by this class of products. Both research and industrial applications are mature enough and provide measurement tools concerning workflow measurable entities [8], [9]. Several kinds of duration measures about activities/processes, waiting queues, produced deliverables and human resource efforts are frequently evaluated and can provide quantitative knowledge about business processes. However, current workflow products do not take into account risk management. Indeed, the workflow log collects automatically raw execution data that can be used for process monitoring and performance evaluation. These log data are invaluable to lay out a process oriented risk management system.

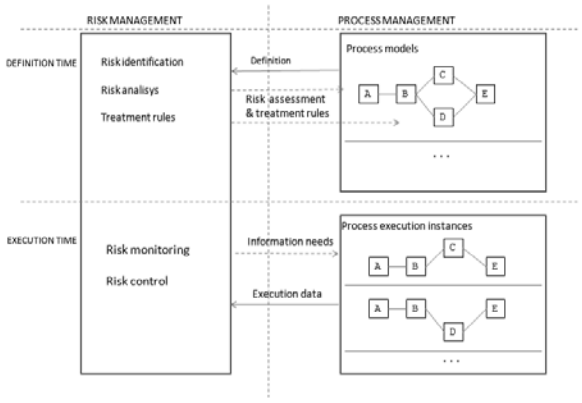


Figure 1. Top level model for process oriented risk management

The premise behind the process oriented risk management system is similar to other widely accepted approaches to assessment and measurement: there exists information need that, when satisfied, increases the decision capability.

A widely accepted approach to project measurements in the field of software engineering is GQM (Goal-Question-Metrics) [10], [11]. The GQM approach is based upon the assumption that, an organization must first specify the goals for itself and its projects in order to measure in a powerful way. Subsequently the organization must trace the goals and the relative operational data and finally provide a framework for interpreting the data according to the stated goals.

Another well-known method for software measurement is PSM (Practical software and systems measurement) [12]. It describes an approach to management based on integrating the concepts of a Measurement Information Model and a Measurement Process Model. A Measurement Information Model defines the relationship between the information needs of the manager and the objective data to be collected, commonly called measures. The Measurement Process Model describes a set of related measurement activities that are generally applicable in all circumstances, regardless of the specific information needs of any particular situation and provides an application.

From the point of view of the risk management system, there exists an information need about process instances that a WfMS can help to satisfy. The left side of the model shown in fig. 1 describes how a risk management system can be integrated with a WfMS. At definition time, when the process model is established, risk data are stated and relied to the process model. Note that the risk statement can

be relied to both process and activity. This choice reflects the different process perspectives that managers and operational staff have on processes. Managers look at the process level and think in terms of risks at this level in order to provide support for continuous monitoring of risks deriving from the execution of workflow instances.

III. WORKFLOW QUANTITATIVE MEASUREMENT

A risk assessment methodology normally comprises a combination of qualitative and quantitative techniques. Management often uses qualitative assessment techniques where risks do not lend themselves to quantification or when sufficient reliable data required for quantitative assessments are not available. Quantitative techniques typically bring more precision and are used in more complex and sophisticated activities to supplement qualitative techniques [13].

Starting from these premises, we build on the top level model for process oriented risk management shown in fig.1 to determine quantitative and qualitative measures inspired by the GQM approach applied to the domain of business processes and in compliance with the three layer PSM measurement model.

Before discussing the method that faces with the quantitative approach, we briefly introducethe open procedure "call for tender" managed by a public agency, that will be used as a case study. This procedure, whose BPMN model is shown in fig. 2, selects the provider of goods and services on the basis of award criteria stated in the tender specifications.

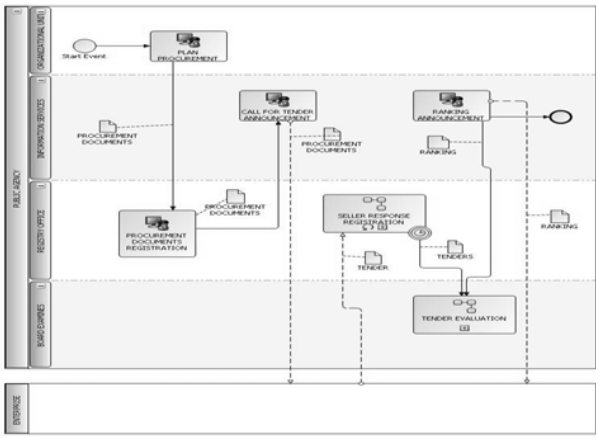


Figure2. Call for tender process BPMN model.

The procurement documents, as contract notice, tender specifications or invitation to tender, are first sent to the Registry Office that proceeds to a formal registration. Then, the Information Services organizational unit publishes the call for tender announcement enabling the interested enterprises to download the procurement documents. The Registry Office awaits the incoming request to participate until the time limit for receipt of tenders is reached. Afterward, the Board of Examiners is involved in the sub-process of "tender evaluation" that produces the ranking to be published by the Information Services.

Considering the case study discussed above and following the GQM approach that defines in a top down fashion Goals, related Questions and Metrics, in the scenario of WfMS supported business processes we could be interested to obtain general goals stated in terms of *efficiency*, *effectiveness* and *control costs*. These goals are then refined into process oriented queries that, in their turn, are related to metric in order to provide a precise evaluation about the degree of goals achievement.

Goals

G1. *efficiency*: the comparison of what is actually produced or performed with what can be achieved with the same consumption of resources (money, time, etc)

G2. *effectiveness*: the degrees to which objectives are achieved and the extent to which targeted problems are resolved.

G3. *control cost*: the application of investigative procedures to detect variance of actual costs from budgeted costs, diagnostic procedures to ascertain the causes of variance and corrective procedures to effect realignment between actual and budgeted costs.

Questions

Some typical questions addressed by an analyst during the process evaluation are [14]:

Q1. What is the duration of a given task instance of "tender evaluation"? (G1)

Q2. What is the global throughput (process stared and completed) over the past year? (G1)

Q3. How many work items has a given employee completed? (G1)

Q4. How many procurements have been done with respect to the procurement plan? (G2)

Q5. What is the exception rate in the WfMS after the deployment of processes? (G2)

Q6. What is the average cost of "call for tender"? (G3)

Q7. How much is the difference between the planned costs and the real costs of a process instance? (G3)

To obtain precise answers to the queries such as those above, we need to develop a measurement framework by means of which numbers can be assigned to the various entities represented within the WfMS. The following examples are representative of a three levels measurement framework: primitive, fundamental and derived measures whose complete definition can be found in [15]. It will be used as a

fundamental model for a risk management system based on workflow execution data.

Two primitive operators for measuring work and time are:

$$\#(x) = |x| \quad (1)$$

the cardinality of a set, and

$$\Delta(e_i,e_j)=\text{abs}(\text{time}(e_i)-\text{time}(e_j)) \quad (2)$$

the length of the time interval between the occurrence times of two events e_i and e_j . Let I be the set of process, task and work item instances and i a generic instance, i∈I. We assume that each instance, at a given time, can be in one among the states: `created`, `running`, `suspended`, and `completed`; furthermore a state transition is a consequence of a suitable event such as `completedInstance` that happens when a task instance is completed or when a process instance completes its last task. The *fundamental* measures arise from the composition of primitive operators. For example, by means of the operator Δ, it is possible to build different fundamental measures such as `instanceDuration` that evaluates the total duration of an instance from its creation to its completion.

```
instanceDuration(i)=
     Δ(event(i,e_type(i,e)=createdInstance)),
       event(i,e_type(i,e)= completedInstance))
```
(3)

`instanceDuration`can be used to answer the question Q1. The operator filter is the standard operator for the choice of elements from a set I, according to a first order predicate p:

$$\texttt{filter}(I,\ p) = I' \text{ with } I' \subseteq I \text{ such that:}$$
$$\forall i \in I \begin{cases} p(x)=\text{true if } i\in I' \\ p(x)=\text{false if } i\notin I' \end{cases} \quad (3)$$

The following example refers to the case study introduced in section 3. According to the predicate p, filter returns all the tasks instances named "`procurement document registration`" in the context of the process "`call for tender announcement`".

```
filter(I,p):
P=i_type(i)=task ∧
  i_name(i)="procurement document registration" ∧
  i_name(father(i))="call for tender announcement"
```
(4)

A frequently used fundamental measure evaluates the workload in the scope provided applying a suitable filter to the set of all workflow instances. Queries of this kind require the capability to isolate within the WfMS the set of objects with the desired properties and then to evaluate its cardinality. By the combination of the operators # and `filter` we define the measure `work`;

```
work(I,p)=#(filter(I,p))
```
(5)

The example below shows how the measure work can be applied to evaluate the question Q3.

```
work(I,p):
P=i_type(i)=task ∧
  i_name(i)="procurement document registration"∧
  i_name(father(i))="call for tender announcement"∧
  actor_name(i)= "Brown"∧
  current_state(i)= completed
```
(7)

The need of a derived measure (the third level of measured framework) becomes evident if we consider the evaluation of contribution that resources, especially human resources, make to the progress of a process. Given a process `P`, the contribution of the generic actor to `P` is considered. The evaluation can be done from the point of view of time overhead, work overhead or cost and is expressed in percentage.

In order to define some kind of contribution measures, it is necessary to introduce the auxiliary function sigma that is itself defined in terms of `sum`and map. `sigma` implements the concept of "summation of measures" where the input parameter measure gets as a value the measurement definition to apply to the elements of a set `X`. The function `sum`, given a set of values, returns the sum of all the members in the set.

```
sigma(measure, X)=sum(map(measure,X))
```
(8)

where `map` is a function that denotes the usual operator for the application of a function to a set of values

$$\texttt{map}(f\{x_1,x_2,\ldots x_n\})= \{f(x_1),f(x_2),\ldots f(x_n)\} \quad (9)$$

$$\texttt{timeContribution}(\texttt{timeMeasure},x_1,x_2)= \frac{\texttt{sigma}(\texttt{timeMeasure},x_1)}{\texttt{sigma}(\texttt{timeMeasure},x_2)}*100 \quad (10)$$

$$\texttt{costContribution}(\texttt{costMeasure},x_1,x_2)= \frac{\texttt{sigma}(\texttt{costMeasure},x_1)}{\texttt{sigma}(\texttt{costMeasure},x_2)}*100 \quad (11)$$

$$\texttt{workContribution}(I,p_1,p_2)= \frac{\texttt{work}(I,p_1)}{\texttt{work}(I,p_2)} * 100 \quad (12)$$

Care must be taken to specify the set $X_1$and $X_2$and the predicates p_1 and $p_2$since a proportion requires that the numerator is less than or equal to the denominator.

Let t_{actor_k}be the working time spent by the generic actor on `P`. In general, to `actor_k` can be assigned more than one work item even in the

context of a single process P. Given a process P, the actor time contribution (atc) of actor_k on P is

$$atc(P) = \frac{t_{actor\_k(P)}}{\sum_{j=1}^{n} t_{actor\_j(p)}} * 100 =$$

$$= timeContribution(workingDuration, filter(I,p_1), filter(I,p_2); \quad (13)$$

$$p_2 = i\_type(i) = workitem \wedge current\_state(i) = "completed" \wedge i\_name(father(father(i))) = "P";$$

$$p_1 = actor\_name(i) = "actor\_k"$$

Let c_k the hourly cost of actor_k; a particular case of(11) provides the definition of actor cost contribution (acc) of actor_k on a process P:

$$acc = \frac{t_{actor\_k(P)} * c_k}{\sum_{j=1}^{n} t_{actor\_j(P)} * c_j} * 100 \quad (14)$$

IV. Workflow Qualitative Measurement

Qualitative analysis is usually pursued relating likelihood and consequences of risks; a widely used model for this kind of analysis is *the priority-setting matrix* [16], also known as *risk matrix* where cells, representing fuzzy *risk exposure values*, are grouped in a given number of risk classes. In the matrix shown in Fig. 3, the risk exposure classes are represented by: **L** means low, negligible risk, **M** indicates a moderate risk, **H** a risk with high impact and probably high loss, and **E** represents the class of intolerable, extreme risk with very likely loss. Obviously, when the impact or likelihood grows, or both, the risk consequently grows; therefore a risk can modify its position from a lower category to an upper category. For each category of risk exposure, different actions have to be taken: values **E** and **H** involve a necessary attention in priority management and a registration in the Mitigation plan; a value **M** requires to be careful during the whole project management; a value **L** falls within ordinary management.

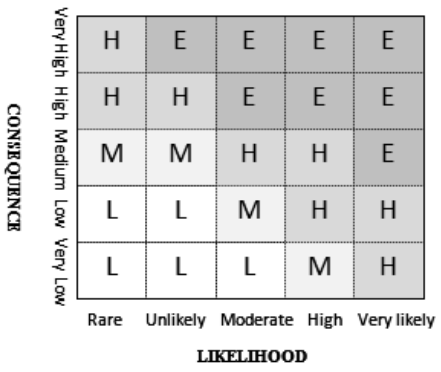


Figure 3. A risk matrix

The qualitative analysis is very useful either when a preliminary risk assessment is necessary or when a human judgment is the only viable approach to risk analysis. However, since a risk state(likelihood and/or consequence) might change continuously, the data collection about it is a time consuming activity often perceived as an unjustified cost. Another problem is the timing; if data are not collected according to a real time modality, they are of little or any value as the actions anticipated by the contingency plan could be no more effective. These considerations inhibit the implementation of risk management systems. The top level model for process oriented risk management suggests how, at definition time, the organization of questionnaires and checklists can be arranged. For example, within the scope of "call for tender", if we are interested in the following goals:

G4. *Transparency:*
- Lack of hidden agendas and conditions, accompanied by the availability of full information required for collaboration, cooperation, and collective decision making.
- Minimum degree of disclosure to which agreements, dealings, practices, and transactions are open to all for verification,

G5. *Impartiality:*
- Impartiality is a principle holding that decisions should be based on objective criteria rather than on the basis of bias, prejudice, or preferring the benefit to one person over another for improper reasons,

G6. *Correctness:*
- Conformity to laws,

then, the related questions and checklists can be:

TABLE I. Quality Assessment Specifications for the Tasksof "Call for Tender"

call for tender: quality assessment			
goal	*question*	*Checklist*	*Task*
G4	Q8.Are the full information available and published on the web site ?	[yes, no]	call for tender announcement
	Q9. Are the evaluation criteria for call for tenders complete and non ambiguous?	[poor, sufficient, good, very good]	plan procurement
G5	Q10. Are all tenders evaluated with the same criteria?	[yes, no]	tender evaluation
G6	Q11.Is the announcement compliant with the current laws?	[compliant, not compliant]	plan procurement
	Q12. Has the call been registered at the registry office?	[yes, no]	procurement registration
	Q13. Does the winner provide the right solution?	[poor, sufficient, good, very good]	tender evaluation

where we associate to each task a set of goals together with the corresponding set of questions (at least one question for each goal, according to the GQM approach) and a checklist that suggests the judgment to be expressed. Generally, the question is aimed at assessing a quality criterion and is evaluated against a list of fuzzy values such as `{compliant, not compliant}` or `{poor, sufficient, good, very good}`. Human judgments collected as soon as possible can feed the risk matrix. In other words, we can define task quality criteria whose satisfaction provides a contribution in the direction of quality goals for the task and in general for the whole process. When given criteria are not satisfied, the risk relied to the task increases and the task monitoring rules react raising the risk status and invoking the appropriate risk treatment. We will return on this point in the next section.

A WfMS usually provides a suitable definition and execution environment that allows with little implementation effort the set up of a subsystem devoted to the collection of qualitative process execution data. Indeed, applications for the exposition of questionnaires and checklist can be easily designed and implemented because usually the WfMS allows the launch of a complementary software application both at scheduling time and at completion time of a task instance, for example, the interaction with a questionnaire. In this case, the answers are collected and then stored in the workflow execution log feeding the part of the risk management system that has the responsibility for the monitoring and control of qualitative risks.

V. WORKFLOW ORIENTED RISK ASSESSMENT

To show how the top level model for workflow oriented risk management allows continuous operational risk management with respect to tasks and processes, consider the phases of a generic risk management methodology that encapsulates the concepts discussed so far:

- *Define the context*: goals, processes, stakeholders, evaluation criteria
- *Identify the risks*: what events can have an impact on tasks and processes?
- *Analyze the risks*: state the likelihoods, consequences, measures, thresholds, prioritization
- *Write the contingency plan*: define the approach – avoidance, minimization, transfer-about risk or a set of a related risks
- *Monitoring*: collect qualitative and quantitative execution data, acquire risk status and record it, evaluate risk indicators
- *Control*: decide for the best reaction when the risk probability increases or when unwanted events happen
- *Communication:* is a cross activity in the sense that data or information handled by a certain task/process can be communicated to the involved stakeholders.

To be useful a sound risk management system must be reactive; in other words, it must provide real time responses to unwished events that might happen in an unpredictable way. To specify the behaviour of a risk management system charged to manage events with a possible negative impact on the correct execution of tasks and processes, we shall use a rule based logic language called RSF [17], [18]. With this language a reactive system can be defined in terms of event-condition-transition rules able to specify systems requirements subjected to temporal constraints. A shown in fig. 4, at risk definition time the risk manager has the possibility to access the process model database in order to link behavioral rules to tasks and processes that state how to react when the risk exceeds a given threshold.

At process execution time, critical task or process attributes are evaluated against the measurement framework and/or the risk matrix discussed in the previous sections. Then, if the current risk state is acceptable the process enactment proceed regularly, otherwise the dangerous situation is immediately notified at the appropriate responsibility role, e.g. the task executor, the process owner or the risk manager.

At each time, the risk management system records a state concerning various kinds of data about risks. When an unwished event with a negative impact on an activity is recognized, the system reacts adjusting the state and eventually taking some risk treatment action.

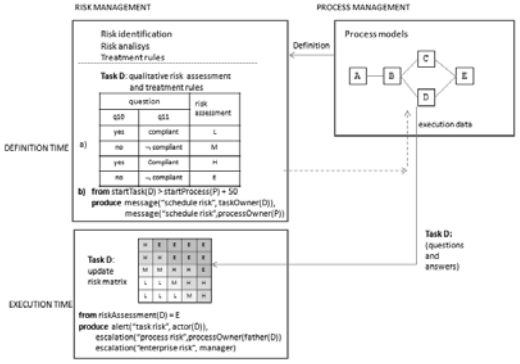


Figure 4. Relations between process management andrisk management

At risk definition time, as shown in Fig. 4, the risk manager defines a questionnaire containing, for example, two questions q10 and q11 (cfr. the case study "call for tender") and establish four risk assessment values for the task D. At execution time, when D completes its execution, the workflow engine presents the questionnaire to the user, collect the answers and send them to the RMS in order to

associate the appropriate risk status for D depending on the collected responses. The rule for the treatment of qualitative risks linked to D states that if the risk assumes the value E, then send an alert to the actor who executed D and activate an escalation procedure. The escalation signals a "process risk" to the process owner (the role responsible for the process instance that provide execution context for D) and an "organizational risk" to the appropriate business manager.

In section 3 we outlined a three level measurement are supported by a WfMS that, during the execution of workflows, stores raw execution data in log files using them to feed the measurement framework.

By the coupling of a WfMS with a RMS we can obtain an additional value in terms of capability to manage operational risks through quantitative techniques. Consider again the opportunity that a risk manager has at definition time to define the reactive behavior of a RMS. The rule b) shows how a reactive behavior can be relied to a task D. It states that when the workflow engine creates an instance of D assigning it to the work list of an actor, a check has to be done. If the instance of D is created 50 time units after the instance creation of its father, (the process P to which D belongs) then two messages highlighting a schedule risk for the task D are produced, one to the actor that is executing the task and the other to the process owner.

The measurement framework can bring more than a reactive behavior. The need to assess the risk relied to the missing process completion is one of the characteristic that one could require to a system that integrates a WfMS with a RMS. Such proactive behavior lays on the availability of execution data automatically collected by the WfMS and on the risk analysis data represented within the RMS.

Let P be a process and i_p an instance in the execution of P. The WfMS can assess the residual duration of i_p by considering the difference between the average duration of already completed instances of P and the current duration of i_p. Remembering that sigma evaluates the sum of measures of instances (filtered by means of P) and that work counts the number of such instances we have:

$$\text{residual\_duration}(i_p) = \frac{\text{sigma}(\text{instance\_duration}, \text{filter}(I,p))}{\text{work}(I,p)} - \text{current\_duration}(i_p) \quad (15)$$

$$p = i\_\text{name}(i) = i\_\text{name}(i_p) \wedge \text{current\_state}(i) = \text{completed}$$

Depending on the value returned by the application of `residual_duration`, the RMS has three possible alternative interpretations of the expected residual duration of P. When the value is equal to 0 we have an indication that from now on delay will be accumulated; if the value is less than 0, the process is late, otherwise, the residual duration represents an assessment of the time needed to complete the process. The measure `residual_duration` should be evaluated by the WfMS at the completion of each task instance in i_p thus providing in real time to the RMS the information necessary to eventually choose the best reaction to the current situation.

Apart from the workflow measurement framework used in this paper, the risk manager can take advantage of other existing set of risk indicators. It is sufficient to plan at risk definition time both: a) the link between expected value of measures and tasks b) the rules for the risk treatment.

In this way standard measures can be used and evaluated locally to put under control potential risks engraving on tasks. Moreover, the set of measures can be enhanced if a distributed process is enacted by one centralized WfMS or different cooperating WfMSs; as an example, in [19] are discussed new measures taking into considerationthe interaction among different Virtual Enterprise (VE) participants. The *contribution of a participant to the global risk exposure* is a measure that requires the acquisition of distributed knowledge.

To formally define such an indicator, let us consider the probability $p_f(T)$ that the generic task T fails (e.g. it never starts, it begins at some point but it never ends, it ends but the customer does not remunerate the activity as expected by the VE), and the completion percentage $p_c(T)$ of the task T.

Given O_j and the set T_j of tasks assigned to O_j. The function assumes the following values

$$p_j(T) = \begin{cases} 1 & \text{when the task } T \text{ has not already starte} \\ 1-p_c(T) & \text{when the task } T \text{ is in progress} \\ 1-p_c(T_j) & \text{at the end of the task } T. \end{cases} \quad (16)$$

Where $p_c(T)$ has values in $[0,1]$ and represents the completion degree of the tasks that are members of T_j. In other words, $p_c(T_j)$ is equal to 0 when do not exist tasks already started in T_j and it assumes value 1 when all the tasks in T_j have been completed and compliant with the customer expectations. In all other cases $0<p(T_j)<1$.

The functions $p_c(T_j)$ and $p_f(T)$ have to be evaluated at the completion of each assigned task.

The value $p_f(T)=1-p_c(T_j)$ expresses a residual risk on a that depends on how many tasks in T_j have been considered compliant by the customer.

During the execution of a distributed process, we can use indicators to compute either the *Economic Risk Exposure* of an organization at time

t_k or the *Global Economic Risk Exposure* of the project at time t_k as follows:

$$ERE_{O_j,t_k} = \sum_{i=1}^{n_j} p_r(T_{j,i}) \cdot c(T_{j,i}) \quad (17)$$

where $n_j < m$ is the number of task of O_j

$$ERE_{G,t_k} = \sum_{j=1}^{n} ERE_{O_j,t_k} \quad (18)$$

where Gis the set of all organizations.

To evaluate the *Expected Contribution* of a participant to the economic global risk exposure we define:

$$\frac{ERE_{O_j}}{ERE_G} \quad (19)$$

Fig. 5 shows how these indicators can be used in the practice. At time t_k, the risk manager assesses the risk project state trying to evaluate the expected global economic risk exposure. Two tasks have been already successfully completed by O_1, three by O_2 and none by O_3. One task for $O_1$is in progress and contributes only for a fraction of its cost to the determination of ERE_{O_1,t_k}. At the beginning of the project the leader organization can easily compute the expected value ERE_{G,t_0} because it is equal to the entire budget scheduled for the project. Once the tasks have been assigned and the project is in progress, the leader organization can evaluate this variable when it obtains from each organization the value ERE_{O_j,t_k}. The evaluation of ERE_{G,t_k} provides a prediction for the future project trend from the perspective of cost risks.

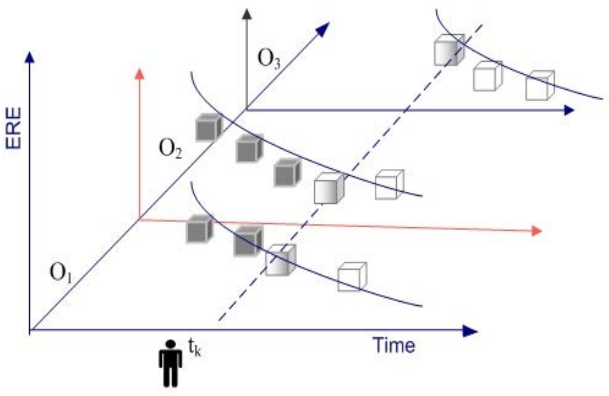


Figure 5. Evaluating the Global Economic Risk Exposure

VI. CONCLUSION

Enterprise risk management is an emergent research field. Apart from application area such as banking, insurance and health where risk management has traditionally been considered a primary management discipline, more and more organization are planning today the introduction of a risk management system. The model for workflow oriented risk management proposed here arise from the consideration that the degradation of process execution in terms of poor performances/ effectiveness, high costs and low quality can cause great difficulties even undermining the survival of organizations. It can be taken as a reference model by process focused enterprises for the implementation of advanced risk management systems. As a matter of fact, from the coupling of a WfMS with a risk management system we obtain an integrated system capable of managing risks that could have an impact on the regular execution of workflows. Any deviation from the prescribed workflow behavior implies either a missed deadline, an increased execution cost or even a danger or an illegal situation. The basic information needs concerning the workflow execution, from the point of view of risk management, can be satisfied by the workflow engine either automatically recording relevant events during the process execution (i.e. creation, completion of work items, task and processes) or collecting qualitative data before or after the examination of each scheduled activity.

Both kinds of measures, qualitative and quantitative are effective tools that help the management to identify threats during the enactment of processes. At risk definition time, the risk manager looks at the definition of activities and processes assigning to them risk monitoring rules that can be automatically managed by the WfMS during the workflow execution.

Even if the implementation of the level model for process focused risk management can contribute to reduce the cost of data collection and to the acquisition of precise data about workflow execution, the model brings its advantages especially in the area of operational risks. A risk manager must be aware of this limitation considering the decision support system provided by the process focused risk management as an important part of a wider RMS that can take advantage also of traditional techniques in order to handle the four risk management areas discussed in the introduction.

REFERENCES

[1] COSO,"Enterprise Risk management - Integrated framework. Executive Summary", Committee of Sponsoring Organizations of the Treadway Commission, 2004 Internal control systems, Available : http://www.coso.org/Publications/ERM/COSO_ERM_ExecutiveSummary.pdf.

[2] C.J.Alberts ; A.J. Dorofee, A framework for categorizing keydrivers of risk, Technical report, CMU/SEI-2009-TR-007, ESC-TR-2009-007, Software Engineering Institute, Carnagie Mellon University, Pittsburgh, PA, April 2009.

[3] M. A. Stankosky, "Creating the discipline of Knowledge Management: the latest in University Research", Elsevier Butterworth-Heinemann, ISBN :0-7506-7878-X, 2005.

[4] D. Neef; G. A. Siesfeld; J. Cefola, "The economic impact of knowledge", Butterworth-Heinemann, ISBN 0-7506-7009-6, 1998.

[5] I. Nonaka, "Knowledge management: critical perspectives on business and management", Routledge, ISBN 0-415-34030-6 vol. 1, 2005.

[6] G. M. Weinber, "An Introduction to General Systems Thinking", Dorset House Publishing Company, ISBN 0-932633-49-8, 2001.

[7] F. Leymann, D. Roller, "Production Workflow", Prentice Hall, ISBN 0-13-021753-0, 2000.

[8] M. zur Muehlen, "Workflow-based Process Controlling. Foundation, Design, and Application of Workflow-driven Process Information Systems", Advances in Information Systems and Management Science, Logos Verlag Berlin, 2004.

[9] Oracle. "Oracle Workflow Guide", vol. 1, release 2.6.2, Oracle Corporation, 2002.

[10] V. R. Basili, G. Caldiera, H. D. Rombach. "The goal question metric approach", in Enciclopedia of Software Engineering , John Wiley & Sons Inc., New York , pp. 528-532, 1994.

[11] M. G. Mendoça, V. R. Basili, "Validation of an approach for improving existing measurement frameworks", IEEE Transactions on Software Engineering , vol.26, no.6,pp.484-499, June 2000.

[12] J. McGarry, D. Card, C. Jones, B. Layman, E. Clark, J. Dean, F. Hall,"Practical Software Measurement" , Addison Wesley. ISBN 0-201-71516-3, 2001.

[13] COSO,"Guidance on monitoring internal control systems", 2009.

[14] R. Aiello, G. Nota, M. Faggini, "Continuous Process Measurement", Economics and Complexity, vol. 3, no.1, 2006, ISSN: 1398-1706.

[15] R. Aiello, A. Esposito, G. Nota, "A Hierarchical Measurement framework for the evaluation of automated business processes", International Journal of Software Engineering and Knowledge Engineering, vol. 12, no. 4, 2002.

[16] D. F. Cooper, S. Grey, G. Raymond, P. Walker,. "Project Risk Management Guidelines: Managing Risk in Large Projects and Complex Procurements", John Wiley & Sons, ISBN 0-470-02281-7, 2008.

[17] M. Degl'Innocenti, G. L. Ferrari, G. Pacini, F.Turini, "RSF: A formalism for executable requirement specifications". *IEEE Transactions on Software Engineering*,vol.16, no.11, pp.1235–1246, 1990.

[18] G. Nota, G. Pacini,"Querying of Executable Software Specifications", IEEE Transactions on Software Engineering, vol. 18, no. 8, August 1992.

[19] G. Nota, M. P. Di Gregorio, R. Carvello, "Distributed Risk Management in a Virtual Enterprise Environment", International Journal of Quality Management, to be published.

[20] D. Hillson, "Earned valued management and risk management: a practical synergy", PMI 2004 Global Congress Proceedings, Anaheim, California, USA.

Session 2: Universal Networking Language

ISBN 978- 974-672-556-9

Preparation of Bangla Speech Corpus for Phoneme Recognition

Md. Shahadat Hossain[1], Mohammed Rokibul Alam Kotwal[1], Manoj Banik[2], Foyzul Hassan[3], Mohammad Mahedi Hasan[4], Mohammad Nurul Huda[1] and Chowdhury Mofizur Rahman[1]
[1]Department of CSE, United International University, Dhaka, Bangladesh
[2]Department of CSE, Ahsanullah University of Science and Technology, Dhaka, Bangladesh
[3]Enosis Solutions, Dhaka, Bangladesh
[4]Blueliner Bangladesh, Dhaka, Bangladesh
shhc495@yahoo.com, rokib_kotwal@yahoo.com, mbanik99@yahoo.com, foyzul.hassan@gmail.com, mmahedihasan@gmail.com, mnh@cse.uiu.ac.bd and cmr@uiu.ac.bd

***Abstract*—In this paper, we prepare a Bangla speech database for a phoneme recognition system of Bangla Automatic Speech Recognition (ASR). Most of the Bangla ASR system uses a small number of speakers, but 40 speakers selected from a wide area of Bangladesh, where Bangla is used as a native language, are involved here. In the experiments, mel-frequency cepstral coefficients (MFCCs) and local features (LFs) are inputted to the hidden Markov model (HMM) based classifiers for obtaining phoneme recognition performance. It is shown from the experimental results that MFCC-based method of 39 dimensions provide a higher phoneme correct rate and accuracy than the method of using LFs with 25 dimensions. Moreover, it requires fewer mixture components in the HMMs.**

Index Terms- Bangla speech corpus; phoneme recognition; hidden Markov model

I. INTRODUCTION

There have been many literatures in automatic speech recognition (ASR) systems for almost all the major spoken languages in the world. Unfortunately, only a very few works have been done in ASR for Bangla (can also be termed as Bengali), which is one of the largely spoken languages in the world. More than 220 million people speak in Bangla as their native language. It is ranked seventh based on the number of speakers [1]. A major difficulty to research in Bangla ASR is the lack of proper speech corpus. Some efforts are made to develop Bangla speech corpus to build a Bangla text to speech system [2]. However, this effort is a part of developing speech databases for Indian Languages, where Bangla is one of the parts and it is spoken in the eastern area of India (West Bengal and Kolkata as its capital). But most of the natives of Bangla (more than two thirds) reside in Bangladesh, where it is the official language. Although the written characters of Standard Bangla in both the countries are same, there are some sound that are produced variably in different pronunciations of Standard Bangla, in addition to the myriad of phonological variations in non-standard dialects [3]. Therefore, there is a need to do research on the main stream of Bangla, which is spoken in Bangladesh, ASR.

Some developments on Bangla speech processing or Bangla ASR can be found in [4]-[11]. For example, Bangla vowel characterization is done in [4]; isolated and continuous Bangla speech recognition on a small dataset using hidden Markov models (HMMs) is described in [5]; recognition of Bangla phonemes by Artificial Neural Network (ANN) is reported in [8]-[9]. Continuous Bangla speech recognition system is developed in [10], while [11] presents a brief overview of Bangla speech synthesis and recognition. However, most of these works are mainly concentrated on simple recognition task on a very small database, or simply on the frequency distributions of different vowels and consonants.

In this paper, we build an ASR system for Bangla phoneme in a large scale. For this purpose, we first develop a medium size (compared to the exiting size in Bangla ASR literature) Bangla speech corpus comprises of native speakers covering almost all the major cities of Bangladesh. Then, mel-frequency cepstral coefficients (MFCCs) and local features (LFs) are extracted from the input speech, and finally extracted features are inserted into the hidden Markov model (HMM) based classifier for obtaining the phoneme recognition performance. For evaluating Bangla phoneme correct rate (PCR) and phoneme accuracy (PA), we have designed two experiments (a) MFCC+HMM and (b) LF+HMM.

The paper is organized as follows. Section II briefly describes approximate Bangla phonemes with its corresponding phonetic symbols; Section III explains about Bangla speech corpus; Section IV provides a brief description about MFCC-based and LF-based methods, while Section V gives experimental setup. Section VI explicates the experimental results and discussion, and finally, Section VII draws some conclusions and remarks on the future works.

II. BANGLA PHONEMES WITH PHONETIC SYMBOLS

Phonetic inventory of Bangla consists of 8 short vowels, excluding long vowels, and 29 consonants. Table I shows Bangla vowel phonemes with their corresponding International Phonetic Alphabet (IPA) and our proposed symbols. On the other hand, the consonants, which are used in Bangla language, are presented in Table II. Here, the Table exhibits the same items for consonants like as Table I. In the Table II, the pronunciation of /শ/, /ষ/ ধহফ /স/ are same by considering the words বিশ (/bi□/), বিষ(/bi□/) and ডিস(/□i□/). On the other hand, in the wordsজাম (/d□am/) and যাক (/d□ak/), there is no difference of pronunciation of /R/ and /h/. Again, there is no difference of /Y/ and /b/ in the words হরিণ (/h□in/) and তিন (/tin/). Phonemes /o/ and /p/ carry same pronunciation in the words পাহাড় (/paha□/) and আষাঢ় (/a□a□/).

Native Bangla words do not allow initial consonant clusters: the maximum syllable structure is CVC (i.e. one vowel flanked by a consonant on each side) [12]. Sanskrit words borrowed into Bangla possess a wide range of clusters, expanding the maximum syllable structure to CCCVC. English or other foreign borrowings add even more cluster types into the Bangla inventory.

TABLE I. BANGLA VOWELS

Letter	IPA	Our Symbol
অ	/□/ and /o/	a
আ	/a/	aa
ই	/i/	i
ঈ	/i/	i
উ	/u/	u
ঊ	/u/	u
এ	/e/ and /æ/	e
ঐ	/oj/	oi
ও	/o/	o
ঔ	/ow/	ou

III. BANGLA SPEECH CORPUS

At present, a real problem to do experiment on Bangla phoneme ASR is the lack of proper Bangla speech corpus. In fact, such a corpus is not available or at least not referenced in any of the existing literature. Therefore, we develop a medium size Bangla speech corpus, which is described below.

Hundred sentences from the Bengali newspaper "Prothom Alo" [13] are uttered by 30 male speakers of different regions of Bangladesh. These sentences (30x100) are used for training corpus (D1). On the other hand, different 100 sentences from the same newspaper uttered by 10 different male speakers (total 1000 sentences) are used as test corpus (D2). All of the speakers are Bangladeshi nationals and native speakers of Bangla. The age of the speakers ranges from 20 to 40 years. We have chosen the speakers from a wide area of Bangladesh: Dhaka (central region), Comilla – Noakhali (East region), Rajshahi (West region), Dinajpur – Rangpur (North-West region), Khulna (South-West region), Mymensingh and Sylhet (North-East region). Though all of them speak in standard Bangla, they are not free from their regional accent.

TABLE II . BANGLA CONSONANTS

Letter	IPA	Our Symbol
ক	/k/	k
খ	/k^h/	kh
গ	/□/	g
ঘ	/□□/	gh
ঙ	/ŋ/	ng
চ	/t□/	ch
ছ	/t□□/	chh
জ	/d□/	j
ঝ	/d□□/	jh
ট	/□/	ta
ঠ	/□□/	tha
ড	/□/	da
ঢ	/□□/	dha
ণ	/n/	n
ত	/t /	t
থ	/t □/	th
দ	/d /	d
ধ	/d □/	dh
ন	/n/	n
প	/p/	p
ফ	/p□/	ph
ব	/b/	b
ভ	/b□/	bh
ম	/m/	m
য	/d□/	j
র	/□/	r
ল	/l/	l
শ	/□/ / /s/	s
ষ	/□/	s
স	/□/ / /s/	s
হ	/h/	h
ড়	/□/	rh
ঢ়	/□/	rh
য়	/e / /-	y

Recording was done in a quiet room located at United International University (UIU), Dhaka, Bangladesh. A desktop was used to record the voices using a head mounted close-talking microphone. We record the voice in a place, where ceiling fan and air conditioner were switched on and some low level street or corridor noise could be heard.

Jet Audio 7.1.1.3101 software was used to record the voices. The speech was sampled at 16 kHz and quantized to 16 bit stereo coding without any compression and no filter is used on the recorded voice.

ISBN 978- 974-672-556-9

IV. SYSTEM CONFIGURATIONS

A. MFCC-based method

Conventional approach of ASR systems uses MFCC as feature vector to be fed into a HMM-based classifier and the system diagram is shown in Fig. 1. Parameters (mean and diagonal covariance of hidden Markov model of each phoneme) are estimated, from MFCC training data, using Baum-Welch algorithm. For different mixture components, training data are clustered using the K-mean algorithm. During recognition phase, a most likely phoneme sequence for an input utterance is obtained using the Forward algorithm.

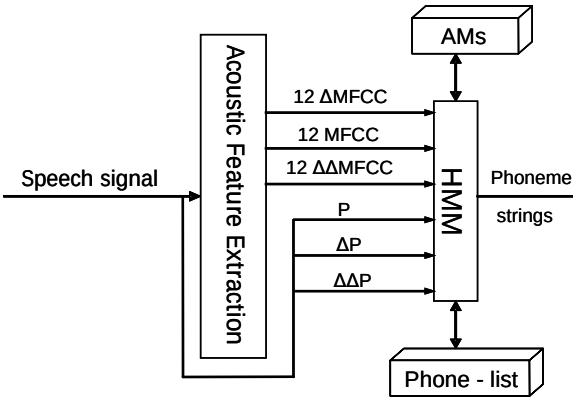


Figure1.MFCC-based phoneme recognition method.

B. LF-based method

At an acoustic feature extraction stage, firstly, input speech is converted into LFs that represent a variation in spectrum along time and frequency axes [14]. Two LFs are first extracted by applying three-point linear regression (LR) along the time t and frequency f axes on a time spectrum pattern respectively. After compressing these two LFs with 24 dimensions into LFs with 12 dimensions using discrete cosine transform (DCT), a 25-dimensional (12 Δt, 12 Δf and ΔP, where P stands for log power of raw speech signal) feature vector named LF is extracted. Then, the extracted LFs are inserted into the HMM-based classifier for obtaining the phoneme strings. The procedure is shown in Fig. 2.

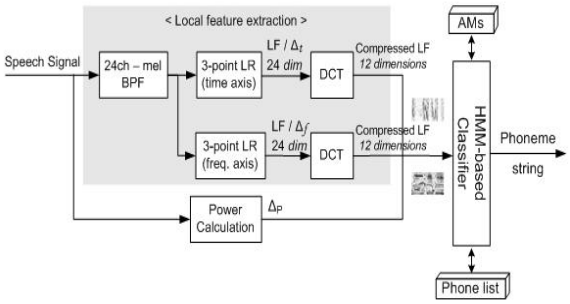


Figure2. LF-based phoneme recognition method.

V. EXPERIMENTAL SETUP

The frame length and frame rate are set to 25 ms and 10 ms, respectively, to obtain acoustic features (MFCCs or LFs) from an input speech. MFCC comprised of 39 dimensional (12-MFCC, 12-ΔMFCC, 12-ΔΔMFCC, P, ΔP and ΔΔP, where P stands for raw energy of the input speech signal). Acoustic feature vector LFs are a 25-dimensional vector consisting of 12 delta coefficients along time axis, 12 delta coefficients along frequency axis, and delta coefficient of log power of a raw speech signal [14].

For designing an accurate phoneme recognizer, PCR and PA for D2 data set are evaluated using an HMM-based classifier. The D1 data set is used to design 39 Bangla monophone (8 vowels, 29 consonant, sp, sil) HMMs with five states, three loops, and left-to-right models. Input features for the classifier are 39 dimensional MFCC and 25 dimensional LF for the MFCC-based and LF-based systems, respectively. In the HMMs, the output probabilities are represented in the form of Gaussian mixtures, and diagonal matrices are used. The mixture components are set to 1, 2, 4, 8, 16 and 32.

To obtain the PCR and PA we have designed the following experiments

1) MFCC+HMM
2) LF+HMM

VI. EXPERIMENTAL RESULTS AND DISCUSSION

Fig. 3 shows the comparison of PCR of training data set between MFCC-based and LF-based systems. It is observed from the figure that MFCC-based system always provides higher PCR than the other method investigated. For an example, at mixture component 32, the MFCC-based system exhibits 79.25% correct rate, while 70.04% PCR is obtained by the LF-based system. On the other hand, Fig. 4 gives corresponding PA for the methods investigated. It is also shown from this figure that

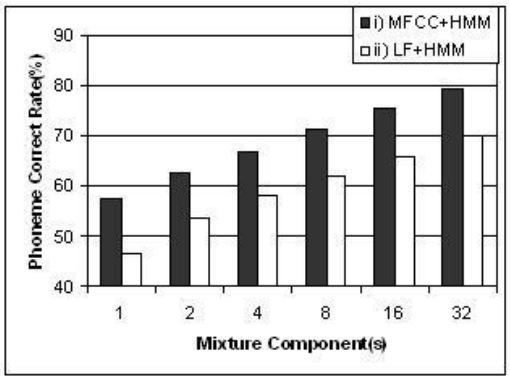


Figure 3.Phoneme correct rate for training data.

similar types of results are obtained. These results exhibit the excellence of MFCC-based systems over

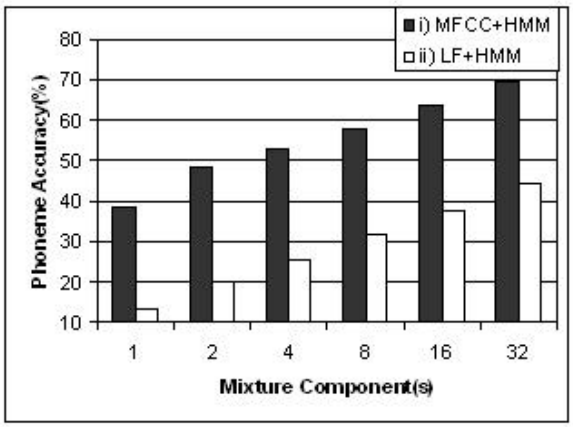


Figure 4.Phoneme accuracy for training data.

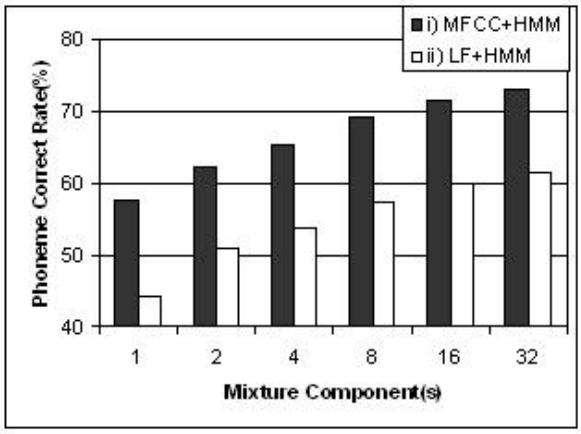


Figure 5.Phoneme correct rate for test data.

the LF-based systems.

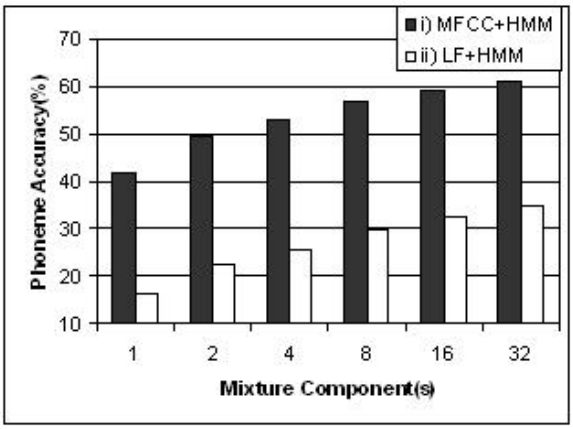


Figure6. Phoneme accuracy for test data.

On the other hand, the PCR and PA for test data are shown in the Figs. 5 and 6, respectively for the investigated methods. The MFCC-based method outperformed the LF-based method for the both evaluation. It is noted from mixture component 32 of Fig. 6 that the MFCC-based system having 60.88% accuracy shows its better recognition performance over the LF-based system of 34.68% accuracy.

The reason for providing better result by MFCC-based system are i) dimensionality (MFCC dimension is 39, whereas LF dimension is 25), where large dimensional feature vector exhibits its excellence and ii) insertion of ΔΔP parameters in MFCC (no ΔΔ parameter in LF).

VII. Conclusion

This paper showed a preparation of Bangla speech corpus and provided some experiments to obtain phoneme recognition performance. The following conclusions are drawn from the experiments:

i) The MFCC-based system provides tremendous improvement of Bangla phoneme recognition accuracy for both training and test data.
ii) A higher Bangla phoneme correct rate for training and test data is also obtained by the MFCC-based system.

The author would like to do further experiments for obtaining phoneme recognition performance after inserting both features into the neural network based systems.

Acknowledgment

This work was supported in part by United International University (UIU), Dhaka, Bangladesh.

References

[1] http://en.wikipedia.org/wiki/List_of_languages_by_total_speakers, Last accessed April 11, 2009.
[2] S. P. Kishore, A. W. Black, R. Kumar, and Rajeev Sangal, "Experiments with unit selection speech databases for Indian languages," Carnegie Mellon University.
[3] http://en.wikipedia.org/wiki/Bengali_phonology, Last accessed April 11, 2009..
[4] S. A. Hossain, M. L. Rahman, and F. Ahmed, "Bangla vowel characterization based on analysis by synthesis," *Proc. WASET*, vol. 20, pp. 327-330, April 2007.
[5] M. A. Hasnat, J. Mowla, and Mumit Khan, "Isolated and Continuous Bangla Speech Recognition: Implementation Performance and application perspective, " in *Proc. International Symposium on Natural Language Processing* (*SNLP*), Hanoi, Vietnam, December 2007.
[6] R. Karim, M. S. Rahman, and M. Z Iqbal, "Recognition of spoken letters in Bangla," in *Proc. 5th International Conference on Computer and Information Technology (ICCIT02)*, Dhaka, Bangladesh, 2002.
[7] A. K. M. M. Houque, "Bengali segmented speech recognition system," Undergraduate thesis, BRAC University, Bangladesh, May 2006.
[8] K. Roy, D. Das, and M. G. Ali, "Development of the speech recognition system using artificial neural network," in *Proc. 5th International Conference on Computer and Information Technology (ICCIT02)*, Dhaka, Bangladesh, 2002.
[9] M. R. Hassan, B. Nath, and M. A. Bhuiyan, "Bengali phoneme recognition: a new approach," in *Proc. 6th International Conference on Computer and Information Technology (ICCIT03)*, Dhaka, Bangladesh, 2003.
[10] K. J. Rahman, M. A. Hossain, D. Das, T. Islam, and M. G. Ali, "Continuous bangle speech recognition system," in *Proc. 6th International Conference on Computer and Information Technology (ICCIT03)*, Dhaka, Bangladesh, 2003.

[11] S. A. Hossain, M. L. Rahman, F. Ahmed, and M. Dewan, "Bangla speech synthesis, analysis, and recognition: an overview," in *Proc. NCCPB*, Dhaka, 2004.
[12] C. Masica, *The Indo-Aryan Languages*, Cambridge University Press, 1991.
[13] www.prothom-alo.com
[14] T. Nitta, "Feature extraction for speech recognition based on orthogonal acoustic-feature planes and LDA," Proc. ICASSP'99, pp.421-424, 1999.

Formation of Bangla Word Dictionary Compatible with UNL Structure

Muhammad Firoz Mridha[1], Manoj Banik[2], Md. Nawab Yousuf Ali[3], Mohammad Nurul Huda[4], Chowdhury Mofizur Rahman[4], Jugal Krishna Das[5]
[1]Department of CSE, Stamford University Bangladesh,
[2]Department of CSE, Ahsanullah University of Science and Technology, Dhaka, Bangladesh,
[3]East WestUniversity, Dhaka, Bangladesh,
[4]United InternationalUniversity, Dhaka, Bangladesh,
[5]Jahangirnagar University, Dhaka, Bangladesh.
mdfirozm@yahoo.com, mbanik99@yahoo.com, nawab@ewubd.edu, mnh@cse.uiu.ac.bd, cmr@uiu.ac.bd, drdas64@yahoo.com

***Abstract*—The usage of native language through Internet is highly demanding now a day due to rapidly increase of Internet based application in daily needs. Universal Networking Language (UNL) addressed this issue in most of languages. But the UNL is unable to convert from any native language to Bangla. Addressing this issue, this paper presents a structure namely, "Formation of Bangla Word Dictionary Compatible with UNL Structure" by integrating Bangla Word Dictionary along with grammatical attributes of Bangla Words into the framework of UNL. The proposed work is theoretically able to format any kinds of Bangla Words with Prefixes and Suffixes that is compatible with UNL.**

Index Terms- Bangla Word Dictionary; DeConverter; EnConverter; Morphological Analysis; Universal Networking Language; Universal words

I. INTRODUCTION

The Universal Networking Language Programme started in 1996, as an initiative of the Institute of Advanced Studies (IAS) of the United Nations University (UNU) [1, 5] in Tokyo, Japan. In January 2001, the UNU set up an autonomous organization, the UNDL Foundation, to be responsible for the development and management of the UNL Programme. The Foundation, a non-profit international organisation, has an independent identity from the UNU, although it has special links with the United Nations (UNs). It inherited from the UNU/IAS the mandate of implementing the UNL Programme so that it can fulfill its mission. Its headquarters are based in Geneva, Switzerland. The mission of the UNU program is to allow people across nations to access information in Internet in their own languages. The core of the project is UNL, a language independent specification for serving as a common medium for documents in different languages. Researchers involved in this project from different countries have been developing UNL system for their respective native languages. The goal is to eliminate the massive task of translation between two languages and reduce language to language translation to a one time conversion to UNL. For example, Bangla corpora, once converted to UNL, can be translated to any other language given UNL system built for that language. The UNL system does this by representing only the *semantics* of a native language sentence in a *hypergraph.* EnConverter [13] (parser) converts each native language sentence to a UNL hypergraph and DeConverter [14] translates from hypergraph to any native language. The main aim of the UNL project is to overcome language barriers. This project currently includes 16 official languages. Bangla is not yet included. We have attempted to demonstrate that we can do similar tasks for *Bangla* as it has been done for other official languages. By this time we have reached two agreements with the UNL Center of UNDL Foundation to incorporate Bangla with the UNL as Bangla is the 4th widely spoken language with 250 millions speakers so that a great number of people can be accessed and shared a vast repository of knowledge through the Internet. They assigned us ID, User Name and Password to use their resources to progress our work. In this paper we present the Bangla Words compatible with UNL structure. The major components of our research works touches upon i) construct the concept of words,ii) development of morphological analysis and construct the concepts of both roots, words and their morphemesand iii) outlining the concept of grammatical attributes of Words .

This paper is organized as follows: the literature review related to the UNL structure and the UNL format of dictionary is presented in Section 2 while the proposed structure together with its constituent parts is detailed in Section 3 and Section 4 shows concluding remarks 2. Literature Review.

The UNL structures and the UNL format of dictionary that are related to the proposed work are detailed in the next sections

II. UNL STRUCTURE

UNL is an artificial language that allows the processing of information across linguistic barriers

ISBN 978- 974-672-556-9

[10]. This artificial language has been developed to convey linguistic expressions of natural languages for machine translation purposes. Such information is expressed in an unambiguous way through a semantic network with hyper-nodes. Nodes (that represent concepts) and arcs (that represent relations between concepts) compose the network. UNL contains three main elements:

- Universal Words: Nodes that represent word meaning.
- Relation Labels: Tags that represent the relationship between Universal Words i.e. between two nodes. Tags are the arcs of UNL hypergraph.
- Attribute Labels: Additional information about the universal words.

These elements are combined in order to establish a hierarchical Knowledge Base (UNLKB) [10] that defines unambiguously the semantics of UWs. The UNL Development Set provides tools that enable the semi-automatic conversion of natural language into UNL and vice-versa. Two of such tools are the EnConverter and the DeConverter. The main role of EnConverter [11] is to translate natural language sentences into UNL expressions. This tool implements a language independent parser that provides a framework for morphological, syntactic and semantic analysis synchronously. This allows morphological and syntactical ambiguities resolution. The DeConverter [3, 12], on the other hand, is a language independent generator that converts UNL expressions to natural language sentences.

A. Universal Words

Universal Words are words that constitute the vocabulary of UNL. A UW is not only a unit of the UNL syntactically and semantically for expressing a concept, but also a basic element for constructing a UNL expression of a sentence or a compound concept. Such a UW is represented as a node in a hypergraph. There are two classes of UWs from the viewpoint in the composition:

- labels defined to express unit concepts and called “UWs” (Universal Words)
- a compound structure of a set of binary relations grouped together and called “Compound UWs”.

B. Relational Labels

The relation [1] between UWs is binary that have different labels according to the different roles they play. A relation label is represented as strings of three characters or less. There are many factors to be considered in choosing an inventory of relations. The following is an example of relation defined according to the above principles.

Relation: There are 46 types of relations in UNL. For example, agt (agent), *agt* defines a thing that initiates an action, agt(do, thing), agt(action, thing), obj(thing with attributes) etc.

C. Attributes

The attributes represent the grammatical properties of the words. Attributes of UWs are used to describe subjectivity of sentences. They show what is said from the speaker’s point of view: how the speaker views what is said. This includes phenomena technically [4, 5] called speech, acts, propositional attitudes, truth values, etc. Conceptual relations and UWs are used to describe objectivity of sentences. Attributes of UWs enrich this description with more information about how the speaker views these state of affairs and his attitudes toward them.

D. UNL Format of Dictionary

The UNDL foundation provides a dictionary format. The Word Dictionary is a collection of the word dictionary entries. Each entry of the Word Dictionary is composed of three kinds of elements: the Headword (HW), the Universal Word (UW) and the Grammatical Attributes. A headword is a notation/surface of a word of a natural language that composing the input sentence and it is to be used as a trigger for obtaining equivalent UWs from the Word Dictionary in enconversion. An UW expresses the meaning of the word and is to be used in creating UNL networks (UNL expressions) of output. Grammatical Attributes are the information on how the word behaves in a sentence and they are to be used in enconversion rules.

Each Dictionary entry has the following format of any native language word [5].

Data Format:

[HW] {ID} “UW” (Attribute1, Attribute2,…) <FLG, FRE, PRI>
Here, HW ← Head Word (Bangla word), ID ← Identification of Head Word (omitable), UW ← Universal Word, ATTRIBUTE ← Attribute of the HW, FLG ← Language Flag(we use B for Bangla), FRE ← Frequency of Head Word, PRI ← Priority of Head Word

Format of an element of Bangla-UNL Dictionary would be:

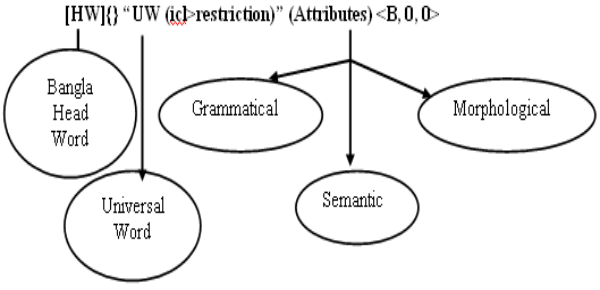


Figure 1. Format of Word Dictionary

Some example entries of dictionary for Bangla are given below:

[শহর] {} "city(icl>region)" (N, PLACE) <B,0,0>
[প্রচুর] {} "huge(icl>many)" (ADJ) <B,0,0>
[পাখি] {} "bird(icl>animal>animate thing)" (N, ANI, SG, CONCRETE) <B,0,0>

Where attributes, N stands for Noun, PLACE for place, ADJ for Adjective, FLG field entry is B for Bangla, ANI for animal, SG for Singular number, CONCRETE for concrete thing respectively.

Now we are concerned how to make Bangla Word Dictionary for UNL. In UNL Knowledge Base (KB) made by the UNL center of UNDL Foundation (Last updated version) on May 26, 2009, there are 21862 formats of Universal Words (UWs) [11]. We can find the UWs for each of the Bangla HW by searching the UNL KB to develop Bangla Word Dictionary for UNL. As per our perception this is not the suitable way to find out the UWs for the Bangla HW. Firstly, it is a long process to build Bangla Word Dictionary for UNL by searching the appropriate UWs from a huge number of words formats in UNL KB. Secondly, a word may have two or more meanings. Such types of words are represented with various concepts in UNL KB. So, which one to choose out of two or more meanings for a Head Word is a hard job and we cannot get out appropriate/accurate words for the corresponding Bangla HWs.

III. PROPOSED WORK

Construct the concepts of simple Bangla Words for UNL and Outlining the Morphological Analysis of Bangla Words along with their grammatical attributes compatible with UNL structure:

A. Construct the concepts of simple Bangla Words for UNL

We have found a new way (easiest and shorten) of searching based on existing works of other languages especially for English. Firstly, we can take some manually translated texts from Bangla to English in different forms and then convert them into UNL expressions (using English-UNL EnConverter). For example,

Assertive sentence: আমি ভাত খাইতেছিin English " I am eating rice."
Interrogative sentence: আমি কি ভাত খাই? in English "Do I eat rice?"
Negative sentence: আমি ভাত খাই না। in English "I do not eat rice."

If we convert the first sentence by the English-UNL Converter [12] we get the following UNL expressions.

I am eating rice.
agt(eat(icl>consume>do,agt>living_thing,obj>concrete_thing).@entry.@present.@progress,i(icl>person))
obj(eat(icl>consume>do,agt>living_thing,obj>concrete_thing).@entry.@present.@progress,rice(icl>cereal>thing))

The same way, if we convert the two other sentences above, we get the same concepts of the words "I", "eat" and "rice" respectively. So, the Bangla Words "আমি""খাই" and "ভাত" can be represented as.

[আমি] {}" i(icl>person)"
[খাই] {}"eat(icl>consume>do)"
[ভাত] {}"rice(icl>cereal>thing)"

These concepts are not enough for representing the words for the dictionary entries. We have to develop grammatical attributes for the words. Grammatical attributes have to be developed by the rules (EnConversion and DeConversion) and Dictionary developers. Now we are representing the above Bangla Words along with their grammatical attributes for Dictionary entries as follows.

[আমি]{}"i(icl>person)"(1P,1SG,PRON,HPRON,SUBJ) <B,0,0>
[খাই] {}"eat(icl>consume>do)"(V) <B,0,0>
[ভাত] {}"rice(icl>grain>thing)"(N,CONCRETE, HF) <B,0,0>

Here we have defined 1P, SG, PRON, HPRON, SUBJ for "আমি", V for "খাই" and N, C, HF for "ভাত" as grammatical attributes. As "আমি" is the first person we have used 1P, SG for singular number, PRON for pronoun, HPRON for human pronoun and SUBJ as the word used as a subject in a sentence. As "খাই" is the root we have used V and as "ভাত" the noun, human food and concrete thing, we have used N, HF and CONCRETE respectively.

In the same way by manually translating the different types of simple Bangla sentences (with variety of words) to English sentences and then English sentences to UNL expressions, we can get the appropriate concepts of thousand of Bangla Words to build the Bangla Word Dictionary for UNL.

Secondly, we would take texts from some reliable translated sources (from Bangla to English) from Bangla Academy Scientific literatures. Then we can convert them into UNL expressions as above sentences and again can get the concepts of thousands of words for dictionary entries.

During Word Formation of Bangla Word Dictionary for UNL we have to resolve many ambiguities. Say, many Bangla Words have two or more English meanings. Similarly, many English Words also have two or more Bangla meanings. For example, we use "সে" in Bangla, but in English it has two meanings "he" and "she". Again, we use "rice"

in English, but in Bangla it has three meanings "ভাত"or "চাউল" or"ধান". "ধান" means paddy in English, when it is in the field. To resolve these ambiguities we can represent them in the dictionary as follows.

[সে(পুরুষ)] {} "he(icl>person)" (PRON, MALE, ANI, 3SG, HPRON) <B,0,0>
[সে(মহিলা)] {} "she(icl>person)" (PRON, FEMALE, ANI, 3SG, HPRON) <B,0,0>
[ভাত] {} "rice(icl>cereal>thing)" (N, C)<B,0,0>
[চাউল] {} "rice(icl>grain>thing)" (N, C)<B,0,0>
[ধান] {} "rice(icl>grain>thing)" (N, C)<B,0,0>
[ধান] {} "rice(icl>paddy>thing)" (N, C, PLANT)<B,0,0>

Same way we have to resolve many other ambiguities while preparing Bangla Word Dictionary.

B. Outlining the Morphological Analysis of Bangla Words Compatible with UNL structure

Morphological analysis is found to be centered on analysis and generation of word forms. It deals with the internal structure of words and how words can be formed [8]. It is applied to identify the actual meaning of the words [6, 7] by identifying the Prefixes (উপসর্গ) and Suffixes(প্রত্যয়)

C. Prefixes (উপসর্গ)

Prefixes are the words that are used before words to express various meanings of the same words. There are around fifty (50) prefixes used in Bangla sentences. In Shangskrit Bangla we use twenty (20) prefixes[2] say ,প্র(প্রকর্ষ), পরা(বৈপরীত্য), অপ(বৈপরীত্য), etc. , in Bangla we use thirteen prefixes (13) prefixes[2]such as বে(বৈপরীত্য), গর(বৈপরীত্য), অন(অভাব)etc., five(5) foreign prefixes[5] such as গর (না), দর (নিম্নস্হ), বদ্ (খারাপ) etc.,four English prefixes[3] such as সাব(অধীন অর্থে), হেড(প্রধান), ফুল(পুরা), হাফ(অর্ধ) etc. and other prefixes[3] saypপুরঃ(সমক্ষে/সামনে),প্রাদুঃ(দৃষ্টিগোচর),বহিঃ(বাহিরে), etc. These prefixes are used before words to make thousands of meaningful Bangla Words.

In our work, we will make separately Word Dictionary entries for all of these prefixes and words, so that they can combine meaningful words by applying rules. For example, if we consider prefix "প্রতি"[10](means like/similar/every/opposite/against etc.) we can make "প্রতিদিন", "প্রতিশব্দ", "প্রতিপক্ষ"etc. Now we can make the word "প্রতি" for dictionary entry. But the word "প্রতি" has two or more meanings so that we have to represent two or more dictionary entries for the word as follows.

[প্রতি]{} "every (icl>thing)" (ABSTRACT THING) <B,0,0>
[প্রতি]{} "opposite (icl>thing)" (ABSTRACT THING) <B,0,0>
[প্রতি]{} "against (icl>thing)" (ABSTRACT THING) <B,0,0>

Now if we want to represent the concepts of the words say প্রতিদিন, প্রতিশব্দ , প্রতিপক্ষ etc., we need not represent the whole words. We have to represent only the words "দিন" ,"শব্দ" and"পক্ষ" in the dictionary entry as per the following format.

[দিন] {} "day(icl>period>time)"(N,ABSTRACT THING, LIGHT)<B,0,0>
[শব্দ] {} "sound(icl>occure>thing)"(N,ABSTRACT THING)<B,0,0>
[পক্ষ] {} "group(icl>person)"(N,CONCRETE)<B,0,0>

If we have the concepts of the prefix "প্রতি" and the root words "দিন", "শব্দ", and"পক্ষ" with their grammatical attributes in the Word Dictionary as above format, the conversion rule will make the concepts of the whole words"প্রতিদিন", "প্রতিশব্দ" and "প্রতিপক্ষ", combining the first, second and third concepts of "প্রতি" respectively. By applying the same rule the EnCo can make all other words used with "প্রতি", which have the concepts of the words in the word dictionary.

Similarly, if we consider Bangla prefix "রাম" we can make "রামছাগল", "রামদা"etc. We can separately represent the concepts of "রাম" , "ছাগল"and"দা"in the dictionary entry according to the following format.

[রাম] {}"big(icl>large>thing)"(ADJ, ABSTRACT THING) <B,0,0>
[ছাগল]{}"goat(icl>animal>animatething)"(N, CONCRETE, ANI) <B,0,0>
দা{}"knife(icl>edge_tool>thing)"(N,C) <B,0,0>

Therefore, if we have the concepts of all the words in the dictionary we can make the dictionary entry of all the complete words combined with "রাম". Here we also can use "ছাগল" and "দা" as separate words for other dictionary entries. Finally, we can infer that conversion rules can be applied to prepare thousands of complete Bangla Words combining with prefixes (mentioned above) and words to represent their full concepts in the Bangla-UNL Dictionary.

D. Suffixes (প্রত্যয়)

Morphological analysis describes that every word is derived from a root word. A root word may have different transformations. This happens because of adding different morphemes with it as suffixes. So, the meaning of the word varies for its different transformations. There are four different types of morphologies [9].

1) *Noun Morphology:* Bangla Nouns have very strong and structural inflectional morphology base on case. Case of noun may be nominative ("ছেলে", boy), accusative ("ছেলে-কে", to the boy) and genetive("ছেলে-র", of the boy) and so on. Gender and number are also important for identifying proper categories of nouns. Number may be singular ("ছেলে", boy or

ISBN 978- 974-672-556-9

"ছেলেটি", the boy, "বই", book, "বইটি", the book) plural ("ছেলেরা", boys "ছেলেগুলি", the boys"বইগুলো", the books etc.). So, from the word "ছেলে" we get "ছেলের", "ছেলেকে", "ছেলেরা" "ছেলেটি", "ছেলেগুলি" etc. and from the word "বই" we get "বইটি", "বইগুলো" etc. Some dictionary entries may look like. [ছেলে]{} "boy (icl>person)" (N, HN, C, ANI)<B,0,0>
Here, "boy (icl>person)" is the UW for "ছেলে" but "র" , "কে" etc. have no UWs. Therefore, they should be represented in the dictionary only with grammatical attributes as follows.

[র] {} "" (3P, SUF, N)<B,0,0>
[কে] {}""(3P, SUF, N, HUMN, SG)<B,0,0>
[রা] {}"" (3P, PL, SUF, N, HUMN)<B,0,0>
[টি] {}"" (N, SG, SUF,3P) <B,0,0>
[গুলি] {}"" (N, PL, SUF,3P) (<B,0,0>
[গুলো] {}""(N, SG, SUF,3P) <B,0,0>

We use 3P, SUF and N as grammatical attributes with "রা", because "রা" is used with third person say "ছেলেরা", N for noun and SUF as "রা" is a suffix. We have to put meticulous attention while defining the grammatical attributes. Because we use HUMN for human noun as "কে", "রা" are used with human being only, say ছেলেকে, তাহাকে, ছেলেরা, তাহারাbut notগরুকে , গরুরাetc. But we can not use HUMN with "র","টি", "গুলি" and"গুলো" because they are used with both human and non human, say পাখির, ছেলেটি, গরুগুলো, etc.

2) *Adjective Morphology:*As Adjective we can consider Bangla words "সাহস", "সুন্দর" and "ভালো" meaning "bravery", "beautiful" and "good" in English respectively. From the first word we get সাহসী (সাহস+ই), সাহসের (সাহস+এর).And from the second and third words we get সুন্দরী, ভালোর, ভালোটাetc. We have to have the dictionary entries for সাহস, সুন্দর, ভালো, ই, এর, র, টাto make the meaningful words সাহসী, সাহসের, সুন্দরী, ভালোটাetc. by combining the morphemes with the root words using analysis rules.

For example, if we consider a Bangla sentence, "সাহসীরা সাহসের সাথে অন্যায়ের প্রতিবাদ করে।"We can represent the sentence as "সাহস-ই-রা সাহস-এর সাথে অন্যায়-এর প্রতি-বাদ কর-এ।". Here, ই, রা, এর, বাদ, এ etc. are morphemes. So, we can see that a number of morphemes are added with the root words to make the full meaning of the new words as well as sentence.

3) *Pronoun Morphology:* Here we can consider the word root "তাহা"(he/she). From this we get তাহা-রা, তাহা-কে, তাহা-দের, তাহা-দের-কে, তাহা-দিগকে etc. So, we have to consider these morphemes রা, কে, দের, দিগকেfor dictionary entries to form words with "তাহা" as above.

4) *Verb Morphology:* Diversity of verb morphology in Bangla is very significant. If we consider 'যা' (means go) as a root, we can represent this root in the dictionary as

[যা]{} "go (icl>move>do)" (V, @present) <B,0,0>
Some transformations based on the persons and tenses are.

- For first person:

[ই]{} "" (SUF, PRESENT, 1P)<B,0,0> for যাই (যা+ই))
[ইতেছি] {}"" (SUF, PRESENT, 1P) <B,0,0> (for যাইতেছি (যা+ইতেছি)) etc.

- For second person:

[ইতেছিলেন] {}""(SUF, PAST)<B,0,0> (forযাইতেছিলেন (যা+ইতেছিলেন))

- For third person:

[বে] {}"" (SUF, FUTURE)<B,0,0> (forযাবে (যা+বে))

For resolving the ambiguities of the words গিয়েছি, গিয়েছিলাম, গিয়েছেন, গিয়েছিলেন, যাইতে থাকবেন, গিয়াছে etc. we have to define them as full words for dictionary entries. For instance,
[গিয়েছিলাম]{} "go(icl>move>do)"(V, PAST, INDEF, 1P). Using the same procedure we can make dictionary entries for different transformations of other roots such as কর্(do), লিখ্(write), দে (give) etc.

Moreover, there are a huge number of Primary (কৃৎ প্রত্যয়) and Secondary (তদ্ধিত প্রতয়) suffixes used with roots and words. Each of them has own meaning [4]. For example, লৌকিক (লোক+ইক), মাসিক(মাস+ইক), দৈনিক(দিন+ইক)etc. Here, ইকis a suffix added with "লোক", "মাস"and "দিন"words to form new words.

*Primary suffixes (কৃৎ প্রত্যয়):*The suffixes that are used after roots to form new meaningful words called primary suffixes [4]. In examples above ই, ইতেছি, ইতেছিলেন,বে are all primary suffixes. In addition to these there are many more primary suffixes like অন (বাঁধ্+অন=বাঁধন, নাচ্+অন=নাচন), আ(পড়্+আ=পড়া) etc.

Secondary suffixes (তদ্ধিত প্রত্যয়): The suffixes that are used after words to form meaningful new words called secondary suffixes. Examples are given above with noun, adjective and pronoun morphologies. In addition, there are many other secondary suffixes like অই (পাঁচ+অই=পাঁচই, সাত+অই=সাতই) আই (মিঠা+আই=মিঠাই, ঢাকা+আই=ঢাকাই, পাটনা+আই=পাটনাই)etc.

We will outline Dictionary entries for all these primary and secondary suffixes along with their grammatical attributes, so that we can prepare thousands of Bangla Words combining with roots and words for building Bangla Word Dictionary for UNL.

IV. CONCLUSION

In this paper we have outlined a dictionary development procedure of simple Bangla Words and roots along with their prefixes and suffixes for UNL. We have also presented morphological structures of the Bangla Words compatible with UNL structures and outlined grammatical attributes of the Head

Words (Bangla Head Words means all the Dictionary entries of Words, morphemes etc.) for developing Bangla Word Dictionary to convert the Bangla sentences to UNL expressions and vice versa. Our future plan in this respect is to outline comprehensive Bangla Word Dictionary of around 50000 words and analysis rules for EnConverter to convert Bangla sentences to UNL document and generation rules for DeConverter to convert any UNL document to Bangla sentence.

REFERENCES

[1] H.Uchida, M. Zhu, and T. C. D. Senta, *Universal Networking Language*. NDL Foundation, International environment house, 2005/6, Geneva, Switzerland.

[2] D. M. Shahidullah, "Bangala Vyakaran", Maola Brothers Prokashoni, Dhaka, August 2003, pp.110-130

[3] D.C. Shniti Kumar, "Vasha-Prokash Bangla Vyakaran",

[4] Rupa and Company Prokashoni, Calcutta, July 1999, pp.170-175

[5] D. S. Rameswar, "Shadharan Vasha Biggan and Bangla Vasha", Pustok Biponi Prokashoni, November 1996, pp.358- 377

[6] H. Uchida, M. Zhu, "The Universal Networking Language (UNL) Specification Version 3.0", Technical Report, United Nations University, Tokyo, 1998

[7] M. M. Asaduzzaman, M. M. Ali, "Morphological Analysis of Bangla Words for Automatic Machine Translation", International Conference on Computer and Information Technology (ICCIT), Dhaka, 2003,pp.271-276

[8] S. Dashgupta, N. Khan, D.S.H. Pavel, A.I. Sarkar, M. Khan, "Morphological Analysis of Inflecting Compound words in Bangla", International Conference on Computer, and Communication Engineering (ICCIT), Dhaka, 2005, pp. 110-117.

[9] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, A.M. Nurannabi, "Morphological Analysis of Bangla Words for Universal Networking Language", Third International Conference on Digital Information Management (ICDIM 2008), London, England.pp. 532-537

[10] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, M. E.H. Choudhury, "Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language", International Conference on Computer and Communication Engineering 2008 (ICCCE'08), Kuala Lumpur, Malaysia, pp. 726-731

[11] Bangla Academy, "Bengali-English Dictionary" Bangla Academy Dhaka 2007

[12] http://www.undl.org/unlsys/uw/UNLKB.htm

[13] http://www.unl.ru/deco

[14] EnConverter Specification, Version 3.0, UNL Center, UNDL Foundation, Tokyo 150-8304, Japan 2002

[15] DeConverter Specification, Version 2.7, UNL Center, UNDL Foundation, Tokyo 150-8304, Japan 2002.

Conversion of Bangla Sentence to Universal Networking Language

Muhammad Firoz Mridha[1], Md. Nawab Yousuf Ali[2], Manoj Banik[3], Mohammad Nurul Huda[4], Chowdhury Mofizur Rahman[4], Jugal Krishna Das[5]
[1]Department of CSE, Stamford University Bangladesh, [2]East West University, Dhaka, Bangladesh, [3]Department of CSE, Ahsanullah University of Science and Technology, Dhaka, Bangladesh, [4]United International University, Dhaka, Bangladesh, [5]Jahangirnagar University, Dhaka, Bangladesh.
mdfirozm@yahoo.com, nawab@ewubd.edu, mbanik99@yahoo.com, mnh@cse.uiu.ac.bd, cmr@uiu.ac.bd, drdas64@yahoo.com

***Abstract*—Conversion from another language to native language is highly demanding due to increasing the usage of web based application. Firstly, the respective sentence of a native language is converted to Universal Networking Language (UNL) expressions and then UNL expressions can be converted to any native language. There is no method to convert Bangla sentence to UNL expressions. This motivates to propose a new model that performs morphological, syntactic, semantic, and lexical analysis synchronously to convert Bangla to UNL. The proposed model will successfully able to convert error-free and ambiguous Bangla sentences to UNL expressions. Hopefully this paper will be helpful for the MT researchers to build and efficient MT System for Bangla Language.**

Index Terms- DeConverter; Enconversion; EnConverter; Natural Language Parsing; Predicate Preservation; Universal Networking Language; Universal Words

I. INTRODUCTION

The Internet today has to face the complexity of dealing with multilinguality. People speak different languages and the number of natural languages along with their dialects is estimated to be close to 4000 [5]. Of the top 100 languages in the world, English occupies the top position. That is why English is the main language of the Internet. But not all people know English. And vast information resources in different languages that are scattered all over the world remains mostly inaccessible due to non machine representation and language barrier. But knowledge and information should be shared globally as much as possible to advance civilization [11]. This is how the problem interlingua translation rose to its new heights worldwide. Among those who did their best to tackle this problem was the United Nations University/Institute of Advance Studies (UNU/IAS) [2, 5]. The institute conducted a review of all internationally available machine translation programs, and finally decided to start devising a better, more efficient and more workable technique via the Internet under the Universal Networking Language (UNL) project. The UNL project is concerned in developing an intermediary language system whereby any written text can be converted to many languages (all languages involved in the UNL programs; so far sixteen official languages are involved, Bangla is not yet included) through UNL and simultaneously, all text written in different languages can be converted to that particular language.

Bangla is the fourthwidely spoken language with more than 250 million speakers, most of whom live in Bangladesh and the Indian state of West Bengle [7]. Therefore it is essential to take into account a conversion process of Bangla to UNL and vice versa, so that a vast people can be benefited to share information in the Internet. We attempt a system to automate the generation of semantic net like expressions from text documents. The objective is to establish appropriate relations between the syntactic units of a sentence to capture the syntactic and semantic attributes of words. The output of the system is a set of *UNL Expression*, which is binary relations among disambiguated words along with speech acts attributes attached to these disambiguated words. The UNL represents information, i.e. meaning sentence by sentence. Each sentence is represented as a hypergraph having concepts as nodes and arcs represent relation between concepts. This hypergraph is also represented as a set of directed binary relations, each between two of the concepts present in the sentence. Concepts are represented as character-strings called Universal Words (UWs). Knowledge within a UNL document is expressed in three dimensions:

A. Universal Words (UWs).

Word knowledge is expressed by Universal Words which are language independent. UWs constitute the UNL vocabulary and the syntactic and semantic units that are combined according to the UNL laws to form UNL expressions. They are tagged using restrictions describing the sense of the word in the current context. For example, *drink (icl>liquor)* denotes the noun sense of *drink* restricting the sense to a type of *liquor*. Here *icl* stands for inclusion and form an is-a relation like in semantic nets [8].

ISBN 978- 974-672-556-9

B. Relation Labels.

Conceptual knowledge is captured by the relationship between Universal Words (UWs) through a set of UNL relations [2]. For example, *Human affect the environment* is described in the UNL expression as,

agt(affect(icl>do).@present.@entry:01,human(icl>animal). @pl)
obj(affect(icl>do).@present.@entry:01,environment (icl>abstract thing).@pl)

where, **agt** means the *agent* and **obj** the *object.* affect (icl>do), human (icl>animal) and environment (icl>abstract thing) are the UWs denoting concepts.

C. Attribute Labels

Speaker's view, aspect, time of event, etc. are captured by UNL attributes. For instance, in the above example, the attribute *@entry* denotes the main predicate of the sentence, *@present* the *present tense* ,*@pl* the *plural number* and *:01* the *scope ID.* The above discussion can be summarized using the following UNL expression and as a UNL graph.

"কামাল যিনি বিশ্ববিদ্যালয়ের অধাপক, তিনি তাহার অফিসে এক আলোচনা সভার আয়োজন করেছে।"

In English, "Kamal, who is the professor of the university, has arranged a meeting at his office."
The UNL expression of the above sentence is:

mod(professor(icl>post):01.@present.@def,university (icl>institution):02.@def)
aoj(professor(icl>post):01.@present.@def, kamal (icl>person):00)
agt(arrange(icl>do):03.@entry.@present.@complete.@pred, kamal(icl>person):00)
pos(office (icl>shelter):04, kamal (icl>person):00)
obj(arrange(icl>do):03.@entry.@present.@complete.@pred, meeting(icl>conference):05.@indef)
plc(arrange(icl>do):03.@entry.@present.@complete.@pred, office(icl>shelter):04)

The graph for the sentence is given in figure 1.

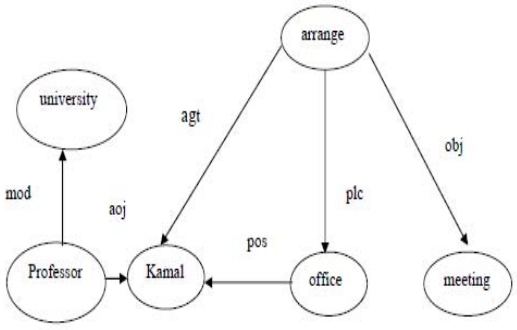


Figure 1. UNL graph

In the figure above, *agt* denotes the *agent* relation, *obj* the *object* relation, *plc* the *place* relation, *pos* is the *possessor* relation, *mod* is the *modifier* relation and *aoj* is the *attribute-of-the-object* relation. The UNL expressions provide the *meaning content* of the text and search could be carried out on this meaning base instead of the text. This of course means developing a novel kind of search engine technology. The merit of such a system is that the information in one language needs to be stored in multiple languages. This paper organized as follows: the literature review related to the analyzer system EnConverter and Lexicon (format of word in the dictionary) is detailed in Section 2 while the proposed model with its constituent parts is presented in Section 3 and finally Section 5 focuses some concluding remarks.

II. LITERATURE REVIEWS

The UNL structures of the EnConverter and UNL format of Bangla Word Dictionary that are related to the proposed work are detailled in the next sections.

A. *Enconverter Machine*

The EnConverter (EnCo) [2, 3, 4] is a language-independent parser provided by the UNL project, a multi-headed Turing Machine [8] providing synchronously a framework for morphological, syntactic and semantic analysis. The machine has two types of windows namely *Analysis Windows (AW)* and *Condition Windows (CW).* The machine traverses the input sentence back and forth, retrieves the relevant dictionary entry (UW) from the Word Dictionary (Lexicon) and depending on the *attributes* of the nodes under the AWs and those under the surrounding CWs and finally generates the semantic relations between the UWs and /or attaches speech act attributes to them.

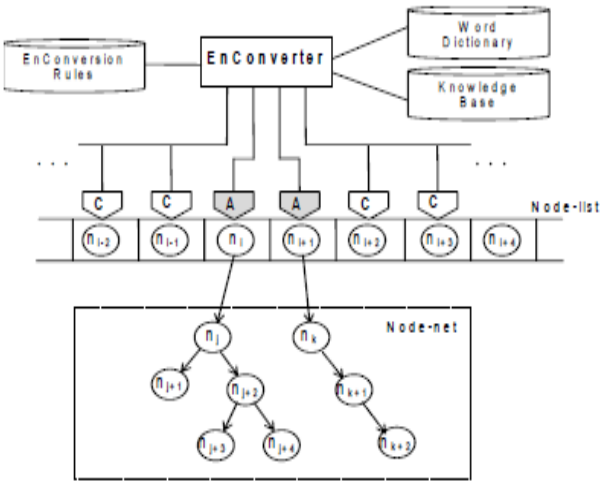


Figure 2. Structure of EnConverter

"A" indicates an Analysis Window, "C" indicates a Condition Window, and "n_i" indicates an Analysis Node. As a result, a set of UNL expressions is made equivalent of UNL graph.

B. Analysis rules theory

EnCo is driven by analysis rules to analyze a sentence using Word Dictionary and Knowledge Base. These rules are *condition-action structure* that can be looked upon as program written in a specialized language to process various complex phenomena of a natural language sentences. The enconversion rule has the following format: [6]

<TYPE>
["(" <PRE> ")" ["*"]]...
"{"|""""" [<COND1>] ":" [<ACTION1>] ":"
[<RELATION1>] ":" [<ROLE1>] "}"|"""""
["(" <MID> ")" ["*"]]...
"{"|""""" [<COND2>] ":" [<ACTION2>] ":"
[<RELATION2>] ":" [<ROLE2>] "}"|"""""
["(" <SUF> ")" ["*"]]...
"P(" <PRIORITY> ");"

Where, characters between double quotes are predefined delimiters of the rule.

The rule means that IF , under the Left Analysis Window (LAW) there is a node that satisfies <COND1> attributes and under the Right Analysis Widow(RAW) a node that satisfies <COND2> attributes and there are nodes to the left of the LAW, between the LAW and the RAW and to the RAW that fulfill the conditions in <PRE>, <MID> and <SUF> respectively, THEN the lexical attributes in the nodes under the AWs are rewritten according to the <ACTION1> and <ACTION2> as specified in rule and new attributes are added if necessary. The operations are done on the node-list depending on the type of the rule shown in the field <TYPE>. There are 15 types of enconversion rules [5]. They are left composition rule (+), right composition rule(-), left modification rule(<), right modification rule(>), left shift (L), right shift(R), insertion/attribute changing(:), backtrack(?), left node assignment backtrack (?L), right node assignment backtrack((?R), node copy(C), Syntactic tree coy(G), node exchange (X), left node deletion (DL) and right node deletion(DR). <RELATION1> describes the semantic relation of the node on the RAW to the node on the LAW and <RELATION2> describes the reverse [6]. <PRIORITY> describes the interpretation order of the rules, which is in the range of 0-255. Larger number indicates higher priority. Matching rule with the highest priority is selected for multiple matching rules [1].

A sequence of such rules get activated depending on the sentence situation i.e. the conditions of the nodes under the AWs. The main task is to creating the UNL expressions of natural language sentences using EnCo by providing a rich lexicon and a comprehensive set of analysis rules. For example like, ‘করিম ভাংগে’ *Karim breaks*, the *animate* attribute of *Karim*, the *verb* attribute of *break* and the adjacency of these two words under the AWs dictate with high probability establishing *agt (agent)* relation between the corresponding two nodes in the UNL graph. The dictionary entries (along with the rich lexicons) of the above two words in the sentence are:

[করিম] {} “Karim(icl>person)”(N, ANI, 1P, SG, CONCRETE, SUBJ)
[ভাংগা] {} “break(icl>do)” (V, ROOT, SORNT, PRES, SIMPL)

Here N stands for noun, ANI stands for animate object, 1P for first person, SG for singular, CONCRETE for concrete object, SUBJ for subject of the sentence, V for verb, ROOT for verb root, SORNT means that the word ‘ভাংগা’ is ended with vowel, PRES and SIMPLE for simple present

C. Lexicon

The lexicon used for the system consists of mapping of Bangla words to Universal Words and lexical-semantic attributes describing the words. Any entry in the Dictionary is put in the following format [2]:

[HW] {ID}“UW”(ATTRIBUTE1, ATTRIBUTE2 . . .) <FLG, FRE, PRI>

Where, HW ← Head Word (Bangla word), ID ← Identification of Head Word (omitable), UW ←Universal Word, ATTRIBUTE ←Attribute of the HW, FLG ←Language Flag (e.g. B for Bangla), FRE ←Frequency of Head Word, PRI ←Priority of Head Word .

Some examples of Dictionary entries for Bangla are given below.

[মাছ] {} “fish (icl>animal>animate thing)” (N, ANI, SG, CONCRETE) <B,0,0>
[সে] {} “he(icl>person)” (PRON, MALE, ANI, 3SG, HPRON) <B,0,0>
[সুন্দর] {} “beautiful (icl>state)” (ADJ) <B,0,0>

Here the attributes, N stands for Noun, ANI stands for animate object, SG stands for singular, CONCRETE for concrete object, PRON for pronoun, MALE for gender is male, 3SG for third person singular, HPRON for human pronoun, ADJ for adjective and FLG field entry is B which stands for Bangla. The attributes in the lexicon are collectively called *Lexical Attribute*. The syntactic attributes include the word category-noun, verb, adjectives, etc. and attributes like *person* and *number* for nouns and *tense* information for verbs.

III. PROPOSED WORK

In this section we describe about the preserving of predicate till the end of the conversion, scanning and encoding Bangla Sentence to UNL expressions.

A. Predicate Preserving Parser

As EnCo scans the input sentences from left to right, the Morphological Analysis and Decision Making are taken place at every step. Morphology is concerned in which words are formed from basic sequences of morphemes. It acts as the crossroads between phonology, lexicon, syntax, semantics and context. Morphology [8, 9, 13] is required to identify the actual meaning of the word. Because the meaning of the word varies for its different transformations. For example, if we consider a Bangla word 'খা' as a root then after adding "বে" we get a word "খাবে" (will eat). Similarly, "খাইতেছে" (is eating), "খেয়েছিলো" (ate) etc. The UWs defined in the lexicon have roots specified as UWs followed by appropriate restrictions while the headword has the *Longest Common Lexical Unit* (LCLU) of all the possible transformations of the roots. For example, the dictionary entry for the verb 'খা' (to eat) is:

[খা] {} "eat (icl>consume>do)" (List of Semantic and Syntactic Attributes) <B, 0, 0>

Where "খা" is the LCLU for "খাবে", "খাইতেছে", "খেয়েছিলো" etc.

Word suffixes along with their attributes for nouns, verbs and adjectives are kept in the lexicon. For example,

[বে]{}""(V,SUF,FUTURE INDEF, PRON) <B, 0, 0>
[ইতেছে] {} "" (V, SUF, PRESENT CONT, PRON) <B, 0, 0>

EnCo selects the longest matched entry from the lexicon, starting from the first character. So, when EnCo consults the dictionary for a particular morpheme say "ইতেছে", will be able to retrieve the LCLU (খা here) present in the dictionary. Then the analysis rules, which look ahead and signal that there is a verb suffix ahead, take control and complete the morphological analysis.

In case of decision making according to the lexical attributes of the nodes under the two AWs, the parser decides whether (i) the nodes are to be combined into a single headword or (ii) a relation is to be set up between them or (iii) an UNL attribute is to be generated. While combining or modifying the two nodes, one of the nodes is deleted from the node-list. Multiple rules may be applied for firing in a situation, calling for assignment of priorities for the rules as in expert systems. The strategy for prioritizing the rules is briefly as follows [8].

1) Morphological analysis rules have the highest priority. Obviously, unless we have the morphed word we cannot decide the part of speech of the word and its relation with the adjacent words.

2) Rules for dealing with specific constructs are given higher priority than those for general sentence structures. For instance, rules for clausal and passive sentences are given higher priority, so that while analyzing clausal or passive sentences a general rule-eligible to be applied-does not fire.

3) Right shift rules which facilitate right movement when there is nothing else to do are given the lowest priority. For example, when the LAW (Left Analysis Window) is on SHEAD (sentence start marker) and the RAW (Right Analysis Window) is on the subject (N), no rule other than the right shift is applicable. This rule, which is very useful, is:

R {SHEAD: : :} {N : : :}P1

4) Composition rules are usually given less priority than modification rules. The former ultimately resolve relations while the latter change the properties of the nodes under the AWs.

B. Scanning a Bangla sentence

Say we convert the Bangla sentence, "আমি ভাত খাইতেছি।" meaning, I am eating rice. We assume that we have all the words and morphemes of the give sentence in the dictionary as follows:

[আমি]{} "i(icl>person)"(1P, SG, SUB, PRON, HPRON)
[ভাত]{} "rice(icl>cereal>thing)" (OBJ, CONCRETE, FOOD)
[খা]{} "eat(icl> consume > do)" (V, ROOT, SORANT)
[ইতেছি]{}(KBIVOK, SORANT, PRESENT, CONT)

EnCo can input either a string or a list of words for a sentence of a native language. A list of morphemes/words of a sentence must be enclosed by [<<] and [>>] [6]. When we input our sentence into EnCo, the Sentence Head (<<) will be on LAW, sentence texts/morphemes/words will be on RAW and the Sentence Tail (>>) will be on Right Condition Window (RCW) shown in figure 1. EnCo uses CWs for checking the neighboring nodes on both sides of the AWs in order to judge whether the neighboring nodes satisfy the conditions for applying an *analysis rule* or not.

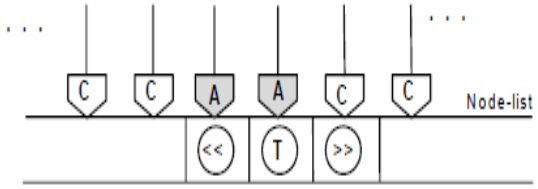


Figure 3: Initial state of the Analysis Windows and the Node-list.

When the sentence is input, the input sentence is scanned from left to right. Here left most word of our sentence is "আমি". When an input string "আমি" is scanned, all matched morphemes with the same string characters e.g. আ, আমি, আম, আমই etc. are retrieved from the Word Dictionary and become the candidate morphemes according to a rule priority in

order to build the syntactic tree and the semantic network (UNL expressions) of UNL for the sentence.

In this case, the following insertion rule is applied to these candidate morphemes to select "আমি" from Word Dictionaryfor the node list.

Rule 1:
: {SHEAD :::} {SUBJ :::}
Rule LAW RAW

{"SHEAD" :::} denotes the condition of the node on the LAW, here sentence head (<<) is on the LAW, {SUBJ :::} denotes the condition of the node on the RAW, here "আমি" is the first word of the sentence which is subject. This rule is applied to insert the subject "আমি" of the sentence into the node-list and word "আমি" is shifted left to the next window which is LAW. Now EnCo analyzes the next word of the sentence "ভাত" like as "আমি" using the following rule to insert the word into the node-list

Rule 2:
: (SHEAD, SANI) {SUBJ :::} {OBJ, FOOD :::}
Rule LCW LAW RAW
Type

This rule can be applied to insert word "ভাত" into the node-list, when Sentence Head(SHEAD) is on the LCW1 and subject is animate(SANNI) , which is on the LAW and the object "ভাত" is on the RAW. After applying the rule the word "ভাত" will be shifted left to the next window and the last word of the sentence" খাইতেছি" will be on the RAW.

Now EnCo starts morphological analyses with the word "খাইতেছি", to find the actual meaning of the word. It first breaks the word into "খা" and "ইতেছি", which are available in the dictionary. Here rule 3 and rule 4 are applied to insert root "খা" and suffix "ইতেছি" in the node list:

Rule 3:
:(SUBJ, SANI) {OBJ:::}{ROOT, SANI:::}

Rule 4:
: (SUBJ, 1P) (OBJ, FOOD){ROOT:::}{KBIVOK, 1P, HPRON}

The analyzer then adds the root (খা) and suffix (ইতেছি) using rule 5 to find out the actual meaning of the word "খাইতেছি" from the dictionary.

Rule 5:
+:+{ROOT, SORANT:::}{KBIVOKTI, SORANT, PRESENT:::}

When the EnCo finds the meanings of all the words/morphemes of the input sentence from the *word dictionary*, it makes two binary relations namely agent (*agt*) relation between the words eat (খা) and I (আমি) and object (*obj*) relation between the words eat (খা) and rice (ভাত) using the rules 6 and 7.

Rule 6: > PRON::agt :}{ V, ^SUBJ :::}

Rule 7: > {N::obj} {V, ^SUB :::}

Where, ^ indicates the negation of the following attributes.

Agent relation: The rule 6 denotes that the UW of the node pronoun, which has the attribute *PRON* will act as the agent indicated by the relation *agt* of the UW of the following node verb, which has the attribute *V*. When this rule is applied, a binary relation of UNL such as "agt(UW of V , UW of Pronoun) is created.

Object relation: The rule 7 describes that the UW of the node noun, which has the attribute *N*, will act as an object relation indicated by the relation *obj*, with the UW of the following node verb, which has the attribute V. When the rule 7 is applied a binary relation of UNL such as "obj(UW of V, UW of N) is created.

When the both rules 6 and 7 are applied the following UNL expressions will be created.

agt(eat(icl>consume>do,agt>living_thing,obj>concrete_thing).@entry.@present.@progress,i(icl>person))
obj(eat(icl>consume>do,agt>living_thing,obj>concrete_thing).@entry.@present.@progress,rice(icl>cereal>thing))

As we have already the UNL expression of the given Bangla sentence by EnCo, this expression can be deconverted to any native language, the member of UNL society.

C. Encoding a Bangla sentence

The encoding process will be performed by shift/reduce parsing [1]. We have observed that Hindi language has syntactic similarities with Bangla and Hindi to UNL [5, 12] system developed at Indian Institute of Technology, Bombay serves as a reference for us. Here we give an example of the analysis of a simple assertive sentence. Assertive simple sentences have only one main clause. The analysis of such sentences is explained below with an example. We assume that analysis rules (Enconversion rules) and the dictionary of Bangla to UNL are given to the analyzer system EnCo.

The Bangla sentence is:
"**বাড়িতেআজখুবআনন্দ**"*(At home today is very joy.)*

In English:
"It is a very joy today at home."

The node list is shown here within "<<"and ">>". The analysis window is within "[" and "]". Steps related to morphological analysis are shown here. The nodes delimited by "/" are those explored and fixed by the system. /<</ [বাড়ি] / [তে] / আজখুব আনন্দ>>/. The pronoun "বাড়ি" of the type *PLACE* and case

ISBN 978- 974-672-556-9

maker "ে**"** are combined and an attribute *PLC* is added to the noun to indicate that *plc* relation can be made between the main predicate of the sentence and this noun. /<</ [বাড়ি ে] **/** আজ খুব আনন্দ**/>>/.**In this step, the system right shifts here because there is no combination or modification rule between noun and adverb. **/<</** বাড়ি ে **/** আজ **/** [খুব] **/** [আনন্দ**]>>/ .** The system recognizes "খারাপ" as a predicate of the sentence. So, it generates *man* relation between the adverb "খুব" and the predicate "খারাপ" and "খুব" is deleted. So we get, /<</বাড়ি ে **/** [আজ] **/** [আনন্দ]/>>/

Then the analyzer looks ahead further right shift beyond the noun phrase "**আজ"** to get the predicate "আনন্দ". A *tim* relation is created between "**আজ**" and "আনন্দ" and "**আজ**" is deleted. Current sentence in the analysis window is as follows: /<</ **[**বাড়ি ে] **/** [আনন্দ />>/. Similarly, by using *PLACE* attribute of the noun "বাড়ি", the system generates *plc* relation between the noun "বাড়ি" and the predicate "আনন্দ" and "বাড়ি" is deleted. /<</ আনন্দ**/>>/**

A right shift at this point brings the Sentence Tail (STAIL) under the LAW and thus signals the end of analysis. This right shift rules also attaches the attribute @entry to the last word left in the Node-list and thus the predicate "আনন্দ" is preserved till the end. The UNL output of the corresponding sentence is:

aoj(joy(icl>emotion>thing,ant>sorrow).@entry.@present,it(icl>thing))
fictit(joy(icl>emotion>thing,ant>sorrow).@entry.@present,very(icl>how,equ>extremely))
tim(joy(icl>emotion>thing,ant>sorrow).@entry.@present,today(icl>how,equ>nowadays))
fictit(joy(icl>emotion>thing,ant>sorrow).@entry.@present,at_home(icl>how))

We have also verified that some other types of sentences such as interrogative and exclamation sentences can similarly be translated.

IV. Conclusion

In this paper a pioneer approach is presented for Universal Networking Language (UNL) expressions. This model uses a language analyzer called EnConverter developed by the UNL organization. The Bangla Sentence will be firstly converted to UNL expressions by integrating morphological, syntactic, semantic and lexical analysis rules into the framework of UNL and then the UNL expressions can be deconverted to any native language. This will increase the application of Bangla Sentence in web based technology.

References

[1] J. Earley, "An Efficient Context Free Parsing Algorithm,"*Communications of the ACM*, 13(2), 1970

[2] H. Uchida, M. Zhu, "The Universal Networking Language (UNL) Specification Version 3.0", Technical Report, United Nations University, Tokyo, 1998

[3] *Enconverter Specification Version 3.3*, UNU Center/ UNDL Foundation, Tokyo 150-8304, Japan 2002.

[4] *Enconverter Specifications, Version 3.3*, UNL Center/UNDL Foundation, Tokyo, Japan 2002.

[5] P. Bhattacharyya, "Multilingual Information Processing Using Universal Networking Language", Indo UK Workshop on Language Engineering for South Asian Languages LESAL, Mumbai, India, 2001

[6] H. Uchida, M. Zhu, and T. C. D. Senta, *Universal Networking Language*, NDL Foundation, International environmenthouse, 2005/6, Geneva, Switzerland.

[7] S. Dashgupta, M. Khan, D.S.H. Pavel, A.I. Sarkar, M. Khan, "Morphological Analysis of Inflecting Compound words in Bangla," International Conference on Computer, and Communication Engineering (ICCIT), Dhaka, 2005, pp.110-117

[8] J. Parikh, J. Khot, S. Dave, P. Bhattacharyya, "Predicate Preserving Parsing," Department of Computer Science and Engineering, Indian Institute of Technology, Bombay.

[9] M. M. Asaduzzaman, M.M. Ali, "Morphological Analysis ofBangla Words for Automatic Machine Translation," International Conference on Computer, and InformationTechnology (ICCIT), Dhaka, 2003, pp. 271-276

[10] S. Dashgupta, M. Khan, "Morphological Parsing of Bangla Words using PC KIMMO," International Conference on Computer, and Information Technology (ICCIT), Dhaka, 2004

[11] M.E.H. Choudhury, M.N.Y. Ali, "Framework for synthesis of Universal Networking Language," *East West Journal*, vol.2, Dhaka, 2008, pp.28-43.

[12] M.N.Y. Ali, J.K. Das, S.M. Al-Mamun, M. E.H. Choudhury "Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language, "International Conference on Computer, and Communication Engineering 2008 (ICCCE'08), Kuala Lumpur, Malaysia, pp. 726-731

[13] M.N.Y. Ali, J.K. Das, S.M. Al-Mamun, M. E.H. Choudhury, "Morphological Analysis of Bangala Words for Universal Networking Language," International Conference on Digital Information Management 2008 (ICDIM2008),London,England,pp.532-537.

Development of Grammatical Attributes for Bangla Root and Primary Suffix for Universal Networking Language

Muhammad Firoz Mridha[1],Md. Zakir Hossain[1], Manoj Banik[2], Mohammad Nurul Huda[3], Chowdhury Mofizur Rahman[3], Jugal Krishna Das[4]
[1]Department of CSE, Stamford University Bangladesh, [2]Department of CSE, Ahsanullah University of Science and Technology, Dhaka, Bangladesh, [3]United International University, Dhaka, Bangladesh,
[4]Jahangirnagar University, Dhaka, Bangladesh.
mdfirozm@yahoo.com, zakir267ju@yahoo.com, mbanik99@yahoo.com, mnh@cse.uiu.ac.bd, cmr@uiu.ac.bd, drdas64@yahoo.com

***Abstract*—The Universal Networking Language (UNL) deals with the communication across nations of different languages and involves with many different related discipline such as linguistics, epistemology, computer science etc. It helps to overcome the language barrier among people of different nations to solve problems emerging from current globalization trends and geopolitical interdependence. In this paper we propose a pioneer work that aims to contribute with morphological analysis of those Bangla words from which we obtain roots and Primary suffixes and developing of grammatical attributes for roots and Primary suffixes that can be used to prepare Bangla word dictionary and Enconversion/Deconversion rules.**

Index Terms- Bangla roots; Bangla words; Grammatical Attributes; Morphology; Primary Suffix; Universal Networking Language

I. INTRODUCTION

In the last few years, machine translation techniques have been applied to web environments. The growing amount of available multilingual information on the Internet and the Internet users has led to a justifiable interest on this area. Hundreds of millions of people of almost all levels of education and attitudes, of different jobs all over the world use the Internet different purposes [1]. English is the main language of the Internet. Understandably, not all people know English. Urgent need, therefore, arose to develop interlingua translation programs. The main goal of the UNL system is to provide Internet users access to multilingual websites using a common representation. This will allow users to visualize websites in their native languages. The UNL system has a growing relevance since the usage of the WWW is generalized across cultural and linguistic barriers. Many languages [10] such as Arabic, French, Russian, Spanish, Italian, English, Chinese or Brazilian Portuguese have already been included in the UNL platform. Lexical knowledge representation is a critical issue in natural language processing systems. Recently, the development of large-scale lexica with specific formats capable of being used by several different kinds of applications has been given special focus; in particular to multilingual systems. Our aim is to introduce Bangla into this system. In order to implement this project with the lowest time and human effort costs, we will reuse linguistic resources already available as much as possible.

In this paper we present the UNL system for Bangla. The major components of our research works touches upon i) development of grammatical attributes for Bangla root and Krit Prottoy to construct *Bangla Word Dictionary* and use of morphological analysis ii) UNL Expression of the Bangla attributes and *iii)* Selecting scheme of attributes. In section 2 we describe the UNL system. In sections 3 and 4, we present our main works that include all the above three components.

II. UNIVERSAL NETWORKING LANGUAGE

UNL is an artificial language that allows the processing of information across linguistic barriers [10]. This artificial language has been developed to convey linguistic expressions of natural languages for machine translation purposes. Such information is expressed in an unambiguous way through a semantic network with hyper-nodes. Nodes (that represent concepts) and arcs (that represent relations between concepts) compose the network. UNL contains three main elements:

- Universal Words: Nodes that represent word meaning.
- Relation Labels: Tags that represent the relationship between Universal Words. Tags are the arcs of UNL hypergraph.
- Attribute Labels: Additional information about the universal words.

These elements are combined in order to establish a hierarchical Knowledge Base (UNLKB) [10] that defines unambiguously the semantics of UWs. The UNL Development Set provides tools that enable the semi-automatic conversion of natural language into UNL and vice-versa. Two of such tools are the EnConverter and the DeConverter. The main role of

ISBN 978- 974-672-556-9

EnConverter [11] is to translate natural language sentences into UNL expressions. This tool implements a language independent parser that provides a framework for morphological, syntactic and semantic analysis synchronously. This allows morphological and syntactical ambiguities resolution. The DeConverter [3, 12], on the other hand, is a language independent generator that converts UNL expressions to natural language sentences.

A. Universal Words

Universal Words are words that constitute the vocabulary of UNL. A UW is not only a unit of the UNL syntactically and semantically for expressing a concept, but also a basic element for constructing a UNL expression of a sentence or a compound concept. Such a UW is represented as a node in a hypergraph. There are two classes of UWs from the viewpoint in the composition:

- labels defined to express unit concepts and called "UWs" (Universal Words)
- a compound structure of a set of binary relations grouped together and called "Compound UWs."

B. Relational Labels

The relation [1] between UWs is binary that have different labels according to the different roles they play. A relation label is represented as strings of three characters or less. There are many factors to be considered in choosing an inventory of relations. The following is an example of relation defined according to the above principles.

Relation: There are 46 types of relations in UNL. For example, *agt* (agent), *agt* defines a thing that initiates an action, agt(do, thing), agt(action, thing), obj(thing with attributes) etc.

C. Attributes

The attributes represent the grammatical properties of the words. Attributes of UWs are used to describe subjectivity of sentences. They show what is said from the speaker's point of view: how the speaker views what is said. This includes phenomena technically [4, 5] called speech, acts, propositional attitudes, truth values, etc. Conceptual relations and UWs are used to describe objectivity of sentences. Attributes of UWs enrich this description with more information about how the speaker views these state of affairs and his attitudes toward them.

III. MORPHOLOGY OF BANGLA WORDS

Morphology is the field of linguistics that studies the structure of words. It focuses on patterns of word formation within and across languages, and attempts to formulate rules that model the knowledge of the speakers of those languages. Thus morphological analysis is found to be centered on analysis and generation of word forms. It deals with the internal structure of words and how words can be formed. Morphology plays an important[2, 8] role in applications such as spell checking, electronic dictionary interfacing and information retrieving systems, where it is important that words that are only morphological variants of each other are identified and treated similarly. In natural language processing (NLP) and machine translation (MT) systems we need to identify words in texts in order to determine their syntactic and semantic properties [7]. Morphological study comes here to help with rules for analyzing the structure and formation of the words. A Bangla morpheme, besides the root word, is supposed to be represented in the Bangla-UNL dictionary using the following UNL format [10].

[HW] "UW" (ATTRIBUTE 1, ATTRIBUTE 2 …) <FLG, FRE, PRI>
HW← Head Word (Bangla Word)
UW← Universal Word
ATTRIBUTE← Attribute of the HW
FLG← Language Flag
FRE← Frequency of Head Word
PRI← Priority of Head Word

The attributes describe the nature of the head word classifying it as a grammatical, semantic or morphological feature. So, we will be especially concerned about representation of morphemes using various attributes.

A. Bangla Roots

Bangla Language contains a lot of verbs. The core part of those verbs is called roots. In another way if we split the verbs we get two parts *Roots* and *Suffixes*. From verbs if we remove suffixes we get roots. For example 'করে' (do) is a verb. Its two parts are: কর+এ; here 'কর' is a root and 'এ' is a suffix. Some other Bangla roots verbs are চল্ , পড়্, ধর্, গড়্, ঘষ্, নাচ্, কাঁদ্ etc.

B. Bangla Primary Suffixes (কৃৎপ্রত্যয়)

We know that the core of the verb is called root and if number of suffixes are added to roots then they form verbs. When sound or sounds [8] are added with roots and form nouns or adjectives then the root words are called root verbs and the sound or sounds are added with root verbs are called Primary Suffixes. For example চল্(Root verb)+অন্(Primary Suffix)=চলন (Noun) and চল্(Root verb)+অন্ত (Primary Suffix)= চলন্ত (Adjective). Some others primary suffixes are অন, অনা, অনি, অক, আ etc.

C. Vowel ended and Consonant ended root

Verb roots that are ended with vowel are called SORANTO root. For example, চা, হ, পা, নী,ধু, খা,ছা etc. And verb roots that are ended with consonant are called BANJANTO root. For example, কর্, চল্,পড়্,নাচ্,চিন্. etc.

D. Morphological Analysis of Bangla verbs

Morphological analysis is applied to identify the actual meaning of the word by identifying suffix or morpheme of that word. Every word is derived from

ISBN 978- 974-672-556-9

a root word. A root word may have the different transformations. This happens because of different morphemes which are added with it as suffixes. So, the meaning of the word varies for its different transformations. For example, if we consider 'কর্'(do) as a root word then after adding 'ই' we get the word 'করি'[6, 8] which means a work done by someone(first person , present tense). Similarly after adding 'আ', we get the word 'করা'. Here, this word represents noun of the root word 'কর্'. Therefore, by morphological analysis we get the grammatical attributes of the main word. Derivational morphology is simple and a word rarely uses the derivational rule in more than two or three steps. The first step forms nouns or adjectives from verb roots. The next steps form new nouns and adjectives [5]. We have examined derivational morphology for UNL Bangla dictionary too.

The UNL format of the word for the dictionary:

[কর্]{} "do(icl>do"(List of Semantic and Syntactic Attribute)<B,0,0>;

In the following we have given morphological analysis of a Bangla verb word. We can select the head words as the Longest Common Lexical Unit (LCLU) of all the possible transformations of the word [8]. We can give the example of the Bangla word "পড়"(means read). The corresponding UW in basic form is "read". The dictionary entry is: [পড়] { } "read (icl>do)", where 'পড়' is the *head word* and (icl>do) is from the knowledge base. Some possible transformations of 'পড়' in the Bangla to UNL dictionary are given as follows [9, 10]:

// For first person:
[পড়]{ }"read (icl>do)" (ROOT, BANJANT)<B, 0, 0>
[ই]{ }"read(icl>do)" (ROOT, BANJANT, PRESENT INDEF)<B, 0, 0>
[ইতেছি]{ }"read (icl>do)" (ROOT, BANJANT, PRESENT CONT)<B, 0, 0>

// For second person:
[পড়]{ "read (icl>do)" (ROOT, BANJANT)<B, 0, 0>
[ছেন]{ }"read (icl>do)" (ROOT, BANJANT, PRESENT CONT)<B, 0, 0>
[বেন]{ }"read (icl>do)" (ROOT, BANJANT, FUTURE INDEF)<B, 0, 0>

//For third person:
[পড়]{ }"read (icl>do)" (ROOT, BANJANT)<B, 0, 0>
[এছি]{ }"read (icl>do)" (ROOT, BANJANT, PRESENT PERF)<B, 0, 0>
[ছে]{ }"read (icl>do)" (ROOT, BANJANT, PRESENT CONT)<B, 0, 0>

Such dictionary order with root word followed by derivations will help in any quick search to find UW and the attributes of a Bangla word.

IV. Methods Of Finding Grammatical Attributes For The Headwords For Bangla Word Dictionary

Representing Universal Words (UWs) for each of the Bangla Head Word we need to develop grammatical attributes that describes how the words behave in a sentence. Grammatical Attributes (GA) have to be developed by the rules (Enconversion and Deconversion) and dictionary developers. They play very important rules for writing Enconversion and Deconversion rules because a rule uses GA in morphological and syntactic analysis, to connect or analyze one morpheme with another to build a meaningful (complete) word and to examine or define the position of a word in a sentence.

When we analyze the head words for representing them in the word dictionary as UWs, we find all the possible specifications of the HWs as attributes named grammatical attributes, so that they can be used in the dictionary for making rules. For example, if we consider "খা" meaning *eat* as a HW, then we can use attributes V(as it is verb), ROOT(as "খা" is a verb root), SORANT(as is vowel ended). So, this word can be represented in the dictionary as follows:

[খা] {} "eat(icl>consume>do)"(V, ROOT, SORANT)
Head Universal Word Grammatical Attributes
Word

Same way we can represent the words ধান (paddy), পাখি (bird) etc. as follows:

[ধান] {} "rice(icl>grain>thing)"(N, CONCRETE)

[পাখি]{} "bird(icl>animal>animate thing)"(N, ANI, SG, CONCRETE)

Where, N for noun, CONCRETE for any concrete thing which is touchable, ANI for animal and SG for singular number.

As we are the initiators of developing rules and word dictionary for Bangla we are proposing some grammatical attributes and their descriptions in table1.

TABLE 1. SOME PROPOSED GRAMMATICAL ATTRIBUTES

Grammatical Attributes	Descriptions	Examples (Here we can use Bangla Words)
ROOT	Only for verb root	পড় (Read), কর(do) etc.
SORANT	Verb roots that are ended with vowel	চা (want), যা(go) etc.
BANJANT	Verb roots that are ended with consonant	পড় (read), ধর (catch) etc.
1P	First person	আমি (I), আমরা (we) etc.
2P	Second Person	তুমি (you) etc.
3P	Third Person	সে পুরুষ (He), সে মহিলা (She) etc.
PROT	For all suffixes	আ, অন, আই etc.
KPROT	For the suffixes that are used after roots to create Nouns, Adjectives etc.	ইক, অন etc.
KBIVOKTI	For the suffixes that are used after roots to create only verbs.	ই, ইতেছি, বে etc.
ADJ	For any adjective	ভাল (good), সুন্দর (beautiful) etc.
N	For any noun	কলম (pen), আম(mango) etc.
MNOUN	For the suffixes that are added with roots to make nouns.	তা etc.
MADJ	For the suffixes that are added	অন্ত etc.

A. Primary suffixes with consonant ended roots (BANJANTO):

Some primary suffixes that are added with roots (which are ended with consonant) to form nouns and adjectives are given in table 2.

TABLE 2. CONSONANT ENDED ROOTS WITH PRIMARY SUFFIXES

BANJANTO root + Primary suffix	Noun and Adjective
কর্+অন্ত	করন্ত,
পড়্+অন্ত	পড়ন্ত,
চল্+অন্ত	চলন্ত
চল্+তি	চলতি
পড়্+তি	পড়তি
কাট্+তি	কাটতি
ঝুল্+তি	ঝুলতি

The dictionary entries for the roots and their suffixes are as follows:

[কর]{} “do (icl>do)”(ROOT, BANJANT)\
[পড়]{} “read (icl>do)”(ROOT, BANJANT)
[চল্]{} “go (icl>do)”(ROOT, BANJANT)
[কাট্]{} “cut (icl>do)”(ROOT, BANJANT)
[ঝুল্]{}“hang (icl>do)”(ROOT, BANJANT)
[অন্ত]{} (PROT, KPROT, BANJANT, MADJ)
[তি]{}(PROT,KPROT,BANJANT, MADJ, MNOUN)

B. Primary suffix with vowel ended roots (SORANTO):

Some SORANTO is added only with *primary suffix* ‘Av’ to form nouns are shown in table 3.

TABLE 3. VOWEL ENDED ROOTS WITH PRIMARY SUFFIX

SORANTO root + Primary suffix	Noun
চা+আ	চাওয়া
পা+আ	পাওয়া
দে+আ	দেওয়া
খা+আ	খাওয়া

So, the dictionary entries would be

[চা]{} “want(icl>do)” (ROOT, SORANT)<B,0,0>
[পা]{} “get(icl>obtain)” (ROOT, SORANT)<B,0,0>
[দে]{} “give(icl>do)” (ROOT, SORANT) <B,0,0>
[খা]{} “eat(icl>do)” (ROOT, SORANT) <B,0,0>

[আ]{}“” (PROT, KPROT, SORANT, MNOUN)<B,0,0>

C. There are some such primary suffixes which have two forms. One form is added with vowel ended roots and other forms are added with consonant ended roots. Those are given in table 4.

TABLE 4. PRIMARY SUFFIX WITH SORANTO ROOTS AND BANJANTO ROOTS

BANJANTO root + Primary suffix	Noun
চা+আ	চাওয়া
পা+আ	পাওয়া
দে+আ	দেওয়া
চল্+আ	চলা
পড়্+আ	পড়া
কাট্+আ	কাটা

The UNL form:

[চা]{} “want(icl>do)” (ROOT, SORANT)<B,0,0>
[পা]{} “get(icl>obtain)” (ROOT, SORANT)<B,0,0>
[দে]{} “give(icl>do)” (ROOT, SORANT) <B,0,0>
[আ]{}“”(PROT,KPROT,SORANT,MNOUN)<B,0,0>
[পড়]{} “read (icl>do)”(ROOT, BANJANT)<B,0,0>
[চল্]{} “go (icl>do)”(ROOT, BANJANT)<B,0,0>
[কাট্]{} “cut (icl>do)”(ROOT, BANJANT)<B,0,0>
[আ]{}“”(PROT,KPROT,BANJANT,MNOUN)<B,0,0>

All the Roots are added with “আ” Krit Prottoy and form a meaningful word. For example,

কর্+আ=করা
হার্+আ=হারা

But Some BANJANTA Roots (দুল্,খুল্) etc do not form a meaningful Bangla word to add with “আ” Prottoy. For example,

দুল্+আ=দুলা
খুল্+আ=খুলা.

ISBN 978- 974-672-556-9

Rather দোল in place of দুল and খোল in place of খুল is added with "আ" Krit Prottoy and form meaningful word. For example, দোল+আ=দোলা, খোল+আ=খোলা.

D. To solve this problem we divide BANJANTA Roots into two categories. One is General BANJANTA that is attributed as BANJNT and another is attributed with URoots.

TABLE 5. EXAMPLES OF GENERAL BANJANTO AND UROOTS

BANJANTO root + Primary suffix	Noun
কর্+আ	করা
পড়্+আ	পড়া
হার্+আ	হারা
দুল্+আ	দুলা
খুল্+আ	খুলা
দোল+আ	দোলা
খোল+আ	খোলা

The UNL form:

[কর্]{} "do (icl>do)"(ROOT, BANJANT)<B,0,0>
[হার্]{} "lost (icl>do)"(ROOT, BANJANT)<B,0,0>
[আ]{}(PROT, KPROT, SORANT, MNOUN)<B,0,0>
[দোল]{} "swing (icl>do)"(ROOT, BANJANT)<B,0,0>
[খোল]{} "open (icl>do)"(ROOT, BANJANT)<B,0,0>
[দুল্] {} "swing(icl>do)"(ROOT, BANJNT ,URoot) <B,0,0>
[খুল্] {} "open(icl>do)"(ROOT, BANJNT ,URoot) <B,0,0>

We make two entries in Bangla word dictionary for URoots. For example, দুল্ and দোল, খুল্ and খোল. In Bangla there are some primary suffixes which added with roots and form a new different words for example, বচ + ক্তি = মুক্তিand মুচ+ ক্তি= মুক্তি. These words are added to word dictionary in special category.

The dictionaries would be:

[উক্তি]{} "speech (icl>do)"(ROOT, BANJANT, SP)<B,0,0>
[মুক্তি] {} "free (icl>do)"(ROOT, BANJANT, SP)<B,0,0>

The suffixes আন্ত, তি,আetc. will be in the dictionary only with grammatical attributes. They will be added with the roots to form verbs, nouns or adjectives using rules. In the above examples we have classified suffixes in the basis of adding either with SORANTO (vowel ended) or BANJANTO (consonant ended) to give them proper attributes so that they can be used to make appropriate rules enconversion and deconversion.

V. CONCLUSION AND FUTURE WORK

In this paper we have generated grammatical attributes of Bangla roots and Krit Prottoy for developing Bangla Word Dictionary for Universal Networking Language (UNL). We have presented some method to select grammatical attribute using morphological analysis Bangla words that can be used to make dictionary for converting the Bangla sentences to UNL documents and vice versa. We have done limited work so far for Bangla words. Our future plan is to build a Bangla language server that will contain a complete Bangla Word Dictionary.

REFERENCES

[1] S. Abdel-Rahim, A.A. Libdeh, F. Sawalha, and M. K. Odeh, "Universal Networking Language(UNL) a Means to Bridge the Digital Divide,"Computer Technology Training and Industrial Studies Center, Royal Scientific Society, March 2002.

[2] M. M. Asaduzzaman and M. M. Ali, "Morphological Analysis of Bangla Words for Automatic Machine Translation", International Conference on Computer, and Information Technology (ICCIT), Dhaka, 2003, pp.271-276

[3] S. Gilles and B. Christian, "UNL-French econversion as Transfer & Generation from an Interlingua with Possible Quality Enhancement through Offline Human Interaction,"*Machine Translation Summit-VII*, Singapore, 1999.

[4] M. E. H. Choudhury, M. N.Y. Ali, M.Z.H. Sarkar, and R. Ahsan, "Bridging Bangla to Universal Networking Language- A Human Language Neutral Meta- Language", International Conference on Computer and Information Technology (ICCIT), Dhaka, 2005,pp.104- 109

[5] M.E.H. Choudhury and M.N.Y. Ali, "Framework for synthesis of Universal Networking Language," *East West University Journal*, vol. 1, no. 2, pp. 28-43, 2008

[6] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, and M. E.H. Choudhury,"Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language,"International Conference on Computer and Communication Engineering 2008(ICCCE'08), Kuala Lumpur, Malaysia.pp. 726-731

[7] S. Dashgupta, N. Khan, D.S.H. Pavel, A.I. Sarkar, and M. Khan, "Morphological Analysis of Inflecting Compound words in Bangla," International Conference on Computer, and Communication Engineering (ICCIT), Dhaka, 2005, pp. 110-117

[8] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, and A. M. Nurannabi, "Morpholoical Analysis of Bangla worfs for Universal Networking Language,"icdim,08.

[9] Bangla Academy.*Bengali-English Dictionary*, Dhaka, 2004

[10] H. Uchida and M. Zhu, "The Universal Networking Language (UNL) Specification Version 3.0", Technical Report, United NationsUniversity, Tokyo, 1998

[11] Enconverter Specifications, version 3.3, UNL Center/ UNDL Foundation, Tokyo, Japan 2002.

[12] Deconverter Specifications, vers ion 2.7, UNL Center/ UNDL Foundation, Tokyo, Japan, 2002

Distinctive Phonetic Feature Extraction for Japanese Language

Manoj Banik[1], Aloke Kumar Saha[2], Mohammed Rokibul Alam Kotwal[3], Mohammad Nurul Huda[3],Chowdhury Mofizur Rahman[3]
[1]Department of CSE, Ahsanullah University of Science and Technology, Dhaka, Bangladesh
[2]Department of CSE, The University of Asia Pacific, Dhaka, Bangladesh
[3]Department of CSE, United InternationalUniversity, Dhaka, Bangladesh
mbanik99@yahoo.com, aloke71@yahoo.com, rokib_kotwal@yahoo.com, mnh@cse.uiu.ac.bd, cmr@uiu.ac.bd

***Abstract*—This paper presents a method for extracting distinctive phonetic features (DPFs). The method comprises three stages: i) acoustic feature extraction, ii) multilayer neural network (MLN) and iii) HMM-based classifier. At first stage, acoustic features, local features (LFs), are extracted from input speech. On the other stage, MLN generates a 45-dimentional DPF vector from the LFs of 75-dimentions. Finally, these 45-dimentional DPF vector is inserted into hidden Markov model (HMM) based classifier to obtain phoneme strings. From the experiments on Acoustic Society of Japan (ASJ), it is observed that the proposed DPF extractor provides a higher phoneme correct rate with fewer mixture components in HMM compared to the other methods investigated.**

Index Terms-hidden Markov model; Japanese language; phonetic feature extraction

I. INTRODUCTION

There have been many researches [1-3] for feature extraction in speech recognition. In the research by Kirchhoff [1], acoustic features were mapped into the distinctive phonetic features (DPFs) using a set of lower level multilayer neural networks (MLNs) of five groups, in which each MLN was trained to extract a corresponding DPF in the group. The DPFs output from lower level MLNs were input to a higher level MLN which produced an acoustic likelihood of subword units. Their work improved the recognition accuracy of spontaneous speech as well as speech with additive noise. Again, Jain et al. [2] also applied a set of MLNs corresponding to each BPF channel to extract DPFs, and then used the DPFs in a higher level MLN as input that is similar to [1].Though these methods provide recognition accuracy up to a particular level, they have some demerits: i) they require higher mixture component to obtain higher recognition performance, ii) higher computational cost is needed and iii) they use mel-frequency cepstral coefficient features.

To eliminate these problems, in this paper we have developed a method based on articulatory features (AFs) or distinctive phonetic features (DPFs). The method comprises three stages: i) acoustic features, local features (LFs), extraction from an input speech, ii) MLN to obtain DPFs from LFs and iii) HMM-based classifier for achieving phoneme recognition performance. This method has some advantages: (i) it uses local features (LFs) instead of MFCCs as input to the MLN and (ii) it provides a higher phoneme with fewer mixture components in HMMs. For evaluation purposes, we have investigated the following methods: i) MFCC+HMM, ii) MFCC+LF+HMM and iii) LF+MLN+HMM (the proposed method).

The paper is organized as follows: Section II discusses the necessity of DPFs. Section III explains the system configuration of the existing methods with the proposed one. Experimental database and setup are provided in Section IV, while experimental results are analyzed in Section V. Finally, Section VI draws some conclusion and some remarks on future works.

II. DISTINCTIVE PHONETIC FEATURES

A phoneme can easily be identified by using its unique DPF set [4-5]. The Japanese balanced DPF set for classifying Advanced Telecommunications Research Institute International (ATR) phonemes have 15 elements. These DPF values are vocalic, high, low, intermediate between high and low <nil>, anterior, back, intermediate between anterior and back <nil>, coronal, plosive, affricate, continuant, voiced, unvoiced, nasal and semi-vowel. Here, present and absent elements of the AFs are indicated by "+" and "-" signs, respectively. Table I shows the Japanese balanced DPF-set for classifying ATR phonemes.

ISBN 978- 974-672-556-9

TABLE I
JAPANESE BALANCED DPF-SET FOR CLASSIFYING ATR PHONEMES

DPFs	a	i	u	e	o	N	w	y	j	my	ky	dy	by	gy	ny	hy	ry	py	p	t	k	ts	ch	b	d	g	z	m	n	s	sh	h	f	r
vocalic	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
high	−	+	+	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	−	−	+	−	+	−	−	+	−	−	−	−	+	−	+	−
low	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−
nil	−	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+	−	+	+	−	+	+	+	+	−	−	−	+
anterior	−	−	−	−	−	−	−	−	+	+	−	+	+	−	+	−	+	+	+	+	−	+	+	+	+	−	+	+	+	+	+	−	+	+
back	+	−	+	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	+	−	−	−	−	−	−	−	−
nil	−	+	−	+	−	+	−	+	−	−	+	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−
coronal	−	−	−	−	−	−	−	−	+	−	−	+	−	−	+	−	+	−	−	+	−	+	+	−	+	−	+	−	+	+	+	−	−	+
plosive	−	−	−	−	−	−	−	−	−	−	+	+	+	+	−	−	−	+	+	+	+	−	−	+	+	+	−	−	−	−	−	−	−	−
affricative	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	+	−	−	−	−	−	−	−
continuant	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	+	+	+	−
voiced	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	−	−	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+
unvoiced	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	+	−	+	+	+	+	+	+	−	−	−	−	−	−	+	+	+	+	−
nasal	−	−	−	−	−	+	−	−	−	+	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−
semi-vowel	−	−	−	−	−	−	+	+	−	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+

III. PHONEME RECOGNITION METHODS

A. *The Existing Methods*

1) MFCC- based Method: Conventional approach of ASR systems uses MFCC of 38 dimensions (12-MFCC, 12-ΔMFCC, 12-ΔΔMFCC, P and ΔP, where P stands for raw energy of the input speech signal) as feature vector to be fed into a HMM-based classifier and the system diagram is shown in Fig. 1. Parameters (mean and diagonal covariance of hidden Markov model of each phoneme) are estimated, from MFCC training data, using Baum-Welch algorithm. For different mixture components, training data are clustered using the K-mean algorithm. During recognition phase, a most likely phoneme sequence for an input utterance is obtained using the Forward.

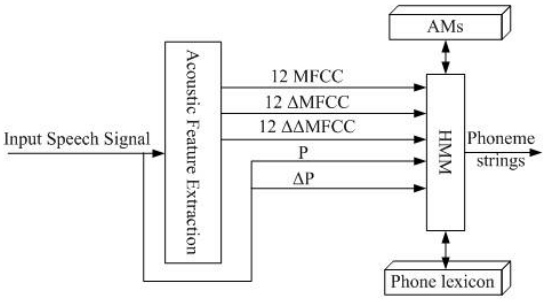


Figure 1.MFCC-based system.

2) MFCC- LF based Method: At an acoustic feature extraction stage, firstly, input speech is converted into 25 dimensional local features (LFs) that represent variations in spectrum along time and frequency axes [6]. These 25 dimensional LFs are combined with MFCC of 38 dimensions (12-MFCC, 12-ΔMFCC, 12-ΔΔMFCC, P and ΔP, where P stands for raw energy of the input speech signal) to obtain a 63-dimensional feature vector. This feature vector is inserted into an HMM-based classifier to obtain phoneme recognition performance.

B. *Proposed Method*

Twenty five dimensional LFs vector extracted through the procedure mentioned in section III.A(2) are then entered into an MLN with four layers, including two hidden layers, after combining a current frame x_t with the other two frames that are three-points before and after the current frame(x_{t-3}, x_{t+3}). The MLN has 45 output units (15x3) corresponding to a set of context-dependent DPF vector, which consists of three DPF vectors (a preceding context DPF, a current DPF, and a following context DPF) with 15 dimensions each. The two hidden layers consist of 256 and 96 units from the input layer. Fifteen DPF elements of Japanese balanced DPF set are used. The MLN is trained by using the standard back-propagation algorithm to output a value one for the corresponding DPF elements with an input phoneme and its adjacent phonemes. This method has several advantages: it provides (a) robust features in different acoustic environments and (b) higher WCR with fewer mixture components in the HMMs. On the other hand, this method still occupies some demerits: it shows some misclassification caused by co-articulation at phoneme boundaries; it cannot solve co-articulation problems because a single MLN has an inability to represent context information.

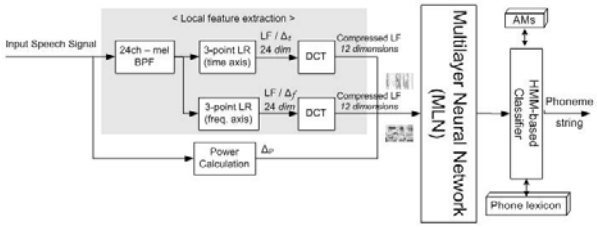


Figure 2.Proposed System.

IV. EXPERIMENTS

A. *Speech Database*

Following two clean data sets are used in our experiments.

D1) Training data set: A subset of the Acoustic Society of Japan (ASJ) Continuous Speech Database [7] comprising 4513 sentences uttered by 30 different male speakers (16 kHz, 16 bit) is used.

D2) Test data set: This test data set comprises 2379 Japanese (JNAS) [8] sentences uttered by 16 different male speakers (16 kHz, 16 bit).

B. *Experimental Setup*

The frame length and frame rate are set to 25 ms and 10 ms, respectively, to obtain acoustic features, MFCCs and LFs, from an input speech.LFs are a 25-dimensional vector consisting of 12 delta coefficients along time axis, 12 delta coefficients along frequency axis, and delta coefficient of log power of a raw speech signal. On the other hand, MFCCs comprises 38 dimensions (12MFCC, 12ΔMFCC, 12ΔΔMFCC, P and ΔP, where P is the power of raw speech signal).

To measure phoneme error rate (PER), D2 data set are evaluated using an HMM-based classifier. The D1 data set is used to design 37 Japanese monophone HMMs with five states, three loops, and left-to-right models. Input features for the classifier are MFCCs of 38 dimensions, MFCC-LF of 63 dimensions and DPFs of 45 dimensions. In the HMMs, the output probabilities are represented in the form of Gaussian mixtures, and diagonal matrices are used. The mixture components are set to 1, 2, 4 and 8. In our experiments of the MLNs, the non-linear function is a sigmoid from 0 to 1(1/ (1+exp (-x))) for the hidden and output layers.

To obtain PER, we have investigated the following DPF based methods.

i) MFCC+HMM
ii) MFCC+LF+HMM
iii) LF+MLN+HMM

V. EXPERIMENTAL RESULTS AND DISCUSSION

This DPF-based method (i) gives robust features to different acoustic environments with fewer mixture components in HMMs, and (ii) it improves the margin between acoustic likelihoods. Figs. 3(a) and 3(b) show the phoneme distances of five Japanese vowels in an utterance /ioi/ that are calculated with a mel-frequency cepstral coefficient (MFCC)-based ASR system and a DPF-based system, respectively. In both the systems, each distance is measured using the Mahalanobis distance between a given input vector and the corresponding vowel set of mean and covariance in a single-state model. The input sequence in the figures,

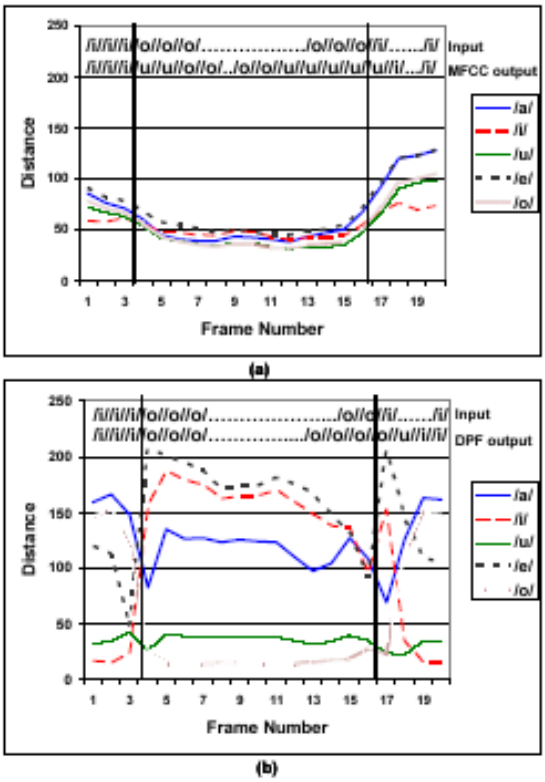


Figure 3.Phoneme classification for a) MFCC- based system and b) DPF-based method system.

/i/../i//o/../o//i/../i/, exhibits phoneme for each frame and has total 20 frames in which first three frames, middle 13 frames, and last four frames are phonemes /i/, /o/, and /i/, respectively. The MFCC-based system (Fig.3(a)) shows seven misclassification of phonemes (/u/ output for /i/ input) for frames 4, 5, 13, 14, 15, 16, and 17, while two misclassification (/o/ and /u/ output for /i/ input) for frames 17 and 18 are exhibited by the DPF-based system (Fig. 3(b)). Therefore, the DPF-based system outputs few misclassifications.

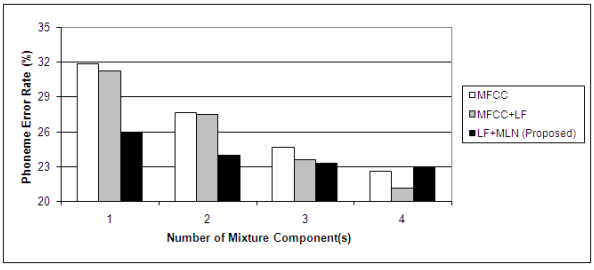


Figure 4.Phoneme error rate for investigated methods.

Fig. 4 shows PER for different investigated methods. From the figure, it is observed that the proposed system provides lower PER for all the mixture components investigated except 8. For an example, at mixture component 2, 23.94% PER is obtained by the proposed method while corresponding values for the methods based on MFCCs and MFCC+LF are 27.6% and 27.47% respectively.

ISBN 978- 974-672-556-9

TABLE II. COMPARISON OF TIME COMPLEXITY BETWEEN MFCC+LF+HMM AND PROPOSED METHODS

	MFCC+LF+HMM	Proposed
PCR=76.43	4 Mix	-
PCR=76.06	-	2 Mix
Required Multiplication	20 k	10 k

Table II is given to indicate the computation time more specifically with the methods based on MFCC+LF+HMM and the proposed one. We have measured the HMM time required by MFCC+LF+HMM and the proposed method using the formula mS^2T where m, S and T indicates number of mixture components, states and observation sequences respectively. For MFCC+LF+HMM, the required time is $4x5^2x200$ (=20K), while the corresponding time for the proposed method is $2x5^2x200$ (=10K) assuming number of observation sequence is 200 frames. Therefore, our proposed method is faster than the method based on MFCC+LF+HMM.

VI. CONCLUSION

This paper has presented a method for extracting DPFs and then evaluated phoneme recognition performance using the extracted features. Findings of the method are given below:

(a) At lower mixture components the proposed method provides better result than the other methods investigated.

(b) The proposed method requires less computation time for fewer mixture components.

In near future, the authors would like to evaluate Bengali articulatory feature extraction using the method proposed in this paper.

REFERENCES

[1] K. Kirchhoff, et al."Combining acoustic and articulatory feature information for robust ASR,"*Speech Comm*., 2002.

[2] P. Jain, et al., "Distributed speech recognition using noise-robust MFCC and TRAPS-estimated manner features," Proc. ICSLP'02, pp.473-476, 2002.

[3] Automatic speech recognition with neural networks: Beyond nonparametric models, *SpringerLink*, pp. 104-121, vol. 745, 1993.

[4] S. King and P. Taylor, "Detection of Phonological Features in Continuous Speech using Neural Networks," *Computer Speech and Language* , vol.14, no.4, pp. 333-345, 2000.

[5] E. Eide,"Distinctive Features for Use in an Automatic Speech Recognition System," Proc. Eurospeech 2001, vol.III, pp.1613-1616, 2001.

[6] T. Nitta, "Feature extraction for speech recognition based on orthogonal acoustic-feature planes and LDA," Proc. ICASSP'99, pp.421-424, 1999.

[7] T. Kobayashi, et. al, "ASJ Continuous Speech Corpus for Research," *Acoustic Society of Japan Trans.* vol.48, no.12, pp.888-893, 1992.

[8] JNAS: Japanese Newspaper Article Sentences. Available: http://www.milab.is.tsukuba.ac.jp/jnas/instruct.html

Session 3: RFID Technology and Applications

ISBN 978- 974-672-556-9

Zoo Application of RFIDTechnology: A Case Study of Chiang Mai Zoo, Thailand

NopponChoosri[1,2], Saisakul Chernbumroong[1,2], Suttipong Intaya[1], Apitchaka Singjai[1], Anthony Atkins [2], and Hongnian Yu [2]
[1]College of Arts Media and Technology, Chiang Mai University, Chiang Mai, Thailand
[2]Faculty of Computing, Engineering and Technology, Staffordshire University, Stafford, UK

Abstract—The research aims to build a technology system that improves service and operation for a zoo using RFID application. A problem investigation has been conducted with a case study at Chiang Mai Zoo, Thailand to propose and develop a prototype solution to overcome some practical issues. The proposed framework utilises RFID technology to store data and deal with other information technologies, there are two major subsystems: e-Ticket system and mobile tracking system. An implementation of the proposed system should provide significant benefits to the zoo as well as other similar businesses.

Index Terms- Chiang Mai zoo; RFID applications; RFID Technology; Thailand zoo

I. INTRODUCTION

Chiang Mai is one of Thailand most popular tourist destinations, located in the North of Thailand. For years ithas been the tourist hub of the north having more than 5 million people visiting per year, bringing in around THB 38 million in revenue [1]. Chiang Mai is surrounded by a variety of famous tourist attractions including Chiang Mai Zoo, which is the first and only zoo in the North of Thailand, established in 1974 C.E. by My Young, an Americanmissionary, and latercontrolled by the Zoological Park Organization in 1977 [2]. Chiang Mai Zoo is a popular eco-tourism and wild life learning attraction for all kinds of tourists, ranking top three most visited tourist attractions in Chiang Mai Thailand [1].

In order to maximise the zoo's assets, increase the number of tourists and customer satisfaction, an investigated into emerging Radio Frequency Identification (RFID) technology, which is an identification technology using radio waves to identify objects, has been undertaken. RFID has normally been used in manufacturing, logistic andsupply chainmanagement domain. Recognition and acknowledgement of the significant potential of using RFID in other areas inspired this investigation into employing this technology in thetourism sector in particular at Chiang Mai Zoo. RFID can be used to improve the zoo's management, quality of services and customer satisfaction. Investigation has been carried out on the zoo's current systems and services. Observations were made to assess whether RFID technology could enhance the existing framework and the problem justification is outlined in Section II. The paper is structured as follows: Section III provides a review of RFID technology and Section IV explains a proposed framework. In Section V there is a discussion of the system architecture. Section VI projects the benefits of implementation and finally Section VII provides the conclusions.

II. PROCEDURE FOR PAPER SUBMISSION

Problem investigations were studied by reviews andobservation at the zoo. From the current system, thefollowing challenges were identified that could be improved by using RFID technology.

A. *Handling a large management system*

Chiang Mai zoo is one of the largest zoos in Thailand with an area of more than 200 acres. Because of its large size, it is difficult for staffs to deliver the highest quality services, for example some of the tourists might encounter problems involving searching and locating different locations e.g. Panda's enclosures, toilets, snack shops, in the zoo etc. The zoo accommodates more than 8,000 animals in which different species require different type of foods and habitats. Veterinarians and zoo staffs have challenges in remembering and identification of factual information for individual animals e.g. on specific illness and treatment, dietary, etc., as currently there are no online devices available to support staff. Recently due to the arrival of a new baby giant panda as there is one ofa few pandas born outside China, the zoo has received more attention from both locals and overseas tourists. The increase in the number of visitors have given further recognition to the case for a new management system being introduced in order to maintain visitors' satisfactions and best care for the animals.

B. Improve Manual operation

Most of the zoo operations are done manually including key operations such as selling and checking tickets. For example, at the entrance, a staff member will check the visitor's ticket. This is prone to human error as the staff visually inspects the ticket. Also, from the observation, only one staff member is allocated to this job which causes a bottle neck during rush hours. As the labour intensive process slow the flow of the overall process and is unproductive and inefficient. The

ISBN 978- 974-672-556-9

safety and security of such a process is low as there is no system to keep process and it is possible that unauthorized people could enter the zoo without valid tickets.

C. Improve guiding aids

The zoo is not just a place to visit for recreation, and many tourists may come with an expectation of also gaining some knowledge related to zoology and ecology. Observations indicated that the current services could be improved, for instance many current animal explanatory boards are not effective enough i.e. some are made fromflimsy materials (Figure 1) and some do not provide information in other languages for foreign tourists.

Figure 1.Ineffective animal information label.

In this research, RFID could be used to improve efficiency at the zoo or at least to overcome the issues mentioned previously. RFID e-ticket would be proposed to tackle the inefficiency problem of paper tickets and associated manual checking operations would offer a safe and secure system as money-related tasks are completed electronically over an automated machine system. The e-ticket would allow visitor information such as adults and/or children, nationality, etc. to be created automatically and used to create tailored services for improved visitors' experience. For example, the animal information could be displayed in different languages according to visitors' nationalities to support moretourist groups. Tourists can also use the e-Tickets for locating themselves within the zoo i.e. e-Ticket will display a map and surrounding locations when it is 'present'at the informationpoint (RFID reader and linked network) at the animal enclosures. Also converting a manual system to a digital system as proposed would allow analysis to provide executive information, statistics and reports to answer 'what if' scenarios such as which service generates the most profit, which customer type visits the most etc. This information could be used to improve future systems and services.

III. RFID TECHNOLOGY

A. Explanation of RFID Technology

Radio Frequency Identification (RFID) is a contact less data identification system using radio frequencies as a carrier andcanreplace barcode technology. The technology allows asmall radio device attached to an item to carry an identification of that item [3]. At least 3 components are required to integrate the system; that is, the reader, tag and middleware respectively [4]. The RFID concept is shown in Figure 2 and a brief explanation of each unit is given asfollows:-

Tag is an electronic chip containing storage memory and it communicates with a reader via an antenna. Tags are roughly divided into 3 types according to the source of power: activetag, passive tag, and semi-active tag. An active tag is a tag that supplies power to itself by using an onboard battery, a passive tag is supplied with power by a reader, while a semi-active is a hybrid of active and passive. A tag also has on-board power supply for its operation, but for transmitting its data, a semi-active tag uses emitted power from reader like passive tag [5]. The tag can be classified into several types based on its radiofrequency usage, for example LowFrequency (LF), HighFrequency (HF), and Ultra High Frequency (UHF).

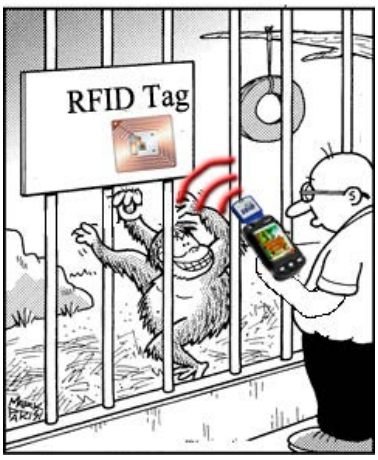


Figure 2. Component of RFID system

Reader or receiver is an RFID device that sends instructions to the tag, and it normally operates with a computer device. Although its primary operation, as its name suggests, is reading the tag, a write operation is possible on some types of tags and there are several types of readers. For example, a stationary reader can be used for desktop computers, while a handheld reader is used for mobile computers such as PDAs.

Middleware or reader interface layer is a unit that controls the reading process and handles the acquired databeforesending it to the backend database system such as SQL, MySQL, Oracle, Postgres, or a similar database for a specific vendor system.

ISBN 978- 974-672-556-9

B. RFID in non-logistic applications

RFID is an alternative technology which is being used in several applications to replace traditional barcodes systems. Primarily RFID is used in logistics operation and this paper outlines a novel application of this technology to a zoo operation. In medical application, Nicholls and Young [6] proposed an *ebed* system to improve in-patient management which can identify bed vacancies in real time; it can also contribute a key identification role to support 'life-or-death' operations in health care. Jiang et al [7] proposed an RFID solution for blood transfusion services in hospitals to prevent misusing of donated blood by medical staff. In livestock, RFID is applied to track the movement of animals without interrupting normal activities [8]. In military's operations, RFID tag can be attached to all weapons and equipments, so that at any time during a given mission, the exact equipment that each individual soldier holds can be known, so a unit leader of the troop can properly assign tasks to specific soldiers. In education, RFID technology can play a role as one of the key components in building the ubiquitous learning environments for school students by verifying current location and environment of a learner.

C. Emerging RFID Technology in Thailand

A 2005 survey found that the RFID market in Thailand was valued THB 857 m (Tag: 38%, Reader: 32%, System: 23%, and Software: 7%) [9]. Anecdotal evidence shows RFID technology has been adopted in several areas, for instance, in the airline industry where the international airport in Bangkok uses RFID tag technology to keep a record of all air freight (including internal luggage logistics) passing through its cargo terminals [10]. With this growing trend, its value is expected to reach $32 m in 2010 [11].

IV. PROPOSED FRAMEWORK

RFID technology applications other than logistics can drive improvements in efficiency for example at a zoo operation. The proposed system is composed of two platforms of application as depicted in Figure 3. An E-ticket is the core system and mobile tracking is a complementary system having potential to improve animal care and improve tourist satisfaction. Each system is described as follows:-

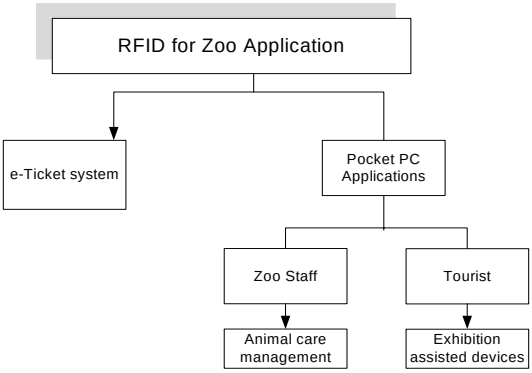


Figure 3. Conceptual framework of the two platform system

An E-ticket system is proposed to substitute the existing conventional paper based ticket system. An e- ticket is technically a RFID tag which acts as a pre-paid cash card. A credit amount is set to the card by a Write operation once a tourist buys a ticket. The card can also be used to attend special events or buy additional services in a zoo. The cardholder is required to operate a Read operation with a machine (RFID reader) at service points. The new system will not only improve the efficiency of ticket validation, but it will also improve tourist satisfaction when travelling though the zoo. With only one ticket, they can pursue any services provided in the zoo, accurately locate their current position at an animal enclosure (this helps to prevent tourists from getting lost) and use the new interactive system to provide a more enjoyable and informative experience. Figure 4 shows a scenario of the proposed system.

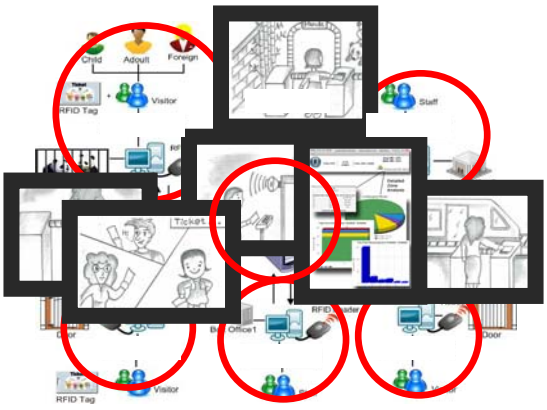


Figure 4.e-Ticket system

The flow of the proposed new system is shown in Figure 5. When tourists arrive at the zoo, they have to buy their electronic ticket and charge a credit for entrance. At the gate, they have to present their ticket to the system for verification. When they are in the zoo, tourists can use the card to access any service provided by the zoo. For particular services a charge may be required, and the system will deduct credit from the ticket. If the credit on the card

ISBN 978- 974-672-556-9

is too low the tourist can recharge the credit at a kiosk in the zoo. Finally, on leaving the tourist has to return their card to the ticket office, and then the card can be reused by a new visitor.

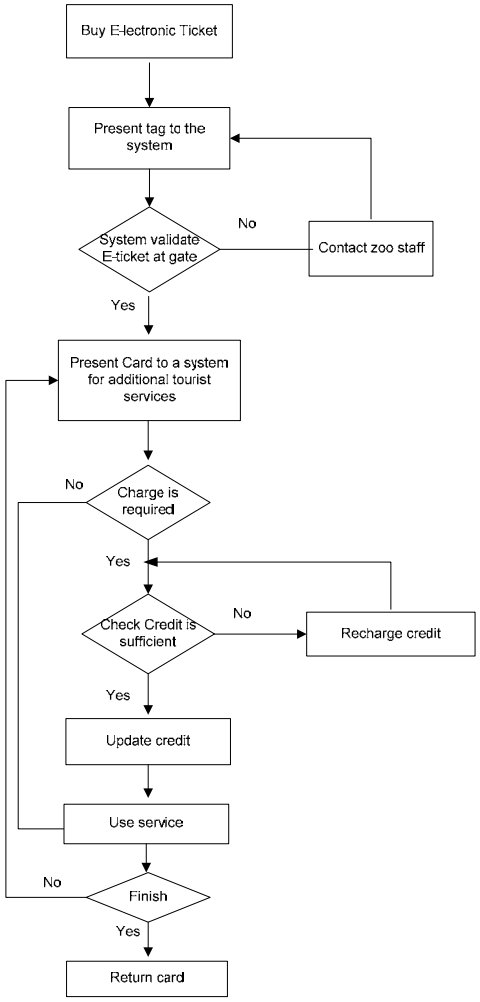


Figure 5. Flowchart of the proposed system

As well as re-engineering the core system using electronic ticket as outlined,mobile RFID system can also be developed for both zoo staff and tourist in the exhibition area. In the proposed system, animal carers will have a device which is Pocket PC associated with RFID mobile reader to help them give improved care to the animals. As each animal behaves differently and has different food requirements, when the device is presented to tag at an animal enclosure it will help by delivering feeding and other relevant information. The RFID mobile device can be used another way by tourist in an exhibition hall, where it can be used for listening to information.

V. System Architecture

The system operates by an integration of 4 technologies as shown in a Figure 6. There are: RFID, Software application, Database, and Local Area Network (LAN). Users interact with an RFID system using software applications which involve retrieval and storage of data with a database with communication among each component supported by the LAN. Functions of each component are described in Figure 6 and as follows:

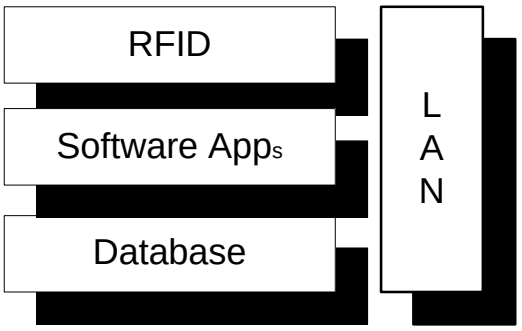


Figure 6. System architecture

1) *RFID*

RFID technology is a key driver of this system. Low Frequency (LF) systems are specifically selected due to the nature of the applications requires a close range identification, i.e. if a long range identification system such as Ultra-High Frequency (UHF) is used, auditing cannot be individually made for ticket inspection at the gate. Also, LF is the most robust for working in a 'rough' environment that includes human bodies and animal-enclosures, in which humidity are typically founded [4]. The proposed interrogation process requires Reader and Tag. The system requires user participation and there are 2 scenarios. First, the e-ticket system requires uses to carry tags with them to use with a reader installed in the zoo, and secondly, a Pocket PC application that requires users to carry a reader to use with an attached tag.

2) *Database*

Since the RFID tags have limited memory and in order to promote a flexibility of data processing, most of data in the system are stored in the relational database to support decision-making. It is proposed that data i.e. animal information; tourist informationand transactional data are stored centrally which also allows holistic decisions to be made.

3) *Local Area Network*

The LAN system will allow client-server information service and the network infrastructure will allow all zoo users, i.e. tourists, zoo staff, decision makers, to access the database in a real time. For example, when a tourist retrieves information at an enclosure, notification will be sent instantaneously to update visiting statistics in a database, thus at any point in time, staffs or decision- makers will be able to retrieve current information.

4) *Software Application*

As previously mentioned in Section III.A), middleware is required for controlling RFID operations, as well as communicating with the user.

ISBN 978- 974-672-556-9

Software applications are developed for working associatively with RFID reader for both Pocket PC and PC platforms. Features of the system will be discussed in Section VII.

VI. PROTOTYPE IMPLEMENTATION

Software was developed to integrate with RFID systems to verify the proposed concepts. As mentioned the system is composed of two major subsystems running on different platforms. These are the core systems i.e. the electronic ticket (e-ticket) system (Figure 7) and mobile tracking system.

Conventional paper tickets are changed to RFID tickets with automatic checking systems (made from Low Frequency Tag). Tourist status together with amount of credit is programmed at a ticket office; tourists can specify their initial credit amount, and they can top-up with more credit later using kiosks in the zoo. With an e-ticket having sufficient credit, tourists can use various services in the zoo e.g. a tram service; in the zoo by simply applying this card to an automatic vending machine (RFID reader) at a service point. Also, tourists will benefit from enjoying a new experience of zoo visit when they can have a more interactive system that can be used easily e.g. pre- programming tourist status will enable the system to know the users, hence appropriatelanguage and style of conversation can be delivered. For example if the RFID indicates astatus of overseas visitor the system will communicate with that tourist in their native language e.g. English. E-ticket systems are operated by the following application:-

Figure 7. E ticket and its reader

1) Ticket vending

Electronic tickets are sold by zoo staff at a ticket office. Tourists can initially buy any amount of credit that reaches a minimum entrance price. Zoo staff operate this system by recording a requested credit from tourist into an e-ticket (RFID tag), also Tourist ID is recorded to indicate tourist status (children and/or adult, overseas visitor or Thai) as shown in Figure 8. Data is also replicated to the zoo's database system for data analysis purposes.

2) Ticket Verification

Before the tourist enters the zoo, the system will check whether the e-ticket is valid or not. Tourists must have their tickets approved by the system. If the ticket is valid, the system will display a welcome message and sound and open the gate. If not, notification will be prompted to ask a tourist to contact a member of the zoo staff.

Figure 8. User interface of ticket vending system

3) Instruction guide

Zoo attendance is available to anyone, and the new framework encourages visitors to interact with the system. In struction guides are provided to help tourists using the system in their own language.

4) Information service

During a visit, tourists can use their tickets for viewing information and using services that are operated by the RFID system. At an animal's enclosure a tourist may present their card to an RFID reader to listen to information about that animal. Stored Tourist ID will direct what language the information should be delivered. Tourists may use their ticketto buy other services ina zoo or to buy a ticket for special events available in the zoo e.g. animal show and internal tram service. The E-Ticket can be used to locate the visitor's position and surrounding when it is used at a service points (as shown in Figure 9). Any service that tourist participate in will be recorded in the zoo's database.

Figure 9. Map information service

5) Zoo data analysis

Transactional data gathered by the system will be sent to the central database where the zoo's data analysis application can retrieve information and allow authorised zoo staff to view the data (as shown

ISBN 978- 974-672-556-9

in Figure 10). Visitor statistics from the system enable planners to making strategic decisions; forinstance 'what-if' query i.e. which month of the year has the most visitors, which animal is the most popular, or the home country of the greatest number of tourists and their nationality etc. This information would be beneficial in enabling complex data analysis technique, like data mining to bedeveloped for future work.

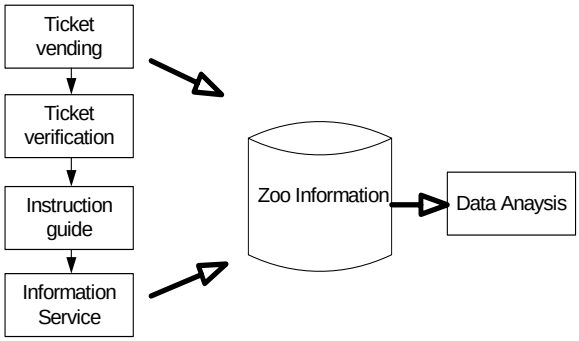


Figure 10. e-Ticket system component

The second application in this framework is a mobile tracking. Unlike an e-ticket system that operates by a tourist carrying an RFID tag which interact with installed readers in the zoo. Tourists who use this application have to carry a Pocket PC with on-board RFID Reader to apply with the installed tag in the zoo. The advantage of this system is that a tourist will have more flexibility in accessing information i.e. equipment is independent from other tourists and since a device is handheld, the system can provide more interactivitywith the user. However, this system may not be practical if they are given to every tourist at the zoogate. This system is proposed to complement the core system; it could be issued to a tourist in a small area, such as the exhibition hall as the return of devicescan be more strictly controlled. The intention of developing this part of the prototype i.e. the mobile tracking system, is to help veterinarian or zoo staff to record and retrieve treatment information for quality animal welfare. An example interface of the mobile tracking system is shown in Figure 11.

Figure 11. Mobile tracking systems

VII. BENEFIT

Implementing the proposed system should contribute to a number of efficiencies and benefits to the zoo as follows:-

1) Improvement to fee collection

Inefficiency from manual ticket inspection would be removed. Using human beings to inspect an entrance ticket at the doorway is a tedious job for zoo staff, also during a peak times, it might be difficult to detect a person who tries not to pay because of crowded conditions. With the e-ticket system, tourists are required to approve their ticket to the system. This would make an inspection more robust, and would make collection of fees from tourists more effective.

2) Tourist satisfaction

The new system will help improve tourist satisfaction by introducing a new service which will assist tourists in learning about the animals' environment. Tourists can also use information services by interacting with the system in their own language, and to confirm their location within the zoo. These new features would help the zoo to extend their niche markets in attracting visitors who do not speak Thai.

3) Asset protection

Animals are the most valuable asset of a zoo. The new system would help staff take care of the animals more effectively. With hundreds of animal species that behave and live totally differently, people who care for them must try andremember every individual treatment e.g. food and vaccine requirements. This proposed technology would provide comprehensive information that would avoid risks of mistakes and errors (particularly when new staffs are recruited). The new system will help reduce that risk by providing behavioral information and feeding instructions that can be retrieved by the handheld reader.

4) Reduce paper used

The current system uses a printed paper ticket which does not create any value and is wasted when tourists enter the zoo. The e-ticket system will allow tickets to be reused by other tourists and/or used as a loyalty card for frequent visitors to the zoo.

VIII. CONCLUSIONS

This research discusses the deployment of RFID technology for a zoo operation. A prototype systemis developed for acase study at Chiang Mai Zoo, Thailand. The core system ofthe proposed framework is an e-ticket system where a Low Frequency Tag as an e-ticket to replace a conventional paper ticket. This new system could improve the efficiency of zoo management, especially for ticket checking, as well as introducing tourists to a new experience. Another application is a

ISBN 978- 974-672-556-9

mobile tracking system operated by RFID mobile reader and a Pocket PC. The portability features of this system will offer more special flexibility of object tracking; however it may not be applicable within a real-life scenario if the devices were to be given to all tourists. It is proposed to use this system for animal caretaking and it might also be suitable for loan to tourists in an exhibition area. This proposed framework should provide both efficiency gains and improved tourist satisfaction at the zoo and it could also apply to similar systems such as fun parks, museums or hospitality services.

REFERENCES

[1] Tourism Authority of Thailand (2007) .Domestic Tourism Statistics[Online] Available: http://www2.tat.or.th/stat/web/static_tst.php [Accessed April 20, 2010]

[2] Chiang Mai Zoo (2010). Chiang Mai Zoo [Online] Available: http://chiangmaizoo.peam.biz/general_information.html/ [Accessed April 20, 2010]

[3] Glover, B & Bhatt, H, 2006, RFID Essentials. California, U.S.A.,O'Reilly

[4] Choosri, N, Yu H, Atkins A S and Hampornchai N; (2008), 'RFIDTechnology in Supply Chain Logistics Application,' International Conference on Software, Knowledge, Information Management and Applications (SKIMA), March 2008, Kathmandu, Nepal Co-sponsored by IEEE pp 227-231 ISBN 9781851432516

[5] Lahiri, S. (2005) RFID Sourcebook, New York, IBM Press

[6] Nicholls, Angela G. and Young, Frank R. (2007) Innovative hospital bedmanagement using spatial technology. Spatial Science Queensland, 2007(2), 26-30.

[7] Jiang, M., Fu, P., Chen, H., Chen, M., Xing, B., Sun, Z.,Deng,P.,Wang, G., Xu,Y.,& Wang, Y. (2005) A Dynamic Blood Information Management System Based on RFID In: The 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference,Shanghai, China,1-4 September 2005, pp.546-549.

[8] Tedjasaputra, A. (2006) Fighting bird flu with RFID technology [Online] Available: http://www.rfidasia.org [Accessed January 4, 2009].

[9] National Electronics and Computer Technology Center (2005)Thailand's RFID Cluster Report 2005, Ministry of Science and Technology,Thailand.

[10] Bacheldor, B. 2007), Thai Airport Tests RFID to Track Cargo, Streamline Customs [Online] Available: http://www.rfidjournal.com/article/articleview/3464/ Accessed April 20, 2010]

[11] Thailand RFID Market. [Online] Available: http://www.researchandmarkets.com/reports/358501/thailand_rfid_markets.html/ [Access September 10, 2009]

Integrated Information Systems Using Emerging RFID Technology in the Hospitality Industry

Benjamin Ohene Adu[1], Anthony S. Atkins[1], Hongnian Yu[1], and Dale L Nix[2]
[1]Faculty of Computing, Engineering, and Technology, Staffordshire University, Beaconside, Stafford, UK
b.o.adu@staffs.ac.uk, a.s.atkins@staffs.ac.uk, h.yu@staffs.ac.uk
[2]Chapman-Nix Ltd, The Orangery at Hawkesyard Hall, Armitage Lane, nr Rugeley, Staffordshire, WS15 1PU UK
info@chapman-nix.com

***Abstract* - The emergence of Radio Frequency Identification (RFID) technology and mobile RFID equipment systems offers the opportunity to provide real time information which can be used in inventory and tracking systems. The paper presents some case study applications for using this technology in the hospitality industry utilizing RFID and mobile technology as a solution in asset management and wine cellar management. The system uses low cost passive tags (costing a few cents) to provide information in real time using a TCP/IP protocol which is internet compatible and can be viewed anywhere in the organization worldwide to provide more effective management control. The use of RFID technology provides both operational visibility and authenticity which is vital in the wine industry.**

Index Terms- hospitality industry; mobile technology; RFID technology; RFID utilization; wine cellar management

I. INTRODUCTION

The hospitality industry globally faces a number of challenges such as the search for cost reduction solutions whilst maintaining quality and profitability (Luo, et al., 2009). Customer acquisition, growth, and retention of loyal clients and provisioning of customized services are but a few of the challenges of the industry. This level of service provision is dependent on captured data and its transformation into business intelligence to bring about the needed level of service quality. Though a high proportion of service providers in the hospitality industry use some form of electronic records, the data capture effort relies on manual inputs into the computer system. Though this method is prone to error, inventory inaccuracies, and may be time consuming, these issues can be resolved by adopting non-interference automatic identification and data capturing using emerging RFID technology (Michahelles, et al. 2009). In such occurrences, having real time operational visibility would be valuable. Contributions on the impact and potentials of RFID have been shared by Kambil and Brooks (2002) and Tellkamp (2006).

Integrating RFID technology provides a source of dynamic data with a higher level of accuracy and time reduction compared to manual processes. The timely use of dynamically generated information is a key element capable of bridging the gap of operational visibility, enabling integration and the control of business functions in a timely manner (Sarac and Aysegul, 2008), (Mahadevan & Barker 2005b).

To achieve a competitive advantage in the hospitality industry, there are three essential aspects: the gathering of data, the management of data, and the utilization of the generated information. (Mahadevan & Barker 2005a). These three key data aspects provide an opportunity for innovative techniques such as emerging RFID and mobile technology.

The integration of RFID solutions into existing systems or new applications is more complex than widely imagined and requires appropriate specification and customisation to facilitate appropriate implementation. A number of physical issues are involved, such as antenna configuration, radio wave absorption, and optimal antenna positioning which need to be addressed to facilitate alignment with business applications (Atkins, Yu 2010). Currently there are no packaged software solutions for RFID implementation since the process model may differ for every organization and therefore require low level software development to handle the data communication from readers with enterprise systems (Asif, & Mandviwalla 2005).

The requirements for technical advice at the feasibility and implementation stages are crucial to the success of the application. This technology would provide useful applications to the management of high value wine cellars in terms of tracking, monitoring and inventory, with downstream impact on labour cost savings and promotional campaigns.

ISBN 978- 974-672-556-9

II. Application Features of RFID System

A. Equipment Specification

The components of such a system basically comprise an RFID reader-writer, RFID tags and a scanning antenna. The tags can be conceptualized as mini databases that can be written to and as such accumulate information. Unlike their barcode counterpart they do not need line of sight, therefore enabling the identification of, for example, a box of wine by scanning it within limits. In its application it can be attached to products for the purposes of identification and tracking (Bravo, 2008; Sarac, et al, 2008).

The two main classifications of the tags are passive and active tags. A **passive** tag's transmission power or energy source is derived from the reader device. The derived energy emanates from the electromagnetic (EM) waves emitted by the RFID reader device. An **active** tag uses some form of energy source to power the circuitry and antenna either partially or fully. Energy sources may vary depending on the application type, as some may be replaceable or even directly connected to an external power source. These have the capability of transmitting data at scheduled times or at predefined locations unlike passive tags which send data only when they are in read range of a reader. Currently, advances in RFID are enabling reading distances of up to 85 meters (ActiveWave Inc., 2009).

B. Reader classification and selection

It should be emphasized that RFID solution implementation is not an out-of-the-box endeavour. This involves careful site planning and survey and appropriate equipment selection. RFID tag **readers come in various** forms, their usage determined by the application environment. These readers fall into two basic categories, fixed readers and mobile readers, for on-the-spot reading of specific tags. **Static** RFID readers as shown in Figures 1a and 1b are suitable for applications where the object under observation is static or in motion relative to the reader. Inventory and security of dispatched goods or delivery of goods from a warehouse is a typical example of where a fixed reader could be used. Current advancements in reader technology enable a combination of readers and network appliances for running RFID applications.

a. b.

Figure 1. a) Intermec Reader (Intermec) b) Motorola Mobile Reader (Motorola)

Another aspect of data capture that gives flexibility to a fixed system is mobile data capture. A Mobile RFID reader shown in Figure 1b reduces the records' visibility gap when an inventory is on the move. It facilitates the deployment of RFID read points at virtually every key junction of movement.

III. Architecture of the RFID Solution

Inventory operations are a key accounting activity in retail, including the hospitality industry (hotels and restaurants). This is commonly performed a number of times a year, normally after a store has ceased trading for the day; however for all-day trading institutions, trading must be interrupted for a period of time (Hajo, 2003). The results of the inventory provide information on the financial status and condition of the business. Since inventory levels are estimated in the financial statements of months when there has not been an "actual" inventory, less frequent inventory taking may result in long periods of inaccurate information. Figure 2 illustrates the traditional workflow in an inventory system. Here, operational bottlenecks associated with physical counting such as errors of omission, duplications, the financial costs of non-trading due to stocktaking activities etc. can be recognised.

Figure 2. Architecture supporting Wine Cellar System

In the concept overview in Figure 2, the application presents the Front Desk Assistant with a real time view, for the simple reason that they are the first point of call for enquiries and order placement. Other users, e.g. Application Support Personnel, manage stock levels and provide direct product locations in real time, reducing search and location times to service customer needs**.** The application as shown in Figure 2 employs a client server architecture with: low cost EPC Gen II tags attached to high value wine products, an Alien ALR 9800 reader, and a wireless router to take advantage of the existing corporate wireless network (Benyo, et al. 2009), as a cost saving incentive. The host computer was utilized as a system configuration point for the RFID readers which periodically communicate data back to the reader. The Alien tags are pre-programmed with customized identification numbers, for example BCD2000 566 6898, to facilitate easy identification. The setup places a pair of RFID

antennae in a wine cellar enclosure in which the interrogators (readers) are deployed in an RFID system. The interrogated data is communicated wirelessly over the existing wireless system to the host application computer. Workers read and write data to the tags and attach them to each wine bottle.

However, this practice of tagging wine bottles in-house will become less popular in the near future as tagging from source is gaining in interest, particularly from French and Italianwineries. This initiative also is supported by RFID tag research and the production of innovative tags suitable for wine bottle tagging. The positive side of the workflow is that, most of the bottlenecks associated with traditional inventory system are eliminated.

IV. Wine Cellar Storage Concepts

A. Wine Cellar Systems

Figure 3 illustrates the two main concepts of wine storage systems within the scope of this discussion: passive and active. The Traditional Natural Cave and Victorian property wine cellars are classed as passive. Passive cellars refers to natural cave environments e.g. deep natural caves, whilst the active are man-made cellars which make use of insulation, monitoring and cooling systems, artificial humidifiers and seals.

Figure 3. Wine Cellar types and concepts compiled by author (Magaux), (Wine Storage Solutions)

B. Economic value of some vintage wine cellars

The results of a market survey conducted by Sotheby in 2008 show the value of some of the most expensive corporate and privately-owned wine cellars in the hospitality industry (Sotheby, 2008). The prices of these wines are affected by their verifiable provenance and excellence of condition. Table I shows the estimated sale value of some privately-owned wine cellars. Having a product of verifiable provenance provides some competitive advantage in this industry.

TABLE I. SALE VALUEOF WINE CELLARSCOMPILED BY AUTHOR (FSN, 2008)

Cellar Owners	*Qty*	*Est. value*	*Vintages*
Thomas Ryder	5477	£ 0.875m	Romanee
Warren Stephens	4765	£ 2.4m	Bordeaux, Burgundy
Magnificent Bordeaux	4400	£ 1.0m	Bordeaux
Cellar of Siriola	27000	N/A	1850 labels

Wine cellars are built with the prime purpose of ensuring the availability of products to meet demand, and furthermore to preserve the storage conditions and quality of these wines. Secondly, the available list of wines conveys a corporate image and its commitment to quality and care, whilst unsold stock represents economic assets and long term investment.

V. RFID Technology in Wine Cellar Management

A. Wine Cellar management – Case Study

A case study is based on a 30-room hotel specializing in food and wine. This establishment also engages in auctioning some of its expensive vintage brands worth millions of pounds.

The hotel holds nearly 24,000 bottles of wine with a yearly wine stock turnover of about 8,900 bottles. Wine plays an important role in the hotel's operations, accounting for about 20 percent of its restaurant costs, according to Paolo Milani (Pasquini, 2009). It is also noted that quite a significant amount of time is spent on stock verification and inventory. Another issue worth noting is the potential to discourage pilferage of these high value wines by using passive RFID tags to track them, from a survey conducted by (Falken Secure Networks, 2008). Too often, records of temperature variation are absent and therefore unverifiable. RFID technology is used as a tool to manage stocks and customer orders initiated from the front desk of the hotel or over the phone. Issues affecting business growth have caught the attention of the hotel management who are seeking innovative solutions to revamp customer satisfaction and profitability.

B. Effect of RFID Integration on business process

Using RFID technology would result in large amounts of data. With such a volume of data, some business processes would have to be strategically changed to cope with integrating this new system. Whilst it makes provision for traceability it simplifies the process of inventory management as well. This view is also shared by some French wine companies who expressed concern both about preservation and handling during transportation and distribution, and counterfeiting (Falken Secure Networks, 2008). The expression of concern by wine makers began the search for an innovative solution to counteract counterfeiting, improper management of barrel profiles, and the absence of storage condition histories for fine and vintage wines.

Figure 4. Tagged wine barrels

Incorporating RFID technology in wine cellar management provides the potential to reduce the occurrence of these problems. In addition, storage temperature visibility is paramount to quality aging, but the records of this within the industry are either erratic or absent. This problem can be solved by using semi-active temperature tags which provide the storage temperature history of the product. RFID eliminates the manual synchronization of data captured using barcodes to the back-end database systems.

VI. IMPLEMENTATION AND TESTING

The application is implemented as a client server architecture with basic components: low cost EPC Gen II tags attached to high value wine products as mentioned in section II, an Alien ALR 8800 reader, a pair of linear antennas and a wireless router links to the existing corporate wireless network system (Benyo, et al. 2009), as a cost saving incentive. A host computer was utilized as a system configuration point for the RFID readers which periodically communicate data back to the reader. Pre-programmed tags with customized identification numbers, for example BCD2000 566 6898, facilitate easy identification. However, detailed information related to the product each tag identifies is stored in a relational database. The radio frequency power setting of readers is regulated between 1 – 2W to ensure operation within health and safety standards. The positioning and orientation of the tagged bottles were also crucial to the optimization of signal propagation reliability of read tags.

Figure 5. Wine Cellar application system

When the application is launched, it searches for readers in active communication with the host computer. On receiving signals from the reader antenna, the identified tags are pre-matched and the data written to the back–end database. This is enabled as the Transfer Communication Protocol (TCP) listener utility of the reader listens for incoming connections from readers over the network and captures any data the reader sends (alien demo guide document). With the configuration in the Smart Shelf mode, as seen in Figure 6, the Alien ALR 9800 reader facilitates automatic reading and logging of product identities designated by the tags into the shelf inventory (database) as the wine shelves are stocked; likewise, the reader logs a stock reduction as products are removed from the shelves.

Assumptions made in the implementation were that tags had reasonable physical protection against the harsh conditions of operation, that the RFID tags were configured to be read-only, and that each wine product was tagged at item level.

Figure 6. Tag id display with product image (Smart Shelf mode)

A. Observation and results

A number of observations were made during the test and implementation. In section VI a number of configuration settings were made to assist in the in the optimization of the system. Reliability and readability of the alien 9640 tag is 100% reliable in a range of 0 – 7m. However, readability decreased dramatically when attached to the wine bottle due to absorption of the alcoholic content. This method was only functional at a short range. This effect varies among the different brands of wine tested.

TABLE II. RESPONSETO RF SIGNAL – DIRECT TAG PLACEMENTON WINE BOTTLES

Content	Signal Strength	Reliability	Distance
Tag (alone)	100 %	100%	0 –7 m
Red Wine (Translucent)	75%	70%	0-2.5m
Dark bottled wine	88%	90%	0–3.5m
Brown bottle	95%	96%	0 – 3.5m
Black bottled wine	unstable	65 %	0–1.5m

Direct placement of the RFID tags was functional only at short range. Sensitivity and readability was improved by adapting simple techniques such as shielding the RF from absorption by placing a thin piece of polystyrene and dry thin cardboard between tag inlay and the bottle. It was observed that applying a hang tag to the bottle around the neck provided a reliable read range and also reduced the problems caused by the liquid content. Though reviews of RFID literature suggest that it does not need line of sight, varying the orientation of tag inlay between 0 - 20_0 resulted in average detection rates of between 70% and 96%. In general, tags facing reader antennas (angular sensitivity) exhibited better detection rates and an improved relative read range.

Read accuracy is of significant to the success of the RFID applications: this implies misreads result in discrepancies between recorded and physical inventory which eventually leads to inefficiencies in the information system (Lin, et al. 2008). Missing

data due to misread may lead to nonvalue added activities such as manual search and find or manual reconciliation of stock for payment or reorder points. The read accuracy (reliability) at receiving portal is dependent on tag quality, reliability and capabilities of the tags and readers used in relation to application domain. To compensate for problem of imperfect reads the setup utilized multiple reader antennas to improve the powering region.

B. Operational Information

Figure 7 and 8 provide a visualization of the software interface used to engage the application in its interaction with the Alien Reader hardware. The application can be used by the Food and Beverage Manager, Restaurant Manager or Waiter to find information about a requested wine product, its location, and the quantity available.

Specific instructions can be issued to the sommelier as to the exact location of the product. In an RFID system a large amount of data is produced, but this is of no use if it is not transformed into a format capable of supporting an intelligent decision-making process.

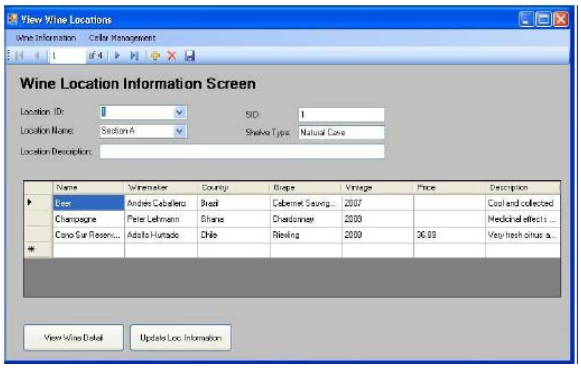


Figure 7. Location with wine information

When the Add / Update Product button is clicked the system enters into a product addition mode; the system reads a tag on any selected antenna, and product names and product IDs are briefly displayed. These products will remain on the screen until removed in the subtract mode. The same screen can also be used for product information updates such as writing notes on a particular brand or marking a bottle as reserved for a certain customer as shown in Figure 7. This application can be extended to provide cross selling in its range of products. Cross selling information, may include food, dissert suggestions to complement that particular wine selection (Wessel, 2009).

VII. Conclusions

This paper has presented an application covering a domain of operations in the hospitality industry using traditional RFID as a solution for wine cellar and asset management which can also incorporate mobile RFID devices to update in real time.

Figure 8. Data capture screen

The integration of RFID solutions in managing wine cellars has the potential to provide faster and better information regarding vintage wine products and increase customer satisfaction. In evaluating the implementation it is evident that time saving is a potential critical benefit of this application. In comparison with common identification technologies such as barcodes, RFID-based solutions present many advantages, such as the reduction of product inventories and improved quality of gathered data with a further improvement in security relevant applications (Vojdani, 2006).

It is perceived that the applications demonstrated offer significant benefits over traditional identification approaches for asset and security management. The implementation costs of these applications are relatively cheap compared to other solutions. In the future, when wine manufacturers utilize RFID from the point of the winery, the opportunity will be created to integrate wine information into the Electronic Product Code Information System (EPCIS). This initiative could be driven and promoted by innovative custom tags e.g. tags embedded in corks (Jonathan, 2005). Smart labels are reusable while the cork e4mbeded tags may not.

It is also intended to develop a knowledge layer on top of operation information to provide business intelligence which presents decision makers with information on inventory usage patterns and financial analysis. Armed with this information, hospitality owners can gain increased insight into their enterprise, leading to increased efficiency and reduced cost.

Acknowledgements

This project has been supported sponsored by an EPSRC Collaborative Training Account (CTA) and Chapman-Nix Ltd. of Rugeley, Staffordshire, in conjunction with Staffordshire University. The authors would also like to thank Dr Emma Price for editing the paper.

REFERENCES

Activewave Inc, 2009. Active wave RFID solutions. [Online] Available: http://www.activewaveinc.com/ [Accessed 14 August 2009].

Alien Technology Cooperation 2007. RFID - From pilots to production: A look at the state of RFID Industry [Online] Available at: http://www.alientechnology.com [Accessed 15 January 2010].

Asif, Z., Mandviwalla M, 2005. Integrating the supply chain with RFID: An In-depth technical and business analysis. Communications of AIS, 15(24), 1-57.

Atkins, A S., Zhang L; and Yu H., 2010. Application of RFID and Mobile Technology in Tracking of Equipment for Maintenance in the Mining Industry' 2010 Underground Coal Operators Conference, February 2010, ISBN: 978-1-921522-16-1, The Australasian Institute of Mining and Metallurgy, pp. 350-358 ISBN 978-1-921522-16-1.

Benyo, B., Vilmos, A., Fordos, G., Sodor, B. & Kovacs, L., 2009. The StoLPan view of the NFC ecosystem. Wireless Telecommunications Symposium, 2009. pp. 1-5.

Bravo, J., Hervas, R., Chavira, G., Nava, S.W. & Villarreal, V., 2008. From implicit to touching interaction: RFID and NFC approaches, Human System Interactions, May 25-27, 2008, IEEE. pp. 743-748.

Daily RFID Co. Ltd, 2009. RFID passive tags for unlimited applications. [Online] Available at: http://www.rfid-in-china.com [Accessed 22 December 2009].

Pasquini, E L ,2009. Italy's hotel wine lists require investment, expertise. [Online] Available at: http://www.hotelnewsnow.com [Accessed 24 January 2010].

Falken Secure Networks, June 2008. RFID for the wine industry. [Online] Available at: http://www.falkensecurenetworks.com [Accessed 12 Jan 2010].

Hajo, A.R., Ed, 2003. Design and control of work flow processes. Netherlands: Springer.

Jonathan, C., 2005. Wine Bottles Get Corked With RFID. [Online] Available at: http://www.rfidjournal.com/article/articleview/2117/1/1/ [Accessed 10 May 2010] RFID Journal.

Karpischek, S., Michahelles, F., Bereuter, A. and Fleisch, E., 2009. A Maintenance System Based on Near Field Communication, Third International Conference on Next Generation Mobile Applications, Services and Technologies, 2009. NGMAST '09. 2009, NGMAST pp. 234-238.

Lin, W., Bryan, A.N., Jayant, R., 2008. Maximizing read accuracy by optimally locating RFID interrogators. In: A. Syed and I. Mohammad, eds, *RFID handbook - Applications, Technology, Security and Privacy*. USA: CRC Press, pp. 204-208.

Kambil, A., J. Brooks. 2002. Auto-id across the value chain: From dramatic potential to greater efficiency and profit. Auto-ID center.

Magellan, 2009. Magellan stack technology [Online] Available at: http://www.magellan-rfid.com/ [Accessed 15 January 2009].

Mahadevan, A. and Barker, J., 2005a. Embracing Radio Frequency Identification (RFID): Introduction and Analysis Using the Five Pillar Foundation Framework.

Mahadevan, A. and Barker, J., 2005b. Embracing Radio Frequency Identification (RFID): Introduction and Analysis Using the Five Pillar Foundation Framework, Asian Conference on Sensors and the International Conference on new Techniques in Pharmaceutical and Biomedical Research, 2005, 5-7 September 2005, pp. 26-30.

Ming Luo, Wei Liu, Leck Leng A W, Tsong Jye NG and Jing Bing Zhang, 2009. Evaluation of enabling RFID technology on the operation visibility in MRO plant. 4th IEEE Conference on Industrial Electronics and Applications, 2009. ICIEA.

Motorola, 2008. Mobile RFID reader. [Online] Available at: http://www.motorola.com [Accessed 14 November 2009].

Vojdani, N., Spiznagel, J and Resch, S., 2006. A process and knowledge-based model to identify and to evaluate the potential of RFID applications. World automation congress, July 24-26. IEEE.

Tellkamp, C., 2003.The auto-id calculator: An overview. Auto-ID center Whitepaper.

Wine Storage Solutions LTD, 2009. Wine cabinets, wine coolers and wine fridges [Online]. Available at: http://www.winess.co.uk [Accessed 17 January 2010].

Sarac, A., 2008. A simulation approach to evaluate the impact of introducing RFID technologies in a three-level supply chain, WSC '08: Proceedings of the 40th Conference on Winter Simulation, 2008, pp. 2741-2749.

Sotheby, 2008. Market review [Online] Available at: http://www.sothebys.com [Accessed 10 January 2010].

McCarthy, U., Ayalew, G., Butler, F., McDonnell, K., and Ward, S. 2009. Impact of reader antenna polarisation, distance, inlay design, conveyor speed, tag location and orientation on the coupling of UHF RFID as applied to modified atmosphere packaged meat. Comput. Electron. Agric. 69, 2 (Dec. 2009)

Wessel, R., Oct. 27, 2009. Interactive Wine Kiosk Wows Customers. [Online] Available: http://www.rfidjournal.com/article/view/5329/ [Accessed 10 May 2010].

RFID Technology in Intelligent Tracking Systems in Construction Waste Logistics Using Optimisation Techniques

Lizong Zhang, Anthony S. Atkins, Hongnian Yu
Faculty of Computing, Engineering and Technology, Staffordshire University,
Octagon, Beaconside, Stafford ST16 9DG, United Kingdom

***Abstract* -Construction waste disposal is an urgent issue for protecting our environment. This paper proposes a waste management system and illustrates the work process using plasterboard waste as an example, which creates a hazardous gas when land filled with household waste, and for which the recycling rate is less than 10% in the UK. The proposed system integrates RFID technology, Rule-Based Reasoning, Ant Colony optimization and knowledge technology for auditing and tracking plasterboard waste, guiding the operation staff, arranging vehicles, schedule planning, and also provides evidence to verify its disposal. It h relies on RFID equipment for collecting logistical data and uses digital imaging equipment to give further evidence; the reasoning core in the third layer is responsible for generating schedules and route plans and guidance, and the last layer delivers the result to inform users. The paper firstly introduces the current plasterboard disposal situation and addresses the logistical problem that is now the main barrier to a higher recycling rate, followed by discussion of the proposed system in terms of both system level structure and process structure. And finally, an example scenario will be given to illustrate the system's utilization.**

Index terms - Intelligent Tracking system; RFID technology; Rule based Reasoning

I. INTRODUCTION

With the growing awareness of the concept of sustainable development in recent years, waste management has became an urgent issue to be addressed by most government initiatives and environmental pressure groups, which seek efficient, secure and environmentally sound solutions for future waste disposal. A typical example is the disposal of plasterboard waste.

Plasterboard is a common construction material widely used in the 1970s and 1980s, but much of this has reached its life end and needs replacement or demolition[1]; it can be anticipated that the volume of plasterboard waste will rise over the next 15 years[1, 2]. However, traditional landfill is not an environmental solution for plasterboard waste, and is even dangerous, as it results in the emission of hydrogen sulphide (H_2S) gas from the reaction between plasterboard waste and organic waste [3].

The reaction is caused by the gypsum and organic carbon, which are the main components of the plasterboard and normal household waste respectively. The reaction causes a large amount of hydrogen sulphide (H_2S) gas emission - 100 tonnes of land filled sulphate can potentially produce 35 tons of H_2S [4]. H_2S gas is malodorous and dangerous in high concentrations [3].

The latest plasterboard waste statistics were published by DEFRA in 2007, and stated that more than 1mt of plasterboard were sent to landfill with only 70,000 tons being recycled per year, which is less than 10% [2]. In fact, plasterboard waste is a valuable waste for recycling, as it contains more than 95% gypsum and about 5% fibre materials [1].

There is no barrier to recycling plasterboard both in the technology and final product utilization aspect, but there are transportation and cost issues. There are many different techniques for recycling plasterboard waste, but in general, they all follow two major processes: separation of gypsum core from the fibre material liner by a screen, and crushing the core to produce a gypsum powder in different sizes [5]. The final products differ, but the gypsum is usually used to make new plasterboard [5].

The two barriers to recycling plasterboard are the cost and the logistical problem. Before the "10% rule" - sulphate content reduced to less than 10% per load – was removed from the regulation in November of 2008 [6], the recycling cost was much higher than mixing plasterboard with household waste and land filling in traditional sites. Even since November 2008, the cost of recycling is still expensive compared to normal waste disposal, and is the cause of illegal disposal of plasterboard waste.

Another problem is the logistical problem: there are only four plasterboard recycling facilities and two warehouses currently running in the UK, but all of them are located in England [7]. This imbalanced geographical distribution results in long transportation times and high costs, particularly from areas such as northern Scotland.

This paper first reviews the reasoning system and optimization methods, and then proposes a concept of a waste management system based on RFID technology, a rule-based system and ant colony optimization for auditing and tracking plasterboard

ISBN 978- 974-672-556-9

waste disposal, and discusses the work process and structure. Finally, an example scenario will be given to illustrate its utilization.

II. Proposed Solution Descriptions

The proposed solution relies on Radio Frequency Identification (RFID) technology to recognize the identity of each object involved in the system. RFID is an automatic identification technology which has potential prospects in asset tracking and, it is widely believed, will eventually replace barcodes. RFID technology involves many electromagnetic and electronic technologies, but in general, the work process can be described as: 1) transmit adequate energy to power up the tag and 2) communicate with the tag to request and receive the identifier [8].

There are many standards of RFID technology published by EPCglobe and ISO. ISO has issued about 50 standards related to RFID technology. However, the most important and widely used RFID standard is published by EPCglobe and called EPC standard, which aims to build the concept of the "internet of things". These two organizations have competed for several years, but they are now collaborating on EPC class 1 Generation 2 standards, introduced into an ISO standard system as ISO18000-6 with a slightly modification[9].

RFID has many benefits compared to other identification technology. Firstly, it is non-contact technology that uses radio waves as a communication media, and is useful for improving efficiency in terms of reduced identification time. Secondly, the RIFD tag can work in a complex environment, such as contaminated by dust or waste. This is important for the waste management scenario, as the tag usually contaminated by waste, and other technology, such as barcodes, finds this difficult. Therefore, EPC G2C1 passive RFID technology is introduced for the major object sensors in this application.

An example structure is shown in Figure 1. The design uses EPC G2C1 passive RFID which works in a UHF band and provides up to 10 meters read range. The RFID tag is designed to be attached to the container (bag, box or bin) of plasterboard waste at the source sites, the tag's ID relates to information about the waste such as weight, type, source site, produced time and current location etc. This information is located in the central server, and can be checked by mobile hand-held devices. Before the waste finally reaches the treatment sites (either the recycling facility or landfill site), the waste may temporarily be stored in the transfer depot, and may be put into larger skips/containers for long distance transportation. Therefore, the RFID equipment's antennas/readers needed to be deployed in all the key places to monitor each stage of the waste transportation. Every movement of the waste container generates a record and information can be updated by the operating staff as appropriate. Consequently, the system can track the waste from source to destination and prevent fly-tipping through complete transportation records [10-13].

The proposed waste management system uses RIFD technology and digital imaging equipments to integrate records including location, volume and weight, container movement, delivery tracking,

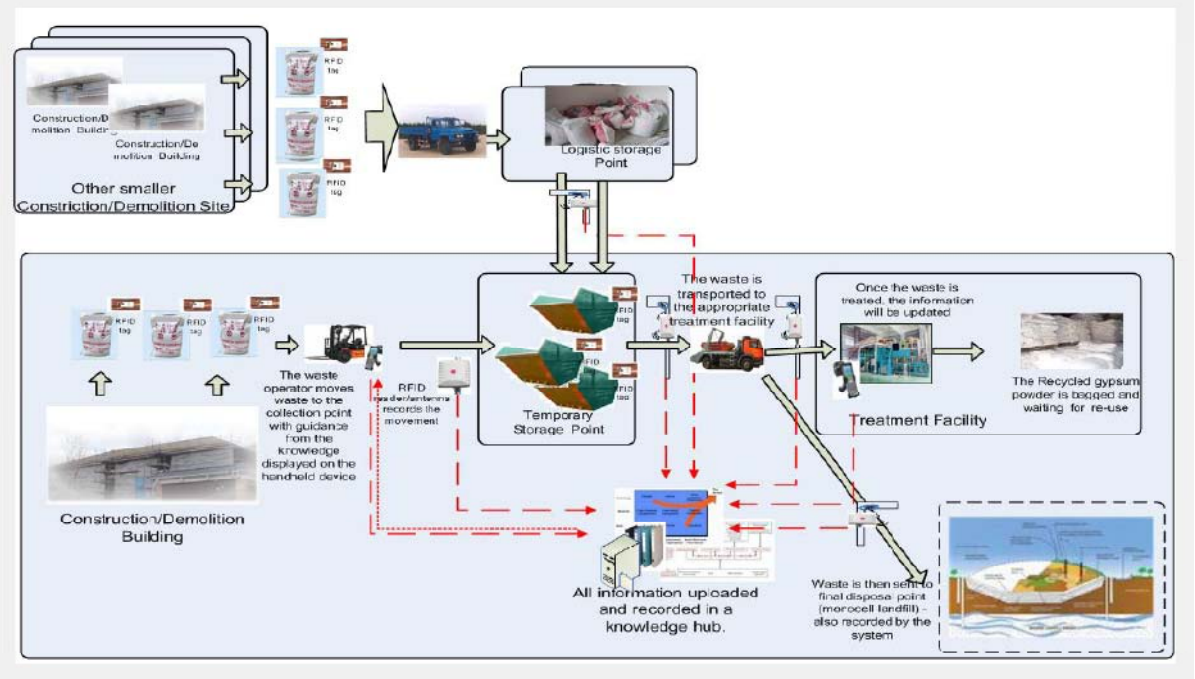


Figure 1. Proposed system structure of the waste management system

inventories and scheduling etc[11]. It works with the support of a technology hub which also includes a knowledge management system to handle and represent the data and helps managers to make decisions on the scheduled logistics of the waste. It could also provide instruction to operating staff via a handheld device, such as PDA. This function is supported by the rule-based reasoning system and Ant Colony Optimization.

III. The Rule-based System

The rule-based system is usually called an expert system, and is the most popular choice for knowledge-based applications. A simplified definition of rule based reasoning is a technology in which knowledge is represented by a set of IF…THEN... production rules and data is represented by a set of facts[14]. The rule will be executed when the fact matches the condition of a rule, and it may add or modified to fact for a new rule execution until the final result is determined [14].

Rule-based reasoning has some advantages compared with other reasoning technology and has been generally accepted as the best option for a knowledge-based system. It typically features natural knowledge representation, uniform structure, separation of knowledge from its processing and has the ability to deal with incomplete and uncertain knowledge. Some features of rule-based reasoning are suitable for the proposed system, and are discussed as follows [14].

Rule-based reasoning technology stores knowledge in IF…THEN structure meaning each piece of knowledge is relevantly independent from other knowledge. This structure is efficient for finding out the target knowledge when the waste regulation is amended. Secondly, the waste management system requires that knowledge should be easy to adopt into the reasoning system without complex transformation. In fact, it is better to input knowledge without any programme skills for ease of use and maintenance/updating purposes. Furthermore, individual knowledge storage is a key required feature that separates knowledge from the system and thus it could be removed without affecting the system design and a new knowledge base which contains the knowledge for other waste management areas could be supplemented.

The system work process structure can be described in 4 layers as illustrated in Figure 2. The lowest layer is the data processing layer, which is the route for acquiring the data and information from the RFID and other sensor equipments into the system. The data gained from the equipment is separately sent to databases in the second layer. The second layer integrates the data from different sensors and combines them into a single main database. This can be accessed and output the data to a long term data warehouse, and an OLAP (Online analytical processing) function can be introduced for better performance. The next layer is called the knowledge reasoning layer, which is responsible for generating solutions for logistical and tracking support. The upper layer is the visualisation layer, and bears the communication function between the system and users. It is designed to represent all the information generated from the system to the users, including the records, real-time tracking data, logistical solution and guidance.

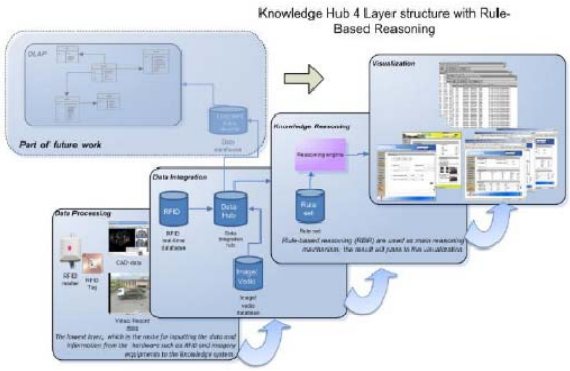


Figure 2. The 4-layer structure of waste management system

The data collected from the lowest layer hardware is stored in the second layer database, and generates the records as „fact“ which will be used in the third layer reasoning for generated the routing plan and guidance. The results of third layer will then be passed to the visualisation layer and displayed to the end users.

IV. Optimization Module Design

The reasoning layer is responsible for the optimized schedule plan, generates the real time guidance and reports on the current situation function, but the optimized schedule plan is the major task of the knowledge reasoning layer.

Normally, schedules include two aspects: the time plan and the route plan. However, considering the application is designed for a waste recycling company and most waste collection times are contracted, therefore the proposed system only needs to generate the route plan and the time schedule has been assumed to be initially confirmed by contract between the waste company and the construction company.

The routing plan of the transportation can be seen as a classic TSP (Travelling Salesman Problem) question, which has the same requirement: the vehicle departs from the recycling facility, visit each site one time, and finally returns to the recycling facility. The major task of the reasoning layer is planning and finding an efficient route. It is also responsible for real-time planning in case of an emergency where a new route needs to be planned.

ISBN 978- 974-672-556-9

TSP is a classic problem and there is much related research on this topic. The start of the problem can be traced back to 1832, but it was first formulated as a mathematical problem in the 1930s by Karl Menger, and Hassler Whitney introduced the name Travelling Salesman Problem [15].

The expansion of research into TSP started in the 1950s and 60s, and many algorithms and methods have been developed to solve the problem efficiently. Furthermore, in 1972, the TSP was proved to be an NPC (Non- deterministic Polynomial Complete) problem, that is, the best result usually takes a long CPU time and is impossible for real applications. Therefore, an optimized algorithm has been proposed to find an acceptable „good" result rather than the best result, such as Ant Colony, Genetic algorithm or Tube Search [15].

The requirement of the proposed system's application area restricts the route plan algorithm to matching the following features: 1) Inherent parallelism, which needs consider more than one route at same time 2) Efficient for Travelling Salesman Problem and similar problems, which the main problem can be seen as a Travelling Salesman Problem 3) Can be used in dynamic applications. Therefore, for this application, ACO will be introduced in the system that is responsible for generating the route plan [16-19].

The ACO module is only dealing with the vehicle routing plan, therefore it needs to be independent from the main rule-base to reduce complications, and thus it does not need convert to the production rule format. It only works when the vehicle type and target site has been decided by the rule based reasoning system; the vehicle and site information will be passed to the ACO module as the initial parameters, then the acceptable result can be generated in limited iterations. The work procedure is illustrated in Figure 4.

The reasoning layer is used not only for the schedule planning, but also provides guidance and deals with some emergency situations. The data collected from the hardware layer represents the progress of the waste collection and it can be compare with the schedule, taking into account delay, vehicle problems or road blocks, and a new plan or other solution may be generated to reduce the effect.

The work procedure of the reasoning layer starts from the time schedule and routing plan. Firstly, the system will check the current time and query the database if there are any sites which need to be visited in this time (day, week or month) and also query the last operation on that site to roughly estimate the tonnage of the waste. The estimating also takes into account the site project, construction progress and even its financial situation.

The next step is to decide the vehicle type and the number. After the site which must be visited in the next period has been decided and the waste tonnage of each site is estimated, obviously the total amount of waste will be known. The vehicle type can then be decided based on this information; the capacity of the vehicle should be larger than the tonnage and depends on the containers used on the sites. The rule-based system will be based on these 'facts' to reason out the vehicle type and number. Planning the details of vehicle routing is the function of the ACO, which

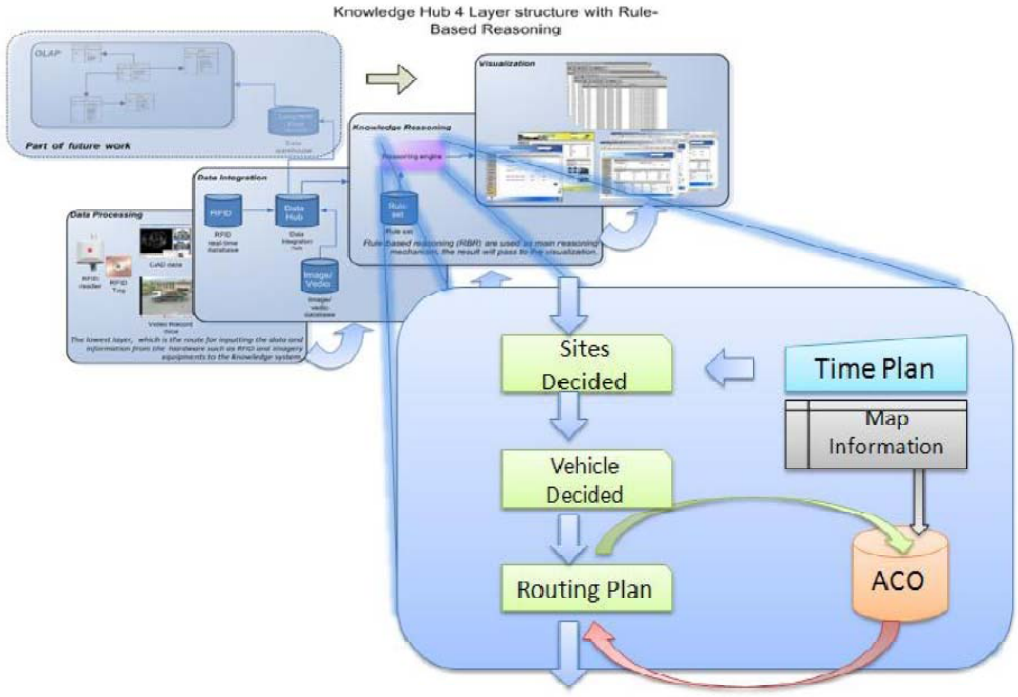


Figure 3. The 4-layer structure with ACO module

for a single trip. Then the exact route will be calculated by the ACO, in the prototype of the waste management system, only the original ACO will be introduced for evaluating purposes. After the routing has been decided, the details will be passed to the visualization layer for guidance.

Another important function of the proposed system is providing guidance to the operation staff to help them deal with the waste. It works as a handbook to remind them when, where and how to collect/transport the waste. Apparently, the transport plan is part of the guidance information that can give clear instruction about route choice and waste collect procedure to the vehicle drivers.

V. The Example Scenario

In this section, a simple example scenario will be introduced to illustrate the system's work process. As Figure 4 shows, there are 10 source sites which produced plasterboard waste, and 1 transfer station for temporary storage. The recycling facility is the final destination of each transportation. The yellow truck can only carry 2 units of waste, and the green can have 3 units. In this scenario, there are 6 source site and 1 transfer station needing to be recycled.

The beginning of the system schedule plan is checking the database, which contains the regular waste collect data. The last records of the waste collecting operation such as tonnage and last collection time can also be retrieved. Based on this data, the system estimates the waste tonnage and classifies it in three different emergency levels. For easy understanding in Figure 4, the colour of each site represents this level: green means no waste on the sites, grey means some waste but no need to collect right now, and red means it need to be collected as soon as possible, as either the container capacity or the contract deadline have been reached. The estimated waste unit is marked as a number in each site.

On the example scenario, there are two sites marked in red, which means they need to be collected soon. The next step is determining the vehicle type for collecting the waste. The system will find the vehicle whose capacity just matches the target site waste tonnage. Therefore, for site 'a' and 'b', the yellow vehicle will be assigned, and two green trucks will go to the transfer station, where 6 units are waiting to be collected.

After the target sites for each vehicle have been decided, the ACO plans the routing. Because the capacity of the trucks are all above the current estimated waste tonnage, therefore, the nearest site will be involved in the route if there is some waste present. The scenario exampled outlined is simple only for proving the idea, to determine the route in an easy way. The green arrow shows the route for site 'a' and the red one for site 'b'. In route a, the nearest two sites which have 0.5 units of waste will be visited by the yellow vehicle.

The proposed system has the ability to deal with emergency changes. For example, if the vehicle reaches site a, and access is denied, the system will rearrange a new route with the start point is site a,

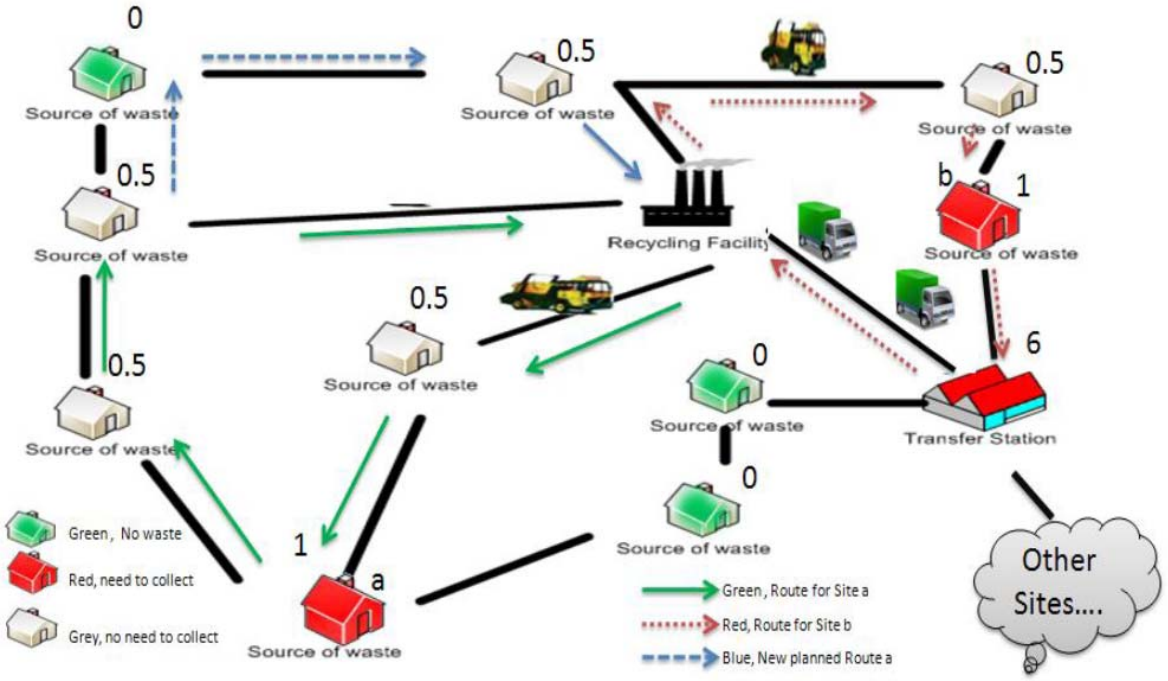


Figure 4. Example scenario of system utilization

ISBN 978- 974-672-556-9

and final destination as the recycling facility. The new route is marked as a blue arrow, and the vehicle will visit another two sites on the route until the capacity is full.

The proposed system will use the original ACO system for simplicity and prove the idea. It usually thus ACO in this application could have a quicker response and avoid complex calculation.

VI. CONCLUSIONS

This paper introduced the current plasterboard disposal situation and addressed the logistical problem which is a barrier to an increased recycling rate. In the UK only 4 known recycling facilities are available, all of which are located in England, and two of them in the London areas. This situation has caused difficulties with transportation, and the recycling fees are even higher than landfill if the source site is far from the facility. This paper also reviewed the reasoning system and optimization methods, then a proposed system for waste management was discussed which uses RFID technology for the main data collection methods, and rule-based reasoning and Ant Colony Optimization for auditing/ tracking the plasterboard waste. It also has the function to make a schedule plan and provide the guidance to the operation staff to ensure each piece of waste is transported to the right location. The system can also handle emergency changes such as road blocks or if site access is denied, as it will re-arrange suitable routes that reduce potential loss. The structure of the waste management and work process are also introduced in this paper; the four layer structure illustrates how it greatly relies on RFID equipment for collecting logistical data and uses digital imaging equipment to give further evidence. The reasoning core in the third layer is responsible for generating schedules and route plans and guidance, and the last layer deliver the results to the users. Finally, an example scenario was given to illustrate the utilization.

REFERENCE

[1] MTP, "BNPB1 Plasterboard -Industry, Product and Market Overview," 2007.

[2] DEFRA, "Waste Strategy for England 2007," Department for Environment, Food and Rural Affairs, 2007.

[3] DEFRA, "Environmental impacts - priority products - Plasterboard," Department for Environment, Food and Rural Affair, 2007.

[4] D. Heguy and J. Bogner, "Cost-Effective Hydrogen Sulfide Treatment Strategies for Commercial Landfill Gas Recovery: Role of Increasing C&D (Construction and Demolition) Waste," *Municipal Soild Waste Management*, 2004.

[5] H. Hamm, R. Huller, and J. Demmich, "Recycling of plasterboard," *Zkg International*, vol. 60, pp. 68-74, 2007.

[6] EA, "Position Statement MWRP 007," Environment Agency, 2008.

[7] P. R. James, E. Pell, C. Sweeney, and C. S. John-Cox, "Review of Plasterboard Material Flows and Barriers to Greater Use of Recycled Plasterboard," The Waste & Resources Action Programme, 2006.

[8] G. Roussos, *Networked RFID: systems, software and services*. London: Springer, 2008.

[9] J. Landt, "The history of RFID," *Potentials, IEEE*, vol. 24, pp. 8-11, 2005.

[10] A. S. Atkins, L. Zhang, and H. Yu, "Application of Knowledge Hub and RFID Technology in Auditing and Tracking of Plasterboard for Environment Recycling and Waste Disposal," in *10th International Conference on Enterprise Information System- ICEIS* Vol. IV, pp.190-196 . ISBN: 978-989-8111-48-7, 2007.

[11] A. S. Atkins, L. Zhang, H. Yu, and W. Miao, "Application of Intelligent Systems Using Knowledge Hub and RFID Technology in Healthcare Waste Management in UK and China," in *International Conference in e-Business (ICE-B 2009)* July, Milan,Italy: pp 44-49, ISBN 978-989-674-006-1, 2009.

[12] L. Zhang, A. S. Atkins, W. Miao, and H. Yu, "Healthcare Waste Management Using Knowledge Hub and RFID Technology in the UK and China," in *3rd International Conference on Advanced Computing &Communication Technologies,November 06-09 2008,APIIT*, Panipat,India,ISBN: 81-87433-68-X. Pp. 14-20 2008.

[13] L. Zhang, X. Zheng, N. Choosri, H. Yu, and A. S. Atkins, "Applications of RFID in Indoor Navigation for The Visually Impaired," in *15th International Conference on Automation & Computing* Luton, UK: In press (19 September 2009), 2009.

[14] J. C. Giarratano and G. Riley, *Expert systems : principles and programming*, 4th ed. Boston: Thomson/Course Technology, 2005.

[15] G. Gutin and A. Punnen, "The Traveling Salesman Problem and Its Variations," in *Combinationial Optimization*. vol. 12 Kluwer, Doedercht, 2002.

[16] A. Colorni, M. Dorigo, and V. Maniezzo, "Distributed optimization by ant clonies," *Proceedings of the 1st European Conference on Artificail Life*, p. 8, 1991.

[17] M. Dorigo and L. M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *Evolutionary Computation, IEEE Transactions on*, vol. 1, pp. 53-66, 1997.

[18] M. Dorigo, G. Di Caro, and L. M. Gambardella, "Ant algorithms for discrete optimization," *Aritificial Life*, vol. 5, pp. 137-172, 1999.

[19] Z. Qiang and Z. Qiuwen, "An Improved Ant Colony Algorithm for the Logistics Vehicle Scheduling Problem," in *Intelligent Information Technology Application, 2008. IITA '08. Second International Symposium on*, 2008, pp. 55-59.

.

Perception of Smart Home Technologies to Assist Elderly People

Saisakul Chernbumroong[1,2] Anthony S. Atkins[1] and Hongnian Yu[1]
[1]Faculty of Computing, Engineering and Technology, Beaconside, Staffordshire University, UK
[2]College of Arts, Media and Technology, Chiang Mai University, Thailand
s.chernbumroong@staffs.ac.uk, a.s.atkins@staffs.ac.uk, h.yu@staffs.ac.uk

***Abstract*—In the last decade, the number of elderly population has increased significantly which affects human in many aspects, especially in healthcare. Many studies have shown increases in expenditures on long-term care. New models of care are needed including supported self-care and home-based services. Advance in sensor and network technologies have made these possible. A smart home which is a residence equipped with smart technologies providing services that enhance human way of living i.e. safety, security, entertainment, etc would allow elderly to maintain living independently in their homes and still in control of their healthcare cost and status. This paper reviews various topics on smart home technologies including smart home projects, smart home network, smart home appliance and sensor technologies for smart home. A successful adoption of smart home technologies requires appreciation of stakeholders' perceptions, needs and concerns. A survey has been carried out at a major hospital, nursing homes and general population to explore the perception of six smart home technologies to assist elderly people and concerns regarding the use of smart home technologies. Overall, the result showed positive feedbacks toward these technologies. Participants were concerned about issues such as lack of human responders, user friendliness of the device and the need of learning new technology. In terms of willingness of adoption of these technologies is still unclear. However, it is suggested that by giving real experiences of smart home technologies, their usefulness and effectiveness would be more appreciated.**

Index Terms - elderly people assistance; technology perception; smart home technology

I. INTRODUCTION

The growing numbers of elderly population and increasing life expectancy have brought enormous challenges to many aspects of human life, especially in health and healthcare. According to the United Nations online database [1], currently the percentage of elderly population is 7.6% which is projected to rise as high as 16.2% in 2050 as depicted in Figure 1. Issues such as increased healthcare expenditure, burden to caregivers and insufficient and inefficient care are more likely to occur as the health of older persons normally deteriorates with increasing age, resulting in more demand for long-term care. The expenditure on long-term care provision in Germany, Italy, Spain, United Kingdom and United States of America is projected to increase significantly [2, 3].

In order to support the demand for healthcare due to population ageing, new models of care will be required. An appropriate balance of settings for long-term care, including supported self-care and home-based services are necessary [2]. The system should reduce the burden on caregivers as well as healthcare costs, while maintaining a good quality of care. Emerging technology can facilitate self-care and extend the self-reliance of the ageing population. The care of elderly could be enhanced through monitoring system, sensor technologies and communication systems.

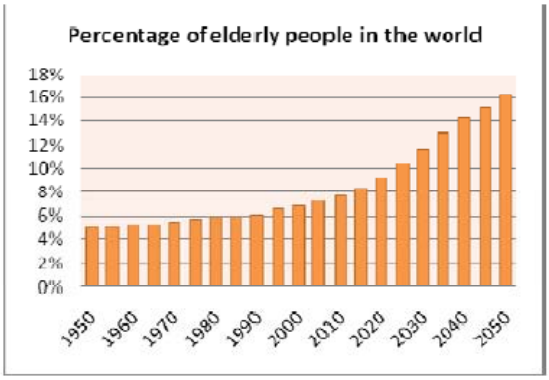


Figure 1. Percentage of ageing population in the world during 1950 and 2050.

Advance in sensor and network technologies have changed the way people live. A number of research projects have been carried out in order to improve human quality of living. Many smart home projects have been developed to explore the use of technology for home automation. King [4] defined smart home as "A dwelling incorporating a communications network that connects the key electrical appliances and services, and allows them to be remotely controlled, monitored or accessed". A smart home is equipped with smart technology and network such as smart lighting system, smart kitchen, energy usage monitoring, security system, etc that enhance people's life in many aspects i.e. security, entertainment, convenience and etc.

Smart home technology can be especially useful for elderly or disabled persons who wish to live independently. Smart home technology has led the idea of ageing in place possible in which elderly person can maintain living independently in their homes and still in control of their healthcare cost and status. Elderly persons can take the advantages of

ISBN 978- 974-672-556-9

smart home technology such as monitoring system, emergency system, dangerous kitchen appliance detection, fall detection and etc, to maintain healthy and safety living while living independently.

The paper reviews smart home technology regarding current smart home projects, smart home networks and appliances, sensor technologies that are used in smart home. The paper also explores the perceptions, needs and concerns of elderly family in regarding using smart home technology for elderly care.

II. SMART HOME TECHNOLOGY

A. Smart home projects

Many research labs have investigated in smart home technology [5]. As the demographic is changing, a field of research in healthcare has resulted in increased interest in the potential of smart home technology for healthcare purposes. Several pilot projects employed smart home concept have been explored and developed with the aim of improving the quality of life and promoted independent living of elderly persons by using advanced sensor and network technology [6, 7] and the paper reviews further elderly care related projects which employ the concept of smart home both in EU and US.

The European Commission has funded a number of research projects to help with the growing number of elderly population in Europe [8]. These projects have been developed to determine how ICTs can meet the needs and maximise the potential of older people. One such project, the Easy line + project [8] examines the use of sensors, neural network and assistive software to develop a control system for white goods such as washing machine, dishwashers, fridges and freezers for easier use of older persons. A variety of sensors such as illumination sensor, temperature sensor, door sensors, radio frequency identification (RFID) and etc have been used in this project. As this project attempt to develop advanced white goods for easier use for elderly, human-machine interface (HMI) must be easy of use and available to any kind of user. The project researched user controls in the market place and found digital television with a remote control to be suitable as elderly people are familiar with and know how to use the basic functions. A touch screen device is preferably for a portable device. Another project called Persona [9], aims to harmonise Ambient Assisted Living Technologies (AAL) by developing sustainable, scalable and affordable services platform for support elderly in activities of daily living, mobility and displacement and protection from health and environmental risk. Persona assesses and evaluates a range of ICTs such as smart textiles, tele-services, bio sensor, smart devices and intelligent software tools. Persona provides basic functionalities and hardware components allowing easier integration to other assisted living/smart home services.

In US, University of Rochester developed the Smart Medical Home [10] which is a controlled environment laboratory setting for concept testing, pilot and prototype testing of technologies and products for personal health system. The smart medical home is equipped with infrared sensors, biosensors, computers and video cameras to collect the data that will augment the data collected by physicians and hospitals. Another project by University of Virginia, Smart In-Home Monitoring System (Figure 2) developed in-home monitoring of residents using a set of low-cost, non-invasive sensors together with data-logging andcommunication module and data management system in order to increase quality of care and provide quality of life indicators [11]. The collected data can be used for the observation of general health and activity level, activities of daily living (ADL), index of well being and the decline in ability over time. The university also developed other projects related with elder care such as gait monitor device, MARC robotic walker, sleep monitoring system, etc.

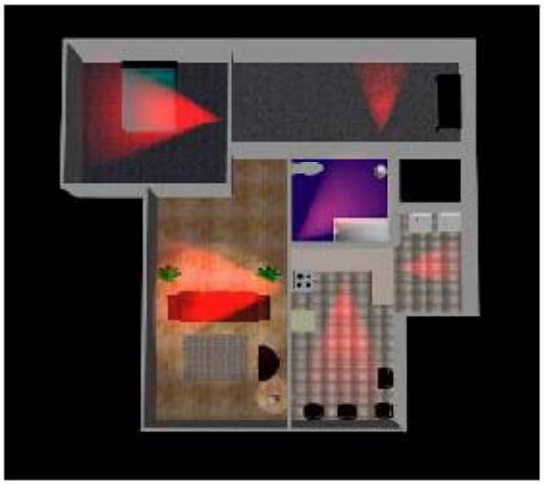


Figure 2: Smart In-Home Monitoring System [12].

B. Smart home network

A smart home incorporates a communication network in order to control or monitor appliances or services within the home. Smart home network technology can be classified by interconnection method into three main types: wire, wireless and both wire and wireless.

1) Wire: Appliances and services in smart home are connected through wiring system such as optical fibre, cable and Powerline. In this system, devices are normally connected directly into the main power supply. The data is sent through the normal wiring to activate or deactivate the appliances/devices. An example of wired communication network for smart home is X10, which is an international and open standard for home automation. X10 allows communication with appliances over standard

electrical wiring. X10 has been developed for over 35 years yet still remains popular as it is inexpensive, easy to set up and widely available. Other technologies are, for examples, European Installation Bus (EIB), Universal Powerline Association, HomePlug and etc.

2) Wireless: Many of new smart home appliances use wireless communication technologies such as infrared and radio frequency (RF). As radio wave can penetrate through walls, floors and cabinets, devices within smart homes can communicate wirelessly. An example of home automation network standard is Z-wave which is a proprietary wireless communication for home automation. It employs low-power RF technology allowing home appliances to communicate with each other such as control of lighting, air-conditioning and security system. Other communication networks for smart home that use RF as transmission medium are such as Bluetooth, ZigBee and Wi-Fi.

3) Both wire and wireless: Some of smart home network standards can work using both wire and wireless technology. For example, INSTEON [13] is similar to X10 but overcomes the limitation of wired network by using RF technology. It is used for home automation such as lighting, appliances, and other home applications control. INSTEON devices uses dual mesh network, RF signals and home electrical wiring to communicate with other devices.

C. Smart home appliances

Smart home appliances are intelligent artefacts that enhance human way of living in term of convenience, safety, etc. A number of appliances have been reviewed in [5]. In this paper, a survey of some of smart home appliances that can be used for elder care as follow:

1) Cooking hob and oven safety control: Many of the elderly people have a condition of forgetfulness in which safety and security of a person is reduced. Hobs and ovens are kitchen utensils regular used for cooking. Elderly person may easily forgets to switch the hob and oven off after finish cooking, especially electrically ones which is harder to notice. Forgetting to turn off these cooking tools create dangerous environment i.e. potential cause of fire, injury from accidentally contact. Hob and oven safety control have temperature sensors to control the heat of the hob or oven and cut off the power if the heat reaches the safety limit.

2) Sleeping pattern monitoring: A bed is equipped with sensors which can detect the presence, respiration, pulse and movement of a person in bed. A sleeping pattern monitoring can be used to detect health condition regarding sleeping such as restless sleep, rapid change in activity level or unusual change in typical routine of a person. For example a person who normally gets up early but on a particular day tends to be lying in. This may indicate that the person may have a serious illness or incident causing the person unable to move. Another example is during the night, a person is detected to leave the bed but has not returned for some times, this may indicate that the person may have accident or is in emergency situation.

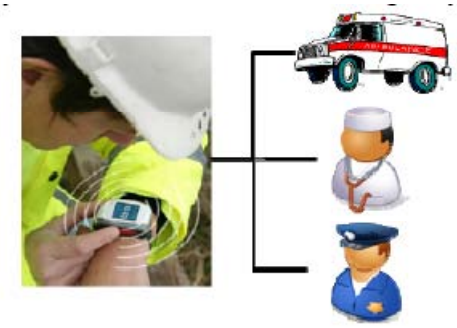


Figure 3. Conceptual diagram of emergency alarm

3) Emergency alarm: An emergency alarm is normally a device which contains buttons that is pressed when a user is in a dangerous or emergency situation and requires immediate help as illustrated in Figure 3. When a button is pressed, the device is automatically sent an emergency message to a designated person or organisations such as family, doctors, polices and etc. The device can be programmed to assign different button to each contact, for example button A for ambulance. The emergency alarm is also equipped with a location sensing sensor such as Global Positioning System (GPS) allowing location information to be sent with the emergency message resulting in quick response to the incident.

4) Automatic lighting system: The automatic lighting system allows any lights within the smart home to be switched on and off automatically when a person is in the area promoting a safer environment. The system uses motion sensor to detect the movement of a person to provide automatic illumination. Moreover, to save the energy, the system also equipped with photocell sensor allowing the system to operate only at night or in a dark environment. The automatic light system can be used in any rooms e.g. bedroom, living room, toilets, etc or at the stairways. An example use of the system is when a person walks downstairs during the night, the system can illuminate the stairways increasing visibility, promoting safety environment.

5) Video monitoring system: The video monitoring system usually composes of video camera and a computer system. It is used to monitor people within the smart home. Video cameras are installed in any places that need monitoring. Example locations are such as living room, kitchen, hallways and etc. The video cameras record visual data which is interpreted into useful information for monitoring purpose later. Video monitoring system is used for security and safety purpose as it can powerfully

ISBN 978- 974-672-556-9

detect human activity and behaviour which resulting in a prompt response in case of emergency and unusual activity i.e. falls and other accidents.

Figure 4: Conceptual diagram of activity monitoring system

6) Activity monitoring system: An activity monitoring system is used to monitor activities of daily living such as walking, standing, cooking and etc of a person and allow unusual activity i.e. fall to be detected for a faster response. The activity monitoring system is composed of a wrist watch which is equipped with various sensors such as accelerometer, radio frequency identification and etc used for activity detection. The sensor data is sent wirelessly and stored in a computer system allowing authorised people such as family and healthcare professionals to monitor the activity of a user anytime anywhere as shown in Figure 4. When abnormal activity is detected that information is sent to any authorised persons to ensure a user receives help as quick as possible.

D. Sensor technology for Smart home

Sensor technology is an important component of smart home. Data collected from sensors are processed to provide intelligent services in smart homes. Examples of sensors are outlined as follow:

1) Radio Frequency Identification (RFID): RFID is a means of storing and retrieving data through electromagnetic transmission to an RF compatible integrated circuit and is now being used as a means of enhancing data handling processes [14]. RFID has been used in various applications across industries such as asset tracking, manufacturing, supply chain management, retail, payment systems, security and access control [15, 16]. Application of RFID technology for a smart home are such as tracking device, activity monitoring system, RFID key card, etc. A basic RFID system, as depicted in Figure 5, comprises three main components: RFID tags, an RFID reader and a middleware.

An RFID tag consists of a microchip attached to a radio antenna mounted on a substrate. The purpose of the tag is to store data which can be altered or retrieved via radio frequency. There are three types of tags: passive tags, semi-passive tags and active tags. The passive RFID tags have no battery and require an external source to invoke a signal transmission. Semi-passive tags also require an external source to activate them, but have a significantly higher forward link capability providing a greater read range. An active RFID tag contains a battery and can transmit signals autonomously. The RFID reader interrogates RFID tags that are in the reading range using radio communication.

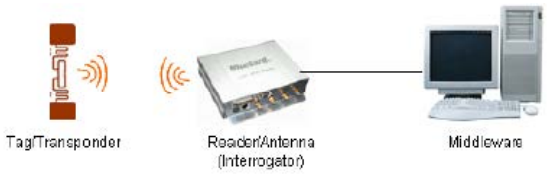


Figure 5: Basic RFID components

The RFID reader works with the antenna emitting a radio wave to activate the tag and to read or write data to the tag. The antenna can send the signal over an area ranging from a few centimetres to 100 metres or more, depending on the frequency used and the output power. Once the tag passes the electromagnetic field, it detects the activation signal and modulates it. The reader decodes the data stored in the tag and sends it to the middleware for further data processing.

Middleware is used for data processing, routing and managing the RFID reader. RFID middleware combines the RFID data with application logic and generates appropriate application events [17].

2) Accelerometer: is an instrument that measures the applied acceleration acting along the sensitive axis [18]. It is widely used for human activity recognition purposes [19 -21] because of its capability to respond to both frequency and intensity of movement, and measure tile as well as body movement. Accelerometers are relatively small and inexpensive which makes them appealing to real-life applications. There are many types of accelerometer for example, piezoresistive, piezoelectric, magnetoresistive, capacitive etc in which different key technologies are used to measure acceleration [22]. Conceptually, a variation of the spring mass system is used. In this system, when acceleration is applied, a small mass inside the accelerometer responds by applying force to the spring, causing it to yield or compress. Measurement of the displacement of the spring is used to calculate the applied acceleration. Examples of accelerometer sensor in smart home appliances are such as fall detection, activity monitoring system, etc.

3) Motion sensor: A motion sensor is used for movement detection. There are 3 sensor technologies that are used to detect motion: passive infrared, ultrasonic and microwave. The passive infrared motion sensor works by detection of the body heat. The infrared radiation cannot be seen by human as its wavelength is longer than a visible light. Any objects that generate heat also generate infrared radiation. Passive infrared motion sensor monitors the temperature and search for changes in infrared spectrum. This type of motion sensor is commonly used in indoor environment. For instance, motion sensor is linked to a light switch and when it detects presence of person, the light is switched on automatically.

ISBN 978- 974-672-556-9

When ultrasonic and microwave technologies are used for motion detector, they are considered as active motion sensors. These two technologies work similarly. They emit optics or sound waves and measure the reflection to detect motion. Ultrasonic acoustic wave cannot be detected by human ear, however may be sensed by certain animals i.e. dogs, fish, etc. An example use is an active motion sensor which emits radar pulses is attached to the gate. When any object enter the area, it disrupts the radar pulses in which the reflection time has changed, the motion sensor could trigger the gate to open.

Other sensors normally found in smart home technologies and appliances/devices are such as pressure sensors, temperature sensors, audio sensors, etc.

III. Perception of Smart Home Technologies to Assist Elderly People

Although smart home technologies demonstrate potential benefits in assisting elderly people, a successful adoption of such technologies would require a thorough assessment of the need, perception, and concerns of related stakeholders i.e. carers, elderly persons, elderly families and relatives, etc. Able to understand their perceptions would allow the development of new smart home technology which meets user requirements.

There are a few studies which have investigated the perception or views of using smart home technologies for elderly care. One of these studies is by Demiris et al [6] who explored the perceptions and expectations of older adults in regard to installation and operation of smart home technologies to improve their quality of life and/or monitor their health status. Devices and sensors in health-related issues such as preventing and detecting falls, assisting with visual or hearing impairments, etc were discussed and the result showed that the seniors, in general, have positive attitude towards these devices and sensors. The study also indicated that older adults were concerned about falls and they perceived technologies that monitor activity levels and sleep patterns as useful. This leads to their follow-up study [7] which investigated in older adults' perceptions in specific smart homes technologies i.e. bed sensor, gait monitor, stove sensor, motion sensor and video sensor. Their finding indicated an overall positive attitude towards those technologies for non-obtrusive monitoring and seniors were concerned about privacy violation, visibility and accuracy of the devices. A study by Steele, Atkins and Yu [23] investigated wireless devices i.e. MDKeeper, pressure sensors, fall detection from acceleration measurement, location tracking devices using RFID and video monitoring and assessed their effectiveness and practicality of real world use for assisted living system and elder care. Based on their questionnaire surveys, these devices were perceived as effective for elder care in home environment. In addition, non-intrusive and mobile devices such as wireless watch are more preferable.

This research aims to explore the perception of people who are associated with elderly such as carers, elderly persons, elderly families, relative and friends in regarding the use of smart home technologies to assist elderly people. The study seeks the perception i.e. usefulness, effectiveness, etc, concerns and willingness of adoption of specific smart home technologies for elderly care. The selected technologies are such as cooking hob and oven safety control, sleeping pattern monitoring, emergency alarm, automatic lighting system, video monitoring and activity monitoring system and are described in Section II-C.

A. Research methods

A survey was conducted on the participants and a questionnaire was designed. The data collection method used is self-administered questionnaires where participants were presented with the questionnaires in persons. The purpose of the inquiry is explained and the participant is left to complete the questionnaires which were picked up latter. This method is selected as it ensures high response rate, accurate sampling and minimum of interview bias [24]. 35 Questionnaires were given out at a major local hospital, nursing homes and the general population to give the questionnaire greater exposure. Each questionnaire is assigned with random participant number which allow the participant if later decided be to withdrawn. The information regarding the survey description and intention, participant's privacy and right were marked on the first page of the questionnaire. Each participant was given a week to complete the questionnaire.

B. Questionnaire design

The core information needed for this research is the perceptions which are usefulness, effectiveness, adoption willingness and the concerns of using smart home technologies for assist elderly people. It was clear that the participants first needed to be informed of the smart home technologies and what their potential usage, therefore a detailed description of each technology is presented alongside the questions. To help non-technical participants to understand each technology better, a conceptual diagram of the usage of the technology is also provided.

TABLE 1. Question Using Likert-type Format

To what extent would the technology be useful in elder care?	Not at all Useful ←==========→ Very Useful				
	1	2	3	4	5

The questionnaire is divided into two sections: Section 1 collects participant's information such as gender, age, IT experiences, etc. The second section detailed the main questions and is divided into 2 parts: Perception of each technology and concerns of the use of smart home technology. In section two, Likert-type format is used where the participant indicates the extent to which they agree or disagree on the statement. A Five level is used where 1 represents the most negative feedback and 5 represents the most positive feedback an example is shown in Table 1. According to previous studies [6, 7, 23], privacy issues are often raised, thus the question about the extent of privacy concern was asked. Additional spaces were also provided for participants to input other factors that may cause concern on certain technology. The next part asked the participant the extent of concern of several factors such as privacy violation, lack of human responder, user-friendliness of the device, the need for training, stigmatisation resulting from using/installing the technology and the cost of technology.

IV. Result

The response rate of the survey was 51.4% and the following data is attained.

A. Participant demographic

A total of 18 participants were involved in this study. Fourteen participants were female, three were male and one participant did not specify. During the survey, it was observed that most of the carers were female including both professional i.e. nurses and non-professional i.e. partner, relatives and friends. The age range of the participants was large, ranging from 26 to over 80 years old and mostly age between 51-59 years old. 27.8% of the participants were not carers, 22.2% were professional carers and 50% were non-professional carers in which mainly were relatives of the elderly as shown in Figure 6. In general, the participants were users of personal computers (PCs), mobile phones and use computer applications such as email, web browsing, and advanced applications i.e. PhotoShop, and Excel, etc.

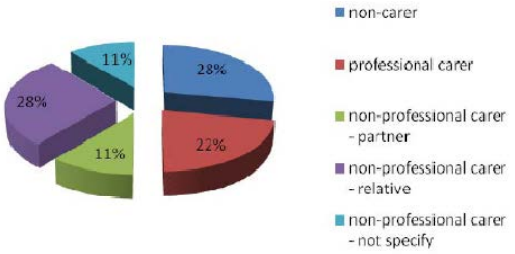


Figure 6: Categories of participants

B. Perception of smart home technologies for elderly care

In the second section of the questionnaire, the perception of six smart home technologies which are cooking hob and oven safety control, sleeping pattern monitoring, emergency alarm, automatic lighting system, video monitoring system and activity monitoring system were asked. The result is shown in Table 2.

Table 2: Participants' IT/Computer Experiences

Use personal computer/laptop	77.8%
Use mobile phone	83.3%
Use email and web browsing	77.8%
Use advanced applications	66.7%

The participants perceived most of the technologies to be useful in elderly care where cooking hob and oven safety control, sleeping pattern monitoring; emergency alarm and automatic lighting system were perceived as very useful while video monitoring and activity monitoring systems were seen as useful. Similarly, nearly all of the technologies were being seen as effective in home care expect video monitoring system which received neutral feedback. Emergency alarm and automatic lighting system were perceived as very effective. Participants strongly agreed that sleeping monitoring pattern, emergency alarm and automatic lighting system would help elderly person in a home environment. The participants agreed all technologies except cooking hob and oven safety control would help carers in nursing homes or hospital environment. Overall, emergency alarm and automatic lighting system received the best feedback while mixed feedbacks were received for video monitoring system and cooking hob and oven safety control, sleeping pattern monitoring and activity monitoring system had positive perception as shown in Table 3.

Table 3: Perceptionof Six Smart Home Technologies

Smart home technology	Usefulness	effectiveness	Help elderly	Help carer	Adoption willingness
	median	median	median	median	median
Cooking hob and oven safety control	5	4	4	3	3
Sleeping pattern monitoring	5	4	5	4	3
Emergency alarm	5	5	5	4	3
Automatic lighting system	5	5	5	4	4
Video monitoring system	4	3	4	4	2
Activity monitoring system	4	4	4	4	3
1-Most negative					5-Most positive

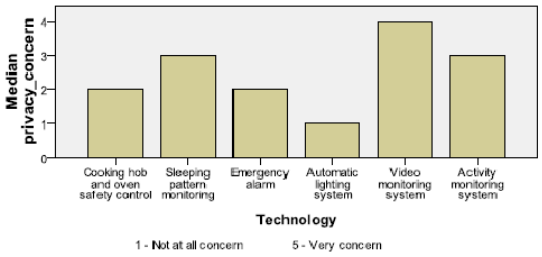


Figure 7: Privacy concern of the six technologies

Privacy concern of each technology was asked in the questionnaire, the result is depicted in Figure 7. All monitoring-based systems were perceived as violation of privacy. The privacy scale of video monitoring, activity monitoring and sleeping pattern monitoring were 4, 3 and 3, respectively, where 1 represents not at all concern and 5 represents very concern. The participants were given additional space to put further concerns of each technology. The issue was raised most about cooking hob and oven safety control was that it may cause confusion to the elderly. For the emergency alarm they were concerned of the false alarm that may cause by an unintentional press or the elderly may press the button to seek attentions. Reliability of the automatic lighting system was the issue that concerned by many participants. One participant used the system and found the system was prone to failure. Most concerns of a video monitoring system were privacy invasion as well as the fear of abuse by those monitoring.

In general, participants were willing to an automatic lighting system to use in their own residences. The decision to adopt cooking hob and oven safety control, sleeping pattern monitoring, emergency alarm and activity monitoring system were neutral. Video monitoring system, on the other hand, was not be willingly adopted by the participants, this may due to the fact that the technology seems to violate the privacy of its users.

C. Concerns regarding the use of smart home technology

Participant also rated concern factors regarding the use of smart home in general. Concerns such as privacy violation, lack of human responders, user-friendliness of the device, the need of training of new technology, stigmatisation resulting from installing/using the technology and the cost of the technology i.e. device and retrofitting were asked. The result showed that the participants were quite concerned with the user-friendliness of the device and the need of training of new technology. When this is looked with IT/computer experiences of the participants, the result surprisingly revealed that most of the participants with computer experiences rated these factors as quite a concern. Other issues were rated as concern were lacking of human responders and the cost of the technology. Unexpectedly, participants were not concerned about the stigmatisation caused by using smart home technologies as illustrated in Figure 8. Participants thought the top five activities that should be monitored were falling, cooking, walking downstairs, walking upstairs and walking, respectively.

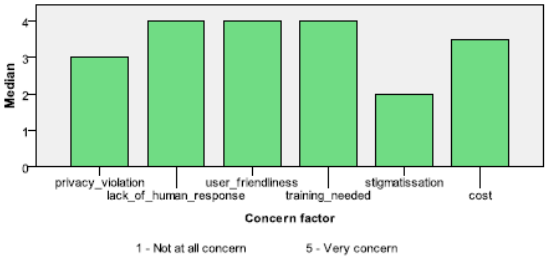


Figure 8: Concerns of the use of smart home technologies

V. CONCLUSION

Over decades the number of elderly population has increased dramatically. This affects human in various aspects especially in healthcare where admission number and healthcare cost will significantly rise. In twenty first century where sensor and network technologies have been advanced, new model of care is possible. Smart home is defined as a place that equipped with technologies that allow people to maintain living independently while in charge of their own healthcare and its cost. It will allow people to age in home, enhance people way of living i.e. safety, security, convenience, etc. In this paper the topics related with smart home technologies including several smart home projects, different type of networks that are used in smart home, smart home appliances and sensor technologies for smart home have been reviewed.

Smart home technologies show great benefits in elderly care and to successfully adopt of these technologies perception, needs and concerns must be thoroughly assessed and understood. A survey has been carried out at major local hospitals, nursing homes, and general population to explore the perception of six selected technologies such as cooking hob and oven safety control, sleeping pattern monitoring, emergency alarm, automatic lighting system, video monitoring system and activity monitoring system to assist elderly people. The result showed positive feedback toward these technologies with emergency alarm and automatic lighting system were the most popular. The adoption willingness of these technologies was still indecisive. Regarding the use of smart home technologies, factors such as lack of human responders, user friendliness of the device and the need of learning new technology were most concerned.

Smart home technologies seem to be favour in elderly care. Although now people may feel uncertainty in the adoption of these new technologies, it is believed that by giving them real

experiences of smart home devices, their usefulness and effectiveness should be more appreciated.

REFERENCES

[1] Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, 2008, World Population Prospects: The 2008 Revision.

[2] W Rachel, B., Doyle, Y., Grundy, E. and McKee, M. (2009) *How can health systems respond to population ageing?*.

[3] Johnson, R.W. and Mommaerts, C. (2010) *Will health care costs bankrupt aging boomers?*, The Urban Institute, Washington.

[4] King, N. (2003) *Smart Home – A Definition*, Housing LIN Intro Factsheet.

[5] Jiang, L., Liu, D.-Y. and Yang, B. (2004) Smart home research, *Proceeding of the Third International Conference on Machine Learning and Cybernetics*, Shanghai, 26-29 August.

[6] Demiris, G., Rantz, M.J., Aud, M.A., Marek, K.D., Tyrer, H.W., Skubic, M. and Hussam, A.A. (2004) Older adults' attitudes towards and perceptions of 'smart home' technologies: a pilot study, *Medical Informatics*, The Internet in Medicine, vol. 29, no. 2, pp. 87-94 .

[7] Demiris, G., Hensel, B.K., Skubic, M. and Rantz M. (2008) Senior residents' perceived need of and preferences for "smart home" sensor technologies, *International Journal of Technology Assessment in Health Care*, vol. 24, no. 1, pp. 120-124.

[8] Boulos, M.N.K., Lou, R.C., Anastasiou, A., Nugent, C.D., Alexandersson, J., Zimmermann, G., Cortes, U. and Young, R.C.M. (2009) Connectivity for Healthcare and Well-Being Management: Examples from Six European Projects, *Int. J. Environ. Res. Public Health* 2009, vol. *6*, pp. 1947-1971.

[9] Amoretti, M., Copelli, G., Matrella, G. Grossi, F. and Baratta, S. (2010) The PERSONA AAL Platform: Deployment in the Italian Pilot Site of Bardi, *AALIANCE Conference on Ambient Assisted Living: Technology and Innovation for Ageing Well*, Malaga, Spain, 11-12 March.

[10] Marsh, J. (2002). House calls. *Rochester Review*, *64*, 22–26.

[11] Alwan, M., Dalal, S., Mack, D., Kell, S., Turner, B., Leachtenauer, J., and Felder, R. (2006) Impact of Monitoring Technology in Assisted Living: Outcome Pilot, *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, January.

[12] Medical Automation Research Center. (2003) Smart In-Home Monitoring System [Online] Available: http://marc.med.virginia.edu/projects_smarthomemonitor.html/

[13] Darbee, P. (2005). INSTEON The detail, smart home technology, 1 August, pp. 1-64.

[14] Liu, H., Darabi, H., Banerjee P and Liu, J. (2007) Survey of Wireless Indoor Positioning Techniques and Systems, *IEEE Transactions on Systems, Man, and Cybernetics, Part* 37, no.6, pp. 1067-1080, November.

[15] Atkins, A.S., Zhang, L., Yu, H., and Miao. W. (2009) Application of Intelligent Systems Using Knowledge Hub and RFID Technology in Healthcare Waste Management in UK and China, *International Conference in e-Business* ISBN 978-989-674-006-1.

[16] Atkins, A., Zhang, L. and Yu, H. (2010) Application of RFID and Mobile technology in Tracking of Equipment for Maintenance in the Mining Industry, *2010 Underground Coal Operators' Conference*, The Australasian Institute of Mining and Metallurgy, February 2010 , pp. 350-358, ISBN: 978-1-921522-16-1.

[17] *Floerkemeier, C. and Lampe, M. (2005) Survey of Wireless Indoor Positioning Techniques and Systems, Proceedings of SOC'2005 (Smart Objects Conference)*, October, Grenoble, France, pp. 219–224.

[18] Mathie , M.J., Coster, A.C.F., Lovell, N.H. and Celler, B.G. (2004) Accelerometry: providing an integrated, practical method for long-term ambulatory monitoring of human movement, *Physiol Meas.*, 25(2).

[19] Bao, L. and Intille, S. (2004) Activity Recognition from User-Annotated Acceleration Data, *Proc. Pervasive*, pp. 1-17, Vienna, Austria.

[20] Ravi, N., Dandekar, N., Mysore, P. and Littman, M. L. (2005). Activity recognition from accelerometer data, *AAAI*, pp. 1541-1546.

[21] Diermaier, J., Neyder, K., Werner, F., Panek, P. and Zagler, W.L. (2008) Distributed Accelerometers as a Main Component in Detecting Activities of Daily Living. *Proceedings of the 11th international conference on computers Helping Prople with special needs*, July, Linz, Austria.

[22] Texas Instruments (2007) *Accelerometers and How they Work*, TI.[online] available: http://www2.usfirst.org/2005comp/Manuals/Acceler1.pdf

[23] Steele, R., Atkins, A.S. and Yu, H. (2009) Wireless Devices to Aid an Aging Population, *15th International Conference on Automation & Computing*, University of Bedfordshire, pp. 235-240, ISBN: 978-0-9555293-4-4.

[24] Oppenheim, A.N. (1992) *Questionnaire design, Interviewing and Attitude Measurement*, Pinter Publisher, London.

Flexible Middleware for Integrating Heterogeneous and Distributed RFID Systems

Pietro Carolla[1], Luca Perrotta[1], Eugenio Zimeo[2]
[1]R4I - Research for Innovation - S.r.l.
[2]Dipartimento di Ingegneria - Università del Sannio

***Abstract*— RFID technology is gaining an increasing interest in many application domains due to a more and more competitive ratio between application flexibility and cost. Although cost plays an important role to decide for the adoption ofsuch a pervasive technology in the mainstream market, the availability of flexible middleware technologies,whichcontribute to burst the creation of software infrastructures for acquiring data from physical environments,represents an important enabling factor for easilyexploiting RFID technology in many application contexts.The paper presents amiddleware for integrating heterogeneous (with more or less computing capabilities) RFID readers that can be placed in distributed locationsto form medium-large logical areas of observation in industrial buildings. The middleware provides configurable components to abstract the functions of heterogeneous readersand a publish/subscribe communication layer for filtering and distributing the events generated by tags with the aim of creating new events with application-oriented semantics. The proposed middleware has been experimented for developing a real application in the working context of a laundry. The paper reports also a performance analysis of the middleware that identifies the operation limits of the proposed infrastructure with reference to a specific configuration setting.**

Index Terms- flexible middleware; heterogeneous RFID readers integration; RFID technology

I. INTRODUCTION

Thediffusion of low-cost, miniaturized tags for radio frequency identification (RFID) is stimulating the idea of abstracting the physical environment that characterizes many application domains. This way, every physical object can be treated as a logical one by information systems, and physical and logical transactions can work together to give rise to pervasive distributed applications, directly deployed in business environments.

Nowadays, the cost of RFID (passive) tags is sufficiently low to stimulate their adoption in many application contexts (especially in substitution of older technologies, such as bar codes).

In spite of its maturity and low-cost, a significant barrier to the widespread adoption of RFID technology is the lack of flexible infrastructures able to rapidly align themselves to the physical constraints of enterprises and the changing conditions of their organizational models.

Existing middleware platforms for handling events generated by RFID technologies are designed for specific classes of readers and applications, andare often unable to adapt their functions to thoseprovided by the heterogeneous hardware thatcharacterize the readers used in complex deployments.

A canonical architecture [1] for an integration middleware for handling events generated by RFID tags is often composed of the following main components: (1) one or more physical readers connected to serial cables or to a local area network; (2) an ALE (Application Level Event) server, for filtering incoming events; (3) an integration bus to route application events towards the subscribed applications.

With this configuration, readers must be physical components with the ability to fire low-level events, whereas the ALE server often becomesa bottleneck in acquiring and filtering events to create new application-oriented ones.

To overcome these limitations, this paper presents a flexible middleware, composed of different software components that can be configured to form distributed logical readers or highly scalable ALE servers. Each reader distributed in a different physical area of an enterprise building is provided with a local filtering system composed of more or less components depending on the physical characteristics of the readers. The resulting virtual readers are connected together through a JMS bus in order to form a distributed, logical reader that aliments an ALE server, or, from a different point of view, a distributed ALE server with differentiated filtering levels.

By adopting this architecture, a logical reader can be built atop a plethora of physical readers whose generated events are filtered or aggregated to form events with additional semantics for the application. The publish/subscribe model allows for a plug&play integration of several readers but might introduce some performance limitations. A performance analysis shows that the middleware is able to handle a good throughput of events, whereas the implementation of the reference application shows the extreme flexibility of the approach.

ISBN 978- 974-672-556-9

The rest of the paper is organized as follows. Section II introduces RFID technology and the related applications. Section III presents the reference application scenario and a flexible middleware for handling a large amount of events coming from a distributed logical reader. Section IV shows how the proposed approach can be exploited to develop a real application. Section V reports the results of the performance analysis conducted over the proposed infrastructure with a specific setting of the underlying hardware. Section VI concludes the paper and introduces future work.

II. RFID TECHNOLOGY AND APPLICATIONS

A. *RFID Technology*

This section provides an introduction to Radio Frequency Identification (RFID) and to the wide variety of applications that can take advantage of this technology. RFID uses radio waves to automatically identify physical objects, either living beings or inanimate items. Therefore, RFID is a technology enabling automatic and remote identification (Auto-Id) of objects without requiring line-of-sight. This technology is opposed to other examples of Auto-Id technologies, including bar code, biometric, voice identification, and optical character recognition systems.

At the highest level of classification, RFID technology can be divided into two classes, based on whether the tag contains an integrated power supply and/or provides support for specialized tasks: passive transponders, active transponders, semi-active/semi-passive transponders.

Passive transponders do not have an on-board power source and exploits the power emitted from the reader to energize itself and transmit its stored data to the reader (during tag-to-reader communication, readers always communicate before tags) [2]. Since there is no battery or sophisticated electronics (mostly sensors), this kind of tag is really simple and has no removable parts. Therefore, it has a long life and is generally resistant to harsh environmental conditions (corrosive chemicals, high temperatures, etc).

A passive tag consists of the following main components: an antenna, a semiconductor chip attached to the antenna, and some form of encapsulation that is necessary to maintain the integrity of the tag and protect the antenna and the chip from damages. A passive tag is generally cheaper if compared to an active or semi-active tag.

Active tags have an on board power source, generally a battery but also other power sources are possible, and electronics for specialized tasks; the on board electronics could consist of microprocessors, sensors, I/O ports, powered by on board power source. Therefore, this kind of transponders could be used to measure the surrounding temperature and calculate the average temperature data (the gathered data could be used to determine other parameters such as the expiry date of the attached goods). The tag can transmit this information to a reader (along with its unique identifier) but, in this case, communication is not started by the reader, because active transponders can broadcast continuously their data to their surroundings (this kind of transponder is also called transmitter, but some other types can enter in a sleep or low-power mode in the absence of a reader).

Semi-active or semi-passive tags have on board power source and electronics for performing specialized tasks, but the battery provides energy to the tag, only for its operations, whereas for transmitting the data it uses the reader's emitted power. The reading distance ranges from a few centimeters to hundreds of meters, depending on tag type, reader power, interference from other devices and so on.

Active and semi-active tags are expensive and typically used for high-value goods or large assets that must be tracked over long distances. In contrast, passive tags are inexpensive and can be used for common materials in very large quantities.

B. *Application Areas*

Due to the technology flexibility and multiplicity of available transponders, a wide variety of applications and business contexts can take advantage of RFID technology; a possible categorization of mainstream RFID applications include security, tracking, authenticity, electronic payments, and entertainment.

RFID-based security applications concern access control (enabling mechanical keys to be replaced by an electronic card), verification, and automatic mechanisms to protect a store's merchandize from shoplifting.

As concerning tracking, there are numerous examples of large organizations that need to track the location of equipment or people to operate efficiently. This is a logistics problem that, on a large scale, can only be tackled by using automatic identification. Automatic tracking enabled by RFID can benefit supply chains, inventory control, people, hospital patients, sports events, cattle, pets, airline luggage, etc.

RFID technology can be used to obtain confidence that any manufactured item of value comes from an authentic source; for example, items of value could be banknotes (tagged banknotes may be automatically tracked between transactions and thus provide information about their use), or drugs (pharmaceuticals often have a high market-value and are therefore a target for forgery).

In the electronic payments field, RFID chips can be applied to auto tolls (by placing a suitably

designed tag in the windshield of a car a tag reader at the tollbooth can automatically scan its ID as it passes by; the appropriate fee can be deducted from the associated prepaid account each time the ID is detected), to electronic tickets, that are prepaid tokens that provide personal access to electronic credit cards.

Even in the entertainment field the invisible nature of RFID communication has been used by some toy manufactures to create toys that appear to magically take on a personality when brought near other objects [3].

C. RFID Extending Sensing

Alarge class of applications uses the ability to convey information that extends beyond an ID stored in an internal memory, and dynamically read data from an on-board sensor [4]. Some transponders can verify that critical environmental parameters remain within a safe range, and as a result can be used to ensure the integrity of perishable goods, and protect the interests of retailers and customers alike.

Typically items such as meat, fruit, and dairy products, should not exceed a critical temperature during transportation, or they may not be safe for consumption at their destination.

An RFID transponder with temperature sensor can serve to both identify and track crates of perishable goods, and ensure their critical temperature has remained within recommended parameters[5][6]. When the tag is read, it will not only respond with an ID, but also provide a warning if the temperature variation have been an issue.

Monitoring the pressure of an automobile's tires from inside the vehicle is another application that is well suited to the unique capabilities of remote RFID sensing. This is a feature that can be of benefit for drivers, as a slow leak often goes by unnoticed before the tire becomes completely flat.

Another physical parameter that can be monitored to useful effect is "acceleration"[3]. If a fragile package is dropped during transportation, it is likely a critical acceleration threshold would have been exceeded. RFID sensing can also be used to support anti-tamper product packaging.

A simple binary switch-based sensor can be incorporated into an RFID tag, such as a thin loop of wire extending from the tag through the packaging and back to the tag. If tampering occurs, the wire is broken and will show-up as a tamper bit when the tag is read during checkout. In this way, at each point in the supply chain, from factory to retail, it is possible to check individual products for tamper activity.

Advanced medical monitoring can also be supported by RFID technology. Some diagnoses can only be made with a direct access to the body's internal organs but RFID transponders can play a valuable role.

RFID sensors can be designed for beingimplanted in the human body during surgery. An external reader can then be used to periodically communicate with the device either during routine visits to the doctor, or as a result of being carried by the patient. A device such as this can provide an on-going and progressive evaluation of the condition being monitored.

III. FLEXIBLE MIDDLEWARE

In applications, such as access control systems, the RFID readers and other devices are not connected together to form a complex infrastructure. An application is mostly monolithic and the gathered data are consumed only by that application; there is a simple association (one-to-one) between the reader and the application. In more complex contexts, such as supply chain management, many readers are distributed across factories, warehouses, distribution centers,therefore RFID data must be processed and consumed by many different applications. This introduces the need for an intermediate software layer, which provides abstraction from protocols; device interfaces, actual paths and locations, and provides functionsfor configuration, management and integration.

From a development point of view, it is also important to abstract from the low-level RFID data features. In fact, RFID data possesses a unique set of characteristics that RFID applications must consider.

A. Application Scenario

In this section, we introduce both a reference scenario to identifycommon requirements foran RFID middleware and an approach to deal with some communications issues among the different nodes.

In particular, we consider the application domain of industrial laundries that provides rental and cleaning services to large institutions (hotels, restaurants, hospitals) that require a constant flow of clean linen. The application logistics (shown in Figure 11) suggests the adoption of three differentsites. One site is delegated to the "soiled side" operations; the other one is delegated to "clean side" operations; the last one is used only for administrative functions.

Each good (manufactured in the same company) is tagged with a UHF transponder, special tags that can withstand high temperatures of the ironing phase. The clean site is used also as the main distribution point and storage area. Here the packaging step takes place.

The processed linen is prepared for delivery. Individual orders (retrieved from the ERP system) are filled (a), based on the needs and requests of the laundry's customers. In this phase, called order fulfillment, we have the association between the instances of the good with the customer order; the

order is processed by the system thanks to an RFID gate (b). This association allows, in real time, for keeping track of goods and automatically updating the availability, the inventory, and other repositories of information.

In the soiled site, the institution's lines are collected by the laundry personnel and returned to the laundry facility through carts and bags. Each item crosses another RFID gate (c) and this phase allows for identifying the goods, the orders, to check the status, for verifying whether the returned order is complete, and for providing other important feedbacks to the company.

The data from the two sites (i.e. the two RFID gates) are collected and sent to a central server for processing, possibly adding some metadata, and delivering them to all the Enterprise Information Systems (EISs) through an Enterprise Service Bus (ESB).

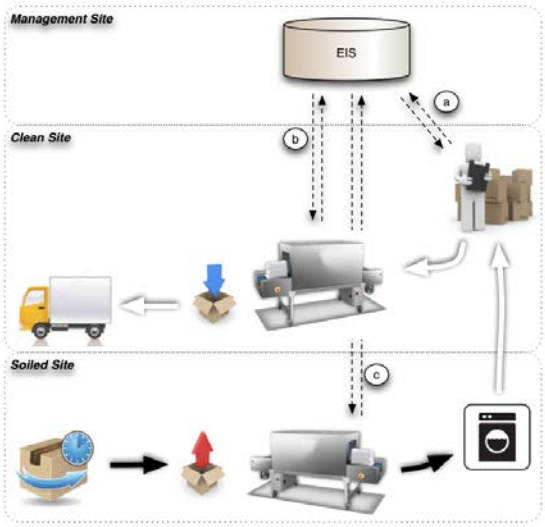


Figure 11. Inner Processes

The groundbreaking feature of the proposed solution is the possibility of a "functional" distribution of the middleware through the different nodes of the infrastructure. Each node hosts a different instance of the middleware, to create a sort of hierarchy of instances communicating with each other and each with a specific role. In this hierarchy, we can identify: some peripheral instances, known as "slave nodes", involved in the management of physical devices; some intermediate instances, known as "intermediate nodes", prevalently involved in aggregation functions, and a single (but not necessarily) "master node" with filtering, aggregation, routing and all the functions that the middleware potentially provides. Using the above terminology, we can identify, in the reference scenario, two slave instances (one for each site), and a single master instance.

Before describing the benefits of this approach, we introduce the technical solution, appropriate to create a communication channel among the different instances, which is reliable, scalable and efficient.

The proposed solution is based on the Java Messaging Service (JMS). The JMS API is a Java Message Oriented Middleware API for sending messages between two or more clients. It is a messaging standard that allows application components to create, send, receive, and read messages [7]. It allows the communication among different components of a distributed application to be loosely coupled, reliable and asynchronous. JMS is composed of the following components:

- *provider*: an implementation of the JMS interface for a Message Oriented Middleware (MOM). Providers are implemented as either a Java JMS implementation or an adapter to a non-Java MOM;
- *client*: an application or process that produces and/or receives messages;
- *producer/publisher*: a client that creates and sends messages;
- *consumer/subscriber*: a client that receives messages;
- *message*: an object that contains the data being transferred;
- *queue*: a staging area that contains messages that have been sent and are waiting to be read. Note that, differentlyfrom what the name queue suggests, messages donot have to be delivered in the order sent. If the message driven bean pool contains more than one instance then messages can be processed concurrently and thus it is possible that a later message is processed sooner than an earlier one. A JMS queue only guarantees that each message is processed only once.
- topic: a distribution mechanism forpublishing messages that are delivered to multiple subscribers.

About the two interaction models, we have a point-to-point and a publish/subscribe model:

- In the *point-to-point* or queuing model, a sender posts messages to a particular queue and a receiver reads messages from the queue. Here, the sender knows the destination of the message and posts the message directly to the receiver's queue. Therefore, only one consumer gets the message, and the producer does not have to be running at the time the consumer processes the message, nor does the consumer needs to be running at the time the

message is sent. Every message successfully processed is acknowledged by the consumer;

- The *publish/subscribe* model supports publishing messages to a particular message topic. Subscribers may register interest in receiving messages on a particular message topic. In this model, neither the publisher nor the subscriber knows about each other. A good analogy for this is an anonymous bulletin board. Therefore, multiple consumers (or none) will be able to receive the message. There is a timing dependency between publishers and subscribers. A publisher has to create a subscription for clients to subscribe. On the other hand, a subscriber must remain continuously active to receive messages, unless it has established a durable subscription. In that case, messages published while the subscriber is not connected will be redelivered whenever it reconnects.

JMS has been used as a connector among the different components used for acquiring data from the environment. We used Apache ActiveMQ as open source messaging and integration patterns provider [8].A JMS provider is integrated in each "slave node", which isa peripheral node involved in the management of physical devices. A component of the middleware collects data from the readers and sends,through a specific connector,data to the JMS Provider; therefore the component acts as JMS Client with the role of Producer. On the other hand, intermediate nodes and the master act as subscribers.

The publish/subscribe interaction model is adopted: the sender (slave node) publishes a message to the topic on its JMS provider; the master/intermediate node at the start-up phase registers interest in receiving messages on that particular message topic; the subscription of the subscribers is always durable so the subscriber has not to remain continuously active and the messages published while the subscriber is not connected will be redistributed whenever it reconnects (data are stored in a local embedded database). The designed platform is robust and can be used in critical environments where reliability, availability, and data integrity is required.

B. Benefits

The designed architecture leads to some benefits that go beyond the specific features of the Java Messaging Service, embracing the needs of designers and developers of complex RFID infrastructures.

The possibility of a "functional" distribution gives the opportunity to use the proposed middleware in physically distributed environments. For example, we can consider large companies that have made a de-localization of production with different sites, factories, warehouses and distribution centers. Our approach allows for abstracting from the actual hierarchy of nodes and the underlying infrastructure; the master instance sees the other middleware instances as local devices (this is true for each node and the nodes directly linked to it). This simplifies the connection between the entire infrastructure and the application layer that processes the business logic.

The core of the middleware is really lightweight; the ability to deploy only some functions gives the opportunity to use the middleware on low-end computers. So a mini-pc, with the middleware on board, can be used to extend a RFID device with limited computing resources. Many cheap readers have a really good RF section but can be considered dummy devices because they have no filtering, aggregation or routing functionalities.

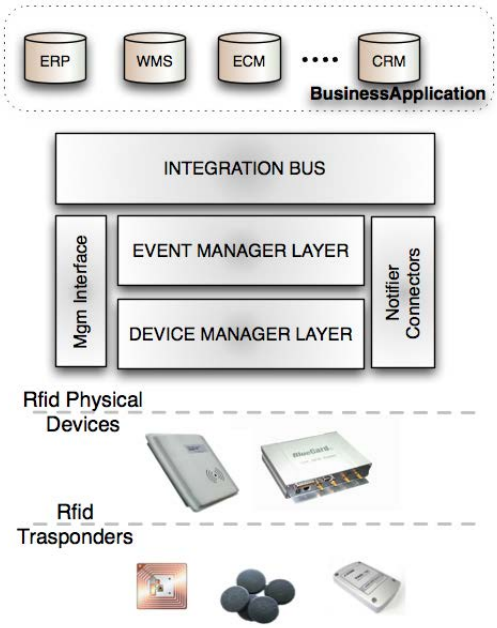


Figure 2 2. Multi-layer Architecture

The middleware can provide all these functions reducing the costs and the hardware complexity. Moreover, on the market of RFID readers, several devices are now available with a mini-pc embedded (e.g. the Favite GF-801); the combination of both, a reader and an instance of middleware, can provide with a low cost and complete unit with features of buffering, persistence, reliability, processing logic, filtering and routing. The same device can be used on a wider array of applications.

The proposed middleware solution consists of three logical layers, a *Device Management Layer*(D.M.L.), an *Event Management Layer*(E.M.L.) and an *Integration BusLayer*(I.B.L.). To better understand the concept of functional distribution introduced at the beginning of this section, we introduce briefly the features of each layer.

The *Device Management Layer* is the software layer responsible for managing the most peripheral tier of a RFID infrastructure. RFID readers can be combined with other input devices (such as temperature sensors, proximity sensors, etc.) to create a complex physical network. Our middleware currently supports a variety of devices from different manufactures. The Device Manager contains, for each device, a different protocol manager, which is the software module that manages the low-level communication protocol used by a particular device. The D.M.L. exports, to the upper layers, an interface with the functionsfor configuration, remote management, monitoring and a basic filtering level (e.g. the Persistence Time of each device).

The E.M.L. defines the interface through which other modules can obtain data (first of all Electronic Product Code - EPC), possibly filtered, from a large variety of devices.The design of this interface meets the needs of any data processing system, for an intermediate software level useful to reduce the large amount of data and transform them in business events.

The Event Manager thus provides:

- Functions to capture data from one or more sources, such as readers and devices;
- Functions to accumulate data in a specified time interval, to eliminate redundant data, to aggregate data according to a specified policy;
- Functions to combine data in a report that is sent to subscribers according to a predetermined schedule (this is a low-level mechanism of routing and reporting that uses *notifiers* for the common protocols).

The I.B.L. provides with an abstraction on top of the implementation that offersimportant services for complex architectures through an event-driven and standards-based messaging engine. It allows developers and integration architects to exploit the value and opportunities of messaging without writing code. In contrast to the monolithic hub-and-spoke architecture of an integration broker, the foundational core of an ESB provides a distributed services architecture.

This architecture has the ability for integration broker functions, such as message routing, data transformation, and application adapters, to be selectively deployed on an as-needed basis [9]. Thus, the concept of functional distribution underlines that, according to the specific role played by a node in the overall infrastructure, the middleware can activate one or more layers.

Figure **33**3 highlights, in dark grey, the inactive layers, in a possible hierarchy of middleware instances. In the slave nodes, the middleware enables the Device Manager (in particular only the plug-in driver for the specific reader connected) and the *notifier* module to the JMS Provider deployed on the same node (as described in the previous section, the slave node is also a JMS publisher); possibly, filtering, pattern matching, aggregation and routing functions are unnecessary,as a consequence the Event Manager and the Integration Bus can be disabled. An intermediate node (a typical configuration includes more than one intermediate node) enables the Device Manager but the plug-in driver is specific to connect it to another middleware instance; thus the middleware instantiates a JMS subscriber to the providers deployed on the remote slave nodes; but the role of the intermediate node is also to send data to its own provider (to make data available to the master node), so the node is both subscriber and publisher; if an intermediate node is connected to many slave nodes, a part of the E.M.L. could be enabled (mostly the aggregation functions).

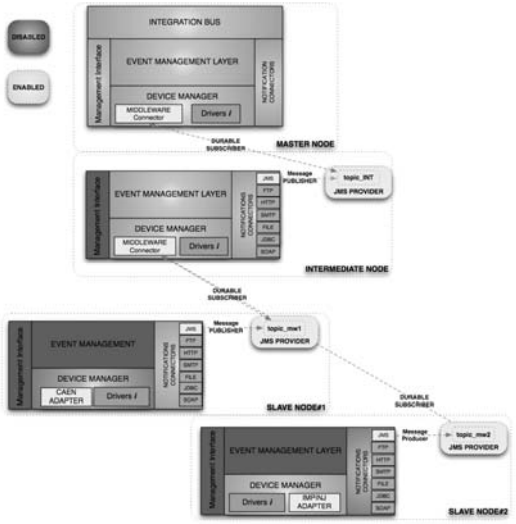


Figure 33. Hierarchy of middleware instances

The master node gathers data from the intermediate instances through the same plug-in that connects the intermediate to the slave nodes. The data are processed and filtered in the E.M.L. and the generated events are propagated to the enterprise applications through the Integration Bus.

IV. REAL APPLICATIONS

In order to illustrate the effectiveness of our middleware in rapid developing applications, we discuss in this sectionhow a real RFID application can be designed and developed. The application implements the scenario described in Section III.

A. Industrial Laundry Management

The application described in the following is related to the management of an industrial laundry that aims at tracking linens. According to the infrastructure described in the previous sections,the proposed middleware is deployed in three sites. An instance is placed in the soiled site and is deployed

on a computer to manage an RFID gate,whose hardware is based on an Impinj's UHF Speedway Reader[10].

In the clean site, an instance is embedded on a Favite FS-GF801[11]reader. It is equipped with Intel ATOM N270 CPU, and memory size of one GB of RAMto meets the needs for a smaller gate with a slower flow of linens.The third instance is deployed on a server placed in the administration site. On the same server, an instance of OpenBravo, a web-based OpenSource Enterprise Resource Planning (ERP) for SMB, is deployed.

The ICT infrastructure includes another enterprise application; a custom application developed for the specific needs of the customer.It is used in the two production sites to handle the orders fulfillment, the clean and the soiled retrieval operations.

We underline a crucial aspect that the middleware has to deal with the two applications require different latencies. The custom application needs to reply immediately to local interactions generated when the linens cross the gates. The ERP, on the contrary, does not need necessarilyshort notification latency; enterprise and legacy applications are not designed to handle streaming data and might need to receive batched updates[12].

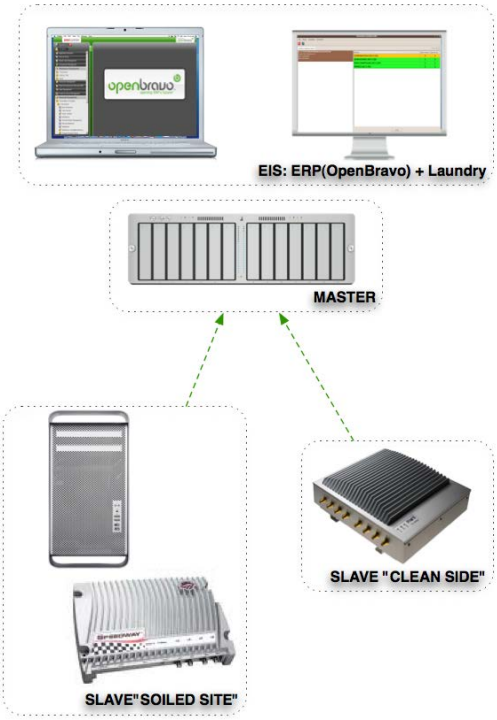


Figure 44. Deployment Diagram

The proposed solution addresses this entire communication issues.The deployment diagram is shown in Figure 44. It shows two slave nodes that are the middleware instances placed in the production sites. The role of these nodes is to manage the physical devices (the two readers), to detect the transit of tags through the gates, to notify data to the master node and to the custom application. The application provides, in real time, feedbacks to the laundry personnel on the correctness of processed goods and activities.

Figure 5 5. An example of a Soiled Retrieval **Operation** shows an example of a soiled retrieval operation; the left column reports customer information and the order details; the right column shows the list of the returned items.

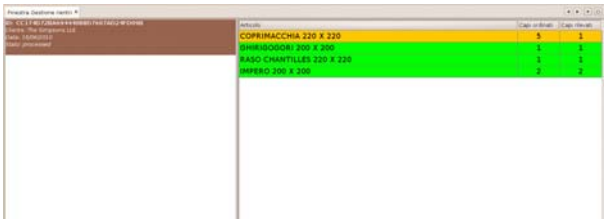


Figure 5 5. An example of a Soiled Retrieval Operation

The third node is a master unit that gathers data from the two slaves. The filtering layer processes the incoming streaming data, normalizing the messages, adding some metadata about their sources. Data are sent to the Enterprise Service Bus and subsequently, by using the associated information, they are delivered to the EISs.

V. Performance Analysis

This section presents a performance analysis of the messaging sub-system of our proposed flexible middleware. The analysis is focused on the features and functions so far discussed, thus the aim is to underline the behavior of the middleware in a distributed context.

A. Test Methodology

We have performed some teststo analyze the performances of the proposed middleware under load, using the management console provided by the middleware itself. Our goal is to evaluate the performance from a designer standpoint. Therefore the test scenarios are stress level conditions for real applications, according to the unique characteristics and capabilities of an RFID middleware [13]:

- Streaming;
- Reactivity;
- Integration.

We have considered and tested the basic pattern/component of a hierarchy of middleware instances; thus,using the well-known terminology, the test cases consist of a scenario with a slave instance and a master instance; the second one has two slave instances and even a single master instance. The performance results and the considerations made for these two simple cases are sufficient to predict the performance of a more complex infrastructure.

On the slave node, we used a simulator of the physical reader to ensure an adequate workload; the amount of resources required by the simulator is very similar to the required resources of a real attached reader with a simple communication protocol, but we can set a specific value of read rate. In order to generate the highest amount of data, (tag_ids) filtering is disabled on both sides. We use the master node to monitor the correctness and the read rate of both instances. Thus, we considered three parameters, the *tag rate generated*, the *tag rate received*, and a parameter called *duration*. The first two parameters are self explanatory; the third one is defined by EPCglobal within the Application Level Event Specification and characterizes an event cycle; it is a time interval used to accumulate reads [14];at the end of this interval a report with all the gathered tag_ids is ready to be notified to the master. Increasing the time interval, the size of the report grows proportionally.

B. Test plan

For each test configuration, we have fixed the tag rate parameter (the tag rate generated by a slave) and varied the duration time. Growing the size of the report (then the size of the exchanged message), more resources are required by the communication system to ensure real-time notification, redundancy and a once-and-only-once guarantee of message delivery. Through the master node, and the processed tag rate in input, we can verify the behaviour and the correctness of the communication.

Each configuration consists of twenty-five test cases. Each test case is executed for a total of thirty minutes and is run three time to get an average value. All performance results are recorded after the connection between slave and master has been created; the two middleware instances were restarted between each test. During the tests, there were not other applications running and the system and the nodes were connected throuh a dedicated 100Mbps Ethernet nework.

C. Hardware Test-bed Equipment

During the tests, the following hardware configurations was used.

Master node configuration:

Intel Core2 Quad Q9300, 2.50GHz Yorkfield M1
Cache L1 4 x 64kb (32 kb code, 32kb data)
Cache L2 2 x 3072kb (On-Die, ATC, Full-Speed)
Chipset Intel P35/G33/G31
5 GB DDR2-SDRAM PC2-6400 (DDR2-800)
Intel(R) 82801 HR/IR/JR(ICH8R/ICH9R/ICH10R)
Disk WesternDigital WD10EADS-65L5B1 1TB 7200RPM
Average seek time 8,9 m/s
Interface SATA-II
Data transfer rate 300 MB/s
Buffer 32MB
Realtek RTL8168/8111C PCI-E Gigabit Ethernet Adapter
Connection Speed 100Mbps
MTU 1500 byte
S.O.: Windows Vista Home SP 2
JVM: Sun JRE 1.6.0.21

Slave node configuration:

DualCore Intel Pentium E2160, 1800MHz Conroe-1M
Cache L1 32kb per core
Cache L2 1Mb (On-Die, ECC, ASC, Full-Speed)
Chipset Intel Broadwater i946GZ
1 GB DDR2-667 DDR2 SDRAM
Controller IDE Intel(R) 82801GB/GR/GH(ICH7 Family) o Disco SAMSUNG HD161HJ – Spin Point 160GB 7200RP
Average seek time 8.9m/s
Interface SATA-II
Data transfer rate 300MB/s
Buffer 8MB
Controller LAN Intel(R) PRO/100 VE Network Connection
Connection Speed 100Mbps
MTU 1500 byte
S.O.: Ubuntu Desktop 9.04 32 bit
JVM: Sun JRE 1.6.0.17

D. Performance Measurement Results

The first test scenario evaluates the actual capacity of message processing of the slave node. Analyzing the results of the first test scenario, as described earlier, we can observe at high values of tag rate and notification time, a general degradation in performance.

For very low values of the duration parameter, the slave is able to process correctly 3000 tags per second, but as shown in Figure 66, some problems arise with a tag rate of 2000 tags per second and a time of notification of 4 seconds. The growing size of the generated report causes a slow down and a delay in the delivery of the messages.

The latter scenario, instead, allows us to assess the limits of the master node. Using two slave nodes we should expect a tag rate, received by the master, twice the rate actually generated by the a single slave node. The results shown in Figure **77** confirm this hypothesis. We can observe that with a value of both slave nodes of 3000 tags per second, thus a total tag rate of 6000 tags/s in input, the processing capacity of the master is saturated, obtaining a maximum tag rate of 5000 tags/s. The machine that gathers data from the intermediate or slave instances, should be a very high performance machine to ensure optimal performance with this high workloads. For this reason a master machine, with hardware characteristics significantly better than the machine used for the slave nodes, was used.

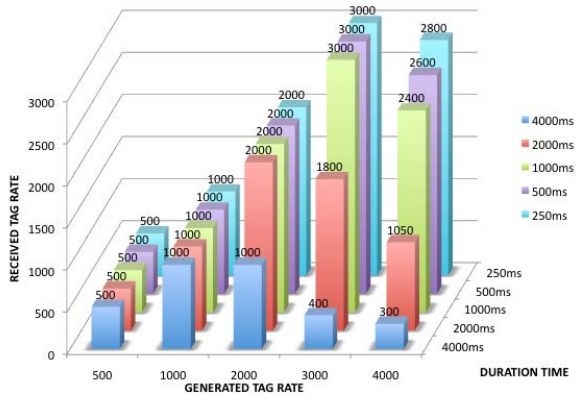


Figure 66. Performance Results Single Slave Scenario

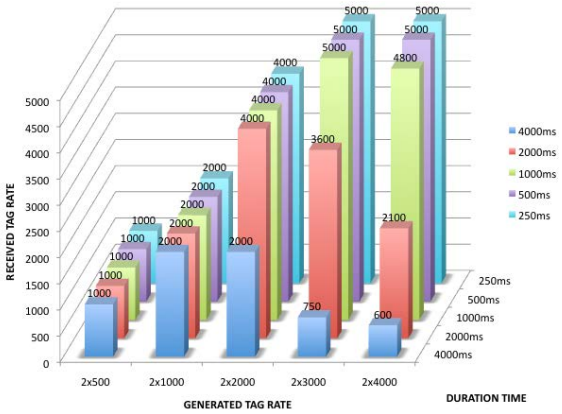


Figure 77. Performance Results Two Slaves Scenario

Since vertical scalability, obtained with high-end machines for the master, is a costly solution especially when the number of slaves and the number of events generated grows, a more effective orizontal scalability approach can be exopoited. To this end, a highly-scalable configuration is possible thanks to the adoption of a publish/subscribe bus:when the master capacity is saturated by a high throughput, an additional master can be considered to extract events from abus topic. This makes it possible also the creation of a more realiable architecture, since the single point of failure of the architecture described and tested is eliminated.

VI. CONCLUSION

The paper has presented a flexible middleware for handling complex configurations of RFID systems. The main features of the proposed middleware are *flexibility* and *high performances*. The former is derived fromsome configurable middleware components deployed directly on the readers or on mini-PCs connected to the readers through RS232 or Ethernet links. The latter is possible thanks to the filtering action that can be enabled in the middleware components working near the readers (slaves), or to the replication of the master node that is connected to the higher level integration bus. Moreover, the reliability levels enabled by the JMS bus ensure a high robustness of the entire infrastructure, since messages are not lost,evenin case of transient disconnections.

Future work will regard a more automatic configuration (adaptation) of the components on the basis of the available features of the underlying hardware and the global settings of the systems. Moreover, the overall system will be improved as concerning the throughput and load balancing among several replicas of the master nodes.

ACKNOWLEDGMENT

This work is framed withinthe Mware project funded by Campania Region under misura 3.17 - metadistretto ICT. We thank to Raffaele Tretola and Antonello Parente for their contributions in the development of the middleware reported in this paper.

REFERENCES

[1] EPCGlobal, “The EPCGlobal Architecture Framework”, Available: http://www.epcglobalinc.org/
[2] Sandi Lahiri, “RFID Sourcebook,” Prentice Hall, 2005
[3] R.Want, “RFID Explained: A Primer on Radio Frequency Identification Technologies”
[4] R.Want, “Enabling ubiquitous sensing with RFID,” IEEE Comput.,vol. 37, no. 4, pp 84-86, April 2004
[5] Microsensys, Temperature Sensing and Logging, www.microsensys.com
[6] K. Opasjumruskit, et al., “Self-powered wirelss temperature sensors exploit RFID technology,” IEEE Pervasive Comput, pp. 54-61, Jan.-March 2006
[7] Sun Microsystem, “Java Messaging Service Specification”
[8] Apache ActiveMQ, http://activemq.apache.org
[9] Chappel, Dave. "ESB Myth Busters: 10 Enterprise Service Bus Myths Debunked. Clarity of Definition for a Growing Phenomenon".
[10] Impinj,Inc. http://www.impinj.com/
[11] Favite,http://www.favite.com/
[12] C.Floerkemeier, C.Roduner, M.Lampe. “RFID Application Development with the Accada Middleware Platform”,IEEE Systems Journal
[13] Zongwei Luo, Ed Wong, S.C. Cheung, Lionel M. Ni and W.K. Chan, RFID Middleware Benchmarking, in Proceedings of the 3rd RFID Academic Convocationin conjunction with the China International RFID Technology Development Conference & Exposition, Shanghai, China, Oct 26-28, 2006
[14] EPCGlobal, “The Applicaton Level Events (ALE) Specification” Available: http://www.epcglobalinc.org/

An Airport Baggage Handling System Using RFID Technology: A Proposed Architecture

Md. Monzur Morshed, Anthony S Atkins, Hongnian Yu and Md. Mostofa Akbar[1]
Faculty of Computing, Engineering, and Technology, Staffordshire University,
Beaconside, Stafford, UK
[1]Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
mw008875@staffs.ac.uk, a.s.atkins@staffs.ac.uk, h.yu@staffs.ac.uk, mostofa@cse.buet.ac.bd

***Abstract* – This paper reports the current results on an ongoing project. We propose a new architecture for airport baggage handling using a combination of technologies such as RFID, the internet, networks, web and mobile communications. In this paper we identify the problems associated with handling baggage in airport and propose a solution to overcome those problems using RFID technology. We also propose some novel ideas to use popular technologies such as web, SMS and interactive television screens for baggage handling systems using RFID technology. We also divided the total baggage path into several zones to track them correctly and to identify where baggage was lost or mishandled if possible.**

Index Terms – airport baggage handling system; proposed architecture; RFID technology

I. INTRODUCTION

Today, airlines are facing a number of challenges. The increasing number of air passengers and hence the increase in checked-in baggage is stressing the world's baggage handling systems. In this process many pieces of baggage are mishandled and lost. The cost of a mishandled bag is a big concern for the airlines, and the problem is growing day by day. In addition to the penalty incurred by the airlines for the mishandled bag, it is also a source of great inconvenience and dissatisfaction for the passenger.

RFID can enhance the automation of baggage handling and can significantly reduce the number of mishandled bags. In the recent years, many airports have initiated the implementation of RFID technology in the aviation industry replacing barcode for the automation of baggage handling and to reduce the mishandling of baggage. The growing number of air passengers and their baggage are creating a big challenge for the airline authorities. Major increases in the reported incidents of mishandled baggage give evidence of this challenge. Increased safety regulations in airports, growing passenger numbers and tight turnaround times are some of the main cause for baggage mishandling [1]. Passengers on U.S. airlines reported more than 4 million mishandled bags in the year 2006 [2]. In Europe, the Association of European Airlines (AEA) also reported that the incidence of mishandled baggage has increased by 1.2 million which is about 14.6 percent more than the previous year. The cost of a mishandled bag is also increasing rapidly. In 2006, approximately 34.3 million bags were mishandled globally, costing the airline industry $3.8B [3]. Giving the importance of missing baggage in airports, the Association of European Airlines (AEA) published a report [1] on 27 major airlines that reported baggage information to AEA. AEA identified the 7 or 8 most affected airlines each year as an "Above average" group in terms of missing bags per 1000 passengers (Figure 1).

Among them were Air France, Alitalia, British Airways, KLM, Lufthansa and TAP Air Portugal. The 'Above average' group has an average in itself that can be as high as 27% (in 2007) compared to the average of all AEA companies. Sometimes anyone reason may cause poorer performance and can cause additional trouble to an airline. TAP airline is an example for this that shows very high rates of bags missing in 2007 due to the extremely busy and exhausted Lisbon Airport. Due to these localized problems, airport baggage mishandling rates may differ significantly among the different airlines as can be seen in the Figure 1.

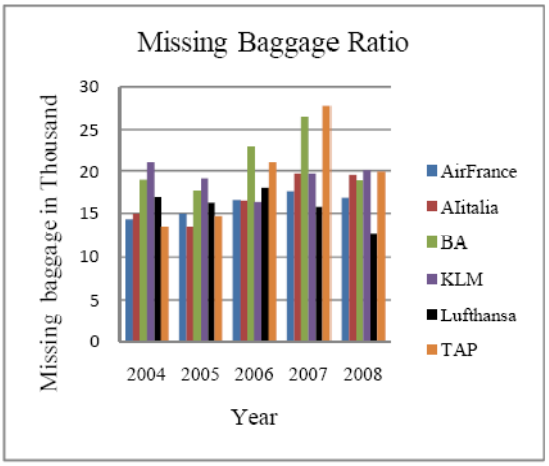


Figure 1. Missing bags per 1000 passengers: airlines with highest rates (Source AEA Consumer Reports, 2004-2009, Compiled by Authors)

ISBN 978- 974-672-556-9

According to the International Airport Transport Association (IATA) survey there are six key issues that are identified for baggage mishandling in air travel [3]. These are shown in Table 1.

TABLE 1. SIX KEY ISSUES IDENTIFIEDFOR BAGGAGEMISHANDLING

Reasons for mishandling	Rate(%)
Transfer bag: late arrival	30
Transfer bag: delay in moving bag	18
Missing baggage sortation message	11
Error at check-in	10
Poor barcode read rate	10
Transfer passenger not checked in	10
Other	11

Source: IATA, RFID business case for baggage tagging, 2007

Most of the major causes can be reduced using RFID technology. The logical characteristics of baggage handling systems are well suited to RFID, which can help make the baggage handling process easier and efficient.

II. RFID TECHNOLOGY

Radio Frequency Identification (RFID) is a technology to identify objects or people automatically. An RFID system typically consists of three components: readers, tags and the back-end database [1].

1) *RFID Reader* (Transceiver): RFID readers with antennas are devices used to read or write data from or to RFID tags. The reader queries a tag to obtain information from the tag.

2) *RFID Tag* (Transponder): An RFID tag is a small microchip designed for wireless transmission. It is attached to designated objects where each tag has a small size of memory to store its unique ID as well as other information. An RFID tag transmits data in the air in response to the interrogation by an RFID reader.

3) *Back-end database:* The main goal of the back-end database is to process data from tags collected by the readers deployed in the RFID infrastructure [5]. The RFID system virtually creates a remote database which travels with the items [6] to exchange information between tags and back-end database.

There are two types of tags: passive and active. In general, passive tags are inexpensive. They have no on-board power; they get power from the signal of the interrogating reader. The main goal of EPC was to make an RFID tag as simple as possible, with the aim of reducing the tag price below 5 cents [7]. Passive tags can operate in many frequency bands and have a read range from two feet to several meters. Although different RFID systems have been in use, popular RFID refers to the Electronic product code. For UHF, the dominant standard will be very likely Class-1 Gen-2. In this standard, tags contain a kill self-destruct feature. When an EPC tag successfully receives a kill command, it becomes inoperable [8].

Active tags contain batteries. They can initiate communications and have read ranges of 100 meters or more. Active tags are expensive, costing some $20 or more. They may also use useful frequencies that may interfere with existing systems and hence are not suitable for large scale implementation in business enterprises

III. RFID IN BUSINESS DOMAINS

Passive tags are low-cost, have no on-board power and work in unused frequency ranges which are suitable for implementation in business enterprises and also in the aviation industry. RFID is widely used in many business areas [9]. It has a significant benefit to supply chain management due to its low cost and flexibility. RFID has comparative advantages over other technologies such as barcodes, because it is, for example, contact-less, has multi-object recognition, does not require line of-sight, and as long-distance reading capabilities. The demand for RFID is increasing day by day. In [10] Harrop points out that the RFID industry will increase by US $2.8 billion in 2006 to $26 in 2016. Large organizations like Wal-Mart, Procter and Gamble, and the United States Department of Defence are employing RFID tags for automated oversight of their supply chains [11]. The U.K. retailer Marks and Spencer has also initiated the use of RFID tagging of individual items of apparel [12]. The potential offered by RFID technology is that all existing physical objects can be managed by a virtual world created by a distributed database in a distributed networked RFID system.

IV. CURRENT USAGEOF RFID IN AVIATION FOR BAGGAGE CONTROL

A number of airports have started to use RFID on a trial basis to handle baggage [1]. Hong Kong International Airport is the first airport in the world to have implemented RFID baggage tagging which has been fully functional since 2005. Bag-tag read success rates were improved to 95% with estimated cost savings of US$ 3.8 million. Later, the read rate increased to more than 97%, much higher than the barcode rate [13]. Beijing, Narita and six other Korean airports have also undertaken successful RFID trials. McCarran Airport in Los Angeles was the first airport in USA to adopt RFID tags in baggage handling. Amsterdam Schiphol Airport was the first airport in Europe to make a large scale attempt to introduce RFID baggage control in 2007. IATA has published some results based on the trials

that have been conducted between airports, airlines and manufacturers during the last few years [17]. The main objective of the trials was to test the read rate success. The results are summarized in Table 2.

TABLE 2. RFID TRIAL READ RATE*

Trials	Date	Read-Rate
Kuala Lumpur Airport	2005-2006	100% (Gen 2) 98% (Gen 1)
Kansai-Hong Kong Airport	2005	95.4% 98.78%
Asiana- Korean Airport Corporation	2004-2005	97.00%
TSA World-wide Trial	2004-2005	~99%
Narita Airport	2004	-
British Airways at Heathrow T1	1999	96.40%

Note * Compiled by Authors (See [17])

The objective of the trial in Kuala Lumpur International Airport was to study the characteristics of UHF tags placed on the test baggage in various situations to identify when reading would become difficult, to study the recognition rate by placing the UHF tags on passenger's baggage in the actual airport environment, and to verify the effectiveness of the baggage tags during operations between airports and the effects on the UHF band by the airport facility materials. In this trial a large amount of RFID materials have been tested and performances analyzed in detail.

The objective of the trial in Kansai Airport- Hong Kong Airport was to carry out a basic performance validation in an operational environment different from Narita Airport and

- to verify the international interoperability of Japan's UHF-band airline baggage tag to confirm the data recognition at Kansai International Airport of the airport baggage tags that were attached in Hong Kong
- to verify the UHF band radio frequency characteristics
- to verify the electric intensity measurement inside the airport.

The trial by the Asiana-Korean Airport Corporation used six Korean airports. Tags for baggage were issued at check-in and were read at a number of points in the baggage process. RFID was used to track the baggage through security, reconciliation and finally to verify its arrival. When the baggage arrived in the claiming section, the passenger received a Short Message Service (SMS) about the location of the baggage. Passengers could also see the information on the Flight Information Display System (FIDS). The trial also focused many new processes. RFID systems enabled the creation of a link between the security screening station and the airline security database, allowing the passenger owning a piece of reported baggage to be identified and security staff notified. This baggage could then be manually searched after an x-ray. RFID was also used for the enhancement of the manual sortation process. RFID systems also gave the baggage loaders the right flight information for the baggage to be loaded. This was a successful trial with a read rate of 97%. The trial also showed the way in which RFID can be used to enhance processes and improve customer service.

The aim of the Transportation Security Administration's (TSA) world-wide trial was to demonstrate the interoperability of UHF RFID Baggage Tag systems between worldwide geographic regions having different UHF transmission regulations.

V. BAGGAGE HANDLING USING RFID: PROPOSED ARCHITECTURE

RFID can help the baggage handling system in aviation in many ways. It can enhance the automation of baggage handling and can simplify the baggage handling system [3]. It can also identify the place and people correctly related to this baggage thus it can significantly reduce the number of mishandled bags. The flowchart of the baggage handling process is given below (Figure2).

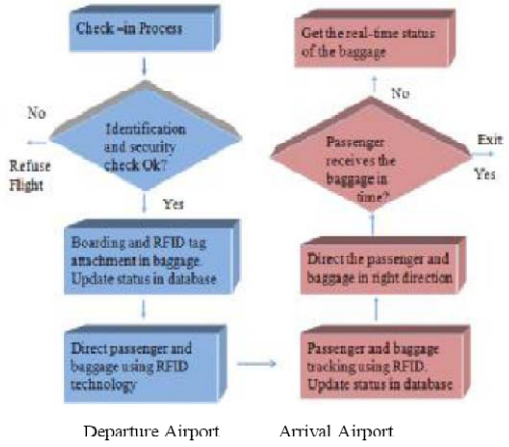


Figure 2. Logical flow of baggage in airport

A. Making zones for identification

During air travel, baggage goes through various paths and places, and it is not possible to track the baggage at every point. For this reason we need to identify various key points of the route. To identify the key points and track the baggage, the area passed by the baggage will be divided into several zones to determine its location, where it is kept or where it may be mishandled. The total area for the baggage can be classified into the following zones [3] at the departure airport (Figure 3):

1. Check-in area
2. Conveyor
3. Distribution area
4. Trolley

Each zone will be implemented with an RFID system in such a way that any baggage passing through will be tracked correctly. The incoming, outgoing and duration of stay of the tags can be determined from the system. All the data will be recorded in the back-end database.

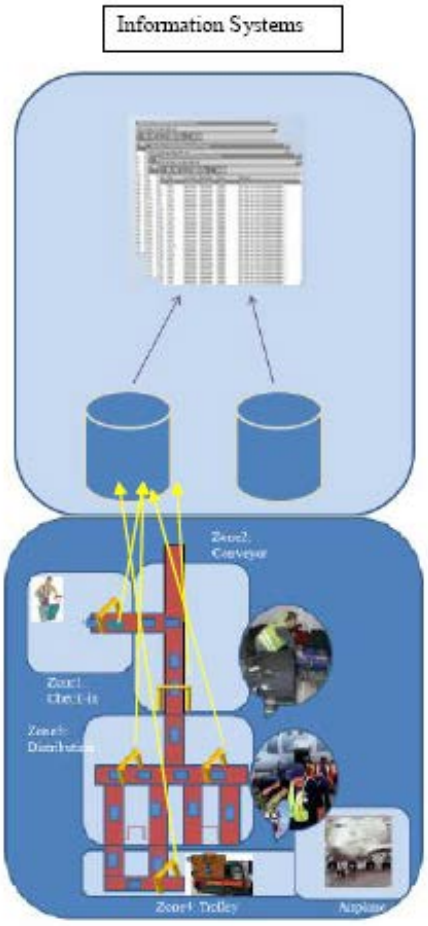


Figure 3. Baggage in different zones

Each member of staffs will carry an RFID tag in his batch and an identity card. Responsibility will be distributed between different zones and employees of each zone will be responsible for their respective zone. So, in each zone the staff will be identified by RFID tag. During duty time, staffs are required to keep their identity card with them and they should wear a uniform with an RFID batch. Staffs are allocated in each zone for different time schedules. In each zone there are RFID readers to scan the RFID tag of the baggage and the staff. The staff will also be responsible for handling and controlling the RFID readers. The total process can be described in several phases of zones:

1. *Check-in area:* In the check-in areas, RFID printers enable the instant encoding of an RFID tag including required information such as flight information and any specific sortation directions. Before placing the bag on the conveyor an RFID reader captures the tag information and stores in the middleware database.
2. *Conveyor:* As baggage moves to the conveyor belt an RFID reader can track the bag through the RFID tag and capture the available information from it, providing the information required to ensure delivery of the right bag to the right gate, right airline and right flight.
3. *Distribution area:* Throughout the routing area, FID readers capture the location of bags at key check points that help to identify the baggage and verify if it is moving in the right direction at real time to ensure timely and correct delivery to the airline.
4. *Trolley:* Here also, RFID readers can capture the information from the baggage to check that it is loaded onto the right trolley for delivery to the right plane. If any bag is mistakenly placed on the trolley it can easily be identified before it leaves, thus saving time.

The zones in the arrival airport are:

1. *Trolley:* Here baggage is loaded from the airplane and shifted to the right distribution area using the help of the RFID system. The arrival information will then be updated to the arrival airport database.
2. *Distribution area:* In the distribution area the baggage will be sorted and distributed to the right conveyor using RFID system. Baggage loaders will be informed if any wrong placements of baggages are done.
3. *Conveyor:* When the baggages are forwarded to conveyor it also takes the help of RFID system to move in a right direction. Databases will be automatically updated with the latest status.
4. *Reclaiming belt area:* Passengers will receive an MS to their mobile phones so that they are sure about their baggage arrival. Passengers can also search in the internet about their baggage with a Information Systems password. If the baggage arrives correctly then systems will automatically send the SMS to the passenger's mobile informing them of the location of their baggage. Also, an email will be sent to the passenger with details of the baggage status. Airport authority will provide computers with internet and browsing facilities near the baggage receiving section. Passengers can also get information from the baggage query section of the information desk. The baggage query section can find information on the baggage from their database or online internet service. The scenario is shown in Figure 4.

ISBN 978- 974-672-556-9

Figure 4. Passengers receive information in various ways

VI. UTILIZATIONOF RFID IN AIRPORT BAGGAGE HANDLING

IATA in [15] reported that the major reason why baggage is mislaid is due to the problem of reading barcodes properly on the baggage. RFID technology can offer a number of significant advantages over barcode solutions.

1. *Flexibility*: RFID does not require line of sight reading like barcodes.
2. *Ability:* RFID can scan multiple bags simultaneously and distinguish from other items; it identifies the object as opposed to bar code, which is used for a single scan at a time.
3. *Reliability and accuracy*: RFID is suitable for a fully automated system and can read reliably with up to 100% accuracy [3].

RFID technology can offer many opportunities in luggage handling in the airport. Here we describe few scenarios and solutions using RFID technology.

1. *Real time tracking and management*. RFID can ensure real time information about the baggage. The current location of the tag is always available on the database. The passenger can get information on his baggage anytime and from anywhere. Using RFID, it is possible to identify exactly which baggage is in which container and in which place. Checking the baggage against the passenger 'aboard aircraft' status, and locating the container which holds the passenger baggage is very important from the point of view of both security and operational efficiency [14].
2. *Cross checking*. Cross checking is very easy and efficient using RFID technology. The movement of both the baggage and the passenger can be tracked. It can easily be checked if a passenger is on the board but the baggage is not and vice versa.
3. *Digital imagery*. It is now possible by using an RFID tag attached to an identity card and air ticket to identify who is travelling. An X-Ray picture of the suitcase with RFID tag can identify what is being carried by different passengers (Figure 5).

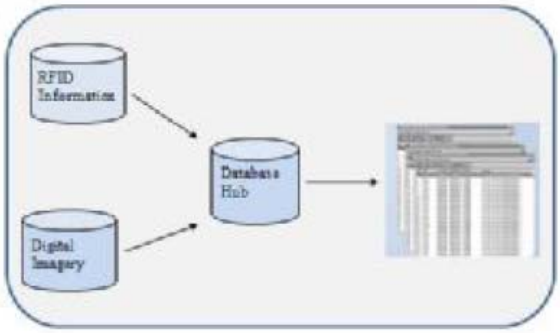


Figure 5. RFID information with Digital image.

4. *Safety against new crimes*. In this era of RFID technology different types of crime are emerging, for example changing or killing the RFID tag information by RFID reader. Some adversaries can change the RFID tag of a bag in order to mislead the owner. This can be done without physical evidence using RFID reader from any covert place such as his pocket. In this case, a video camera cannot identify the adversary. RFID systems can trace this event if the tag seems absent due to unauthorised killing.
5. *Identify who handled the baggage last* Using RFID technology it is possible to identify the zone and the people group who handled the baggage last. From the database it is possible to identify approximately who was near the baggage in that location. This information can be used, with video cameras, to identify who mishandled or stole the baggage.

VII. CONCLUSION

RFID technology is attracting important consideration in airports for baggage handling purposes. RFID can be used for sortation, identification, automation and location tracking. The read rate of RFID is much higher than the barcode read rate. RFID technology will reduce labour costs and strengthen automation. It can reduce the time for baggage sortation and distribution. The different activities required in the baggage handling process are very well suited for RFID technology. If RFID technology is implemented properly then passengers and airlines as well as airports will benefit. The passenger can get the information about their baggage's current location, expected arrival time etc. Moreover latest technologies like Internet web facilities, SMS, databases and interactive televisions can be used to identify and enhance the performance of the system to better handle the baggage in airport. Airlines can reduce their costs if they eliminate the

ISBN 978- 974-672-556-9

charges for mishandling and mismanagement of baggage. It is possible to identify the employees responsible for handling baggage before it is lost or mishandled. Airports also enhance their efficiency and thus ensure better customer services. The trials in different airports also gave very positive results and opened up a window for the future. There are also some challenges in the implementation of RFID in airports. Many airports may not take this technology in the near future due to the initial investment and proper initiative needed. So, profit and benefits cannot be maximized until most of the airports use a common RFID-based network system. Most RFID technologies need to be integrated with existing systems like barcodes. Another challenge of RFID is the privacy and security of the tag, because the content of the tag is vulnerable to an adversary. In the future we will do further work on the security and privacy of RFID tags in airports.

REFERENCES

[1] *White paper by AeroAssist. RFID in Aviation:airport luggage control.* Available at www.AeroAssist.pt

[2] Reuters, Edition: UK, Available at http://uk.reuters.com/article/idUKN0741544820070207

[3] *INDUSTRY BRIEF: Baggage Tracking RFID Solutions.* Available at www.motorola.com

[4] R. Want,"An Introduction to RFID Technilogy," *IEEE Pervasive Computing*, vol. 5, pp. 25 – 33, 2005.

[5] B. S. Prabhu, Xiaoyong Su, Harish Ramamurthy, Chi-Cheng Chu, Rajit Gadh WinRFID – A Middleware for the enablement of Radio Frequency Identification (RFID) based Applications *UCLA -Wireless Internet for the Mobile Enterprise Consortium (WINMEC)420 Westwood Pl., Los Angeles CA 90095.*

[6] E. C. Ozelkan, Y. Sireli, M. P. Munoz, and S. Mahadevan, A Decision Model to Analyze Costs and Benefits of RFID for Superior Supply Chain Performance, presented at *Technology Management for the Global Future*, 2006. PICMET 2006, Istanbul, Turkey, July 2006, pp. 610-617.

[7] A. Jules,S. Garfinkel, and R. Pappu, RFID privacy: an overview of problems and proposed solutions.*IEEE Security and Privacy*. 3(3): 34-43, May/June 2005.

[8] A. Jules. RFID security and privacy: A research survey. *IEEE Journal on Selected Areas in Communication*, 24(2), February 2006.

[9] T. Zhang, Y. Ouyang, Y. He, Traceable Air Baggage Handling System Based on RFID Tags in the Airport, *Journal of Theoretical and Applied Electronic Commerce Research*, ISSN 0718–1876, Electronic Version VOL 3 / ISSUE 1 / APRIL 2008 /106-115 © 2008 Universidad de Talca – Chile.

[10] P. Harrop. (2006, August). *RFID in the Air Industry and Land Transport.* [Online]. Available at http://www.idtechex.com/products/en/articles/00000486.asp

[11] A. Jules, *RFID Privacy: A Technical Primer for the Non-Technical Reader*Draft: 23 February 2005.

[12] J. Collins. Marks & Spencer expands RFID retail trial. *RFID Journal*, 10 February 2004. Available at http://www.rfidjournal.com/article/articleview/791/1/1/.

[13] *Hong Airport Press release.* Available at http://www.hongkongairport.com/eng/media/pressreleases/pr_914.html

[14] A. Cerino, and W. P. Walsh, Research and Application of Radio Frequency Identification (RFID) Technology to Enhance Aviation Security, presented at the IEEE 2000 *National Aerospace and Electronics Conference*, Dayton, OH, October 10-12, 2000, pp. 127-135.

[15] R. Das. (2007, May). *The Myth and Reality of Baggage Tagging.* [Online]. Available at http://www.idtechex.com/products/en/articles/00000534.asp

[16] A. Juels, R.L. Rivest, and M. Szydlo. The blocker tag: Selective blocking of RFID tags for consumer privacy. In V. Atluri, editor, *8th ACM Conference on Computer and Communications Security*, pages 103– 111. ACM Press, 2003.

[17] "RFID Business case for baggage tagging", IATA Simplifying the Business, 2007.

Session 4: Knowledge Management and Applications

ISBN 978- 974-672-556-9

Comparative Analysis on Feature Vectors for Printed Bangla OCR

Saima Hossain, Nasreen Akter, S.M. Faqruddin Ali Azam,
Hasan Sarwar, Chowdhury Mofizur Rahman

***Abstract*—Feature Extraction of Bangla characters is very important during recognition. Efficient feature extraction method speeds up recognition process. In this paper, a comparative analysis is done between two different algorithms which are feature extraction using chain code and shadow, centriod and longest–run features.**

Index Terms- algorithm; Bangla language; Bangla characters; feature extraction; feature vectors

I. INTRODUCTION

Paper [1] illustrated freeman Chain code which is based on the observation that each pixel has eight neighborhood pixels. The 8 transitional positions defined by freeman chain code are then divided into 4 transitional zones in order to keep the correct order of searching. Fig. 1 describes the freeman chain code.

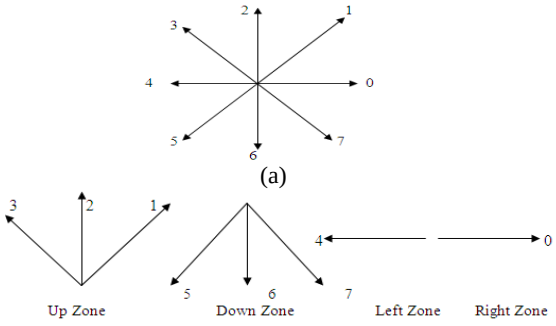
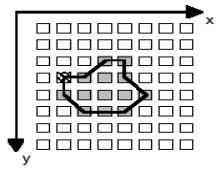


Figure 1. (a) Slope Convention for Freeman Chain code, (b) 8 directional slopes divided into 4 direction zones for searching.

Maintaining an anti clock wise order of searching, zonal information is used to modify the chain coded position of the next selected pixel. The algorithm selects the next pixel if it fulfils all of the following criteria:

- The pixel is Black, i.e., it is a part of the character.
- The pixel is within the bounded rectangle of the connected component.
- The pixel is still not visited.
- The zone of the current pixel.



(a)

Zone	Searching Order
DOWN	3,4,5,6,7,0,1,2
RIGHT	5,6,7,0,1,2,3,4
UP	7,0,1,2,3,4,5,6
LEFT	1,2,3,4,5,6,7,0

(b)

Figure 2. (a) Chain code generation for an image, (b) Searching order in the four zones

Fig. 2 (a) shows the chain code generation of an image marked by gray pixels. When the algorithm starts from the hatched pixel (absolute coordinate, x=1, y=3), it marks the current black pixel as visited and initiates its directional zone as DOWN zone. So it searches for an unvisited black pixel in the directional order: 3,4,5,6,7,0,1,2 (Searching order is shown in Fig. 1.2 (b) for each zone). In this way the process continues and finally produces the chain code, 06700132454.

The frequency of each directional slope at each region is recorded and updated during the traverse. A total of 32 directional slopes or local features for each component are found. Then they are normalized to 0-1 scale. In bangla, as there are more than one components in a character, the normalized 64 features are then averaged to produce 32 features. The calculation of normalized slope distribution is as follows:

If a_1, a_2, a_3, ……,a_8 are 8 directional slopes in region 1, then normalizing constant for region 1 is,

$$N_1 = \sqrt{(a_1*a_1 + a_2*a_2 + \ldots\ldots + a_8*a_8)} \quad (1)$$

So, normalized slope in region i = $S_{i,j}/ N_i$,
Where, i = 1 to 4 and j = 1 to 8
$S_{i,j}$ = Frequency of j'th directional slope in i'th region.
N_i = Normalizing constant in i'th region.

Paper [2] has used a set of 76 features which includes 24 shadow features, 16 centriod features

and 36 longest- run features are computed taking 64 x 64 pixel size binary images. Shadow features are calculated by dividing the image into 8 octants within minimal square. Finally, lengths of all projections on each of the 24 sides of all octants are summed up to produce 24 shadow features, shown in fig. 3

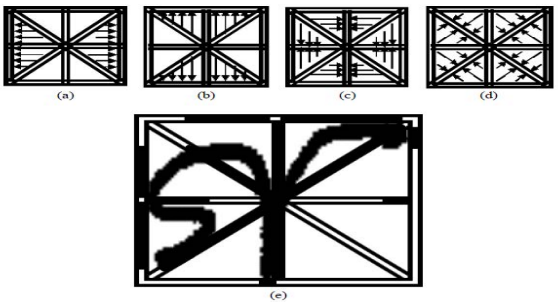


Figure 3. An illustration for shadow features. (a-d) Direction of fictitious light rays as assumed for taking the projection of an image segment on each side of all octants. (e) Projection of a sample image

Coordinates of centriods of black pixels in all 8 octants of a digit image, shown in fig. 4, are considered to add 16 centriod features.

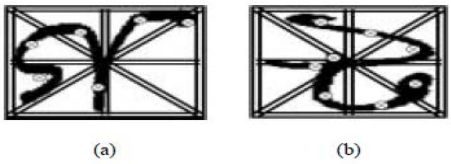


Figure 4. Centriod features of two different characters (a)-(b)

Longest- run features are computed dividing the square into 9 overlapping regions and for each, 4 longest-run features are calculated respectively by row wise, column wise along 2 of its diagonal. Thus 36 features are produced (fig. 5).

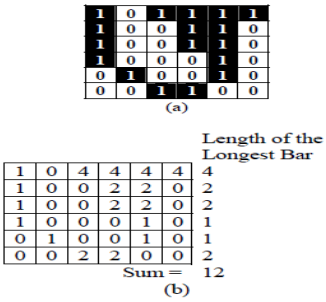


Figure 5. An illustration for computation of the row wiselongest-run feature.(a) The portion of a binary image enclosed within a rectangular region.(b) every pixel position in each row of the image is marked with the lenght of the longest bar that
fits consecutive black pixels along the same row.

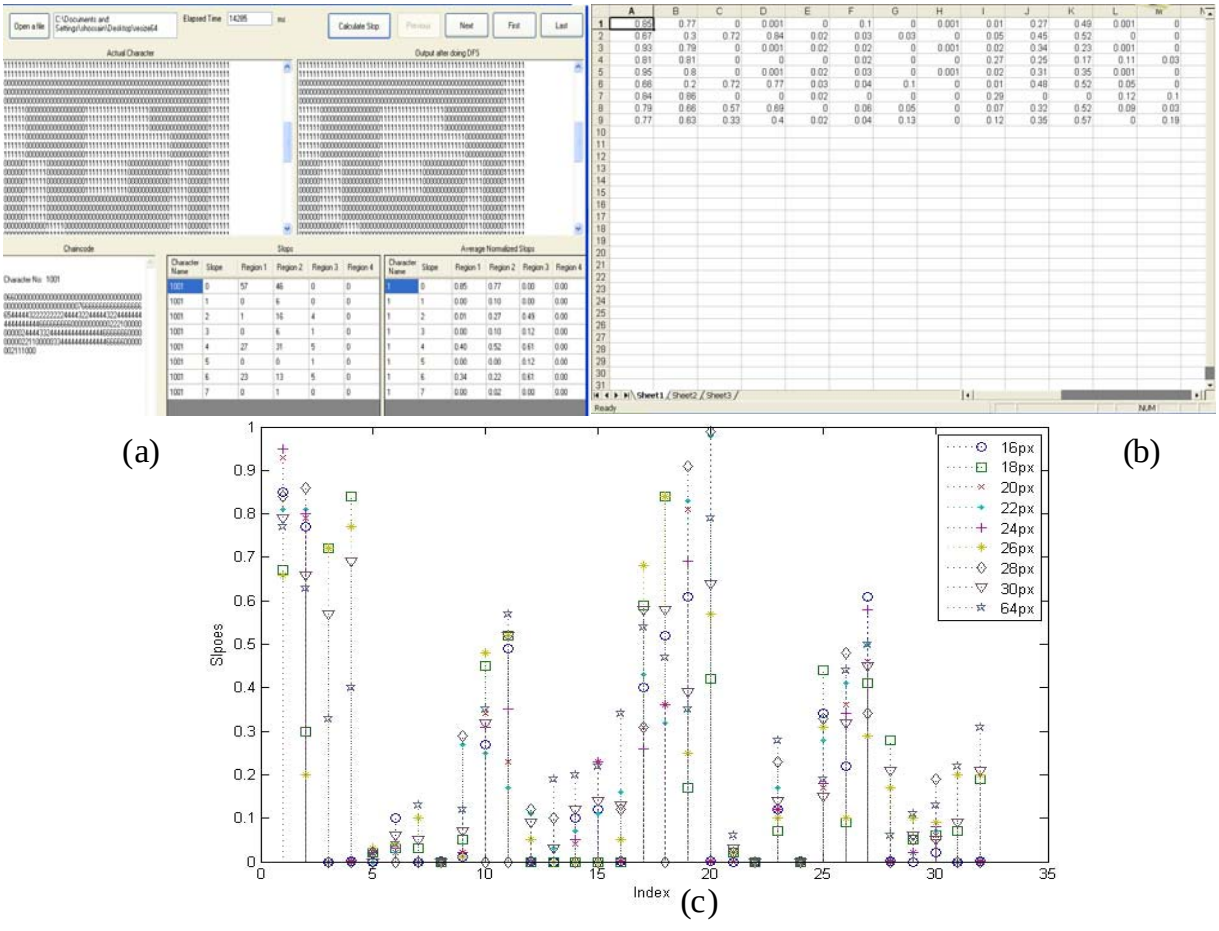


Figure 6. (a) Slope generation using Chain code, (b) 8 sample Aa, each produces 32 slopes, (c) graphical representation for the 8 sample Aa using (b)

ISBN 978- 974-672-556-9

COMPARATIVE ANALYSIS BETWEEN THE TWO FEATURES METHODS

Text document images are constructed with different font sizes. We resize our sample characters into 64 x 64 pixels to make the feature extraction process font size independent. We got these samples from a text image paper [3] and stored the same characters to do the comparison.

A. Feature Extraction Using Chain code

Applying the feature extraction technique using chain code we got the following slopes and corresponding graph for character Aa, অ (fig. 6).

B. Feature extraction using Shadow, Centriod and Longest-run features

Using this feature we have got the 16 shadow features (last 8 shadow features are not considered here), 16 centriod features and 36 longest–run features and their corresponding graph for the character Aa (fig. 7).

C. Comparative Analysis

From the fig. 6 (c), it is seen that the difference between each point of each font sizes compared with each point of 64px font size is larger than fig. 7(c). Using chain code, about 20-30% points are close or even same for a particular character whereas 60-70% points are close or same using shadow, centrio and longest-run features. The differences are shown in Table I for feature extraction using chain code and Table II for shadow, centriod and longest–run features.

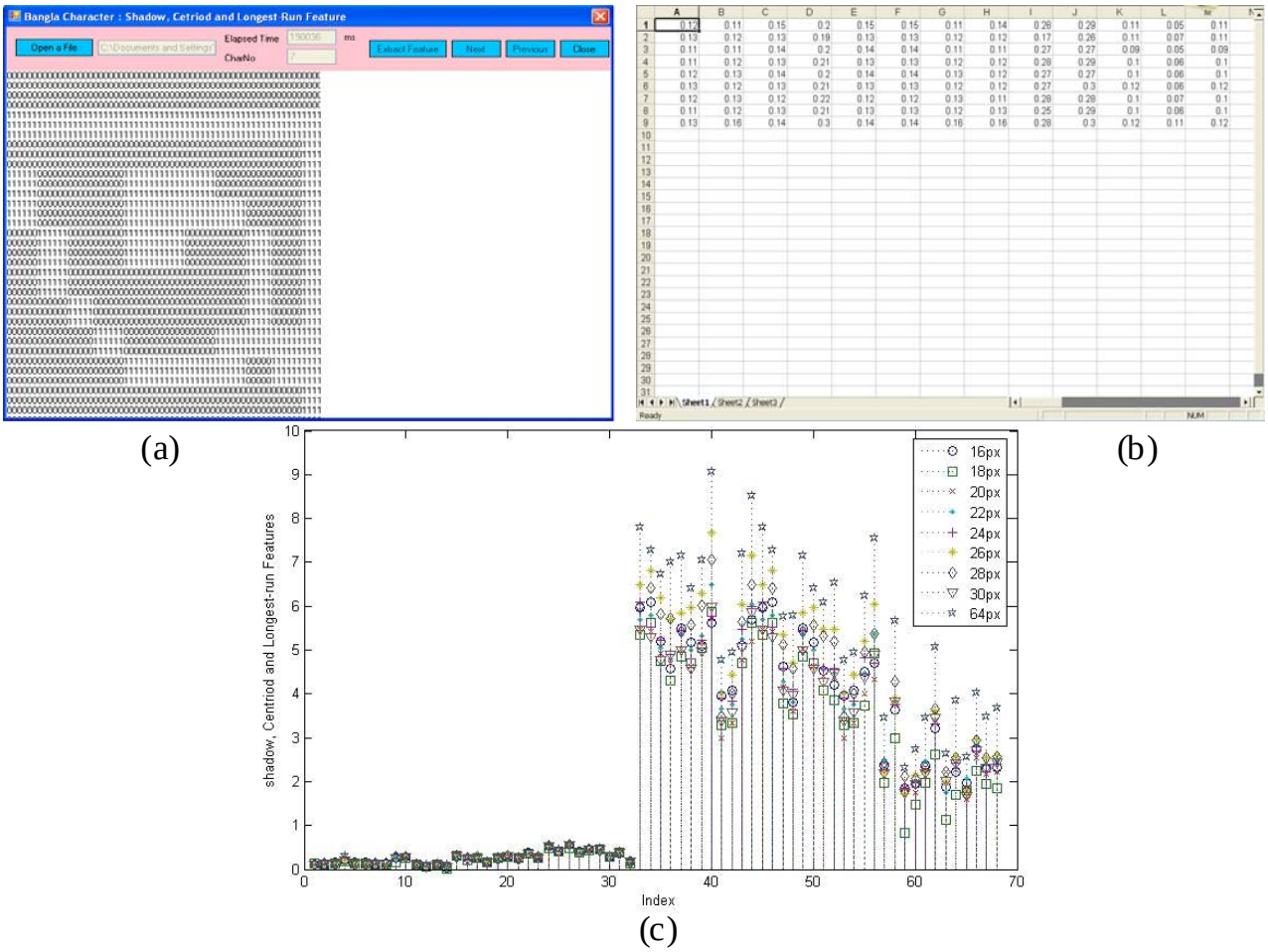


Figure 7. (a) Slope generation using Chain code, (b) 8 sample Aa, each produces 32 slopes, (c) graphical representation for the 8 sample Aa using (b)

ISBN 978- 974-672-556-9

TABLE I. DIFFERENCES AMONG DIFFERENT FONT SIZES WITH 64PX FONT SIZE FOR THE FEATURE EXTRACTION USING CHAIN CODE

Index	Font 16px	Font 18px	Font 20px	Font 22px	Font 24px	Font 26px	Font 28px	Font 30px
1	0.08	0.1	0.16	0.04	0.18	0.11	0.07	0.02
2	0.14	0.33	0.16	0.18	0.17	0.43	0.23	0.03
3	0.33	0.39	0.33	0.33	0.33	0.39	0.33	0.24
4	0.399	0.44	0.399	0.4	0.399	0.37	0.4	0.29
5	0.02	0	0	0.02	0	0.01	0	0.02
6	0.06	0.01	0.02	0.02	0.01	0	0.04	0.02
7	0.13	0.1	0.13	0.13	0.13	0.03	0.13	0.08
8	0.001	0	0.001	0	0.001	0	0	0
9	0.11	0.07	0.1	0.15	0.1	0.11	0.17	0.05
10	0.08	0.1	0.01	0.1	0.04	0.13	0.35	0.03
11	0.08	0.05	0.34	0.4	0.22	0.05	0.57	0.05
12	0.001	0	0.001	0.11	0.001	0.05	0.12	0.09
13	0.19	0.19	0.19	0.16	0.19	0.19	0.09	0.16
14	0.1	0.2	0.16	0.13	0.15	0.2	0.2	0.08
15	0.1	0.22	0.01	0.11	0.01	0.22	0.22	0.08
16	0.339	0.34	0.339	0.18	0.339	0.29	0.22	0.21
17	0.14	0.05	0.23	0.11	0.28	0.14	0.23	0.04
18	0.05	0.37	0.11	0.15	0.11	0.37	0.47	0.11
19	0.26	0.18	0.46	0.48	0.34	0.1	0.56	0.04
20	0.789	0.37	0.789	0.19	0.789	0.22	0.2	0.15
21	0.06	0.04	0.06	0.04	0.04	0.04	0.04	0.03
22	0	0	0	0	0	0	0	0
23	0.16	0.21	0.16	0.11	0.16	0.18	0.05	0.14
24	0.001	0	0.001	0	0.001	0	0	0
25	0.15	0.25	0.02	0.09	0.01	0.12	0.14	0.04
26	0.22	0.35	0.08	0.03	0.1	0.34	0.04	0.12
27	0.11	0.09	0.04	0	0.08	0.21	0.16	0.05
28	0.059	0.22	0.059	0.06	0.059	0.11	0.06	0.15
29	0.11	0.06	0.09	0.09	0.09	0.01	0.06	0.05
30	0.11	0.07	0.07	0.06	0.05	0.04	0.06	0.08
31	0.22	0.15	0.22	0.22	0.22	0.02	0.22	0.13
32	0.309	0.12	0.309	0.31	0.309	0.11	0.31	0.1

TABLE II.DIFFERENCES AMONG DIFFERENT FONT SIZES WITH 64PX FONT SIZE FOR SHADOW, CENTRIOD AND LONGEST-RUN FEATURES

Index	Font 16px	Font 18px	Font 20px	Font 22px	Font 24px	Font 26px	Font 28px	Font 30px
1	0.01	0	0.02	0.02	0.01	0	0.01	0.02
2	0.05	0.04	0.05	0.04	0.03	0.04	0.03	0.04
3	0.01	0.01	0	0.01	0	0.01	0.02	0.01
4	0.1	0.11	0.1	0.09	0.1	0.09	0.08	0.09
5	0.01	0.01	0	0.01	0	0.01	0.02	0.01
6	0.01	0.01	0	0.01	0	0.01	0.02	0.01
7	0.05	0.04	0.05	0.04	0.03	0.04	0.03	0.04
8	0.02	0.04	0.05	0.04	0.04	0.04	0.05	0.03
9	0.02	0.11	0.01	0	0.01	0.01	0	0.03
10	0.01	0.04	0.03	0.01	0.03	0	0.02	0.01
11	0.01	0.01	0.03	0.02	0.02	0	0.02	0.02
12	0.06	0.04	0.06	0.05	0.05	0.05	0.04	0.05
13	0.01	0.01	0.03	0.02	0.02	0	0.02	0.02
14	0.07	0.09	0.06	0.07	0.06	0.06	0.06	0.05
15	0.03	0.02	0	0	0	0	0	0
16	0.03	0.07	0.05	0.01	0.06	0	0.03	0.05
17	0	0.01	0.01	0.01	0.01	0	0.01	0
18	0.01	0.01	0.01	0.01	0	0	0.01	0

ISBN 978- 974-672-556-9

TABLE II. Differences Among Different Font Sizes with 64px Font Size for Shadow, Centriod and Longest-run features (cont.)

Index	Font 16px	Font 18px	Font 20px	Font 22px	Font 24px	Font 26px	Font 28px	Font 30px
19	0.01	0	0	0.01	0	0.01	0.01	0.01
20	0	0	0	0.01	0	0.01	0.01	0.01
21	0.01	0	0	0.01	0.01	0	0.01	0.01
22	0	0.01	0.01	0.01	0	0.01	0.01	0.01
23	0.01	0	0	0.01	0.01	0.01	0.01	0
24	0.02	0.03	0.02	0.01	0.03	0	0.01	0.02
25	0.03	0.04	0.03	0.03	0.03	0.04	0.03	0.03
26	0	0.08	0.01	0.01	0.01	0.01	0	0.01
27	0	0.01	0.02	0.01	0.01	0.01	0.02	0.01
28	0.04	0.05	0.06	0.04	0.04	0.04	0.04	0.04
29	0.05	0.03	0.05	0.05	0.03	0.05	0.04	0.04
30	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.02
31	0.02	0	0	0	0	0	0	0
32	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01
33	1.84	2.47	2.329	2.119	1.73	1.329	1.819	2.34
34	1.21	1.67	1.86	1.5	1.77	0.49	0.88	1.99
35	1.52	1.970	1.80	1.69	1.510	0.54	0.91	1.960
36	2.43	2.69	2.29	2.2	2.09	1.31	1.28	2.11
37	1.68	2.31	2.15	1.8	1.73	1.33	1.76	2.18
38	1.23	1.72	1.77	1.42	1.61	0.43	0.83	1.82
39	1.949	2.02	1.89	1.739	1.829	0.76	1.04	2.079
40	3.45	3.22	3.36	2.59	3.31	1.39	2.03	3.08
41	0.82	1.48	1.8	1.12	0.86	0.77	1.29	1.39
42	0.89	1.63	1.62	1.2	1.12	0.53	0.95	1.38
43	2.11	2.49	2.430	1.96	1.72	1.16	1.55	2.13
44	2.829	2.89	3.319	2.47	2.529	1.359	2.039	2.659
45	1.84	2.47	2.329	2.119	1.73	1.329	1.819	2.34
46	1.21	1.67	1.86	1.5	1.77	0.49	0.88	1.99
47	1.14	1.97	1.6	1.48	1.19	0.41	0.64	1.68
48	1.99	2.27	2.18	2	1.7	1.109	1.21	1.82
49	1.68	2.31	2.15	1.8	1.73	1.33	1.76	2.18
50	1.23	1.72	1.77	1.42	1.61	0.43	0.83	1.82
51	1.57	2.02	1.69	1.53	1.51	0.63	0.77	1.8
52	2.33	2.68	2.16	2.07	1.95	1.06	1.34	2.05
53	0.82	1.48	1.8	1.12	0.86	0.77	1.29	1.39
54	0.89	1.63	1.62	1.2	1.12	0.53	0.95	1.38
55	1.73	2.49	2.23	1.75	1.4	1.03	1.28	1.85
56	2.869	2.63	3.22	2.18	2.85	1.52	2.199	2.659
57	1.09	1.48	1.01	0.98	1.22	1.23	1.03	1.27
58	2.02	2.68	1.87	1.8	1.9	1.75	1.390	1.83
59	0.48	1.51	0.47	0.66	0.49	0.6	0.2	0.59
60	0.8	1.27	0.98	0.74	0.77	0.6	0.67	0.8
61	1.09	1.48	1.15	0.98	1.22	1.23	1.13	1.27
62	1.85	2.46	1.68	1.68	1.75	1.46	1.4	1.62
63	0.77	1.5	0.43	0.9	0.68	0.64	0.43	0.62
64	1.63	2.14	1.58	1.52	1.42	1.29	1.3	1.43
65	0.6	0.78	0.97	0.52	0.74	0.77	0.79	0.9
66	1.29	1.79	1.48	1.18	1.35	1.06	1.08	1.32
67	1.19	1.52	1.3	1.09	1.27	0.94	0.94	1.17
68	1.38	1.83	1.48	1.21	1.3	1.09	1.13	1.25

ISBN 978- 974-672-556-9

II. Conclusion

In sum, from the above experiment, it can be said that shadow, centriod and longest-run feature methods are better than the feature extraction method using chain code. Though there is a limitation using chain code to extract feature which is, if any noise exits in in the binary image, it does not produce good result but this method is much faster than shadow, centriod and longest-run feature technique. Although the later one produce good result but it is very time consuming.

References

[1] Jalal Uddin Mahmud, Mohammed Feroz Raihan & Chowdhury Mofizur Rahman, *A Complete OCR System For Continuous Bangali Characters*, IEEE,2003, pp. 1372-1376.

[2] Subhadip Basu, Nibaran Das, Ram Sarkar, Mahantapas Kundu, Mita Nasipuri & Dipak Kumar Basu , *Handwritten 'Bangla' Alphabet Recognition Using an MLP Based Classifier*, NCCPB, Bangladesh, 2005, pp. 285-291.

[3] Nasreen Akter, Saima Hossain, Md. Tajul Islam & Hasan Sarwar , *An Algorithm For Segmenting Modifies From Bangla Text*, ICCIT, IEEE, Khulna, Bangladesh, 2008, pp.177-182

ISBN 978- 974-672-556-9

Harmonised and Contextual Based Driven Shape Grammar

Bernadette Sharp
Staffordshire University, UK
b.sharp@staffs.ac.uk

Arus Kunkhet
Chiang Mai University, Thailand
arus.kunkhet@gmail.com

***Abstract* - This paper proposes a novel shape grammar dedicated to humanoid characters, derived from the field of natural language processing. A shape grammar consists of a set of shape rules and a generation engine which selects and processes rules. Most shape grammars applications focused on architectural designs and engineering design. Recent studies applied shape grammar in computer graphics, computer games and movies. This research is concerned with developing a shape grammar which takes into account context and harmony, such criteria have not been addressed by the researchers.**

Index Terms- artificial intelligence techniques; harmonized and contextual based characters; natural language processing; shape grammar

I. Introduction

This paper proposes the integration of natural language processing, artificial intelligence techniques and shape grammar for the generation of harmonized and contextual based 3D characters. A shape grammar consists of a set of shape rules and a generation engine which selects and processes rules. Such a grammar is primarily concerned with defining rules to transform geometrical entities and spatial relations, however there is a need to ensure that the output generated by these rules are harmonized and can take into account not only the context of the already existing shapes but also the context of the domain of application. This paper proposes a new approach to the generation engine in order to ensure the new generated shapes are semantically and contextually viable.

This paper will start with a brief review of shape grammars and its applications. It will outline some of the limitations of current approaches to shape grammars, and discusses the proposed new generation engine mechanism derived from the fields of natural language processing and artificial intelligence.

II. Review of Shape Grammar

Shape grammar, which was developed by Stiny and Gips in 1972, begins with a vocabulary of shapes and spatial relations between shapes [1]. The vocabulary elements of a shape consist of points, lines, planes or volumes. Krishnamarti and Earl define shape as "… a finite set of maximal straight lines of finite, nonzero length, where each line is specified by the coordinates of its end points" [2]. A shape is generated by beginning with an initial shape and recursively applying various shape operations of addition and subtraction and spatial transformations such as shifting, mirroring and rotating. Rules specify which and how the particular shapes should be replaced by applying transformations that permit one shape to be part of another.

Since their inception researchers focused on developing shape grammar interpreters in order to automate the application of shape rules, and to generate networks of designs [3]-[6]. Recent work by Jowers and Earl [7] have extended shape grammar interpreters by developing a method for shapes composed of parametric curve segments in 2D and 3D spaces.

The main foundation of shape grammar lies in the clear understanding of the rules structure, diagrammatic and parametric rules. Both rules are found quite similar in their principles, however producing distinct results in different situations.

Figure 1. Diagrammatic Rules Shape Grammar Model

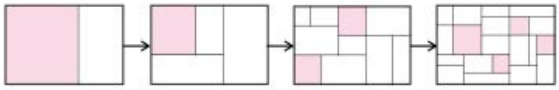


Figure 2. Diagrammatic Shape Grammar Rules in 2D Drawing

Diagrammatical shape grammar rules are based on a generic 2D (X and Y axes). The process starts by applying a rule to a vocabulary, one rule at a time. The applied rule(s) can be repeated. The structure is simple, as the vocabulary will be formulated until the satisfied shapes are achieved (Fig. 1). The example, given in Fig. 2, shows shape grammar is used to generate an evolved representation based on Mondrian paintings [8]. It starts with the rectangle as the initial shape and applies recursively a set of rules as follows.

- New rectangles are obtained by drawing a new line dividing from any side of the previous rectangle across perpendicularly from one side to the opposite side of the rectangle by one third.

ISBN 978- 974-672-556-9

- The rule is repeated to any new obtained rectangle but cannot apply to the same rectangle more than once, even from the different sides of that rectangle.
- The obtained rectangles are considered as new rectangles and are available for further rule application.
- Any obtained rectangle that becomes a square must be masked in pastel red colour.
- Any obtained rectangle that loses the square properties from the previous step must be unmasked the colour.

Parametric Shape Grammar allows variation of parameters, for example changes in lines and angles of shape [9]. The new vocabulary created by the rules is defined by parameters extending the parameter concept to all design elements. Being parametric, a greater variety of forms can be created. Derivations can be used as a new vocabulary, and the process is repeated again to generate a new shape or form (Fig 3). In Fig. 4 the vocabulary of the initial shape consists of eight lines arranged to form an octagon at the centre, and a set of rules are applied as follows.

- A new vocabulary is obtained by rotating the previous vocabulary clockwise or counter clockwise, where the rotation origin is at the centre of that vocabulary.
- A set of new vocabulary is generated by mirroring horizontally or vertically the previous vocabulary.
- The mirror origin must be at the centre of that vocabulary.

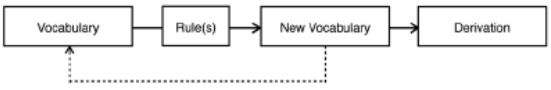


Figure 3. Parametric Shape Grammar Model

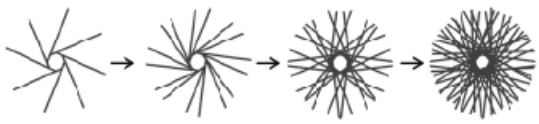


Figure 4. Parametric Shape Grammar Rules in 2D Drawing

III. Applicationsof Shape Grammars

Applications of shape grammars have been used in many areas such as painting, sculpture, architecture, design education, engineering design, product design and computer graphics. [10] applied shape grammar to analyze the art of Tibetan Tangka paintings who used shape grammar to recreate a floor plan of the Villa Malcontena, have demonstrated how a number of plans could be generated in the style of Palladio through the application of shape rules [11]. Flemming [12] introduced 3D geometry into shape grammar to his implementation of Queen Ann houses. Aksamija*et al.* [13] have integrated knowledge bases and shape grammars to study the characterisation of existing vernacular row-house and high-rise apartment buildings in Baltimore. As part of his PhD dissertation Duarte [14] used shape grammar to generate novel designs for customizing Alvaro Siza's mass houses at Malagueira, and in his more recent work he has developed a parametric urban shape grammar for the Zaouiat Lakhdar quarter of the Medina of Marrakech in Morocco [15]. In the last decade shape grammars were applied to product design such as the work of Agarwal and Cagan's implementation of a coffee maker [9], and the development of aU13 shape grammar by Chau*et al.* [16] to support rectilinear and curvilinear basic elements in 3D space which was tested on two case studies, a Coca-Cola bottle grammar and a Head & Shoulder bottle grammar. Shape grammars have also been shown to be an effective engineering design tool. McCormack and Cagan [17] have developed a parametric shape grammar to design the inner hood panel of cars. Lee and Tang [18] has developed an interactive system that uses parametric 2D and 3D shape grammars for digital camera design, incorporating an evolutionary algorithm for exploring product forms at the early stage of design process. Fiedler and Ilčík [19] have extended the application of shape grammar to the procedural modeling of humanoid characters which play an important role in computer graphics and computer games.

Early work in shape grammar was carried out manually; later on many researchers developed a program to implement shape grammar leading to a generation of shape grammars interpreters in 2D and 3D spaces. Some interpreters have focused on design and either generate shapes in the language or are guided by the user who selects the rule to be applied and where in the current shape to apply it. Others have been more analytical in their approach in determining whether the shape is in the language generated by the shape grammar and can generate the sequence of rules it applied to produce the shape. A third type of interpreters have applied grammatical inference strategy to generate shapes in the same style from a given a set of shapes. Table I is a summary of shape grammars implementations as provided by Chau and included in Gips [20].

IV. A Novel Paradigm of Shape Grammar

In his book Stiny [21] discusses how grammar of visual mathematical arguments can be used to describe and construct shapes by means of a formal algebra. He compares shape grammars with Noam Chomsky's verbal grammars. In addition to aesthetics and visual appeal shape grammars focuses on syntax just like any grammar. The shapes are the vocabulary (or lexicon) of a shape grammar, and the grammar rules consist of a set of spatial design transformations that are applied to produce new shapes. In most

ISBN 978- 974-672-556-9

TABLE I. SHAPE GRAMMARS IMPLEMENTATIONS (Gips, 1999)

NO	*NAME*	*REFERENCE*	*TOOL (S)*	*TYPE*
1	Shepard-Metzler Analysis	Gips, 1974	SAIL*	2D
2	Simple Interpreter	Gips, 1975	SAIL	2D, 3D
3	Shape Grammar Interpreter	Krishnamurti, 1982	X	2D
4	Shape Generation System	Krishnamurti, Giraud 1986	PROLOG**	2D
5	Queen Anne Houses	Flemming, 1987	PROLOG	2D
6	Shape Grammar System	Chase, 1989	PROLOG Mac	2D
7	Genesis (CMU)	Heisserman, 1991	C, CLP	3D
8	GRAIL	Krishnamurti, 1992	X	2D
9	Grammatica	Carlson, 1993	X	X
10		Stouffs, 1994	X	2D, 3D
11	Genesis (Boeing)	Heisserman, 1994	C++, CLP	2D, 3D
12	GEdit	Tapia, 1996	LISP*** Mac	2D
13	Shape Grammar Editor	Shelden, 1996	AutoCAD, Auto LISP	2D
14	Implementation of Basic Grammar	Duarte Simondetti, 1997	AutoCAD, Auto LISP	3D
15	Shape Grammar Interpreter	Piazzalunga Fitzhorn, 1998	ACIS, LISP****	3D
16	SG-Clips	Chien, 1998	CLIPS	2D, 3D
17	3D Architecture form Synthesis	Wang, 1998	Java, Open Inventor	3D
18	Coffee Maker Grammar	Agarwal and Cagan, 1998	Java	2D, 3D

Note: *SAIL - Stanford Artificial Intelligence Language, **SeeLog developed at EdCAAD, ***Macintosh Common LISP (MCL), ****ACIS Scheme

applications it is left to the user to guide the selection of the rules in order to meet the design objectives. However there are many choices of rules creating different emerging properties responding to different conditions or objectives. To resolve these problems researchers have applied artificial intelligence techniques to control the selection of the rules. Shea and Cagan [22] have applied shape annealing to produce optimally directed designs. Explicit domain knowledge is placed within the grammar through rule and syntax whilst design interpretation is used to select forms that fulfill functional and visual goals. O'Neil *et al.* [23] have combined genetic programming with shape grammar to encode human domain knowledge to rediscover known benchmark target structures.

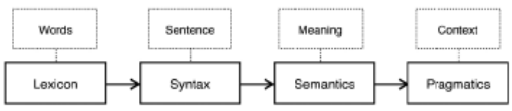


Figure 5. Natural Language Processing Model

Although these approaches can handle complex 3D objects, they are better at generating free-forms or unpredictable shapes rather than fulfilling specific harmonised requirements. These approaches may be appropriate to engineering design and product design; however in computer games and movies applications, they cannot provide a rich semantic and harmonized family of humanoid characters. The generation engine of the shape grammar requires a deeper level of analysis to combine syntax with both, semantics and context, in order to generate a harmonized set of humanoid characters. The harmonisation will bring a realistic feeling to the audience that characters came from the same world or same story; characters should be compatible and exhibit agreeable set of features. Any non-harmonised character must be considered as an error in the design set and should be rejected. Similarity among the characters is different from harmonisation. There are many methodologies that can create similarity in design. The term harmony in design could be achieved differently by applying contextual design grammar and rules. In order to achieve harmonization all characters must exhibit equivalence or correspondence among constituents of a humanoid or between different humanoids.

In this research project we propose a humanoid shape grammar paradigm based on the field of natural language processing which identifies four important levels of analysis: lexical, syntactical,

ISBN 978- 974-672-556-9

semantic and pragmatic levels which can interact in a variety of orders (Fig. 5). In linguistics the lexical level requires a lexicon which may contain the words and their part(s)-of-speech (e.g. determiner, noun, verb), and contain information on the semantic class of the word, and definitions of the sense(s) in the semantic representation. The syntactic level is concerned with analyzing the words in a sentence so as to uncover the grammatical structure of the sentence. The output of this level of processing is a representation of the sentence revealing the structural dependency relationships between the words. Semantic processing determines the possible meanings of a sentence by focusing on the interactions among word-level meanings in the sentence [24]. For example, amongst other meanings, 'file' as a noun can mean either a folder for storing papers, or a tool to shape one's fingernails, or a line of individuals in a queue. To disambiguate the meaning of polysemous words, this requires consideration of the local context, applying pragmatic knowledge of the domain.

We propose to develop a humanoid shape grammar consisting of four levels: *Vocabulary*, *Rules*, *Derivation*, and *Context*. To have a harmony in character design, the four levels must be embedded in the generation engine. The *Vocabulary* of a shape grammar is a lexicon consisting of points, lines, and planes. *Rules* define a set of syntactic structures which constrain the possible spatial and functional transformations specific to the humanoid body characters; these transformations will be interpreted by a semantic model embedded in *Derivation* to ensure legitimacy, consistency and compatibility. In *Context*, legitimate shapes and elements of the characters must adhere to certain contextual properties and roles of the humanoid family design to ensure harmonisation with other humanoid family characters (Fig. 6).

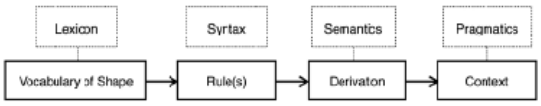


Figure 6. Humanoid Shape Grammar Model

An ontology for the humanoid shape grammar will be required to capture the body of knowledge about the humanoid world consisting of a repository of primitive vocabulary used in meaning representation, and the hierarchical and semantic relationships between various elements of the vocabulary. In this application, the ontology will supply domain knowledge to the four levels of the generation engine.

To create a humanoid family, *Rules* will apply a set of transformations on the primitive vocabulary to create components of a body (e.g. head, body, arms, legs) consistent with the definition of what constitutes a humanoid body. The assembling of these components is governed by a semantic model which dictates their spatial relations, size, weight and height, and personality (e.g. aggressive, friendly). Derivation will check that the humanoid is harmonious with other members of the humanoid family, inheriting/sharing similar and consistent features (e.g. same number or eyes, position of eyes, fingers growing from hands, ability to walk and move). Context will define their family relations among other characters (e.g. parent, child) and ensure that their physical and emotional characteristics are harmonious with their family position, role and function.

V. Conclusions

This paper has proposed a novel approach to shape grammar designed to generate a set of harmonious humanoid characters which will play an important role in computer graphics, computer games and movies. Whilst early work in shape grammar focused primarily on architecture and engineering applications very little research has been done evaluating its usefulness in the domain of computer games. This research is an attempt at addressing this gap.

The proposed humanoid shape grammar, which aims at generating well formed syntactically, semantically and contextually set of humanoid characters, is based on the four levels of analysis of natural language processing. These levels are embedded into the generation engine of shape grammar to ensure the production of harmonious family of humanoids.

Acknowledgments

This research project, which is carried out as part of a doctoral programme carried out in collaboration with two universities: Chiang Mai University in Thailand and Staffordshire University, UK, is funded by the ERASMUS e-link project.

References

[1] G. Stiny, and J. Gips, "Shape Grammars and the Generative Specification of Painting and Sculpture," *Proceedings of IFIP Congress*, vol. 71, 1972, pp. 1460-1465.

[2] R. Krishnamurti, and C. F. Earl, "Shape Recognition in Three Dimensions," *Environment and Planning B: Planning and Design*, vol. 19, 1992, pp. 585-603.

[3] R. Krishnamurti, and Giraud, "Towards a Shape Editor: The implementation of a shape generation system," *Environment and Planning B*: *Planning and Design*, vol. 13, 1986, pp. 391-403.

[4] R. Krishnamurti, "SGI: A Shape Grammar Interpreter," *Research report*, Centre for Configurational Studies, The Open University, Milton Keynes, England, 1982.

[5] S. C. Chase, "Shape and Shape Grammar: From mathematical model to computer implementation," *Environment and Planning B: Planning and Design*, vol. 16, 1999, pp. 215-242.

[6] M. Tapia, "A Visual Implementation of a Shape Grammar System," *Environment and Planning B: Planning and Design*, vol. 26, no.1, 1999, pp. 59-73.

[7] I. Jowers, and C. Earl, "The Construction of Curved Shapes," *Environment and Planning B: Planning and Design*, vol. 37, no.1, 2010, pp. 42-58.

[8] T. Schnier, and J. S. Gero, "From Frank Lloyd Wright to Mondrian: Transforming evolving representations," in I. Parmee, Ed., *Adaptive Computing in Design and Manufacture*, Springer, London, 1998, pp. 207-219.

[9] M. Agarwal, and J. Cagan, "A Blend of Different Tastes: The language of coffee makers," *Environment and Planning B: Planning and Design*, vol. 25, no.2, 1998, pp. 205-226.

[10] W. Zhang, and S. Lin, "Research on Tibet Tangka Based on Shape Grammar," *Computer-Aided Industrial Design and Conceptual Design*, CAID/CD, 2008, pp. 373-376.

[11] G. Stiny and W. J. Mitchell, "The Palladian grammar," Environment and Planning B, 1978, vol. 5, pp. 5-18.

[12] U. Flemming, "More than the sum of its parts: the grammar of Queen Anne houses," Environment and Planning B: Planning and Design, 1987, vol. 14, pp. 323-350.

[13] A. Aksamija, K. Yue, H. Kim, F. Grobler, R. Krishnamurti, "Integration of knowledge-based and generative systems for building chracterization and prediction," Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 2010, vol. 24, pp.3-16.

[14] J. Duarte, "Shape Grammar for Álvaro Siza's patio houses in Malagueira", PhD Thesis Project, MIT, 1973.

[15] J. Duarte, M. Rocha G. D. Soares, "Unveiling the Structure of the Marrakech Medina: A shape grammar and an interpreter for generating urban form," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing Archive*, vol. 21, no. 4, 2007, pp. 317-349.

[16] H. H. Chau, and X. Chen, "Evaluation of a 3D Shape Grammar Implementation," JS Gero, Ed., *Design Computing and Cognition'04*, 2004, pp. 357-376.

[17] J. P. McCormack and J. Cagan, "Designing inner hood panels through a shape grammar based framework, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 2002, vol. 16, pp. 273-290.

[18] H. C. Lee, and M. X. Tang, *Evolutionary Shape Grammars for Product Design*, Hong Kong, Design Technology Research Centre, School of Design, Hong Kong Polytechnic University, 2004.

[19] S. Fiedler, and M. Ilčík, *Procedural Human Posing Using CGA Grammars*, Institute of Computer Graphics and Algorithms, Vienna University of Technology, November 2, 2009.

[20] J. Gips, "Computer implementation of shape grammars", *NSF/MIT Workshop on Shape Computation*, MIT, Cambridge MA, 1999.

[21] G. Stiny, *Shape:Talking about Seeing and Doing*, London: MIT Press, 2006.

[22] K. Shea, and J. Cagan, "Languages and Semantics of Grammatical Discrete Structures", *Artificial Intelligence in Engineering Design, Analysis, and Manufacturing*, Vol. 13, 1999, pp. 241-251.

[23] M. O'Neill, J. M. Swaford, J. McDermott, "Shape grammars and grammatical evolution for evolutionary design", *GECCO'09*, July 8-12, pp. 1035-1042.

[24] J. Allen. "Natural Language Understanding," University of Rochester, The Benjamin Cummings Publishing Company, Inc., Redwood City, CA 94065, The United States of America, 1995

An Integration of Domain Ontology-based Knowledge Management and SquaRE for Thai Local Software Companies

Pornpen Lertthasanawong, Tirapot Chandarasupsang and Nopasit Chakpitak
College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, 50200, Thailand
pornpen@neo-netsoft.com

***Abstract*-The serious problem of Thai Software Industry which has happened for long time is *Requirement Quality* which becomes a big hindrance on business expansion because knowledge workers (i.e. Project Manager, System Analyst) cannot perform productivity and apply their knowledge in work place. How to improve their competency? Do they know what should be known? Learning and implementing the new knowledge management tool (i.e. Software Quality Model, Knowledge Creation and sharing and Domain Engineering) will be the way to develop their skill, ability and change their behavior. There will become improve the coordination of project teams and increase consistency of interaction with customers and also reduce processing errors and time delivery.**

Index terms- Domain engineering; Knowledge Creation and sharing; Ontology-based Knowledge Management; Software Quality Model

I. INTRODUCTION

Thai Software export (2008) was only 4,000 million Baht of which are very low comparing with other developing countries in Asia and the domestic software market 6,000 million Baht. Therefore there must be something that some other countries know and do it right while there must be some wrong doing in Thailand and block the ICT development in Thailand. From the observation, there is a similar key success factor of those countries like Singapore, Malaysia, and Vietnam in develop the local ICT industry and that is to partnership with major ICT international companies and the national strategy.[1]

From the survey of National Statistical Office of Thailand (2008) there is found that the barriers of Thai software industry can be summarized as following (1) Personnel Issue (2) Domestic market (3) International market (4) Source of fund (5) Technological Change (6) Copyright Infringement. The serious problem that has happened for long time is *Requirement Quality* which is come from the knowledge of project manager or system analyst who got high salary but low quality for real competitive situations. Most of Thai local software is very small enterprise (VSE) [2] which develop and/or maintain software that is used in larger systems, therefore, recognition of VSEs as suppliers of high quality software is often required. VSEs have common problems related to the management of risk and quality of software projects. This generates cost overruns, time delay and cancelled projects. From studies and surveys conducted, it is clear that the majority of ISO/IEC standards do not address the needs of VSEs. Conformance with these standards is difficult, if not impossible. Subsequently VSEs have no, or very limited, ways to be recognized as entities that produce quality software in their domain. Therefore, VSEs are often cut off from some economic activities. Some model like the CMMI is not affordable for the small organizations.It has been found that VSEs find it difficult to relate ISO/IEC standards to their business needs and to justify the application of the standards to their business practices. Most VSEs can neither afford the resources, in terms of number of employees, budget and time, nor do they see a net benefit in establishing software life cycle processes. To rectify some of these difficulties, a set of guides have been developed according to a set of VSE characteristics. The guides are based on subsets of appropriate standards elements, referred to as VSE Profiles. The purpose of a VSE profile is to define a subset of ISO/IEC standards relevant to the VSE context, for example, processes and outcomes of ISO/IEC 12207 and products of ISO/IEC 15289.

So Thai Industry Standard Institute (TISI) setting up WG 24S setting up WG 24S in ISO/IEC SC7 to developing ISO 29110 Standard Model for VSE at the beginning of 2009but not much research mentioned about Personnel Issue which is main factor of success for software industry. So there should be some tools for decreasing people problems (Internal and External Organization). [3]

II. BACKGROUND AND MOTIVATION

It was known that knowledge management efforts typically focus on organizational objectives by capture, codify and transfer knowledge across people to improved performance and achieve continuous improvement for competitive advantage. It involves using the ideas and experience of employees, customers and suppliers to improve the

ISBN 978- 974-672-556-9

organization's performance and also using technology to form the new knowledge which is now importance for creating the company values. There are four types of knowledge which are Tacit, Implicit, Explicit and Embedded. Is there a way of integration of using knowledge management with software standard (SquaRE) for Thai local softwarecompanies?

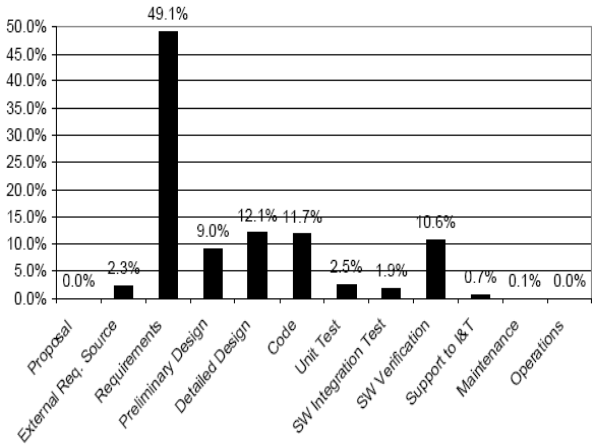


Figure 1. Software Industrial Benchmark 2001.

How do they manage the problem? Some new concepts and tools of knowledge management will be the way out to reach Thai local. How they manage the problem? Some new concepts and tools of knowledge management will be the way out to reach Thai local software companies objectives such as Quality in Customer Requirement and Product Requirement, enhance organizational learning anticipate impending threats and/or opportunities, competitiveness and finally Intellectual capital(IC) with their customer satisfaction**.**

III. LITERATURE REVIEW

A. Ontology and Software Knowledge Management

It is believed that good software document would help software developers make good software project. Organization knowledge can be categorized as individual and group-base knowledge. Individual knowledge is mostly in tacit form and some is in explicit way, but the scale of the later knowledge is very small compare to tacit form such as a huge part of the detail during a meeting with customers are unrecorded and only resided in the developer's mind. The ontology in Software Engineering to provide a source of precisely defined term that can be communicated across people, organization and application (information system or intelligent agents) and also offer a consensual shared understanding concerning the domain of discourse. The most value was to render all hidden assumption concerning the objects pertaining to a certain domain of knowledge.

B. Software product Quality Requirements and Evaluation (SquaRE)

SquaRE is the ISO25000Series which is a process including three type of requirements.[3]

- Customer Requirement (Quality in Used)
- Product Requirement (External Quality Attributes)
- Product-Component Requirement (Internal Quality Attributes)

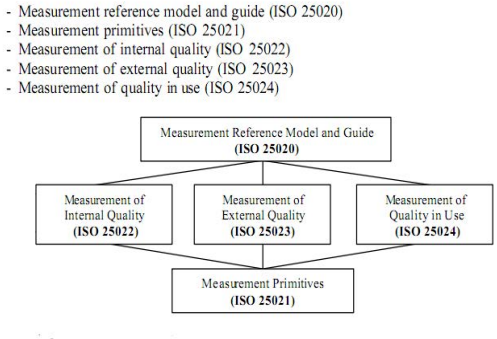


Figure 2. Software product Quality Requirements and Evaluation (SquaRE)

The standard ISO25000 Software Product Quality Requirement has relation with ISO9126 Evaluation of Software Quality as External and Internal Quality Attributes.[4]

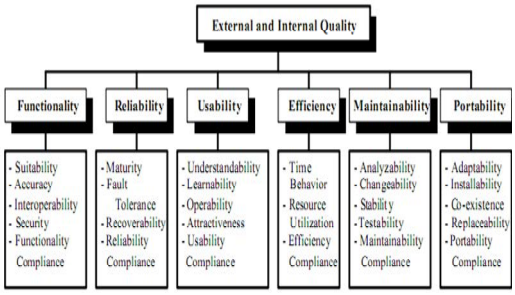


Figure 3. Quality in Software Life Cycle ISO 9126-1[4].

C. Related Researches

Alain, Abran and Niji Habra[5]analyzed the measurement primitives and quality and mapping Quality Model to the measure information model by three different sections which are: data collection, data preparation and data analysis They revisit the ISO working group six (WG6) in order to recommend the proper mapping of concepts to the related methodology terms between the ISO 9126 Quality model and the ISO 15939 Measurement Information Model. They set the measurement model into three parts: Data Collection, Data Preparation and Data Analysis.

Thanasankit [6] survey for understanding the management in requirement engineering used in

Thailand to produce software. The paper suggests that the management process is infused with culture practices which generate divergence from rationality and from structured management models created from "western" studies. The problems encounter by the system analysts came from both technical and social areas such as

- Lack of methodology for develops information systems for staff to follow. No method for eliciting requirements.
- Lack of cooperation when constructing the TOR. Thai culture is built on relationships and trust. Therefore the process of dealing with discrepancies between what was desired and what was possible to deliver. Lack of cooperation from users the users did not have enough spare time to assist the system analyst, therefore they only provided the system analyst with requirement and necessary documents only the system analyst asks them.
- Uncertainty and fear of failure and responsibility. The users felt the statement they gave could be as evidence if the systems did not come out right.
- Conflict between users. In the Thai context, the superior will make decision and every one must accept them without question.
- Constant change of requirements. If the change were not significant then system analyst often allows the clients to change, because they needed to build good relationship with the clients.
- Constant change of requirements. If the change were not significant then system analyst often allow the clients to change, because they needed to build good relationship with the clients

IV. METHODOLOGY

The research aims to test and validate the integration of technical standard (ISO25000 Software Product Quality Requirement & ISO9126 Evaluation of Software Quality) in term of knowledge engineering and management to capture, identify and suggest for the new concept and tool for Thai local software companies. So the research objects are:

1. To investigate the current uses of knowledge engineering system and management of Thai local software companies by using interview technique for gathering the information and data in case of process and human management.
2. To deploy and integrate technical standards (ISO25000 Software Product Quality Requirement & ISO9126 Evaluation of Software Quality) with ontology-based Knowledge Management then create knowledge company and Knowledge sharing for pilot project.
3. To investigation the result of using the tools and techniques (2.) from document and report such as non conformity report (QA), customer satisfaction.
4. To reused Software asset for new projects and to compare the results of using the tools and techniques (2.) between small and medium project

The assumption of this research, *"Will the Domain Ontology-based Knowledge Management and SQuaRE be effective tools for Thai local software companies?"* Because Thai culture is not used to show and share idea and opinion. Most of Thai trusts in Seniority and the Hierarchy system.

V. INITIAL FINDING FROM VSE

Most of Thai local software is very small enterprise (VSE: Very Small Enterprise) which has employee less than 20 people.VSE has common problems related to the management of risk and quality of software projects. This generates cost overruns, time delay and cancelled projects. Some model like the CMMI is not affordable for the small organizations.From the survey of the Association of Thai Software Industry (ATSI) in 2008, there was found that the most factor of software project fail was ***"Requirement Quality"*** even using software standardization in production. The model like the CMMI is not affordable for the small organizations.

Figure 4. Different Characters of Small and Medium Enterprise. [7]

The problem from Survey of VSE: Case Study Software company in Samut Sakhon. The historical data was collected for three years for finding the problem and survey the other evidences.

ISBN 978- 974-672-556-9

TABLE 1. VSE PROJECT DETAIL OF VERY SMALL ENTERPRISE (VSE)

Detail	2006	2007	2008	2009
Production Employee	8	14	10	8
Total Project	13	23	20	24
Total SW Sale (Million Baht)	3.33	5.60	4.20	2.80
Production Process Stanard	No	ISO 9001:2000 CSPM	ISO 9001:2008 ISO/IEC15504 (TQS)	ISO 9001:2008 ISO/IEC15504 (TQS)
Total Problems	32.0	50.0	34.0	33.0
Average Man/Pro.(Resauce)	1.6	1.6	2.0	3.0
Average Prob./Man(Performance)	4.0	3.6	3.4	4.1
Average sale/Man(Productivity)	0.4	0.4	0.4	0.4
Customer Satisfaction	2.8	3.76	3.85	3.85

From the data, there are found that

1. The Optimized man power per project should be two developers for each module because it gives the best performance, high productivity, biggest customer satisfaction and also generates the least problems of the projects for VSE.
2. Add On is the major problems of the project which is from customer requirement and cause many effects to project management especially cost and time.
3. Most requirements problem come from not clear in requirement specification, no method for eliciting requirements and unexpected impact from the changing and not good planned Design Interface.
4. The data shows that even use the ISO15504 as Software Process Assessment and ISO9001:2008 as management systems but some problems still exist. It is shown that the root causes are needed to validate and verify.

When take more information in dept interview with the project manager/system analyst, they said that "*some problem come from the conflict between users. In the Thai culture, the superior will make decision and every one must accept them without question. Uncertainty and fear of failure and responsibility. The users felt the statement they gave could be as evidence if the systems did not come out right.*"

When take a look on customer side, they said that "*it is hard to explain all detail and the expected need because they don't know how the software is going to work and look like. When see the blue print (ER Diagram and other documentation) they seem to be right. But when test the software, there will be some little add on that both side never talk and it will bring many problem to the project.*"

The above initial findings motivate for new concept for VSC in term of knowledge engineering and management to capture, identify and suggest for the new concept and tools for Thai local. There are some specific tool such as ontology-base and SQuaRE which don't mention for requirement quality in use

TABLE II. INITIAL FINDING OF THAI LOCAL SOFTWARE

Detail	2006	2007	2008	2009
Total Project	13	23	20	24
Packaged Software	12	21	18	23
Customized Software	1	2	2	1
QA Bug	18	13	14	11
NC Bug	13	30	20	16
Add ON	32	50	34	33
User Error	4	16	12	7
SUM Problems	67	109	80	67
Problem : Project	32.0	25.0	17.0	33.0

The proposed methodology for software pilot project composes of four steps which are

1. Ontology technique for knowledge gathering, understanding and sharing between customers and project manager/system analyst for requirement.
2. SquaRE and Software Quality model for analyzing and design phase.
3. Set up pilot project by using the analyzing and design from step 2.And then using knowledge creation and Knowledge sharing for software knowledge management during production phase.
4. Test run the project to see the results and customer satisfaction.
5. Domain implementation (*Domain ontology*) by reused assets for new project. The domain will be in two different target: *Neo SME (small project)*and *Neo Cold Storage(medium project)*

From the research plan, the expected results should come with some template and reports as see below.

TABLE III.EXPECTED OUTCOME FOR THAI LOCAL SOFTWARE.

Steps	*Tool&Technique*	*Result*
KM , SQuaRE Implementation	• Ontology Base • Requirement Collecting	Standard Template
Pilot Project Analysis , Design & Production	• ISO 25000 • TQS	• KPI Report • Project Validation & Verification
Pilot Project Implementation	• NEO SME • Neo Cold Storage	• CRM Report • Std. Report • Domain Based Component
Reused Asset (Domain Engineer)	• Demo Project	Develop reusable assets (include code, test cases, documentation.)

ISBN 978- 974-672-556-9

VI. Discussion and Conclusion

The assumption of this research is ***"****An Integration of Domain Ontology-based Knowledge Management and SQuaRE will be effective tools for Thai local software company"* because Thai culture is not used to share idea, opinion and afraid of new decision making. Most of Thai trusts in Seniority and the Hierarchy system. After investigation the data, the project manager/system analyst does need tools and methods for gathering the information which made the connection between the software developer and customer. Without a well-written requirements specification, developers do not know what to build and users do not know what to expect. There should be some identified techniques and tools for enhancing elicitation in quality. Most activity between the customer side the Thai local software side use face-to-face interaction and formal process were avoid or ignored while there were delays in decision making from management so the product implementation will be "bottle neck" by software developers. But in global case, where electronic communication is necessary, formal approved are enforced and team member always provide clarification of requirements and prevent further misinterpretations prior to development**.**

It is known that ontology in Software Engineering can provide a source of precisely defined term that can be communicated across people or organization and also provide shared understanding concerning**.** So this research aims to set up new model for requirement quality for VSE which will be value for both Thai local software house and the customers.

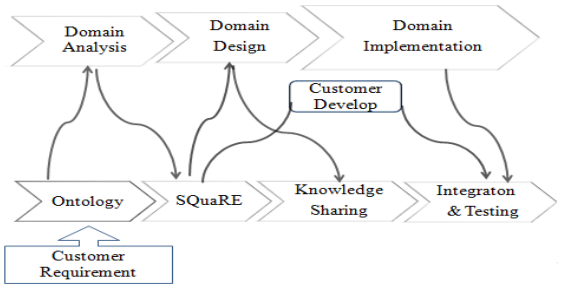


Figure 5. Research Model: Integration of Domain Ontology-based Knowledge Management and SQuaRE for Thai local softwarecompanies.

The integration of Domain Ontology-based Knowledge Management and SQuaRE for Thai local software companies for stage of customer requirement should has three steps which are

1. Data Preparation: Plan and cooperation for constructing the requirement specification and also the agreed set of project requirements and expected products is delivered to the customer.
2. Data Collection: Ontology technique for knowledge gathering, understanding and sharing between customers and project manager/system analyst for requirement.
3. Data Analysis: Identification of quality related requirement for quality evaluation (using framework of SQuaRE) such as
   - Review customer requirement and constrain.
   - Check function and capability of software.
   - External and internal interface.
   - Safety, security and human factor requirement.
   - Requirement evaluation criteria.
   - Verification and validation of requirement.
   - Version control etc.

VII. Limitation

The elicitation of requirement is a difficult process. The most common problems that hinder the identification/definition of the user's need are

- Poor communication of both verbal and written communication.
- Resistance from user to give time, information and responding help.
- Articulation/expertise from using technical term from both side (customer and developer).Taking times to open up and learn from each other will bring a number of benefits.
- Perspective differences because of different background and domain.

The limitation of this stage is that both customers and software house should have the point of view that knowledge sharing of individuals are most commonly rewarded for what they know and what they share. If knowledge is not shared, negative consequences such as isolation and resistance to ideas occur. To promote knowledge sharing and remove knowledge sharing obstacles, the both organizational culture should encourage discovery and innovation. This will result in the creation of organizational culture by using ontology technique for knowledge eliciting and gathering.

Acknowledgement

The study was conducted as a first year project at the College of Art, Media and Technology (CAMT), Chiang Mai University, Thailand. The author would like to thank all CAMT colleagues and staff who support this work and more thank for cooperation and information from the project managers and system analyst of software houses in Samut Sakhon.

References

[1] National Electronic and Computer Technology Center(2008),"*Thailand Information and Communication*

Technology Marketing Survey 2007," ISBN 978-974-229786-2,pp.22-31.

[2] The Association of Thai Software Industry meeting "*ISO29110Intro-Outsourcing Software Engineering Life Cycle Profle,"* 2008, Bangkok.

[3] Henry Caballero, Jose Calvo-Manzano, Gonzalo Cuevas, Tomás San Feliu.(2009), "*Quality and Productivity Analysis a very small enterprise using TSPI*" Retrieved August 01, 2009 from http://www.foibg.com/ibs_isc/ibs-10/ibs-10-p16.pdf2003.

[4] David Zubrow, "*Measuring Software Product Quality: the ISO 25000 Series and CMMI "* Retrieved January 20,2010, from http://www.sei.cmu.edu/library/abstracts/ presentations /esepg.cfm/

[5] Dalkir, K. (2005). "*Knowledge Management In Theory And Practice. Jordan Hill,"* Oxford: Elsevier Inc: 132-133.

[6] Alain Abran and Niji Habra, (2005), *"An information model software quality measurement with ISOStandard,"* from www.info.fundp.ac.be/~nha/Monsite/Rek(Square).

[7] ThreeasakThanasankit (2000), *"Toward understanding mapping requirements: A case study of a Thai software house"* from *http://www.engineering A portal.acm.org/ citation.cfm?id=644417.*

[8] Software Process Improvement for small and medium enterprise: Technique and case study ISO29110 Process Engineer

Sustainable Agricultural in Thailand by GMM

Chatchawal Telavanich, Tirapot Chandarasupsang and Nopasit Chakpitak
College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, 50200, Thailand
info@qualitygreen.com

***Abstract*—In this paper will do research on solving problem of Thai agricultural by try to solve poverty problem by sustainable way. Many theories have been applied to Thai agricultural but in this paper will concentrate on Maturity Model, which has successfully and widely use in many domains, especially in software engineer. The maturity model will be the reference model for grower maturity model (GMM). It will be a guideline for the growers to improve their process, knowledge, and task for their sustainable crop. The model will descript the maturity level of the grower, the necessary goal and task will be described. The theory of knowledge creations will be used to capture the knowledge from the growers then sharing and learning. The behaviors of grower should be changed once they have learned from other growers, who have success and has higher maturity level. The grower maturity model expected to be as a guideline of development for sustainable of Thai agriculture.**

Index Terms-Agricultural, Behaviors, Grower Maturity Model, Knowledge Management, Maturity Model, Sustainable Agriculture

I. INTRODUCTION

Thailand is land of agricultural. Since, agricultural is real sector of Thai economy generated 11.64 % of GDP (OAE 2009). The trend of agricultural area is declined and the usage of agrochemical is increasing by try to boost the crop yield rate. To increase crop yield rate by using agrochemical is polluted to the environments, risk of the farmer health, and increased of crop expends. Appling agrochemical to boost the yield rate is difficult to keep the same yield rate level. It is not sustainable for Thai agricultural sector. The most of farmer lack of knowledge on effected from using of agrochemical and poor management.

Improving quality of life and keep agricultural business sustainable many theory and method had been applied to Thai Farmer. The organic farming is one of the idea has been raised but it is not immediately accepted by most of the farmers. Since, the information farmer has no enough information how to implement the organic farming and good knowledge of crop management. Therefore this paper will do some research on Maturity Model for Thai fresh producer. If the growers have some model how to develop sustainable farming. The maturity model is also one of alternative guideline of agricultural improvement. The higher maturity of the grower is also the key of success factors in sustainable agricultural farming. In this KM project paper will do research on only small part of grower maturity model.

To improve maturity level of grower, we need to analysis and find out characteristic of grower, learning behaviors and how to manage the know ledge of the farmers. In this paper, we analyses learning behaviors of the farmer, who grow the horticultural crop especially for okra grower in Samut Sakhon province, Thailand. By using interview tool, it is to acquire the problem and current situation. Then the problem be analyzed and synthesis. Furthermore, the knowledge map will be created by using KADS knowledge model. The grower maturity model will be analysis and assessment with GAP standard (good agricultural practices)

II. LITERATURE REVIEW

From literature review on learning behaviour, we found out that the learning behaviour of knowledge worker is importance to teach them how why and how to learn. Referring to KADs and Common KADK theories, the process of knowledge audit have been used to identify the problem and opportunities. Then the business flow will support on analysis and synthesis the task knowledge, inference knowledge and domain knowledge. The synthesis of business flow will have the step of capture, analysing, Modelling, Utilization. Then the knowledge after synthesis will be created knowledge map, which could be expert system, decision support system, or case base reasoning. In this paper for concept of grower maturity model will be on decision supporting system. The grower maturity model should have manual for the grower to develop their crop on framework of maturity model.

The knowledge in maturity model will be acquired from tacit, explicit knowledge, the combination of knowledge internalization, socialization, externalization will be adept and to be competency of each level in grower maturity model.

The step of learning by Gagne will be suitable for motivated Thai fresh producer to learn and accept the concept of grower maturity model. Thai fresh producer should be motivated by success case base

ISBN 978- 974-672-556-9

reasoning. To present the success case and out come from base practice could motivated to producer to interested and accept the concept idea. Then fresh producer can be memorize the concept and task it will be ready to be use and applied for better performance and able to evaluated by value of outcome.

III. METHODOLOGY

The methodology is classified in to four stages which are as follows:

1) Selected the target group.

When we analysis the stakeholder of fresh produce supply chain, which as grower, processer or exporter, wholesaler, retailer, and consumer. We found out that exporter is one of the importance chain which could drive the demand and create more value for fresh produce. Moreover, the production of fresh produce is wild rage then we selected only the okra export. Since the okra is market is very unique. It is about 95% of Thai Okra is export to Japan, which has very strict on chemical residual and the supply chain is shorter than other produces. The okra exporter had been selected to do In-Dept interview in the next step.

2) In-Dept interview.

Once the target group had been selected, we applied transcription technique. The questions will be open questions and have the sample of question as follows:-

Probe Code	Question Template	Effect
P1	What are the key factors to select the farm?	Attitude, Environment
P2	Why the growers prefer to do contact farming?	Stability on demand and pricing
P3	Do any standard required for growing fresh produces?	Domain Knowledge, and strategy
P4	Do the relation between exporter and growers are necessary to be considered?	Relation, Collaboration, Culture
P5	What are next that the necessary steps that growers should do?	Vision, Goal

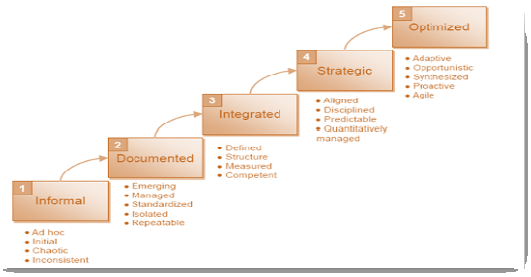
3) Analysis.

Analysis the need and potential of capability of improvement of okra producer in framework of maturity model to be knowledge modeling.

The information from interview will be analysis and have found point of problem that necessary and need to be solved or improved. Moreover, the annotation technique is applied for Knowledge Mapping.

According to reviews on maturity model, it seems that maturity model will be match to be a tool of improve on the knowledge and skills of the grower. Furthermore, it could be change affective on attitude, ethical, and culture of the growers. Since the level of maturity model is have five levels. When it had been applied to use with the grower of fresh producer, we need to use it five levels. Since, in each level can be clarified the ability of the fresh producer growers and can be the guide line for their improvement. The maturity model has the concept of TQM by Deming, Juran, Crosby, and others as pioneer. Which have concept of planning, doing, checking and acting? The process can be measured, it can be understand. Once it can be understood it can be controlled. Once it can be controlled, it can be improved.

Then the necessary knowledge such as task, skill, and attitude of the fresh producer will be classified in to each level of maturity model. The five stage of maturity model will be as below figure.



- *Informal (ad hoc, chaotic, inconsistence)* In this level the fresh okra produces or growers will be using their own experience to grower their crop. The grower will have not enough information. In this level is usually difficult to retain the capability of the crop yield. They have no planning for setting the result. Their success is seems came from luck.
- *Documented (manage, standardize, repeatable)*In the level of document, okra fresh producer still not ready to be standardize. Since they still learning and accumulate knowledge and bring some explicit knowledge to improve and work.
- *Integrated (defined, measured, competent)* At this level the stable practice activities support from individual knowledge. The activities are integrated
- *Strategic (predictable, quantitatively managed)* The strategic level the okra producer must be able to develop the framework and standardized. The fresh producer can manage the crop quantitatively.
- *Optimized (adaptive, synthesized, opportunistic)* The optimize level okra producer is focus on continuous improvement. The okra producer will use the information and knowledge from their quantitatively manage in previously to be apply and guide to be in the optimized level.

ISBN 978- 974-672-556-9

Then behavioral characteristic, process area, key task, and measurement for maturity model have been designed.

4*) Verify the result with the exporters.*

Once we have the model of the grower maturity model, it will be re-interviewed with okra exporter for their comments once again. It is not only exporter but also the other member of stakeholder like grower, or customer will be present the idea and bring advice and comment for further work and improvement.

IV. EXPERIMENT

Any Capability Maturity Model is derived from Humphrey's original maturity framework. In this paper is also applied from maturity framework to be Grower Maturity Model (GMM). The GMM also modified from other maturity model framework like People Capability Maturity Model (P-CMMI), Process Management Maturity Model (PMM) and Knowledge Management Maturity Model (KMMM) from Siemens AG / Corporate Technology.

The GMM will be concreted on the behavioral of grower in different level. The necessary tasks of grower is also mentioned, what kind of task that the grower should have knowledge. Furthermore, the grower in different level may have different of domain knowledge.

The concept of grower maturity model will point out the step of development of grower. Tacit knowledge of grower will be developed to explicit knowledge. Internalize will be externalized. The learning style will use the theory of knowledge creation of Nonaka is been applied. The grower will be check with their tacit knowledge and to be concreted their individual tacit knowledge in the documents or form which will be explicit knowledge of the grower. Then the loop of learning of grower will be the combination and go on form internalization, socialization, externalization. The learning process will be move by loop of learning by using the maturity model to be a tool or guide line of their process and capability improvement.

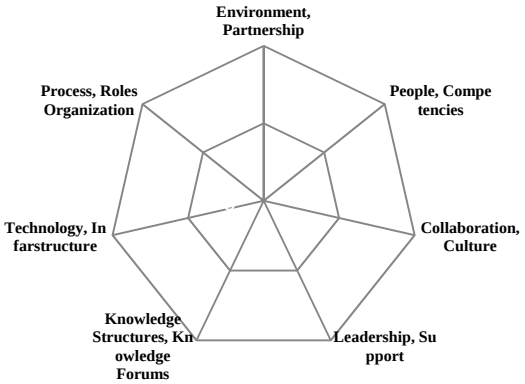
A. The Key Area in the Analysis Model

The Grower Maturity Model will be working on eight key areas of Knowledge Management. (KMMM, Siemens 2002) The Key area will be check at different perspective which is as below table.

Perspective	Key Distinction	Key Area Pair
Time	Strategic vs. Operation	Strategic, Knowledge Goals vs. Leadership, Support
Knowledge	External vs. Internal	Environment, Partnership vs. Knowledge structure, Knowledge forums
Actor	People vs. Technology	Staff, competencies vs. Technology, Infrastructure
Rules	Informal vs. Formal	Cooperation, Culture vs. Process, Roles, Organization

Source: Holistic Development of Knowledge Management with KMMM

The above four perspectives are descript activities which will be transformed to be eight key area of knowledge base management. These eight key areas are based on the EFQM (European Foundation Quality Management). The mode model will be representing to different Knowledge Management specific aspects.



Source: Holistic Development of Knowledge Management with KMMM

The above key area is indecated the realtionship between beside key area the key areas with the opposite side of direction is means of conflicness of the key area.

- Strategy, Knowledge Goals: This topic will be discrip the coorperate vison and goal-setting. The beheavioud of top mandagement, budget policy are analized
- Environment, Partnerships: The topic is related to participants from out side of the organization and aspected also covered to customer and stakeholders. Comparison with other enterprices and problem of using external knowledge.
- People, Competencies: This key area deal with knowledge management which concern individual or personal topic like responsibility management and self management
- Collaboration, Culture : It is significant influence on knowledge management of an oganization It will discrip on cooperate culture, communication

and team structure or noter realtionship structure. The collaboration and culture is less explicit by manger or other leading actor.

- Leadership, Support: The key ares of leadership is such as management model and agreement or targets. It is roles palyed if manager and other staff.
- Knowledge Srtuctures, Knowledge forms: In this area is will discrip aobut specific domain knowledge classification criteria for knowledge and documents.
- Techonology, Infrastructure: This key area will deal with aspect of information and support domain system and process.
- Process, Roles, Organization: This is to discoverhow knolwedge mangement activities ca be added to the specifict business process.

According to each key area we need to designed the process area for each key are and the process area must be able to measure and evaluated which maturiy level that they are in. The processs are will adapt and applied from the framwork of People CMM. The People CMM have the following components.

- Maturity Level
- Process Areas
- Goals
- Practices

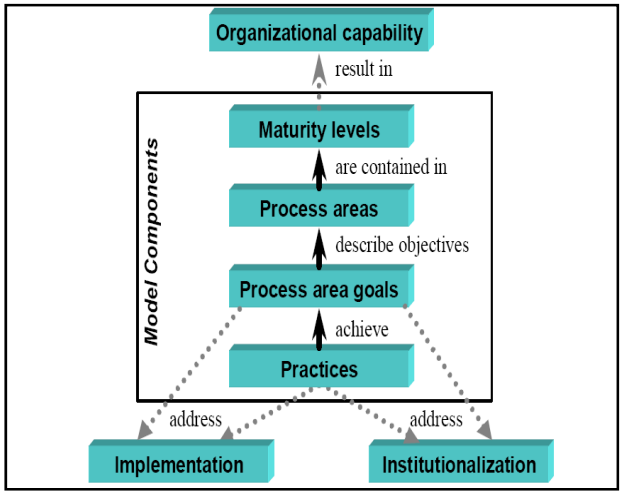
Process Area

The process are is a related practices that when satisfy to the set of goals that contributed to achive next maturity level.

Process Area Goal

The process area goal is that when the organization have achived implementing the practices of a proceas area.

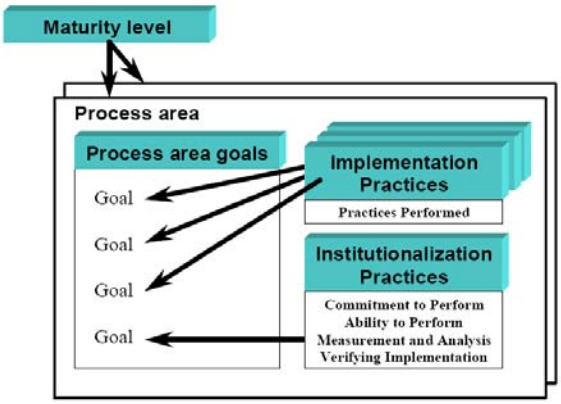
The following table will show the sample of the concept of grower maturity model with process area.



Maturity Level	Focus	Process Area
5 Optimizing	Continously improvement and align personal, workforce organization capability	Continous workforce innovation. Organization Performance Improvement Continous Capability Improvement
4 Predictable, Managed	Empower and Intgrate workforce competency quantitatively	Mentoring Organizational Capability Management Quantitative Performance Management Competency base assets Empower Workgroup Competency Intregration
3 Defined	Develop workforce competencies and workgroups then alliance with strategy and objectives	Partipatory Culture Workgroup Development Competency base practices Career Development CompetencyDevelopment Workforce Planing Competency Analysis
2 Repeated	Responsibility for managing and improveing their people	Compensation Training and development Performance Management Work Environment Communication and Coordination Staffing
1 Initial	Workforce practises applied inconsistently	

B.The Insituationalizeation Practices

The Institutionallization Practices are practices that help the insitutionaliz implement the practize is priactise in the organize culture for effective, repreateable, and lasting. These is will driscrip what should be done to support insituationalization process area.



The Insittuationalization practices are base on four catogatries. The commitment and ability to performed practice and discrip the requirement for impremenring in each area. Measurement and Analisis, and Verifying Impremenrtation. Then finally we will have poractices stateemtn to be as a gideline for the grower for improve their maturity level for sustainable fresh produce. Anyhow, Initialy we may develop only one process area and chek verify the result and commetn from other stakeholder as last stage of methodology.

ISBN 978- 974-672-556-9

V. The Assessment Model

The Model will be assess by using the model of framwork of ISO15504-2 and ISO12207 to eb as refrence model for Grower Maturiy Model Assessment Model. Moreover, the Grower maturity model assessment model will chek whether is could be complied with Good Agricultural Pracice (GAP) or not.

VI. Limitation

At this level we are still unable to clarify necessary task due to limitation of time. The information from stakeholder is limited. We need to do more interview with other parties of stakeholder, who has profit and lost from the fresh produce supply chain. Moreover to deal with fresh producer or grower if the procedure is too much and too details, it might not be interested for the growers or producers. It might be need to comply with current standard like Good Agricultural Practice (GAP) or GLOBAL GAP.

VII. Result / Analysis

According to our previous interview and analysis, we found couple point that to confirm that the poverty of Thai Okra Fresh producer is came from many factors. Lack of Knowledge is one of the key success factors. To solve problem on knowledge we need to analysis knowledge worker or fresh producer from learning behaviour and synthesis the behaviour to be the task knowledge, inference knowledge, and domain knowledge. The most difficult part will be to convert behaviour to process are and then determined necessary task, inference, and domain knowledge. Moreover, the key performance indicator of each maturity level is not been done. Without key performance indicator of each level of grower maturity model, we will have no evident to classify the maturity level.

Furthermore, improve business process of fresh producer according to the framework of grower maturity model, we need to have competency base of knowledge worker. We need to find out competency of each level Otherwise; the knowledge worker will have no idea how to improve to higher level. They may have only concept but have no details how to go. Therefore, the grower maturity model should provide *competency dictionary or competency manual for* fresh produce for their business improvement as the objective of the grower maturity model.

VIII. Discussion

The Maturity Model is theory of performance and process improvement (CMM) Capability Maturity Model of Software Engineer Institute, Carnegie Mellon University. The CMM Capability Maturity Model is had been applied into many business and processes. The model is help to guide and solve conflict, miscommunication, or misunderstanding. The model is can reduce the cost. The concept of grower maturity model is applied from the concept of TQM and CMM. Since, most of the process area can be improved by concept of planning, doing, checking, and action. The planning stage the grower will be able to plan for process, strategic, and goal. Once, the producer has goals and strategy. They will have target know where to go and how to achieve in higher maturity level. Then doing is the procedure will guide the producer how to process or how to do to complete the requirement of current stage. Next stage is checking. This stage is will conform that all process had been done correctly. Then with finding of the previous stage of checking will be bring to action stage. If the result from checking is not conforming to the reference model or target, the producer will consider to the loop of planning step and doing, checking once again. It will be continuous improvement.

The grower maturity model will clarify the step of improvement and guide the producer to know how to increase their maturity level. The initial level of grower maturity model will be on individual expertise. The producer will difficult to retain their talent

In the next maturity stage will be repeated stage, at this stage the produce will have action plan and process will be done according on the manual. The producer will concentrate on their people. The workforce will be trained and compensation base. The working environment will be managed properly.

Next stage is defined. This stage will focus on the competency and producer will have strategic and goal. The process area will participate on culture and working group. The worker will work on competency based practices. The working process should comply with Total Quality Management (TQM) and Good Agricultural Practices (GAP) code of conduct. Competency of organization, personal, function will be analysis.

Next stage is managed level, the will be focus on empower and integrated workforce competency and manage performance quantitatively. The process area in this level will be on increase and measure capability of their worker or able to use knowledge and experience or feedback to develop their capability and measure their capability.

The highest level of maturity model is optimizing stage. It focuses on continuous improvement and innovation. The processes are will be concentrated on created new process and able to develop new concept idea or products. Once the fresh produce be able to develop their maturity according to grower maturity model, we wish that the producer will have better condition of leaving and have sustainable agricultural for Thai fresh producer.

ISBN 978- 974-672-556-9

The maturity model seem step be suitable for Thai fresh producer but have to be modified on behaviour characteristics of fresh producer. Five level of maturity concept should be remained. Since in each level have the meaning and definition of their level.

CONCLUSION

The grower maturity model should be developed in to details. The other part of stakeholder should be interview and analysis to have the result cover from the whole supply chain of fresh produce. The government section also should have concern about the model. Since, the government section will be one of key drive of using the model. If the government accepts the concept and theory of maturity model, it might be able to propose to the committee of Thai Gap to add grower maturity model as one of the key performance of maturity of the grower.

Referring to the behavior of each level will be analysis and synthesis descript to process area of each level. The process area should be clarified all necessary should be mention in every level of grower maturity model. The key performance indicator and competency of each level should be also clarified.

ACKNOWLEDGEMENT

This paper is conduct as the first year project paper for Collage of Art Media and Technology (CAMT). The author would like to thank you for all support from the staff and advisor for this paper.

REFERENCES

[1] Curtis,B.;Paulk,M.;Weber,C.;Chrissis,M; "The capability maturity model: guidelines for improving the software process" Software Engineering Institute, Carnegie Mellon University

[2] Curtis,B.;Hefley,B.;Miller,S ."People capability maturity model (P-CMM)" Version 2.0 CMU/SEI-2001-MM-01, Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, July 2001

[3] EFQM Excellence Framework (available at: www.exchange2improve.com/about/efqm-excellence-model-framework/)

[4] Fundamental Data 2009, Office of Agricultural Economics (OAE), ISBN 978-974-403-626-1

[5] Greenpeace report "Agrochemicals unmasked: fertilizer and pesticide use in Thailand and its consequences to the environment" Feb 1, 2008 (available at www.greenpeace.org/seasia/th/press/reports/agrchemicals-in-thailand-eng)

[6] Kennedy, P. Lynn, Rosson. (2002) Impacts of globalization on agricultural competitiveness: The case of NAFTa, Journal of Agricultural and Applied Economics

[7] Karsten Ehms, Dr. Manfred Langen, Holistic Development of Knowledge Management with KMM, Siemens AG / Corporate Technology, Knowledge Management & Business Transformation

[8] Tattanakorn M, & Kumar P. "Baseline survey export okra production in Thailand" Inter-Country Program for Vegetable IPM in South & SE Asia Phase II, Food & Agriculture Organization of United Nations June 2004

A Tourism English Teaching Model for GMS Universities

- A Knowledge Management Perspective

Fu Jing,
PhD Student
College of Arts, Media & Technology
Chiang Mai University, Thailand

Nopasit Chapitak, PhD
College of Arts, Media & Technology
Chiang Mai University, Thailand

Wided Batat, PhD
University of Lyon 2, Lyon, France

***Abstract* - The Greater Mekong Subregion, short for GMS, has been proved the fastest growing tourism destination since 2004. Consisted of six important developing countries of Southeast Asia with abundant tourism resources in this area, GMS has been attracting more and more travelers worldwide. In order to have a foot in the global tourism market, English, as a universal working language turns to be a crucial factor to success. Since personnel function is always one step ahead other business functions, tourism English education in local universities must unavoidably take responsibilities to provide qualified tourism workers. This paper is a conceptual framework with initial findings, which will focus on tourism English education, one of the most important factors in influencing the internationalization of tourism, hoping to find the gap between what tourism industry are requiring and what tourism education are providing from both industrial and educational perspectives. A case study will be placed on hospitality, then, a teaching model will be proposed for GMS universities from a knowledge management perspective.**

Index Terms— Cooperative Education; GMS; Knowledge Management; Tourism English Education

I. INTRODUCTION

The Greater Mekong Subregion, short for GMS, comprises six countries namely Cambodia, the People's Republic of China, Lao People's Democratic Republic, Myanmar, Thailand, and Viet Nam, which covers an area of 233.19 square kilometers with about 240 million populations. Studies showed that GMS is the fastest growing tourism destination since 2004, bringing in much needed foreign exchange, creating jobs and contributing substantially to economic growth. GMS countries have a strong comparative advantage in tourism with abundant natural resource, unique cultural diversity, rich tangible and intangible heritage and national policy priority which provide significant opportunities for tourism development. Unlike other domestic or international tourism, GMS tourism has been increasingly multi-country with subregional cooperation. It gathered tourism resources and human power from 6 countries, which provided tourism products as one large scale of regional tourism package and showed up as a single destination to the whole world. In 1992, since the inauguration of GMS economic cooperation, six GMS countries endorsed to place primary priority on GMS sustainable tourism. In order to have a share in the international tourism market, it is urgent for GMS countries to improve the quality of tourism products and services by transferring "the place for tourists" to "the place of tourists"(Maggi & Padurean, 2008) . Today's tourist are not "old tourists" any more, who often take package holidays with lower ability in computer and language, but increasingly belong to the kind of customers who favor online booking and are normally well-educated with more disposable income and buying power. These customers are called as "New Tourist" by Poon (1993) and Buhalis (2003). Poon defined new tourists as people who are seeking meaningful experiences by discovering new places and cultures; 10 years later, Buhalis then summarized 5 characteristics, named as "5S" for "New Tourist"- namely Segmentation, Specification, Sophistication, Satisfaction and Seduction. Buhalis stated that what new tourists need is gradually transforming from the interests of summer holiday, urban tourism to a mixed demand of technology, edu-tainment, eco-environment and multi-culture. They need more individualized, authentic, and enriched travel experiences to meet their educational, cultural, intellectual and sporting interests. To transfer "the place for tourists" to "the place of tourists" needs a better understanding of the requirement and expectation of these new tourists by providing a "unique and memorable experience" (Pine & Gilmore, 1999) which is a new dematerialized commodity that will generate increasing returns (Stamboulis & Skayannis, 2003). Study showed that it is five or six times more effective to attract repeat customers than to gain new ones (Ritticinuwat, Mongkonvanit & Qu, 2002), this appeals to a higher developed tourism mechanism and well educated tourism working people. The ability of understanding to anticipate the needs of customers along with the skills of communication to find "Unique Selling Point" (USP, Wood & Jayawardena, 2003) will be largely decided by the arts of expression. English, as the most important working language in international tourism will no doubt improve competitiveness and productivity of tourism industry. Maggi & Padurean (2003) made a thorough investigation on how

ISBN 978- 974-672-556-9

important English acts for international tourism, they pointed out that all the top 20 tourism countries laid huge emphasis on tourism English education, on the other hand, international competition pushed universities to set up internationalized strategies which pay more attention to English education.

However, today's tourism, as a dominant tertiary industry in GMS countries suffered more from a shortage of qualified employees who can make full use of English as their working language to appropriately express, convey and communicate their knowledge and experience etc. meet the sophisticated requirement of international travelers. Current tourism education emphasizes in the supply of a labor force but less attention is given to the development of the quality of human resource for this service industry. As the cradle of the talents at a higher educational level, local universities should take this responsibility to provide competent tourism workers with abilities and skills of working in English. Then these people will continue to contribute to a real sustainable tourism industry for GMS countries.

China (South-western China ad hoc) and Thailand (Northern Thailand ad hoc), two important countries of GMS area, are playing irreplaceable role in the three economic corridors of GMS area, namely East-West Economic Corridor (EWEC, including Thailand, Vietnam and Lao PDR), North-South Economic Corridor (NSEC, including South-western China, Lao PDR, Myanmar and Vietnam), Southern Economic Corridor (SEC, including Thailand, Cambodia and Vietnam). Both of these two countries agreed on cooperation and strategies to optimize the benefits of sub regional tourism in a more environmentally and socially responsible manner. Both of the two countries placed great efforts and policy priorities on tourism English education by intensifying college English reform and strengthening English for Specific Purpose (ESP) program. However, studies showed that a gap still exists between what tourism industries need and what universities provide. This paper will focus on identifying this gap and then design a teaching model for tourism English with the aid of theories and methods of knowledge management.

This research is a conceptual framework with qualitative and observational efforts, which will choose hospitality English as a purposeful sampling, trying to explore the gap between what hospitality industries need and what hospitality English programs provide by a comparison between how presently English is taught and learnt in Chengdu University, South-western China and Chiang Mai University, Northern Thailand. The reason why these two universities are chosen is mainly because the writer has been teaching English to tourism students for ten years in Chengdu University, and has observed and investigated how tourism English is taught in Chiang Mai University since 2009. The methodology of the initial findings of the research will be based on a large scale of interdisciplinary literature review and documental study on syllabus and curriculum in tourism English from both industrial and educational perspectives. A number of discussions aiming at the gap identification and analysis for the research will be drawn with the aid of knowledge management as the direction of designing a model for tourism English teaching, and further study will be planned from a managerial perspective.

II. Literature Review

The output of tourism English education should meet both academic and practical requirement of higher education and demanding industry. Therefore, to design a teaching model for tourism English in a more complete and systematic way, perspectives from industry, education, and management must be considered.

A. From an Educational Perspective-Tourism Education

Tourism education originally emerged as training programs for staff in specific sectors, and then has grown to undergraduates and postgraduates courses in college curriculum to equip graduates at a middle managerial level (Morgan, 2004). The goal of tourism education may be different from the goal of tourism industries. Lundberg (1998) identified two configurations of the strategic objectives of tourism education, namely managerial ideology and humanistic ideology. The former aims to achieve business success by applying a commonly accepted body of knowledge and practice to the given fact; while the latter focuses on individual personal development by the use of distinct sets of knowledge and practice to inform a subjective choice. Although Jenkins (1999) stated that academic research seldom influences the real world of practice, yet the goal of higher education is not to teach students how to survive their jobs, but to educate students to think deeply and seriously about their industry so as to be able to plan for the future as well as managing the present. Tribe (2002) stated that tourism education at a higher education level should not only aim at enabling students to become operationally effective in the industry, but more importantly should aim at encouraging freedom and depth of thought about the nature of the truth and the good life. Since the academic world has been ahead of the industry in emphasizing the social and environmental impact of tourism and the need for sustainable development, what tourism education should provide is the "brain power" to what industry need and get ready for that. When designing a teaching model for tourism English, an objective of developing student's sustainable ability in language using must be considered. And leaning in action is absolutely one of

ISBN 978- 974-672-556-9

the efficient ways to gain the ability of lifelong learning.

B. From an Industrial Perspective -Action Learning

The notion of action learning was first conceived by Reg Revans during the late 1940s, he developed the idea that colliery managers who were experiencing organizational problems might learn from one another by talking through their difficulties and personally taking action. Then, Johnson (1998) developed Reg's idea by stating that learning is about changing. In order to adapt to a continuously environmental change and uncertainty, researchers believed that the keys lie in the ability to learn how to learn from, and during experience. (Senge, 1990; Watkins & Marsick, 1993). Action learning is a process that can be used to help individuals acquire these new learning skills (O'Neil, 1996). Debates never stop on what skills are needed in the changing tourism industry, however, communication in verbal, written and IT forms which can be transferred in every possible job location are always on the top of the list. Communicative skills include narrative, story-telling and creating atmosphere along with developing empathy with customers and anticipating customer's needs. All these can be achieved through the use of language and practice in action. The way to remain ahead is by continuously learning and adapting to the environment. (Jameson, 2000) In designing a teaching model for tourism English, concept of action learning from Albert Einstein can be a guide: "I do not teach my pupils, but I provide conditions in which they can learn."

C. From both Industrial and Educational Perspectives-Cooperative Education

Cooperative Education was considered as a "strategic alliance" between universities and corporations by researchers like Elmuti, Abebe & Nicolosi (2005) and Wheelen & Hungar (2000), they believed cooperation education has brought a more complete and systematic model for on-the job training. Law (1970) defined cooperative education as a work program associated with a school education in which student-learners receive supervised payroll experience as part of the school curriculum. Studies (King, 1994 & Santoro, 2000) showed that cooperative Education will foster the work experience so the students will attain the necessary skills to supplement their theoretical training by realizing four transfers namely research support, cooperative research, knowledge transfer and technology transfer. These four transfers will contribute to a maximum of knowledge of tourism needed for both academic and industrial level. Among the factors which will mostly lead to the improvement of tourism training, industry engagement tops the list (Smith & Kemmis, 2010). Garvin (1998) explained the need of the stakeholders of tourism education, namely students, employers and government. He said all these three stakeholders have been influential in moving the focus of higher education from the content determined by the teacher to the process by which the student learns and develops key skills. The key point is how to balance the needs of the three parts by collaboration between universities and industries. When designing a tourism English model, cooperative education is a crucial and indispensible part.

D. From a Managerial Perspective -Knowledge Management

To meet the requirement of both education and industry is not easy, because it needs an efficient management which is a process of knowledge accumulation and creation, and benefits not only universities or industries, but the whole society as well.

The Emergence of knowledge management as an academic field dates back from 1980s. Knowledge is a real experience of knowledge workers especially experts in executing a specific task. Knowledge Management is the making use of working people and useful business information for core business and innovation. (Nonaka & Takeuchi,1995). Many researchers have paid a lot of attention to apply knowledge management to tourism, because they believe knowledge will change supply environment and the nature of consumer's behavior (Cooper, 2006); they stated that destinations can change due to the response of knowledge creating and using. As a competitive tool, the essentials of the process involve identifying relevant knowledge and capturing it, transferring and sharing it, and ensuring that organizations are engineered to optimize flows and to manage them effectively (Bahra, 2001). Senge (1994) proposed learning organization in his "*The Fifth Discipline"* to call for more focus on recognizing everyone's uniqueness which will help people work together more effectively. All these theories left priceless treasure and challenge to universities on how to reform the knowledge in their teaching content and methods from a managerial perspective to meet both educational and industrial requirement. With more qualified and well-educated graduates, both students and universities will be more efficient, competitive and productive.

III. METHODOLOGY

This research will take hospitality industry as a purposeful sampling, two cities from GMS countries, namely Chengdu, an important city of Southwestern China and Chiang Mai, a centre of Northern Thailand are chosen as the initial findings of the research. Grounded theory design is applied in the open-ended interviews to collect data. Field notes, as well as audio-tape record aided to analyze data. Based on the findings and supportive theories mentioned in Part II, a teaching model will be proposed for hospitality English and hopefully will be contributed to a deeper

ISBN 978- 974-672-556-9

understanding of tourism English teaching. Three steps are taken as below:

A. Comparison on Requirement from Hospitality Industry between Chengdu and Chiang Mai

- *Investigation on the Requirement from Hospitality Industry in Chengdu*

Located in Southwestern People's Republic of China, Chengdu is the capital of Sichuan province and one of the most important economic centers, transportation and communication hubs in southwestern China. On February 2009, Chengdu was awarded as one of the "Best Tourist Cities in China" by China National Tourism Administration (CNTA) and United Nations World Tourism Organization (UNWTO). With the predominance of the hometown of the Giant Panda and hub of six UNESCO World Heritage Sites in or near Chengdu (namely Dazu Rock Carvings, Huanglong Scenic and Historic Interest Area, Jiuzhaigou Valley Scenic and Historic Interest Area, Mount Emei Scenic Area, including Leshan Giant Buddha Scenic Area, Mount Qincheng and the Dujiangyan Irrigation System, Sichuan Giant Panda Sanctuaries), Chengdu is undoubtedly one of the most important tourism cities in southwestern China. As a provincial capital city, Chengdu has about 104 hotels with upper 3 stars in the city which accommodated half a million inbound tourists in 2008 and 2009.

Investigation was made through email and telephone interviews among 10 hotels with 4-5 stars in Chengdu, like Holiday Inn and Jin Jiang Hotel, which are more trusted by international travelers. 20 questions were proposed in the open-ended questionnaire to the managers of human resource departments who are gatekeepers with rich information in the research. Through the interview, it is found out that most of international tourists in Chengdu are from America and Thailand with package tour. The number of tourists from Japan, Singapore and other countries from GMS are increasing. With gradually demanding market from European countries, English is still the major working language. When asked to evaluate the English level of new employees who just graduate from the university, HR managers graded 6 points out of the maximum 10 points. They expected and are satisfied more from the on-the-job training in the hotels than what students learned in the universities. They believed that practice in the real situation is the most efficient way to improve English level and hope more courses in listening, speaking, and skills of communication can be emphasized in English programs in universities. All of them are interested in cooperative education, which was misunderstood as internship, and they expect more cooperation with universities.

- *Investigation on the Requirement from Hospitality Industry in Chiang Mai*

Chiang Mai city, the largest and most culturally significant city in Northern Thailand, and is the capital of Chiang Mai Province. In recent years, Chiang Mai has become an increasingly modern city and attracts approximately one million visitors each year. The city has long been a major center for handcrafted goods, umbrellas, jewelry (particularly silver) and woodcarving. Chiang Mai then slowly grew in cultural, trading and economic importance to its current status as the unofficial capital of Northern Thailand, second in importance only to Bangkok. Data from Tourism Bureau of Northern Thailand shows, there are 341 hotels and guest houses in Chiang Mai in the year 2008, and the number increased to 418 one year later. These hotels and guest houses have accommodated 4,405,720 and 4,181,878 travelers in 2008 and 2009 respectively both inland and overseas.

Investigation was made among 10 hotels with 4-5 stars in Chiang Mai, like Mandarin Oriental Hotel and Centara Duangtawan Hotel, which attracted tourists with more buying power and need more tailor-made services. 20 questions were proposed in the open-ended questionnaire to the managers of human resource departments, the most information-rich people. Due to the mature international environment and convenience of access, it is found out that most of international tourists are from European countries and America both as FIT (Foreign Independent Tourist) and with package tour. The number of tourists from Japan, Singapore, Korea and other countries from GMS are increasing. When asked to evaluate the English level of new employees who just graduate from the university, HR managers graded 7 points out of the maximum 10 points. They described an interesting phenomenon in English learning in Thailand, that is to say people are shy to express themselves especially to speak a foreign language. More expectations are from the on-the-job training in the hotels than what students learned in the universities. They believed that external motivation is the key to force these new staffs to use and improve their English level. They hope more courses in listening, speaking, and skills of communication can be emphasized in English programs in universities. They mentioned class atmosphere and learning culture, hoping English teachers in the universities could adjust their way of teaching to encourage students to be brave to open their mouth to express. All of them are familiar with cooperative education, and welcome cooperation with universities.

ISBN 978- 974-672-556-9

- *Comparison and Conclusion*

The questionnaire laid an emphasis on what basic English skills are needed in hospitality industries in Chengdu and Chiang Mai, and how the HR managers evaluate English level of their newly graduated employees. These points directly relate to the requirement of English skills in an industrial context. From the investigation, a comparison was summarized in Figure 1 as followed:

Comparison on Requirements of English in Hospitality Industry		
Items	Chengdu	Chiang Mai
Basic Skills of English needed	listening & speaking,	Listening & speaking,
Basic skills of communication in English needed	way of expression, response to a certain topic	way of expression, presenting a certain topic
Main function of English n work	tool of communication and symbol of international image	tool of communication and selling
Expectation from English teaching in universities	provide students more chances to practice	provide students more chances to practice
Attitude towards cooperation with universities	positive, welcome internship	positive , welcome cooperative education

Figure 1. Comparison on Requirements of English in Hospitality Industry

From the comparison, we can conclude that listening, speaking and way of communication are greatly needed in hospitality industries. Expectation from the industry lies in more practice for students and cooperative training programs with universities.

B. Comparison on Tourism English Programs between Chengdu University and Chiang Mai University

- *Tourism English Program in Chengdu University (CDU).*

Chengdu University (CDU), run by the Chengdu Municipal Government, is a comprehensive university offering regular undergraduate education and benefits from the system of educational provision jointly administered by relevant competent authorities of either Sichuan Province or Chengdu City.

School of Tourism & Culture Industry, founded in 1994, provides Tourism Management (Bachelor), Forest Resources Conservation and Recreation (Bachelor), Landscape Architecture (Bachelor), Tourism Management (Diploma) and Hospitality Management (Diploma).

English Discipline for tourism-major students in hospitality section includes two parts-College English Course (CE) and English for Specific Purpose Course (ESP). CE program aims at improving students' comprehensive English ability and is taught in the first and second year as a general English course by lecturers from School of Foreign Languages & Culture; ESP program aims at enhancing students' ability of applied English and is taught as a specialized course by lecturers of School of Tourism & Culture Industry. The key function and content of the two courses are described in Figure 2 as followed:

English Programs in Chengdu University		
Items	CE	ESP
Semester	1-4 semester	5-7 semester
Objectives	Basic English skills	Applied English skills
Teaching Hours	216 hours/4 semesters	207 hours/3 semesters
Class Size	70-90 students/class of the same major	70-90 students/class of the same major
Lecturer Qualification	Chinese Lecturers from linguistic background	Chinese Lecturers from tourism background
Teaching Materials	Decided by universities, the same as all non-English majors	Decided by lecturers with different specialties
Way of Teaching	Teacher-centered	Teacher-centered
Field work	No	30%
Evaluation	Final exam & CET	Final exam

Figure 2. English Programs for Hospitality Majors in Chengdu University (CDU)

From the figure, we can see, English programs in CDU focus on comprehensive abilities which include listening, speaking, reading and writing. The way to evaluate students' English proficiency is final exams and National College English Test (CET) which both focus on a general competency of English. Problems were found here:

- Few special courses for listening, and speaking practices are limited.
- A lack of foreign teachers from English native speaking countries leads to a lack of oral and aural practice in real situation.
- Both Chinese teachers for CE program from linguistic background and ESP program from tourism background are weak in hospitality English teaching.
- Teacher-centered instruction and big-sized class limited students' participation and proactivity.
- Field work for practice in the real situation is far from sufficiency.
- Few guide for extra-curriculum activities and independent study from teachers
- Terminology intervened expression and communication

- *Tourism English Program in Chiang Mai University (CMU)*

Chiang Mai University (CMU) was established as the first institution of higher education in Northern Thailand, and as the first provincial university in Thailand. CMU aims at becoming a comprehensive institution in Northern Thailand for

the sake of social and economic development of the region and the country as a whole.

Tourism discipline in CMU was set by Department of Tourism of Faculty of Humanities which is responsible for many basic humanities courses offered to students of all faculties since 1964 when CMU was established. English programs for tourism-major students in hospitality section includes two parts-General English (GE) and English for Specific Purpose (ESP). GE program for tourism majors is the same as students from other faculties in CMU, which aims at improving students' basic English ability and is taught from the first to the third year as a general English course by Thai lecturers from Department of English of Faculty of Humanities or English native speakers hired as part time teachers by Department of English; ESP program aims at an ability to apply English to students' major and is taught as a specialized course by lecturers of Department of English and Department of Tourism. Some ESP courses are taught by foreign teachers who are English native speakers. The key function and content of the two courses are described in Figure 3 as followed:

English Programs in Chiang Mai University		
Index	GE	ESP
Semester	1-6 semester	5-7 semester
Objectives	Basic English skills	Applied English skills
Teaching Hours	270 hours/ 6 semesters	180 hours/3 semesters
Class Size	40 students/class of different major	40 students/class of the same major
Lecturer Qualification	Thai lecturers from linguistic background or English native speakers	Lecturers from tourism background and English native speakers
Teaching Materials	Decided by universities, the same as all non-English majors	Decided by lecturers with different specialties
Way of Teaching	Teacher-centered	Teacher-centered
Field work	No	30%
Evaluation	Final exam	Final exam and TOEIC

Figure 3. English Programs for Hospitality Majors in CMU

From the figure, we can see, English programs focus on comprehensive abilities which include listening, speaking, reading and writing. More attention has been put in applied English in ESP Program. The way to evaluate students' English proficiency is final exams and TOEIC (Test of English for International Communication), which tests students specified English in a certain career. Problems were found here:

- Few courses are set for listening and speaking
- Thai culture of being too polite and modest influenced the way of English learning and teaching
- Both Thai teachers for GE program from linguistic background and ESP program form tourism background are weak in hospitality English teaching.
- Although there are foreign teachers from English native-speaking countries, a lack of teaching experience and knowledge of pedagogy have been found
- Grammar intervened expression and communication

- *Comparison and Conclusion*

According to the investigation on the English program comparison between Chengdu University and Chiang Mai University and open-ended interview with teachers of tourism English, what presently tourism English education provided is clearly seen, as it is shown in Figure 4 in the following:

Comparison on What Present Hospitality English Education Provided		
Items	Chengdu University	Chiang Mai University
Basic English Skills provided	listening, speaking, reading and writing	listening, speaking, reading and writing
Basic skills of communication in English provided	involved in conversation drills, not specified in the program	involved in conversation drills, not specified in the program
Teacher Qualification	Chinese teacher, Linguistic background or tourism background	Thai teacher and foreign teachers , Linguistic background or tourism background
Teaching Method	Teacher-centered	Teacher-centered
Content of Textbooks	General English Skills and English for certain situation	General English Skills and English for certain situation
Features of English Program	General	General
Evaluation of English Level	Final Exam & CET	Final Exam & TOEIC
Output of English Level	not satisfied, weak in listening and limited in vocabulary	not satisfied, weak in speaking and limited in grammar

Figure 4: Comparison on What Present Hospitality English Education Provided.

From investigation and interview with English teachers and tourism students, both the sides are not satisfied with the output of English level. Major factors are found as below:

- A lack of special program for each skill in English learning, especially for listening and speaking
- Skills of communication have not been separated as an independent and important course
- As the main source of reading, textbooks are not up-dated and close to tourism according to the changing industry
- Teacher-centered instruction is less efficient to encourage and promote students' participation and creativity
- A lack of guide of how to apply English to future career and how to conduct independent study as a life-long learners

ISBN 978- 974-672-556-9

C. Gap between the Supply and Demand of Hospitality English

The initial finding of the gap between what hospitality industries need and what hospitality English education provides will be a direction of the design of the teaching model for hospitality English. Comparison is seen in Figure 5 as below:

Gap Analysis on Supply and Demands of Hospitality English		
Items	**Requirement from Industry**	**Supply from Universities**
Basic Skills of English needed	focus on listening & speaking,	general skills
Basic skills of communication in English needed	a specific course on skills of communication	involved in conversation drills, not specified in the program
Role of English	tool of a working language	compulsory course
Expectation	trained in the universities in order to reduce training cost	trained in the work place as a facilitator of employability
Output	be able to understand, provide, promote, and satisfy customer's needs by communicating in English	be able to understand and communicate with customers in English
Attitude towards cooperative education	positive	positive

Figure 5. Gap Analysis between the Supply and Demand of Hospitality English

From the gap analysis, it is clearly seen that what hospitality industries need is not only the ability to speak English, but also response in English by understanding and predicting customers' need, and then to provide and promote tourism products and services to satisfy customers and maximize industries' benefits. Obviously, these requirements are not fully met and reached by hospitality English education.

D. A Teaching Model for Hospitality English Education

Four English skills, namely listening, speaking, reading and writing, can be divided into language input (listening and reading) and language output (speaking and writing). However, language communication combines the two-a good quality of input and a proper way of output. Requirements from both academic and industrial perspectives need to link theoretic work with practical sectors. Theories of linguistics and knowledge of tourism and hospitality need to be integrated into the consideration of teaching methodology, learning atmosphere and uniqueness as well as sustainable learning competency of students. A teaching model is then proposed from a managerial perspective which involves all the factors listed above as it is shown in Fig.6.

In this teaching model, each part is indispensible and has irreplaceable function:

- The whole process of hospitality English teaching and learning is put in a Cooperative Education program, an *Advisory Panel* consisted of mentors from hospitality and lecturers from universities will be set up to implement the whole program which is the key to the success of this teaching model

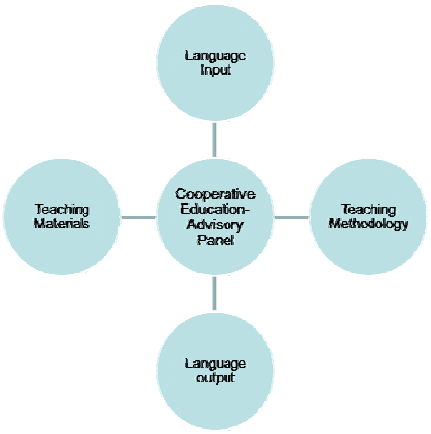


Figure 6. A Teaching Model for Hospitality English Education

- Members of advisory panel will regularly meet, discuss and identify what teaching content and materials will be appropriately introduced to students to enhance language input by listening and reading. Knowledge from linguistics and tourism and hospitality are captured, transferred, shared and codified into teaching planning and course syllabus. For example, the terminology of hospitality, the sentence structure and grammar usually used in the working context. Advisory panel will also decide which way is the most efficient to make students learn and generate new knowledge and skills. Learning in action will be helpful to reach the goal.
- Teaching methodology is very important to create a comfortable atmosphere and learning organization for a better output in speaking and writing. Imitation from real situation in the real workplace and repetition with the guide and instruction of advisory panel will speed up the quality of language output. Learning in action through cooperative education program will facilitate the process of learning and improve the effectiveness of learning.
- The most important function of advisory panel is to train students with system thinking about how to be independent life-long learners, and qualified knowledge workers with high-level English proficiency. Students should be taught to capture new knowledge into their own intellectual capital in English and then to accumulate and generate new knowledge and skills to work better.

ISBN 978- 974-672-556-9

IV. RESULT AND BENEFICIRIES

The research has potential benefits for both academic and social sectors in GMS area. For hospitality industry, this research will be a guide for tourism workers to pay more attention to the gap between how their English was taught in universities and what they need to improve their English efficiency in work. It will also improve the English level by making them know the importance of learning in action and raise an awareness of promoting the quality of services by communicating in English. The effectiveness and competitiveness of hospitality organization will be improved due to reasonable and efficient management in knowledge in English and tourism between knowledge creators and users. For universities, this research will integrate the needs of the industry with universities' English programs under the supervision of adversary panel by implementing cooperative education program. With the help of the proposed teaching model for tourism English in this research, the understanding of English teaching and learning will be deepened; students will be encouraged and motivated to be independent learners, then become life-long learners in their future career. Meanwhile, more qualified lecturers and students will be trained who will bring about qualified tourism workers to tourism industries and a brand loyalty to universities through cooperative education program. For GMS community, this research will bring an innovative thinking about the way of teaching and learning to improve tourism English at both academic level and practical level; so that more tourism working people will better apply knowledge of English and tourism in their routine work and be involved into a sustainable tourism development.

V. CONCLUSION

The ultimate goal of English teaching and learning is to apply English into a specific career. Only when English is used; it is a tool of communication. Communication makes a whole world a global village, with the aid of better understanding by the use of English, tourism destination is not a place for tourists, but of tourists. Only in this way can knowledge be transferred and shared between tourism providers and receivers, and then can knowledge be accumulated and created through cooperation between tourism industry and tourism education. Although this research picked hospitality English as a case study, it can be surely generalized into other fields of tourism. And it can be an indicator to GMS universities who needs a revolutionized way of tourism English teaching. Co-research and collaboration are sincerely welcomed and encouraged to contribute to tourism English education and a sustainable tourism development.

REFERENCE

[1] A. Aksu, "Defining Training Needs of Five-star Hotel Personnel: An Application in the Antalya Region of Turkey", *Managerial Auditing Journal*, Emerald Group Publishing Limited, 2005.

[2] C. Cooper, "Knowledge Management and Tourism", *Annals of Tourism Research*, Vol.33, 2006

[3] D. Elmuti, M. Abebe, & M. Nicolosi, "An Overview of Strategic Alliances between Universities and Corporations", *The Journal of Workplace Learning*, Vol.17, 2005.

[4] D. Gavin, *Managing Quality: The Strategic and the Competitive Edge*, Free Press, NY, 1998.

[5] M. Jameson, "Recruitment and Training in Small Firms", *Journal of European Industrial Training*, Vol.24, 2000.

[6] C. Jenkins, "Tourism Academics and Tourism Practitioners: Bridging the Great Divide", *In Contemporary Issues in Tourism Development*, London: Routldge, 1999.

[7] C. Johnson, "The Essential Principles of Action Learning", *Journal of Workplace Learning*, Vol.10, 1998.

[8] B. King, "Cooperative Education for Hospitality and Tourism Students: An Australian Case Study", *Australian Journal of Hospitality Management*, Vol.1, 1994.

[9] G. Law, *Cooperative Education: Handbook for Teacher-Coordinators*, American Technical Society: Chicago, 1970.

[10] C. Lundberg, "A Prolegomen to Course Design in Hospitality Management: Fundamental Considerations", *Journal of Hospitality and Tourism Education*, Vol. 10, 1998.

[11] R. Maggi & L. Padurean, "Higher Tourism Education in English-Where and Why", *Tourism Review*, Vol.64, Emerald Group Publishing Limited, 2008.

[12] I. Masami, "Evaluating the Effectiveness of GMS Economic Corridors", *IDE Discussion Paper*, 2007. http://hdl.handle.net/2344/632.

[13] M. Morgan, "From Production Line to Drama School: Higher Education for the Future of Tourism", *International Journal of Contemporary Hospitality Management*, Vol.16, Emerald Group Publishing Limited, 2004.

[14] J. O'Neil, "A Study of the Role of Learning Adviser in Action Learning", *Employee Counselling Today, The Journal of Workplace Learning*, Vol.8, 1996.

[15] P. Petrova& P. Mason, "Case Study-The Value of Tourism Degree: A Luton-based Study", *Education + Training*, Emerald Group Publishing Limited, 2004.

[16] J. Pine & H. Gilmore, *The Experience Economy*, Harvard Business School Press, MA, 1999.

[17] A. Poon, *Tourism, Technology and Competitive Strategies*, CAB, Wallingford, 1993.

[18] R. Revan, *Action Learning-New Techniques for Management*, Blond & Briggs, London, 1980.

[19] B. Rittichainuwat, H. Qu& C. Monknonvanit, "A Study of the Impact of Travel Satisfaction on the Likelihood of travelers to Revisit Thailand", *Journal of Travel & Tourism Marketing*, Vol.12, 2002.

[20] M. Santoro, "Success Breeds Success: the Linkage between Relationship Intensity and Tangible Outcomes in Industry-University Collaborative Ventures", *The Journal of High Technology Management Research*, vol.11, 2000.

[21] P. Senge, *The Fifth Discipline: The Art and Practice of the Learning Organization*, Doubleday, NY, 1990.

[22] P. Senge, *The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization*, Doubleday, NY, 1994.

[23] D. Sinclair & C. Jayawardena, "World-wide Hospitality and Tourism Trends: Viewpoint", *International Journal of Contemporary Hospitality Management*, MCB UP Limited, 2003.

[24] E. Smith & R. Kemmis, "What Industry Wants: Employers'Preferences for Training", *Education + Training*, Vol.52, 2010.

[25] Y. Stamboulis& P. Skayannis, "Innovation Strategies and Technology for Experience-based Tourism", *Tourism Management*, Vol.23, 2003.

[26] J. Tribe, "The Philosophical Practitioner", *Annals of Tourism*

Research, Vol.29, 2002.
[27] E. Watkins & J. Marsick., *Sculpting the Learning Organization*, Jossey-Bass, CA, 1993.
[28] L. Wheelen& J. Hungar, *Strategic Management and Business Policy*, Addison-Wesley, Reading, MA, 2000.
[29] P. Wood & C. Jayawardena, "Cuba: Hero of the Caribbean? A Profile of Its Tourism Education Strategy", *International Journal of Contemporary Hospitality Management*, MCB UP Limited, 2003.
[30] "GMS Tourism Development", Greater Mekong Subregion, ADB, http://www.adb.org/GMS/default.asp
[31] "GMS Tourism Strategy", Greater Mekong Subregion Environment Operations Center (GMS EOC), http://www.gms-eoc.org/index.aspx
[32] Mekong Tourists, http://www.mekongtourism.org/site/tourism-strategy/
[33] National Tourism Administration of the People's Republic of China, http://www.cnta.gov.cn/
[34] Chengdu Travel, http://www.cdta.gov.cn/app/travelreport/RC_Report.asp?tid=54
[35] China Tour Online,http://www.chinatouronline.com/china-travel/chengdu/chengdu-tourism/index.html
[36] Chengdu, http://en.wikipedia.org/wiki/Chengdu
[37] Chiang Mai, http://en.wikipedia.org/wiki/Chiang_Mai
[38] Chiang Mai Tourism Bureau, http://www.cmcity.go.th/aboutus/economic.php

About Authors

1. Fu Jing, PhD candidate of Knowledge Management in College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
2. Wided Batat, PhD, Associate Professor of Faculty of Tourism, University of Lyon 2, Lyon, France
Nopasit Chakpitak, PhD, Dean of College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand

Session 5: AI, Decision Support and Optimization Techniques

ISBN 978- 974-672-556-9

Structural Model Approach of Expertise During Industrial Feedback Experience

H. Jabrouni1[2], B. Kamsu[1] and L. Geneste[1]
[1] Production Engineering Laboratory, National Engineering School of Tarbes
47, Avenue Azereix, BP 1629, 65016 Tarbes Cedex
[2]Alstom Transport, Rue du Docteur Guinier - BP 4 - 65600 Séméac
{jabrouni, bkamsu, laurent.geneste}@enit.fr

***Abstract* - During the Problem Solving Processes, Intellectual investment of experts is often considerable. The opportunities for exploitation of expert knowledge are numerous: decision making, problem solving under uncertainty, expert configuration, etc. It is then necessary to assist experts in their tasks of solving problems by dispensing them to produce new work involving a high level of expertise. This can be structured in the process of feedback experience. In this paper, we present experience feedback as an alternative solution to usual knowledge management systems. We propose a structural-model approach for reasoning with Root Cause Analysis. Based on the Fuzzy and probabilistic Theory, our approach presents a general framework that enables the representation of uncertainty in a structure of causality.**

Index Terms – Industrial feedback experience; Root Cause Analysis; structural model

I. INTRODUCTION AND PROBLEMATIC

Industrial products currently developed are more and more complex and make use of several technologies at the same time. Moreover, design times are reduced, bringing new constraints during pre-industrialization phases. Companies have to solve many problems by involving experts who have a partial knowledge of product limited to their field of specialty. These new constraints are rarely taken into account in traditional problem solving methods.

The process of problem solving are generally cumbersome to implement and are often triggered for solving complex problems (requiring a high level of expertise) and critical (with a very negative impact on the client, safety or performance of the company for example). However, one major inconvenient of these processes is the inability to reuse knowledge devoted to solving a past problem, especially that of experts using in analysis phase.

Thus, the ability to capitalize and reuse this knowledge represents a powerful way for optimizing and streamlining the process of problem solving. We call experience the fragment of knowledge capitalized during the experts' activity to solve a problem and we call experience feedback the process that allows to organize the capitalization and exploitation of these experiences.

In the following, this paper will be focused on the proposed relevant mechanisms of analysis promoting efficient reuse of knowledge capitalized with better management aspects of similarity and uncertainty.

II. POSITIONINGSTATE OF THE ART

A. Positioning in relation to problem solving approaches

In the context of continuous improvement of products, services and processes in enterprises, the establishment of the experience feedback process aims to provide a practical solution for accelerating the resolution of problems already encountered and their non-repetition.

In the case of complex problems, an approach of problem solving is often helpful. We recall the definition of a problem solving approach presented by [6]: a problem solving process is a set of planned and systematic activities that can address complex problems. This approach is usually based on the use of rules, principles, expert knowledge. It can mobilize, in a structured and logical, a set of tools and techniques. Whatever the chosen approach [2], we find the same steps of reasoning:

- The composition of problem solving team;
- The description and assessment of the problem highlighted by events;
- The analysis of events to identify their root causes and to validate this analysis;
- The formulation of a solution to the problem and the verification of its application (corrective actions);
- The suggestion of actions to prevent a new occurrence of the problem (preventive actions, lessons learned, etc.).

There are two main approaches for problem solving process [6]:

- The theoretical approach is also called deductive method. This approach is used to solve problems by applying inference mechanisms based on research algorithm (e.g. simulated annealing, the spread of constraints, etc...).

ISBN 978- 974-672-556-9

- The process of problem solving using inductive mechanism. This approach allows identifying the causes of problems which are identified by applying a series of tools allowing tracing the source of problem starting from the observed facts.

In response to our study, we retain the latter approach. It is often manipulated by tools that vary according to the method used. In all cases, we find the Deming PDCA cycle, regardless of the method used, as shown in Figure1.

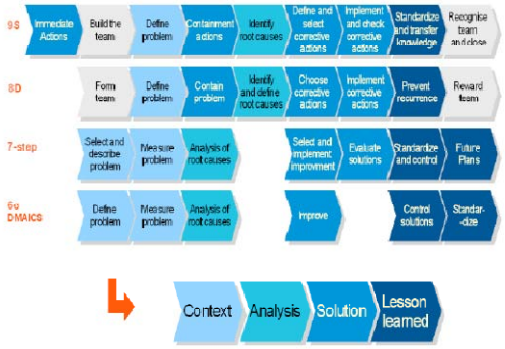


Figure 1. Standard problem solving processes

The nature of information capitalized was characterized in particular in [10]-[14]-[15] during the application of traditional approaches of problem solving. The methods involve four major categories of information: context, analysis, solutions and lessons learned. As a result, this paper will be focused to propose a mechanism ensuring a relevant analysis promoting efficient reuse of knowledge capitalized with better management aspects of similarity and uncertainty.

During problem solving phase, the identification of causes, often called the "analysis phase" is the most important step in a methodology of solving problems. Whatever the problem solving method adopted, the analysis phase is always guided by the process of Root Cause Analysis (RCA). This process relies on the fact that it is judicious to treat the root causes of a problem than to treat the immediate symptoms. First, analyzing the causes of the problem, and second, choosing the most important causes to solve, and thus preventing its repeated again.

RCA can be considered as an element contributing to the iterative process of continuous improvement and represents a common approach for solving problems in a rational and methodical way [6]. We often notice that these specific actions are conducted with lack of knowledge of the main cause that occasions unsatisfactory work performance, that is, a failure to perform the resolution of the problem. Therefore, the real causes must then be diagnosed at the stage of root cause analysis to initiate appropriate corrective actions.

Generally, the main steps of the RCA are:

- Studying the relationship between cause and effect, based on past experience and technical data then summarize them in cause– effect diagram.
- Gather factual information using appropriate tools (e.g. sheet statements).
- Investigate the relationship between cause and effect using the methods of quality, analyze past experimental data, organize data to observe daily and analyze using graphs, histograms, control charts, analysis of variance, regression analysis, etc.
- Synthesize causes to retain those who have been validated. This validation is usually done by contextually appropriate tools (experimental design, Pareto chart, industrial testing ...).Depending on the level of expertise involved, three possible situations can be identified during the analysis phase [29]:
- The group of experts knows, without ambiguity, the real causes of the problem. In this case, the group may propose directly the solution.
- The group of experts has some doubts about causes. In this second case, the group checks the presumed causes before going in search of a potential solution.
- The group of experts has no specific ideas about the real causes of the problem. In the third case, the group does a search of all possible causes. It collects facts and clues that will identify the real causes of the problem. We consider this last possibility to be the most likely situation, in the rest of this document.

B. Positioning in relation to the techniques of reuse experiences

The exploitation process of feedback experience consists of activities to disseminate and use capitalized knowledge in an organization in order to make possible expert's knowledge reusability [8]. In the same time, exploitation of experience stored may be done by experience feedback techniques. Indeed, initially the expert knowledge stored in the form of experience can save both the context of emerging knowledge and accurate information on its explanation that from the point of view and knowledge of the expert concerned. Ina second step, techniques for the re-use of experience can use past experiences to assist the expert in the resolution of problems. These include Case Based Reasoning, noted CBR, which we present in the next section. Other methods, derived from CBR incorporate more elements of the experience. We can cite here the trace-based reasoning (RAPT), which based on temporal recording of units of information called "Trace" of the reasoning process [25]. Finnie and Sun [24] propose a more general idea of Experience

Based Reasoning (EBR). This research approach aims to formalize any type of reasoning based on experience from one or more rules of inference. However, the paradigm of Case-Based Reasoning [23] remains the most popular of these tools previously mentioned. The guiding principle of CBR is to provide a basis for reasoning through previous cases already tested and validated. This approach consists in retrieving an existing solution that enabled, in the past, to solve a similar problem to that which is being addressed. From the description of the current problem to solve, it is a matter of provide sufficient information to enable a function of similarity to access the cases in the proximity of the current case described. Specific tools allow adapting solutions of retrieved cases in the database. We can note that this method is based on the fact that two similar problems have similar solutions.

We propose a new alternative of Case-Based Reasoning as showing in Fig. 2. We are primarily interested in the classical approach of CBR in which the inference mechanism based on the reuse of the solution of solved cases with a context (described by a set of attributes) similar to the new case. However, this type of reasoning is not particularly appropriate to the context of continuous improvement in which our problematic is inscribed. Two problems occurred in similar circumstances often have different solutions especially in the context of complex problems and the adaptation process becomes increasingly difficult. As for the new alternative, the description of an event and its context can serve as an input element to the expert analysis in which the analysis is described by a hierarchy of attributes.

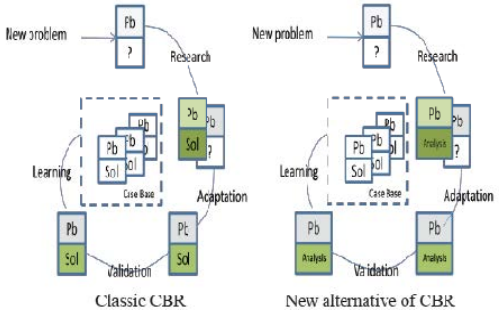


Figure 2. Illustration on the use of a case for resolution of problems

III. Modeling Analysis

A. Representation of the causal analysis by quality tools

In the next section we briefly introduce some tools from the quality approaches dedicated to the representation of the analysis phase. This allows us to position the Experience Feedback from these methods.

1) Cause-effect diagram / Ishikawa diagram

The cause-effect diagram [2] provides a simple way to visualize all potential causes for the finding of an effect regardless of the nature of the concerned problem. It comes in the form of fishbone whose head oversees the effect that we want to know the causes (see Figure 3). Causes are arranged according to their level of importance or detail, resulting in a depiction of relationships and hierarchy of events.

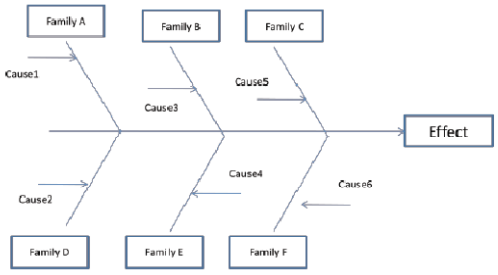


Figure 3. Cause-effect diagram or Ishikawa diagram

After identifying the issue in terms of effect, causes are frequently arranged into four major categories (Manpower, Equipment, Material, Environment), but they can also be replaced with other classifications specific to particular context.

2) ACE Diagram

The ACE diagram (Action on the Causes of Errors)[2] is a variant of a cause-effect diagram method process. It can handle a problem using the cause-effect diagram as support for monitoring the action plan (see Figure 4). At each step of the process. Each step of the process is associated with a Pareto chart of non-conformities updated periodically. In addition, the effect is associated with a temporal chart representing the percentage of defective units observed during the last period of time. The figure below shows this type of diagram.

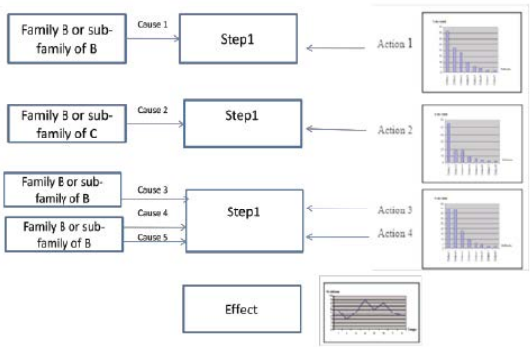


Figure 4. ACE Diagram

3) Cause tree / fault tree

The cause tree (fault tree) is generally used in the field of occupational hazards. The method is to construct a graphical representation of sequences of potentially hazardous events that led to the main fault. The tree branches are built and based on the Why Why Analysis, also called "5 why", which identify the origin of a problem (root issue) by bringing the experts to wonder about the problem by asking gradually the question “why?” in several times. The graphical representation of the list of

potential causes can be presented as arborescence. Figure 5 shows this diagram.

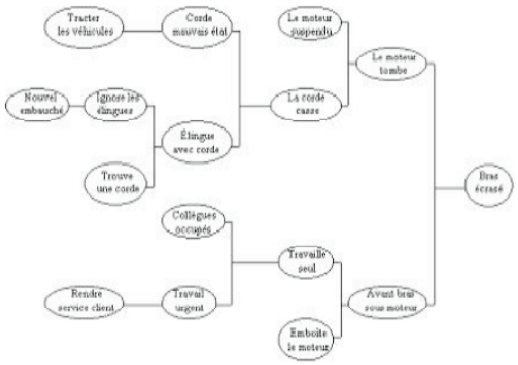


Figure 5. Cause tree / Fault tree

B. A formal model of analysis

In the proposal approach, we propose a simple method of analysis. It is a filter that allows considering only the relevant information. Indeed, if events are not represented in an appropriate manner (modeled precisely), the application of reuse tools is likely to be too limited in some cases and not specific enough in others. This is the reason why we propose a mechanism of the analysis phase to better describe the quality of data and modeling expertise appropriately.

1) The quality of data

In the domain of feedback experience, the quality of data is crucial because it impacts directly on the reliability of results and interpretation. Before any study or analysis of feedback experience, a checking data quality must be done. Three checking criteria allow an analyst to secure
data quality management:

- Their consistency.
- Their validity, representativeness and homogeneity in the case of multi-expertise.
- Exhaustiveness.

Data quality must be evaluated on two levels:

- At the level of collection, before the introduction of information in priori analysis of event.
- At the level of processing or statistical analysis of information in posteriori analysis of event.

All these constraints must be incorporated as part of the formalization of an analysis model.

2) Modeling expertise

We propose a functional diagram of the analysis phase which includes the following steps based on the model using tree causes:

- The description of the main problem is divided into many contextualized hypothesis.

- During the analysis phase, these assumptions must be detailed in the form of several elementary assumptions (H11, H12, H13, etc.). To better understand the problem, each of these assumptions can be more finely detailed in other hypotheses (H11, H12, H13, etc.) by using appropriate tools such as "5 why".

- These sub-assumptions include uncertainty [30] which can be represented by a measure reflecting the degree of confidence or certainty of the expert.

- The experts should naturally validate the potential assumptions in priority (hypothesis with the highest degrees of plausibility). This validation phase consists in applying a filter to determine the assumptions considered as the most relevant root causes of the issue.

Fig. 6 summarizes these steps of analysis described previously.

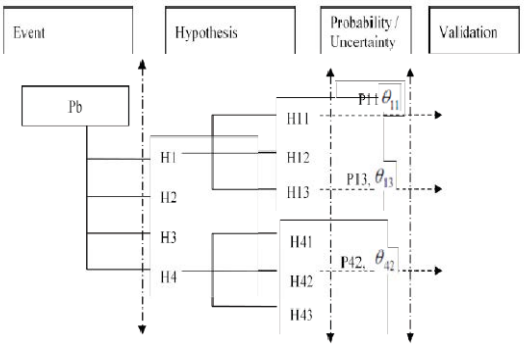


Figure 6. Block diagram of the proposal analysis phase

The experience can be considered as a collection of information that reflects a context in which information is rarely known with precision. They may be totally unknown in some cases [5]. Formalism to model experiences should allow representing the imperfections of information and integrating them in different stages of functional diagram of the analysis phase in order to provide adequate results. The uncertainty must then be spread from the original data until the final result.

The information is rarely given as reliable and perfect data. Many defects, such as uncertainty, vagueness and incompleteness are often associated with them. The unreliability of information can be translated in several forms [7]:

- The uncertainty is related to truth of information, and characterizes its degree of conformity to reality. It refers to the nature of the subject or the fact concerned, its quality, its essence or its occurrence,
- The vagueness regarding the information content and thus indicates its lack of quantitative knowledge,
- The incompleteness that characterizes the absence of information provided by the source on some aspects of the problem.
- The ambiguity reflects the ability of information to allow different interpretations.

ISBN 978- 974-672-556-9

This brief description allows us to better understand the different facets that may take on imperfect information[26] from the real world. We will now introduce several families of means representation of these limitations:

- The Bayesian approach provides a framework for a priori subjective probabilities. The Bayesian inference allows to calculate (or revise) the probability of a hypothesis. The probabilities are based on the weight distribution of a trust unit on singletons in the field of possible values. The Bayesian formulation introduced conditional probability and a priori probability revision. This framework is largely based on the concept of probability [13].

- The possibility theory [22] provides mathematical tools that allow representing incomplete, imprecise or fuzzy information. It is an extension of decision-making of Fuzzy Set Theory [27] that models categories of natural language. The possibility distributions [19] were meant to provide a graded semantics to natural language statements and imply ‘‘judgment’’ in the feeling of ‘‘possibility’’, ‘‘achievability’’, ‘‘acceptability’’ and “capacity of the events to occur’’.

- The theory of belief functions [7] provides, in turn, of the mathematical tools to process information with random and imprecise nature. It is a theoretical framework that generalizes the two previous frameworks. The developments are based on the weight distribution of a trust unit on any sub-set of the domain of possible values. There are several variants including the Dempster-Shafer [27] and the Transferable Belief Model(TBM) [28].

In this paper, although the third approach seems promising and generic, we will deal only with the first two approaches. Indeed, we have not yet improved the mechanisms for proper implementation of the theory of belief function. Thus we propose below an approach combining quantification of expertise by fuzzy concepts and conditional probabilities.

3) Quantifying the uncertainty of the expert by fuzzy concepts

The knowledge of the expert is difficult to quantify and is also uncertain [1]. It is therefore necessary that the analyst can model a probability distribution from the information collected.

Several approaches exist [4] and we can quote the most common of these approaches:

- The simplest method is to ask the expert to express the probability of validating a hypothesis by choosing a number between 0 and 1: value 0 means that the hypothesis is impossible, however, the value 1 means that the hypothesis is absolutely certain.

- If the experts are not able to quantify a subjective value on the scale [0, 1], they can generally express their uncertainty using fuzzy concepts such as "probable”, “very likely" or with symbols like "+ +" "+", "-". These fuzzy concepts are often used in industrial environments and their implementation is easy [11]. [18] define a verbal–numerical probability scale corresponding to these concepts. We present here an extract from a table (see Table 1) estimate the correlation between declarative and probabilistic expressions.

TABLE1: CORRESPONDENCE BETWEEN DECLARATIVE AND PROBABILISTIC ESTIMATION

Expression of fuzzy concept	Average probability associated	Uncertainty interval
Highly probable	0.89	0.60 – 0.99
Probable	0.79	0.30 – 0.99
Likely.	0.71	0.10 – 0.99
Possible	0.37	0.10 – 0.99
Impossible	0.12	0.10 – 0.99
Unlikly	0.11	0.10 – 0.99
Highly improbable	0.06	0.10 – 0.99

Finally, if the experts are familiar with statistical methods, they can apply directly an average value or standard deviation of a distribution. These values are generally considered as the parameters of log-normal distribution, normal or those of Student rule [12].

4) Estimation of expertise by using probabilistic approaches

In general, the analysis phase, we seek to determine an uncertain quantity associated with each plausible cause, by interviewing persons having knowledge of this quantity [4]. These people, in the language of psychology, are "subjects" but they are more commonly called “experts". They are represented in the following by E_1,E_2, E_3,… E_n.

The Analyst *D*, identified as the "decision maker" in the problem solving process must combine expert opinions with his own knowledge [21], in order to evaluate the distribution or the most likely a priori value of θ. That is the person who makes the final decision and therefore stands as surety. He will rely for that on the opinion of several experts. Each expert E_i according to his knowledge must therefore provide to the decision maker *D*, the needed information to develop a probability density θ. This knowledge may evolve, especially when additional information is acquired or when confronted with those of other experts [1].

The expert E_i will therefore provide information conditioned by its own knowledge, C_i represented by the probability p_i (θ/C_i). Symmetrically, a priori opinion of the analyst, *D*, is represented by distribution p (θ/C_D). To simplify, we suppose that

ISBN 978- 974-672-556-9

we want the probability of a future event *A*, and there is only one expert. The extension to *n* experts doesn't pose any particular problem. The analyst *D* has knowledge C_D on *A*, can express a priori estimate of the probability of *A*:$p\ (A\ C_D)$principle of conditional probability[31]). Conversely, the expert E_1, with knowledgeC_{E1} will provide an evaluation $p_1\ (A/C_{E1})$. The analyst has now two estimates of probability *A*. He will seek to evaluate the a posteriori probability of *A*, given C_D and C_{E1},$p\ (A\ C_D, E1)$.

The accuracy of results depends on the consistency of the expert himself. This will make sense in cases where two experts offer two different values for the same assumption. To overcome this problem several alternatives exist [16]-[17], especially the assignment of a confidence weight to each expert characterizing the reliability of his knowledge. Therefore, the formalization proposed for the analysis ensures the following aspects:

The use of the tree causes based on the "5Why" tool, can provide a simple way to represent the logical sequence of events (cases) involved in the occurrence of the defect or main problem. This satisfies the need to ensure the best representation of the real world by experts.

The use of probabilistic and fuzzy techniques allows taking into account the uncertain nature of information provided by experts in the inference mechanisms. This quantification must be provided by each expert for each plausible value of hypotheses.

Both techniques for estimating expertise have been implemented on two platforms that we have specified. The first application named T-REX (figure 7) incorporate in its analysis phase provides a completely interactive module which allows to elaborate the tree causes and to estimate the knowledge of experts with respect to degrees of plausibility of the cause. To help the experts to focus on the causes with the highest plausibility, we implement a filter to refine the list of potential causes. While the second application called PST (Problem Solving Tool)which is implemented at ALSTOM Transport Corporation (figure 8) allow the expert to estimate the average probability associated with each issue proposed.

In addition, both of this applications based on Web architecture, and a centralized database, provide many features:

- The representation of different components of an experience described previously (event, context, analysis, solutions, lesson learnt)
- The structuring process of problem solving organized in clearly identified steps.
- The implementation of appropriate tools to facilitate the achievement of different phases of the process (eg "Is/ Not," "5 whys", "Ishikawa diagram," " action plan management »...).
- Taking into account the lack of information on the causes of events,
- Improvement of mechanisms for the re-use of expert analysis
- The improving the search-engine by setting up specific algorithm allows finding relevant past experiences (by keywords or by similarity).

Figure 7. T-Rex, developed by CRC-IDCE. ENIT Tarbes

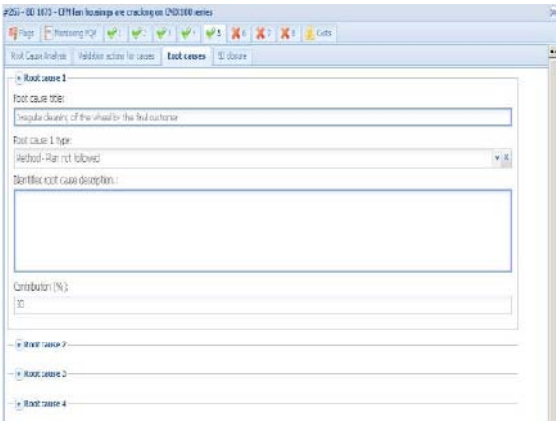


Figure 8. PST implemented at ALSTOM Transport

The capitalization process is now stabilized via the platform HLP web installed throughout the organization of ALSTOM Transport sector (71 sites in the world). The analysis module of T-Rex has given some satisfactions and early outcomes assessments by end-users are very encouraging. This module has facilitated the experts' tasks, notably in the validation process of the most plausible root cause of an event. A preliminary evaluation of the T-REX application has resulted in encouraging results with respect to both increasing the speed of problem solving processes by experts, compared to a previous tool developed in Excel, and enhancing theaccuracy of the analyses. Enhancing models analysis ofroot causes allows development of systematic improvements and assessment of the impact of corrective programs with respect to the top management quality objectives. Note that we are working on the integration of the module T-Rex in

the platform PST in order to combine the two approaches.

IV. CONCLUSION

The problem of cognitive experience feedback systems addressed in this paper is a major challenge in the context of continuous improvement in industrial organizations and in complex socio-technical organizations. However, there are few tools treating the experience feedback as a cognitive entity and not as simple information which may enrich the statistical analysis. Assuming that the knowledge of experts, whether it is included in the experience during the resolution of problems, can offset the low number of experiences that prevents a gross statistical analysis. We made a comparison between cognitive feedback experience and knowledge-based systems, the cognitive feedback experience whose the significant feature is that knowledge is developed incrementally, by adding analysis gradually on capitalized cases. The objective of this work was to propose a formal framework for representation of experiences, taking account of the imperfections of information presented in expertise of problems solving.

We studied the precise nature of experiences that must be capitalized as: event, context, analysis, solutions and lessons learned. We have also shown that the experience is in part the result of a work of interpretation (context), it is interesting to take into account the uncertainty associated with this expertise. Therefore we have proposed to incorporate, explicitly, the uncertainty associated with information on the analysis performed by experts using probabilistic approaches. However, several investigations are required:

Expression of a more precise analysis: we believe that other opportunities for analysis expression to explore. Indeed, the generalized version of Bayes' theorem is compatible with the transferable belief model and the definition of belief networks could be very interesting in the perspective of an approach to risk prevention through experience.

The definition of a multi-expert analysis system [3]: a major interest of belief functions is the possibility to merge information from various sources. We want to further authorize the analysis of similar cases by several experts in the input phase of the experience and analysis.

The effective and targeted spreading of information on lessons learned and theirs results for innovations: this involves integrating the profiles of different actors or to rely on the skills management of the experts when it is effective [9].

The coupling of the proposed experience feedback model with a service-oriented architecture may be useful in order to facilitate interoperability with business applications.

REFERENCES

[1] F.Coolen, Statistical modeling of expert opinions using imprecise probabilities. ISBN 90-386- 0482-3, PhD-thesis, Eindhoven University of Technology, 1994, pp.121

[2] A.-M. Chauvel, Méthodes et outils pour résoudre un problème, 3ème edition, édition Dunod, 2000.

[3] M. D'Acquin, J. Lieber and A. Napoli, Représentation multipointsde vue des connaissances pour l'adaptation, Atelier deraisonnement à partir de cas, RàPC Paris, 2002.

[4] De Fenetti, Theory of probability. Wiley Classics Library.1990.

[5] R.Dieng-Kuntz, D. Minier, F. Corby, O. Corby and L.Alamarguy, Building and using a medical ontology for knowledge management and cooperative work in a health care network. Computers in Biology and Medicine, Volume 36, Issues 7-8, 2008.Pp. 871-892.

[6] K. Hosotani The QC Problem Solving Approach –Solving by Quality Resources, 1992.

[7] D.Dubois and H.Prade, Représentations formelles de l'incertain etde l'imprécis, Dans : Concepts et méthodes pour l'aide à ladécision - outils de modélisation, 2006. Lavoisier, (3), pp. 111-171, vol. 1, Traité IC2.

[8] B. Fuchs, Représentation des connaissances pour le raisonnementà partir de cas. Le système ROCADE, Thèse de Doctorat :Informatique, Saint- Etienne. 1997.

[9] J. Hermosillo Worley, H. Rakoto, B. Grabot and L. Geneste, A competence approach in the experience feedback process. In Integrating Human Aspects in Production Management, IFIP International Federation for Information Processing series, Volume 160, Springer-Verlag, New York, USA. 2005. pp 220-235,

[10] H. Jabrouni, L. Geneste, B. Kamsu, Le retour d'expérience industriel: une aide à la décision basée sur les connaissances.FEES'08. Morocco. 2008.

[11] H.S. Jeffreys, Theory of probability. 3ème edition, oxfordUniversity Press, 1961.

[12] S. Lichtenstein and J.R. Newman, Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Science n° 9. 1967.

[13] P. Naïm, P-H. Wuillemin, Ph. Leray, O. Pourret and A. Becker,Réseaux bayésiens, 2007. Eyrolles; Édition ISBN-10: 221211972.

[14] H.Rakoto, 2004. Intgration du Retour d'Exprience dans lesprocessus industriels – Application à ALSTOM Transport (in France). PhD thesis, National Polytechnic Institute of Toulouse.

[15] R.P. Srivastava, "Value judgments using belief functions", JAIPress, Greenwich, CT, 1996. pp.109 - 130.

[16] S. Moral, J. del Sagrado.. Aggregation of imprecise probabilities.In: Aggregation and Fusion of Imperfect Information (B.Bouchon-Meunier, ed.) Physica Verlag, Heidelberg. 1997, pp.162-188.

[17] W. Wilke and R. Bergmann, Techniques and knowledge used for adaptation during case-based problem solving, 11th International Conference on Industrial Engineering Applications of Artificial Intelligence and Experts Systems, IEA- 98-AIE, Lecture Notes in Computer Science, Springer- Verlag, Berlin Heidelberg NewYork. 1998.

[18] C. Witteman and S. Renooij. Evaluation of a verbal–numericalprobability scale. International Journal ofApproximate Reasoning,Volume 33, Issue 2, 2003, Pages 117-131.

[19] D. Dubois and H. Prade. Resolution principles in possibilisticlogic. International Journal of Approximate Reasoning, Volume 4,Issue 1, 1990, pp. 1-21.

[20] D. Dubois and H. Prade. Bayesian conditioning in possibility theory. Fuzzy Sets and Systems, Volume 92, Issue 2, 1997, pp.223-240.

[21] D. Dubois and H. Prade, Claudette Testemale. Weighted fuzzy pattern matching. Fuzzy Sets and Systems, Volume 28, Issue 3,1988, pp. 313-331.

[22] A. Aamodt, and E. Plaza, Case-based reasoning: foundational issues, methodological variations, and system approaches, AI Communications, vol. 7 (1), pp. 39--59, 1994

[23] G. Finnie and Z. Sun, A logical foundation for the CBR Cycle, IntJ Intell Syst, vol. 18 (4), pp. 367-- 382, 2003.
[24] P.A. Champin, J. Laflaquière, A. Mille and Y. Prié, MUSETTE : a framework for Knowledge Capture from Experience, 12èmeAtelier de Raisonnement à Partir de Cas, 2004.
[25] Smets P., Imperfect information: Imprecision - Uncertainty, Uncertainty Management in Information Systems, From Needs to Solutions. A. Motro et P. Smets (eds.), Kluwer Academic Publishers, pp. 225--254, 1997
[26] L.A. Zadeh. Fuzzy Sets as a basis for a theory of possibility, Fuzzy Sets and Systems, (1), pp. 3--28, 1978.
[27] G. Shafer, A Mathematical Theory of Evidence. Princeton University Press, 1976.
[28] P. Smets, et R. Kennes, The transferable belief model, Artificial Intelligence, vol. 66, pp. 191--234, 1994.
[29] J-Y. Halpern and J. Pearl, Causes and explanations: A structural model approach. 2001. In Proceedings IJCAI-01. pp 27-34.
[30] J. Tian and J. Pearl, Probabilities of Causation: Bounds and Identification. Annals of Mathematics and Artificial Intelligence,2000. pp. 589-598.
[31] G. Weidl. Condition Monitoring, Root Cause Analysis and Decision Support on Urgency of Actions. In Book Series FAIA, Soft Computing Systems - Design, Management and Applications.2002, pp. 221

Decision Support System for Strategic Outsourcing Risk Reduction

Anthony S. Atkins, Bernadette Sharp, and Lyn Atkins.
Faculty of Computing, Engineering and Technology, Staffordshire University, Stafford, UK.
a.s.atkins@staffs.ac.uk, b.sharp@staffs.ac.uk, l.atkins@staffs.ac.uk.

***Abstract*—Today's outsourcing decisions are made in complex and rapidly changing environments, which exist both within organizations and in their external environments. The dynamic nature of environments increases risk and there is a need to support outsourcing decisions with a strong and structured approach. Furthermore, ongoing support is needed to monitor decisions for their continued aptness or inaptness in a rapidly changing world. This paper considers current outsourcing activity, recognizing potential benefits as well as the various strategic and business risks. It describes the holistic approach for outsourcing decisions provided by the HABIO framework and the recent development of a decision support web tool based on the HABIO approach. The web tool development offers an easily accessible, user-friendly application of HABIO principles within organizational procedures and processes as well as supporting outsourcing decisions in dynamic socio-economic environments.**

Index Terms –decision support system; HABIO framework; outsourcing decision; risk reduction.

I. BACKGROUND

There are many definitions of outsourcing, for example "a conscious business decision to move internal work to external suppliers" [1], or "the strategic use of outside resources to perform activities traditionally handled by internal staff and resources" [2]. Where companies take a decision to outsource it may be to a provider in the host country or it may be overseas.

In recent years outsourcing projects have often been in the form of Business Process Outsourcing or Offshore Outsourcing. Business Process Outsourcing may remove routine or specialist tasks from the parent company. It is now seen as a service where total, defined business processes are given to expert service providers who manage the process and ensure the total integration between outsourced business processes and in-house processes [3]. Offshore outsourcing involves moving business activities to companies outside the country of the parent company. One of the main drivers for offshore outsourcing has been lower labour costs. Predictions from Forrester Research Inc. indicated that over 3 million white-collar jobs will move from the U.S. to low-cost countries by 2015 [4]. Examples of outsourced activities are found in both the private and public sectors and in both manufacturing and service industries. A 2004 report suggested outsourcing already accounted for one third of IT spending globally and was a growing market [5].

There is much evidence to demonstrate the potential benefits and risks of decisions to outsource. Outsourcing can free up assets and reduce costs in the immediate financial period [6]. Outsourcing arrangements can help companies respond effectively to peaks in demand for products or service, and allow the parent company to focus on its core business activities. However, effective planning and management of outsourcing strategies are essential. Although outsourcing is considered to be a powerful management tool [7], an incorrect decision can have a devastating effect on the organisation, both economically and politically.

In its 2004 report, Gartner indicated that as many as 80% of outsourcing deals are unsuccessful and that European businesses wasted £4bn on poorly managed contracts [8]. Since outsourced activity particularly at strategic level, often represents large investments, there is an obvious financial risk, but there are also other potential hazards such as those associated with people and management, trade union relations, internal expertise, knowledge spill-over, quality of service, customer perceptions and communication.

People and management factors are closely interrelated. Where outsourcing decisions bring change in the work place or to job security people may be fearful and resistant, and change management skills are essential to ensure success. People and management factors have been recognised [9]. The management risk is a particular concern, unless the organisational structure has been designed specifically to accommodate increased communication, testing, and control requirements that are involved in the distributed development of advanced technology. Another factor closely linked with people and management is trade union relations in that opposition from a strong trade union is a potential threat to outsourcing decisions.

Loss of internal expertise is a possibility, particularly if an activity is provided by an external organisation for a prolonged time. There is further risk where the external provider is updating or developing the product or service in response to market demands or technological opportunities. Beyond the direct loss of in-house skills and

ISBN 978- 974-672-556-9

expertise, there is another potential loss of knowledge. One underreported risk of outsourcing is the loss of knowledge spill-over to and from the local area [10]. Local research and development provide immediate benefits for technological organisations, which may be located near universities and other research and development institutions. As a source of innovation, research and development, is linked to maintenance and development of market share and thereby competitive advantage.

There are potential dangers to quality of service when an activity is outsourced, for example if there is a high level of labour turnover in the outsource company it may be difficult to maintain targets or agreed service levels. Customers' perception of a company can be adversely influenced if there are issues around the quality of products or services. More subtly, if customers are dissatisfied at the loss of local employment or concerned that offshore workers may be paid low incomes these factors may also have a negative effect on their perception of the company.

Communication is a more obvious consideration for offshore outsourcing decisions, where there may be differences in time zones, culture and languages. More subtle, but no less important, is the lack of face-to-face communication in distributed organisational structures [11]. Communication is an essential part of business activity both within the organization and with external customers and suppliers.

Beyond the level of individual firms there are risks at the level of whole industries and localities where large scale outsourcing can have a detrimental influence on employment levels and prospects. There can also be a knock-on effect for other businesses in the industry or locality. Furthermore, at a national level, large-scale offshore outsourcing can increase domestic unemployment and lead to loss of knowledge and skills.

The results of a recent survey by the Engineering Employers Federation (EEF), the industry body for engineering and manufacturing employers, and accountants BDO highlight the risks that UK companies and the national economy can experience from outsourcing outside the country. In the last decade although large numbers of UK firms have moved production to countries in Eastern Europe or Asia, where lower labour costs are attractive, the EEF's survey of 300 manufacturers showed that firms were now returning production to the UK. Reasons for return included cost savings not meeting expected savings, poor product quality and slowness in getting products to market. Almost 70% of companies said the UK was a competitive location for manufacturing, compared with 43% two years ago. Lee Hopley, the EEF's chief economist, explained that manufacturing was "absolutely central" to rebalancing the economy [12].

Strategic decisions in respect of outsourcing are complex and need to take account of various considerations, benefits and risks which may be interdependent, and which exist in a dynamic socio-economic environment. In response to these challenges previous work at Staffordshire University suggested and developed the HABIO framework (Holistic Approach Business, Information Organisation) [13]-[16].

II. THE HABIO FRAMEWORK

The HABIO framework acknowledges that outsourcing is both a business and technical issue [17]. The HABIO Framework offers a holistic and systematic approach to outsourcing decisions that places these decisions within three strategy perspectives, namely business, information and organisational. As such, it is based on the Information Systems Strategy Triangle (ISST) by Pearlson [18]. The triangle suggests that the strategies are interrelated and that changes to one will require reappraisal of the other two.

Fig. 1 shows the HABIO framework concept presented within the ISST. Using this framework each of the three strategic perspectives of the triangle is weighted according to the needs of a specific organisation, and has a screen showing cards that can be selected or deselected according to perceived relevance. The cards define key considerations to be taken into account and evaluated to support sound outsourcing decisions. Behind each card is a scale or grid, as illustrated in Fig. 2, which enables the user to evaluate and score the outsourcing opportunity (company or project) against each consideration represented. As an example, Fig. 2 shows an organisational perspective with grids to evaluate internal expertise, legislations, risk management and union pressures, and scales to evaluate historical precedent and internal policies. Fig. 2 also illustrates additional cards above each of the three screens. These are ready to be given customised definitions and dragged into the screen if required. Enlarged examples of further "spare card" scales and grids are shown away from the screens ready to be used if needed under any of the perspectives. Scores for scales and grids within each perspective are summed and factorised against the weighting attributed to that perspective, then the overall score is calculated by addition of the weighted scores for the three perspectives, expressed as a percentage of the maximum possible weighted score. The following formula illustrates how the total weighted score (*Tws*) is calculated. This allows comparison of the total weighted scores for different companies and projects, as well as monitoring of the scores over time.

$$T_{WS} = \sum_i \left(w_i \cdot \sum_j s_{i,j} \right) \quad (1)$$

where w_i is the weight for perspective i and $s_{i,j}$ the score number j in perspective i.

The HABIO Framework can be updated to respond to changes in the organization and its environment in two ways. First the model lends itself to routine updates though reappraisal of scores and weights, and secondly by the addition or deletion of cards it can accommodate completely new environmental factors that could emerge in a dynamic business world.

The Framework has been applied in two large commercial organizations utilizing outsourcing in customer service and IT infrastructure functions [19]. This paper indicates that the approach was found to be effective and that drivers behind the use of outsourcing are increasingly considerations of strategic factors rather than cost reduction alone. Sharp *et al.* (2005) describe application of the HABIO framework for evaluating outsourcing decisions with respect to seasonal call centre operations for a large UK retailer company, using an intelligent based system approach which provides a series of what-if scenarios for evaluating tender documents[20].

III. HABIO Decision Web Tool

The HABIO Framework's innovative contribution to support outsourcing decisions and reduce associated risk is to offer a holistic approach that takes quantifiable account of all relevant considerations. Recent work has further developed the HABIO Framework concept by creation of a decision support web tool. The added strength of the web tool is to offer the application of HABIO principles via a user-friendly, accessible interface which will appeal to businesses and other organizational users by facilitating the use of HABIO as part of routine processes and procedures around outsourcing decisions. The web tool supports both initial evaluation of potential projects and regular monitoring to maintain strong decisions amid changing socio-economic and technical environments. It is comprehensive and maintains the flexibility of the HABIO Framework in that cards can be added and customized or deleted. It also provides calculations and maintains a history, thereby assisting both usage and auditing. In facilitating use of HABIO principles by businesses and organizations the web tool supports minimizing the risks and maximizing the benefits associated with outsourcing.

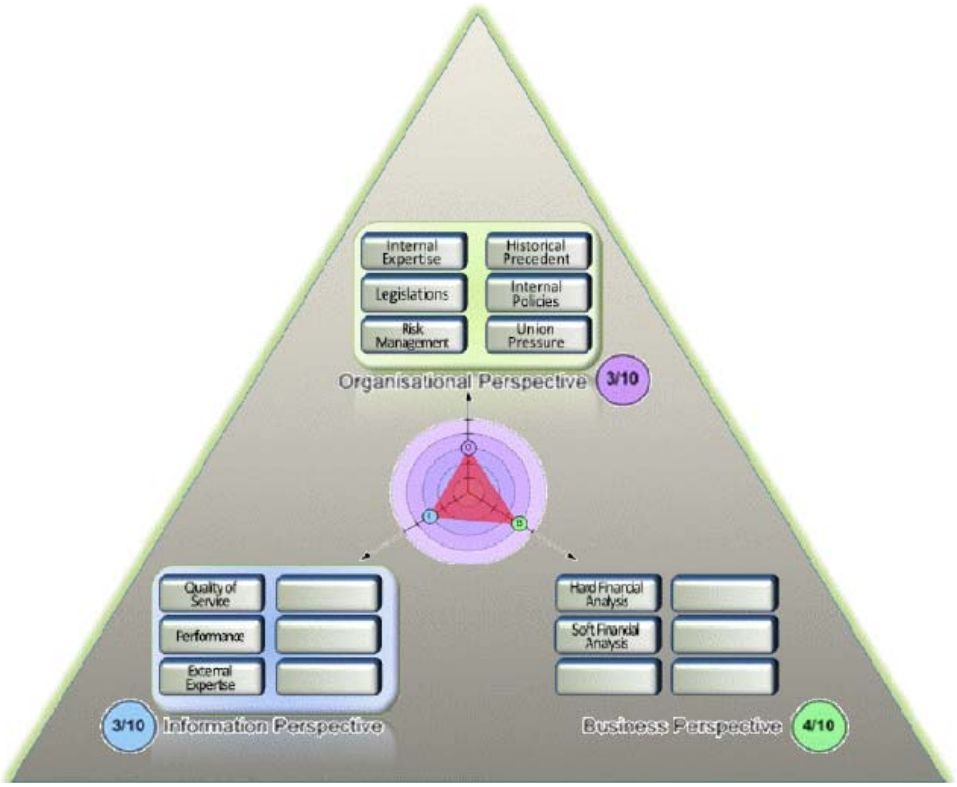


Figure 1. HABIO Framework with in the ISST

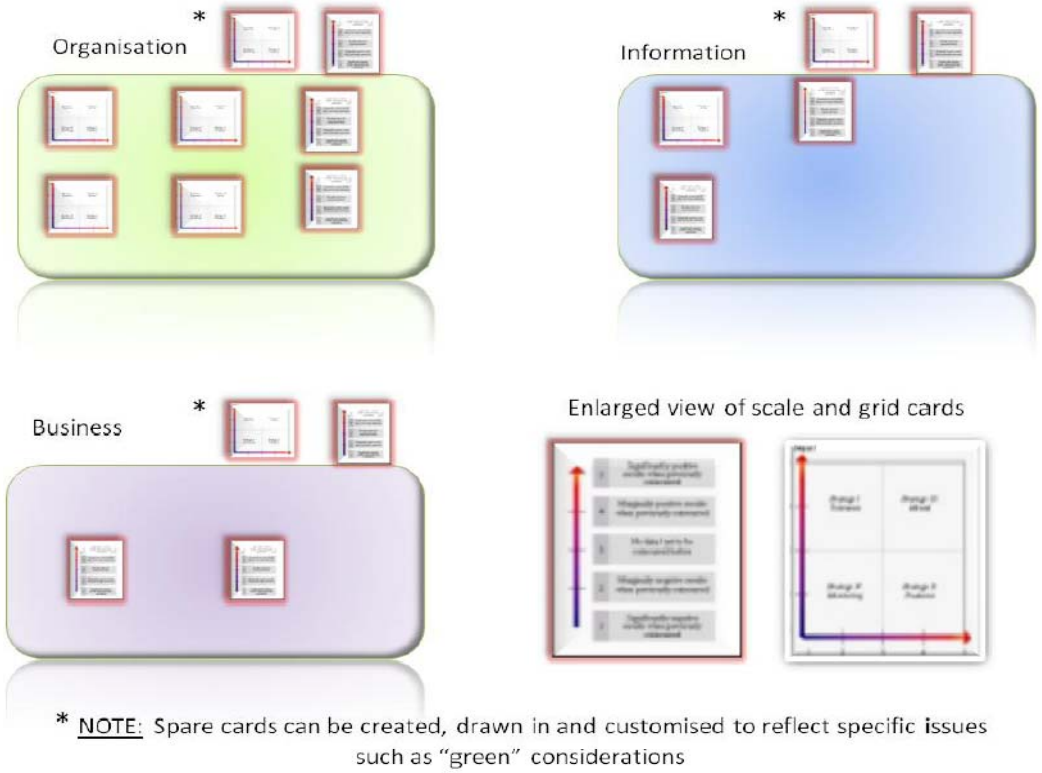


Figure 2. HABIO Scales and Grids

The web tool is implemented in PHP interfacing with MySQL and ArtiChow which is a software library for creating charts and graphics in PHP. It has two modes of operation, namely evaluation and monitoring. The evaluation mode is used to evaluate and compare potential outsourcing projects or companies systematically in consistent and quantifiable terms so that decisions are objective, soundly based and auditable. The monitoring mode includes the purpose of allowing review of the original outsource decisions at later dates to check whether they remain valid against changes, opportunities or threats that emerge in the organization or in its external socio-economic environment.

Figure 3. HABIO Decision Support Web Tool Screen showing Perspectives and Cards

ISBN 978- 974-672-556-9

Fig. 3 shows how the web tool captures the three dimensional perspective of the HABIO Framework. Each perspective consists of a set of cards that can be added to or deleted according to need.

Data is input to HABIO Framework scales or grids that sit behind each card. Fig. 4 gives an example of a scale (quality of service) and an example of a grid (external expertise) used to score the HABIO cards. Scores for both scales and grids are from 1 to 5. Where the measure is a scale there are five descriptive levels and the score is the level that best describes the comparison between the outsource opportunity and the base case. Where the measure is a grid, the user is able to input criteria coordinates and then the overall score and legend are automatically calculated and displayed - the overall score on the grid translates to a value from 1 to 5 so that all scales and grids produce similar measures.

The decision support web tool allows weighted scores for different companies and projects to be exported to a Microsoft Excel spreadsheet. The scorecard shows scores for each card in each of the three strategic perspectives, and the weighting for each perspective. It also calculates the total weighted score for each project or company as illustrated in Fig. 5.

Overall the web tool provides a user-friendly vehicle that supports a holistic view taking into account the complex facets of outsourcing decisions. It thereby offers business and organizational users a convenient means of application of HABIO principles to minimize strategic and business risks.

Figure 4. Example Scale and Grid

HABIO Scorecard	Organisational Perspective						Information Perspective			Business Perspective		Weighted Score
	Weight: 3/10						Weight: 3/10			Weight: 4/10		
	IE	L	RM	HP	IP	UP	QoS	PB	EE	SFA	HFA	
Outsourcing Project A	4	3	4	3	2	4	4	4	3	3	5	74%
Outsourcing Project B	2	4	3	4	2	2	3	4	2	3	3	59%
Outsourcing Project C	4	3	5	3	2	4	5	5	2	3	2	65%

Organisational Perspective
IE: Internal Expertise
L: Legislation
RM: Risk management
HP: Historical Precedent
IP: Internal Policies
UP: Union Pressures

Information Perspective
QoS: Quality of Service
PB: Performance Benchmarking
EE: External Expertise

Business Perspective
Soft Financial Analysis
Hard Financial Analysis

Figure 5. HABIO Scorecard

IV. SUMMARY

The HABIO Framework and recently developed decision support web tool offer holistic, systematic and auditable evaluation of outsourcing opportunities and then ongoing review of these decisions in the context of dynamic socio-economic environments. The complex strategic choices between in-house, various domestic and offshore outsource opportunities highlight the need for decisions that are sound, objective, comprehensive and auditable. The web tool provides a user-friendly and accessible means for the application of HABIO concepts within organizational and business procedures. Outsourcing risks and the dynamic nature of business environments have been discussed in this paper. The decision support web tool development is seen as facilitating structured evaluation and review of outsourcing decisions by business and technical managers. It also supports routine maintenance of records, availability of history and auditable decisions. Effective monitoring and review of outsourcing is vital to support timely action in response to new opportunities or business threats perceived or anticipated. The process of evaluating outsourcing opportunities can reduce the risks associated with outsourcing decisions, offer the added benefit of helping the parent organization identify its own strengths and areas where in which it needs to improve, and recognize the impact of outsourcing decisions on the three ISST perspectives.

The internal and external environments in which private and public sector organizations operate will continue to be subject to technical, social, economic and political changes and opportunities, and strong decisions based on systematically evaluated criteria are essential in facing these challenges. The HABIO Framework and decision support web tool are flexible to support decisions in various sectors and to various activities.

ACKNOWLEDGMENT

This work was conducted under the auspices of the Research Informed Teaching (RIT) project, Staffordshire University. We would like to thank Bruno Caussin for his contributions to the development of the web tool.

REFERENCES

[1] T. Elliot, and D. Torkko, "World Class Outsourcing Strategies", *Telecommunications*. American addition, vol. 30, pp. 47-49, August 1996.

[2] D. Griffiths, (10 January 2010). "The Theory and Practice of Outsourcing", 2001 [Online]. Available: http://www.outsourcing-research.com/kud-002.html

[3] G. Kruse, and C. Berry, "Outsourcing: The How-to Guide". In *IEE Manufacturing*, Aug/Sep 2004, pp. 36-39.

[4] P. Engardio, A. Bernstein, and M. Kripalani, "The New Global Job Shift", *Business Week*, February 3, 2003, pp. 1-6.

[5] Capgemini, *The Outsourcing Report: Making Choices – Insights into Successful Outsourcing*, 2004.

[6] C. Harland, L. Knight, R. Lamming, and H. Walker, "Outsourcing: assessing the risks and benefits for organisations, sectors and nations", *International Journal of Operations and Production Management*, vol. 25, no. 9, 2005, pp. 831-850.

[7] M. Johnson, *Outsourcing in brief*, Elsevier Butterworth-Heinemann, Oxford, 1997.

[8] B. Glick, (16 February 2010). "Does your Outsourcing agreement measure up?" March 2004. [Online]. Available: http://www.computing.co.uk/computing/features/2072394/does-outsourcing-agreement-measure?vnu_lt=ctg_art_related_articles

[9] E. Mumford, S. Hickey, and H. Matthies, *Designing Human Systems – An agile approach to ETHICS*, Morrisville: Lulu Press, 2006.

[10] M. Kyle, "Does Locale Affect R and D Productivity? The Case of Pharmaceuticals". In *FRBSF Economic Letter, vol.* 32, 12 November 2004.

[11] N. Alster, (7th November 2008). *Customer Disservice*. 2005. [Online]. Available: http://www.cfo.com/article.cfm/4390954

[12] BBC NEWS, (15 January 2010). [Online]. Available: http://news.bbc.co.uk/go/pr/fr/-/2/hi/business/8434458.stm

[13] A. S. Atkins, B. Sharp, D. S. Hickey, and L. Ho, "HABIO Management of Outsourcing: Systems, Capabilities and Cybernetics". In *Technia- International Journal of*

Computing Science & Communication Technologies, 2010, vol. 2, no. 2, pp. 407-413.

[14] L. Ho, and A. S. Atkins, "IT Outsourcing: Impacts and Challenges". In B. Walters and Z Tang (Eds.) *IT Enabled Strategic Management – Increasing Returns for the Organization*, United States of America: Idea Group Inc., 2006a pp. 244-274.

[15] L. Ho, and A. S. Atkins, "Outsourcing Decision-Making: A Review of Strategic Frameworks and Proposal of a Multi-Perspective Approach". In H.S. Kehal and V. Singh (Eds.) *Outsourcing and Offshoring in the 21st Century: A Socio-Economic Perspective*, United States of America: Idea Group Inc.2006b, pp. 165-196.

[16] L. Ho, and A. S. Atkins, 2009," IT Portfolio Management: A Holistic Approach to Outsourcing Decisions". In A. Tan (Ed.) *Handbook on Strategic Information Technology and Portfolio Management*, pp.100-112, IGI Global.

[17] J. Dibbern, T. Goles, R.A. Hirschheim and B. Jayatilaka, "Information Systems Outsourcing: A Survey and Analysis of the Literature". In *The DATA BASE for Advances in Information Systems* 35(4), 2004, pp. 6-102.

[18] K. E. Pearlson, *Managing and using Information Systems: A Strategic Approach*, New York: John Wiley & Sons, 2001.

[19] L. Ho, A. S. Atkins, I. Prince, and B. Sharp, "Alignment of a Strategic Outsourcing Framework to Practical Case Studies". *In Proceedings of the European and Mediterranean Conference on Information Systems (EMCIS)*, Costa Blanca, Alicante, Spain, July 6-7, 2006.

[20] B. Sharp, A. S. Atkins, L. Ho, H. Kothari, and D. Paul, "Intelligent Agent Concepts for Outsourcing Decision-Making in Customer Service Operations". In *Proceedings of the IADIS International Conference WWW/Internet (ICWI)*, October 2005, Lisbon, Portugal, pp. 215-221.

Supply Chain Design in the Agricultural Sector to Reduce Poverty

P. Khuntonthong[1], G. Neubt[2] , and N. Chakpitak [1]
[1] College of Arts, Media and Technology, Chiang Mai University, Thailand.
[2] LIESP / ESC ST Etienne, 51-53 Cours Fauriel, 42009 St Etienne Cedex –France

***Abstract*— Most of the poor people in the world are farmers whose livelihoods are based mainly upon agriculture; therefore, agriculture can provide an effective means for reducing poverty and accelerating economic growth. As poverty is the main cause of hunger, an increase in self-reliance with regard to food in the households of farmers in remote areas with scarce resources is as important as increasing income. In order to resolve the problems associated with this situation; however; access needs to be gained to the entire value chain system for goods. As revealed from a previous survey, an alternative to agricultural practices called 'the New Theory Agricultural' process, can help farmers to gain a sufficient income, and become more self-sufficient in food. This paper aims to present a qualitative research method used for analyzing supply chains and identifying effective supply chain redesign strategies. By applying this process, the main goal is to propose a supply chain management strategy for agriculture that can solve the problems arising from external uncertainties and constraints. The process consists of three stages: self reliance - sufficient food for consumption within the household, a small-scale farm's ability to generate sufficient income from a surplus supply and minimizing the total costs of the supply chain.**

Index Terms – agricultural sector; reducing poverty; supply chain design; Sufficiency Economy Philosophy; supply chain management.

I. INTRODUCTION

Seventy-five percent of poor people in the world live in rural areas, and for the most part, their principal sources of livelihood depend on agriculture, forestry, fisheries and related activities [1]. They depend on agriculture for their earnings, either directly as producers or hired workers or indirectly in sectors which derive their existence from farming. While the lack of sufficient income to purchase food is clearly a major factor causing household food insecurity, hunger itself contributes to poverty by lowering labor productivity, reducing resistance to disease and restricting educational achievements.

Agriculture is the most likely, potential source of significant economic growth in many developing countries. Historical experience suggests that agricultural growth and increases in agricultural productivity may be a prerequisite to broader based sustained economic growth and development (DFID, 2002: p. 9, cited in Humphrey, 2006)[2], and agricultural development is normally used as a tool to reduce poverty in developing countries. Many documents reveal that poverty reduction is an effective method for tackling hunger and promoting development [3]. In agriculture based countries, increasing agricultural production boosts rural incomes, which helps to reduce poverty and assists in development, since people depend on agriculture as their main source of income.

In order to sustain benefits for small-scale farmers, some different points of view argue that to solve the problem of poverty and hunger, the problem that should be addressed is the gaining of access to the high-value market chains for goods as well as for labor, not just setting up farmers' cooperatives and leaving the farmers to fend for themselves in a complex marketplace. The new rural reality requires more comprehensive, market-oriented and context-specific approaches, in which rural stakeholders, private investors and donors actively participate.

The aim of this paper is to address the issue of poverty and hunger reduction by proposing alternative ideas that help the rural poor gain access to value chain markets, as well help improve the level of production for household consumption and for generating cash income. Based on the "New Theory Agriculture" process developed in Thailand, a process designed to make small farmers more self-reliant, this paper proposes a methodology for designing the food supply chain in this context. This methodology is based on the identification of all the uncertainties and constraints within the supply chain for small farmers, including products, production, demand management, sourcing and distribution.

II. AGRICULTURE, FOOD SUPPLY CHAINAND POVERTY

Agricultural markets play an important role in economic growth, but the main problem is the welfare of small-scale farmers, those who live in remote areas and have access to scarce resources. Since marketing system of agricultural products has changed due to the globalization, as Taylor (2007) said that globalization and market integration are a fact of life that will continue regardless of whether or not the rural poor are prepared for it. Globalization and integration, marked by the WTO and regional trade agreements, have increased agricultural producers' access to foreign markets while exposing them to greater competition from abroad. This

exposes agricultural producers to international price gyrations but brings potential benefits as well as threats. In most countries agricultural markets are being transformed because of the new demands of exporters and supermarkets.

As Dixon (2001) reveals [4], agricultural development provides an effective means for both reducing poverty and accelerating economic growth, due to it being the principal driving force of the rural economy in developing countries, and its aim is not just to increase the incomes of producers and farm workers, but also to create a demand for non-tradable goods, services and local products. With a large population of small-scale farmers, agricultural growth can reduce urban poverty more rapidly than urban growth itself. The main reasons for this is are the reduction in food costs in urban areas and the lower rates of outward migration from rural areas. There is strong evidence to suggest that productivity based agricultural growth, especially by small producers, has an overall positive economic impact on rural areas. In order to gain these advantages though, small-scale and artisanal food processing techniques may need to give way to more organized, capital intensive processing plants, and a similar scenario may apply for other activities up- or down-stream of the primary agricultural base. However, the "new agriculture" requires high levels of skill and capital inputs. The rapid change in agricultural marketing systems, from traditional markets to markets associated with large-scale supermarket retailing and wholesale operations, is the cause of the reduction in outlets available for small-scale producers. As a result, the risks of falling into poverty for these producers and in the rural community as a whole are increasing.

This change affects the value chain of agricultural products for consumers, retailers, wholesalers, processors and producers, and with enormous implications for the competitiveness and future viability of small-scale producers. Obstructions can be overcome with the right sort of support and the presence of good business partners; small-scale producers can be efficient and reliable providers of quality products. In order to enhance the opportunities available for small-scale producers in the new business world, the ability of them and their organizations to access modern and traditional markets has been a concern, as has the anticipatory public policy framework and the involvement of the agri-food business in the development of the market (Vermeulen,2008)[5].

As stated by Humphrey (2006), Gustavo Anríquez and Kostas Stamoulis (2007)6, and Anríquez et al., agriculture is an engine for growth and an effective means to reduce poverty in rural areas. Agricultural growth is most effective at eradicating poverty, particularly in countries that are not characterized by high levels of income inequality, because it provides a direct link to the poor. The phenomenon of hunger and famine might occur even there is no food shortage, and as Sen (1981) states "not by food shortage[s] but by the shortage of income and purchasing power" which, in turn, depends on "a person's ability to command commodities in general and food in particular". Also for Sen (1999), certain groups of people, if they lack purchasing power, can starve in a society, even though there is no problem with the aggregate availability of food, the markets are well stocked and prices are low. (Sen, 1981 and 1999, cited by Dowlah, 2006)[7]. As hunger and poverty are closely related; therefore poverty eradication is one of the methods that can be used to help solve the problem of food insecurity for the poor.

For small-sale farmers, improving production for household consumption and income generation purposes is crucial (Taylor J. E et al., 2009).8 Similar to Taylor, as Vermeulen (2008)9, Cobb et al. (1999, p. 209)10, and Cobb et.al. (1999, p. 209) recognise "the food chain as a whole is the ultimate framework for a scrutiny of sustainability". The emphasis on sustainable food chains is an important departure; because, in this context, sustainability is seen in terms of the wider agro-food system, especially the vertical linkages in the food supply chain, including the farm.

For the food industry, the supply chain for agrifoods, as in any other supply chain, is a network of organizations working together on different processes and activities in order to bring products and services to the market, with the purpose of satisfying customer demands(Christopher, 2005 , cited in Ahumada, 2009)[11].

To improve performance of the supply chain in the food industry, Vorst (2006)[12], identifies uncertainty and several improvement principles in the food chain in which a simulation model help quantify the effects of alternative configurations and operational management concepts. The identified sources of uncertainty in his study are: order forecast horizon, input data, administrative and decision processes, and inherent uncertainties. The results of the case study suggest that uncertainty reduction can help to improve service levels significantly, although current supply chain configurations restrict the possible benefits. In addition, Sternet al. (Stern et al., 1996, cited by Vorst, 2002)[13] propose a method to redesign the value chain by laying out a marketing channel planning approach that permits there orientation of distribution systems in order to be more responsive to customer needs. Stephen J. New[14] outlines that not only the domain of supply chain management research (such as, how many links are there in the chain?) and the range of methodologies deployed, but also the scope and social/ideological constitution of supply chain innovations and their ethical and political

implications; such as perspectives on the social, political and ethical aspects of supply chain management, compressing advantages into "efficiency", doing better with the same or less amount of investment in resources and driving the logic of the supply chain, suggest that a holistic approach is more efficient than the alternatives.

III. CASE STUDYOF SMALL-SCALE FARMERSIN THAILAND

In Thailand, 40.05% (25 million) of the population works and lives in the agricultural sector (in 2006) and most of them live in the northeast and northern regions of Thailand, where the percentage of those below the poverty line is equal to 18.9% and 9.8%, higher than other regions. An article in the Human Development Report of Thailand states that the highest percentage of poor people, by comparing the occupation of household heads (1988 to 1996) is the farmer. Furthermore, the debt per household of Thai farmers has increased dramatically, from 43,415 Baht in 2002 to 68,158 Baht in 2008, a rise of 36%.

In recent years there has implicitly been only one blanket policy for rural or agricultural development for the whole of Thailand; the farmer has to specialize and to produce commercial crops in response to market demand. Under the market-oriented economic policies, the rural sector in Thailand has contributed to economic growth by generating many agricultural commodities which can compete on the world market, but not all farmers have received these benefits, especially the small-scale farmers in rural areas. Due to a variation in farm characteristics, the correct approach to developing the agricultural sector should be to tackle poverty problems, and not merely depend on the same blanket growth oriented policies. There are two main types of farmland in Thailand: the irrigated fertile areas and the rain-fed infertile areas.

The minority group of farmers lives in irrigated fertile areas, and enjoys a high productivity of exported crops, have lower supply risks and experience lower transportation costs due to their short distance from a major port. The ideal approach to developing this group should be to focus on the competitiveness of the exported crops on the world market. However, the majority group of farmers who live in rain-fed infertile areas have difficulty in producing and marketing due to high risks on the supply side, such as weather uncertainties and long distances from a major port. The aim of a development approach for such a group of farmers should be to focus on the agro-system, shielding it from instability by spreading and minimizing the risks involved [15]. Furthermore, as every farmer is unique, an approach which takes the diversity of opportunities, the problems facing small farmers and their resources into account, is crucial when dealing with poverty and hunger. Economic growth is not the only factor to be considered when looking at poverty reduction. Shifts in the amount of production made available for consumption (shifts in the Consumption/GDP ratio) and shifts in the distribution of consumption among a population (Lorenz curve shifts), can also have a large impact on poverty levels [16].

An alternative approach to agricultural development, one which aims to sustain and balance the development of farmers in Thailand, has been guided by His Majesty, King Bhumibhol, and is called the "Sufficiency Economy Philosophy". In 1974**,** The Sufficiency Economy Philosophy[1]concept was first mentioned, calling for balanced and sustainable development at all levels. The Sufficiency Economy concept is a philosophy which serves as a "guide for the way of living/behavior of people at all levels, toward the middle path". The system of agriculture under the Sufficiency Economy Philosophy, commonly known in Thailand as 'New Theory Agriculture', suggests for small-scale farmers a new sustainable style of agriculture which moves towards self reliance for the rural household. The main purpose of New Theory Agriculture is to make farmers more self-reliant through holistic management of their land, while living harmoniously with nature and within society.

This is different from previous agricultural development strategies, which put more emphasis on specialization and mono-culture cropping, producing in response to market demand. New Theory Agriculture has three stages: (1) sufficiency at the household level, allowing farmers to be at least self-sufficient in terms of food, and creating a proportionate income from selling extra crops and products beyond the necessary consumption of the household; providing a basic self-immunity to farmers against a diverse range of adversities, (2) sufficiency at the community level by encouraging people to seek benefits by cooperating with neighbors in order to reduce the costs of raw materials and to carry out marketing activities, as well as to increase the community's bargaining power when selling surplus products to the market. These activities should lead to more self reliance

[1]"Sufficiency" means moderation, reasonableness and the need for a self immunity mechanism against the impacts arising from internal and external changes. To achieve this, the application of knowledge with due consideration and prudence is essential. In particular, great care is needed in the utilization of theories and methodologies for planning and implementation at every step. At the same time, it is essential to strengthen the moral fiber of the nation, so that everyone, particularly public officials, academics and businessmen at all levels adheres first and foremost to the principle of honesty and integrity. In addition, a way of life based on patience, perseverance, diligence, wisdom and prudence is indispensable to create balance and be able to cope appropriately with critical challenges arising from extensive and rapid socioeconomic, environmental and cultural changes in the world.

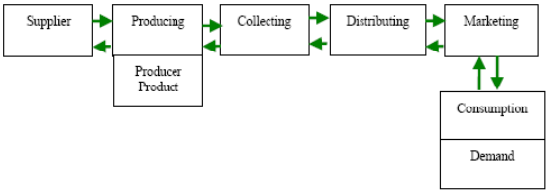
ISBN 978- 974-672-556-9

among communities, while enhancing the capacity of community members to reduce living costs, increase income, and create a community social safety net. This approach can be compared to a cluster of businesses developing in the same locality, with similar activities, in order to achieve economies of scale as well as economies of scope, and (3) sufficiency at the national level; communities are encouraged to expand their activities by reaching out to co-operative firms, banks and other outside sources. The expansion across different levels of organizations or activities can be compared to developing a value-chain in production [17]. The aim of this New Theory Agriculture is for farmers to become self-supporting by dividing the land into four parts, according to the proportions 10-30-30-30. The first 10% of land will be used for housing, perhaps with a small amount of livestock and poultry. The remaining parts, at 30% each, are: paddy fields for self-consumption, with a likely surplus that can be sold on the market, reservoirs, with a depth of four meters to be used for water supplies and with the possibility of raising fish, and other crops such as vegetables, fruit and flowers.

A study from 59 farm households has shown the benefits of the New Theory Agriculture process when compared to extensive mono-cropping. The challenge, when wishing to implement the New Theory Agriculture efficiently, is how to identify specific agricultural and rural development needs and opportunities, and to focus investment in areas where the greatest impact on food insecurity and poverty will be achieved. This identification and resource allocation process can be facilitated by analyzing farming systems, the linkages to markets, as well as the welfare of the farmers, in order to develop an understanding of local factors and linkages.

IV. Research Methodology

This study suggests a re-design of the value chain for small-scale producers in the agricultural sector, in order to solve the problems of poverty and hunger, through implementation of the New Theory Agricultural approach in rural communities. As part of its full scope, the New Theory Agricultural consists of three stages. The first stage aims to create self-reliance and self-sufficiency at the household level, addressing food security for small-scale farmers and their farming systems. The purposes of the second and third stages are to create sufficiency at the community level and expand activities across different organizational or activity levels, a concept in line with the value chain model. This chain can be shown as follows:



The functions of the chain include: economic aspects such as farmers' incomes and costs, social perspectives in the community such as food security, nutrition security, poverty and debt, and environmental aspects such as natural resources and chemical residues in the environment.

The core methodology of this approach:

1) State the value chain objectives
2) Map the existing value chain; the flows of products to provide an accurate description of the current distribution system
3) Define the uncertainties and constraints
4) Define the causes of supply chain uncertainties and constraints, drivers and trends, current constraints and threats posed by internal and external factors. To do this use cause and effect analysis, and
5) Redesign the chain by applying the New Theory Agriculture; to give an ideal distribution system.

Although the definitions of poverty vary substantially from one point of view to another, this study focuses on the hunger experienced in rural areas, as this is the locus of food production and the definition followed by the poor themselves in Thailand, as when asked about poverty, the most frequent responses among the poor are: not enough to eat, low income, no land for agriculture, indebtedness, illness, poor general health or disability (Jitsuchon, 2001)[18].

V. Objective

As stated at the beginning of this paper, this study focuses on the hunger experienced in rural areas, despite the fact that they are the locus of food production. Therefore, the objectives of this study are in line with the objectives of the sufficiency economy, which are; self reliance - sufficient food for consumption within the household and community, the ability to generate a sufficient income from sales of surplus supplies, and to minimize the total costs of the supply chain.

VI. Production, Marketing and Social Situation of the Case Study

The case study area is in Nan Province, located in the north of Thailand, where a development project has been set up to improve the livelihoods of the small-scale farmers. The area is a sub-tropical and temperate environment, with a hill landscape of moderate to steep slopes. The local soils are low in fertility, shallow and susceptible to erosion.

ISBN 978- 974-672-556-9

Forested areas are scattered throughout the area, and deterioration of natural resources and the overall environment has occurred, a result of the increased price of maize due to the high demand for bio-fuels, which has led to the extensive cultivation of fragile slopes without the adoption of appropriate soil and water management practices, as well as the legacy of chemical residues in the water.

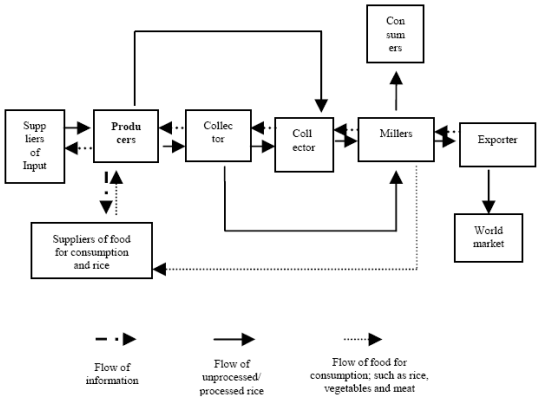
Households are vulnerable to natural disasters and crop failures and the local infrastructure is poorly developed. Due to it being a remote area, links to the market and other systems are poor, and moderate to severe poverty can be found throughout the area. Most agricultural crops are grown under rain-fed conditions. Data collected by the project's staff show that the crops grown are upland rice (14.96% of total arable land), maize (47.21%) and fruit (34.94%), with rice as the staple crop; grown for household consumption, while maize and fruit are grown for cash income. The rice yield is very low and so the household food situation is insecure.

The households have an annual average income of only 101,691.95 Baht; considerably below the international poverty line. The debt per household is 108,873.76 Baht, higher than the average farm operators' debt in the country (which is 84,922 baht per household (2007)19). The households raise pigs and cattle for meat and cash income. Livestock growth rates and production levels are low because animals are raised under conditions of poor animal husbandry and animal health practices. Farms operate semi-subsistence production systems with only the limited sale of products to meet livelihood needs. Furthermore, they are net food buyers; the proportion of expenditure on food in the area is equal to 39.4%.

A. The present value chain

The main agricultural products in the community are rice and maize. Maize is grown for income cash and the community used to benefit from this crop, but its prices have fallen recently due to low product quality, the result of a lack of access to quality inputs and land degradation. The rice is grown mainly for consumption; however, the yield varies due to the crops vulnerability to natural disasters and crop failures, and a lack of access to quality inputs. Food insecurity occurs in the area due to the limited number of crops grown for consumption. The rice output is sold outside the village but very rarely as the sales chain for rice is long. Some households have to buy rice imported from outside the village, and all the vegetables and most of the protein sources such as eggs, pork and chicken are delivered from a town outside the village – a long distance away.

The Present Flow of Goods, Service and Information in the Rice Supply Chain



B. Supply chain uncertainties and constraints

The main aim of this study is to present a qualitative study analyzing the supply chain, in order to identify effective supply chain strategies to solve the problems arising from exit uncertainties and constraints in the study area. The New Theory Agricultural process has been applied within the supply chain, the main purpose being to make farmers more self-reliant through the holistic management of their land, while living harmoniously with nature and within society. The three stages of the New Theory Agricultural; sufficiency at the household level, sufficiency at the community level by encouraging people to seek benefits from cooperation with neighbors, and sufficiency at the national level, though an expansion of activities by reaching out to co-operate with other outside sources, have been fulfilled through the redesign of supply chain management.

The uncertainties and limitations that exist in the case study area are elaborated upon and an overview given of the supply chain redesign strategies used, in the following table.

ISBN 978- 974-672-556-9

Supply chain uncertainties and constraints, and strategies to support change

	Uncertainty and constraints	***strategies for supporting changes***
Product	Low yield of rice and maize	• Intensification of production patterns. • Under-sowing wetland rice with leguminous fodder species before the basin dries out (improve livestock feed supply and also provide green manure for the following crop.) • Emphasis on improving nutrition, stall feeding and the collection of dung. • Development of irrigated land for rice
	Limited sales of products to meet livelihood needs	• Crop diversification, − Finding suitable cash crops for remote areas may be difficult. There may be opportunities for the production of high-value, low-bulk commodities, such as vegetable seeds, spices or medicinal plants. − Diversification requires investments in marketing, transport infrastructure, research, extension and other support services for public and private organizations.
	Small scale	• Organize cooperatives to create a value chain for niche, high value products.
Supply	Lack of power re: agricultural inputs	• Create a value chain of actors involved in input supplies, via a cooperative. • Reforms can be made along the market chain to make it more competitive and efficient. • Poor farmers need to become empowered within the market chain in order to reap the benefits from such changes.
	Poor access to complementary assets	These include: • human capital (education and know-how), • natural capital (ownership of good quality land or access to other households' land via local land markets), • physical capital (the ownership of assets such as livestock and machinery), • public capital (access to public services and infrastructure, such as good roads, schools, health clinics and electricity), • social capital (participation in organizations and associations, and links to other individuals and households, both within and outside the community), • financial capital (access to credit, insurance), and • geographic capital (proximity to markets and favorable agro-climatic conditions).
	High price of inputs.	• Reinforcing the basis for nutrient recycling through crop-livestock integration. • Variety of biological integrative activities − Integrated crop and livestock − Introduce legume crops for cash income and soil improvement. • Development of Integrated Crop Management (ICM) systems for pest control, to reduce the cost of chemical pesticides.
	No adoption of appropriate soil and water management practices	• Reinforcing the basis for nutrient recycling through crop-livestock integration. • Variety of biologically integrative activities − Integrated crop and livestock − Introduce legume crops for cash income and soil improvement.
	Chemical residues in the water.	• Development of Integrated Crop Management (ICM) systems for pest control to reduce chemical use.
	High value of imported food from outside the community	• Improve local market functions • Crop diversification; crop selection for consumption as a first priority and to provide surplus cash income
Demand	Instability of the food price	• Create chain value with the producers or other actors to set the equilibrium interest.
	Unpredictable demand	• Create chain value with producers or other actors to set the equilibrium interest. • Improve market information
	High Consumption Prices	• Create chain value with producers or other actors to set the equilibrium interest.
Producing	Low production of livestock Local infrastructure is poor	• Support from Government
	Vulnerable to natural disasters and crop failures	• Development of Integrated Crop Management (ICM) systems for pest control
	Crops are grown under rain-fed Conditions	1. Improvement of irrigation security 2. Adoption of short-duration crop varieties 3. Improve soil moisture conservation by employing a range of techniques such as mulching.
	Soils are low in fertility	• Increased use of manure

	Uncertainty and constraints	***strategies for supporting changes***
Producer	Food security and inability to diversify income sources	• Crop diversification – Crop selection for consumption as a first priority, and a surplus of high value crops for cash income. – Finding suitable cash crops for remote areas may be difficult. There may be opportunities for production of high-value, low-bulk commodities such as vegetable seeds, spices or medicinal plants.
	High proportion of Food Expenditure	• Crop diversification, crop selection for consumption as a first priority and use the surplus for cash income.
	Poverty traps	• Reduce the consumption costly food through self sufficiency • Reduce risk and increase incomes by diversifying crops • Increase off-farm income • Reduce costs of input by recycling within the farm. • Highlight debt issues • Access to markets for high return crops. • Access to credit to get rid of a liquidity-constraints • Access to agricultural inputs and output
	High debt	• Increase income • Reduce costs of consumption
	Malnutrition	• Outputs used on chicken, eggs, fish and the production of fresh vegetables have grown rapidly, suggesting that these crops should be grown for household protein sources and cash income.
Marketing	Weak market Linkages	• Create chain value with producers or other actors to set the equilibrium interest. • Create chain value via cooperatives or groups, to access input and output markets.
	A lack of power in agricultural output markets	• Organize cooperatives • Create value chains for niche and high value products
	Poor access to complementary markets	• Complement farm production and raise the incomes of the rural poor from off-farm income. This invariably means helping poor households gain better access to, not only financial, but also input and output markets on favorable terms. • Local markets, particularly for labor, need to work in order for poor rural households to reap these indirect benefits • Use traditional systems for exchange labor • Having access to efficient agriculture and nonagriculture markets is fundamental.
	Market isolation	• Introduce non-perishable and high value-to-bulk ratio food crops • Geography is often pointed to as a determinant of chronic poverty. – Support from the public
	Poor access to market information	• Roads carry not only people and goods, but also information. • Develop efficient and low cost communications systems – Cell phones can enable farmers and intermediaries to obtain information about markets, without leaving the village.
Distribution	High Transportation costs	• Create a value chain for high value products.

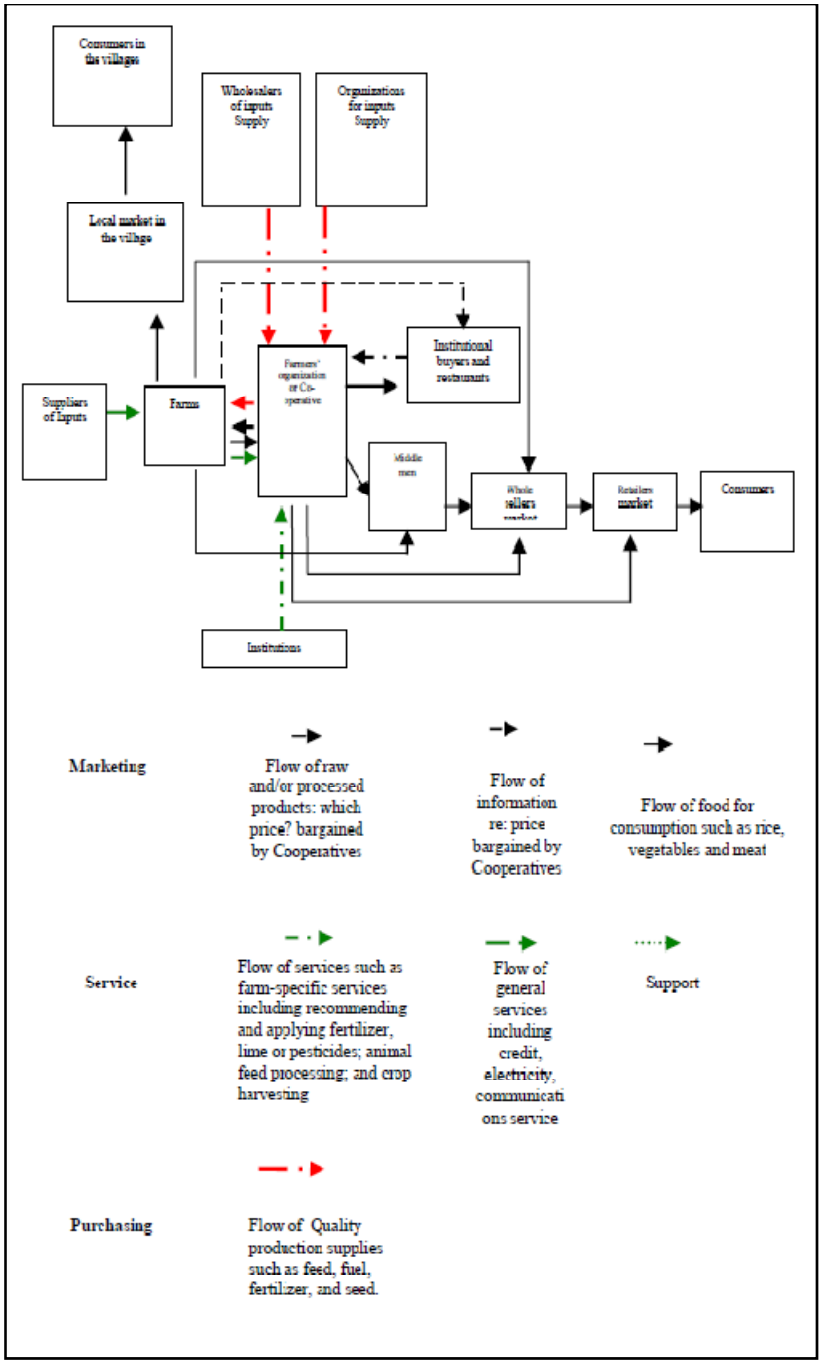
The study area is a worst-case for the poor in Thailand. The uncertainties and constraints which could be overcome by implementing each stage of the New Theory Agriculture process and with support from organizations such as the Government, or from the private sector, are as follows:

	The first stage of the New Theory	***The second and third stages of the New Theory***	***Areas of support from the organizations***
Product	• Limited sales of products to meet • livelihood needs	• Small scale	• Low yield of rice and maize
Supply	• Chemical residues in water. • High value of imported food from outside the community • High consumption prices	• Lack of power in agricultural input • High price of inputs.	• Poor access to complementary assets • No adoption of appropriate soil and water • Management practices
Demand			• Unpredictable Demand
Producing			• Soils are low infertility • Local infrastructure is poor. • Vulnerable to natural disasters and crop failures • Crops are grown under rain-fed conditions
Producer	• Food security and inability to diversify income sources		

	The first stage of the New Theory	***The second and third stages of the New Theory***	***Areas of support from the organizations***
	• High proportions of food expenditure – Poverty traps – High debt • Malnutrition		
Marketing		• Weak market linkages • A lack of power in agricultural output markets • Poor access to complementary markets	• Market isolation • Poor access to market information
Distribution		High Transportation costs	

C. The re-designed value chain

Re-designed Flow of Goods, Services and Information in the Supply Chain



VII. CONCLUSION

The re-designed value chain has been created in order to solve the problems arising from exit uncertainties and constraints. The number of actors in the new value chain is much more important than in the current chain, and it should be note here that, at the beginning of the re-engineering process, the role of the development organizations was very crucial. In order to achieve success it will take some years and will require financial support from public organizations. The farming system needs to change to become multiple cropping, to include at least enough crops to provide the carbohydrates and protein sources needed to meet the minimum nutrient requirements. A local market has to be established, to act as a place to exchange the necessary goods. Farmer cooperatives are an important source of lower price inputs and of knowledge for farmers, as well as acting as an intermediary for welfare issues.

The role of marketing could be handled through farmer cooperation or by middlemen, because the success of these activities ultimately depends on the business skills of those involved. However, the farmers have to obtain reasonable prices from selling their products. Whether collecting and distributing the products is efficient or not partly depends on the production plan in place, due to the farms being scattered across the village. The institutional buyers are good sources for selling the products, due to the consistency in their demand and prices. Poverty and hunger are closely related to each other; increased productivity does not guarantee an increased income and problems of hunger still exist for producers that have scarce resources. There are enormous obstacles for the poor to face in the rural areas, so, in order to solve the problem of poverty and hunger, an increased ability for the household to produce its own food is as important as an increase in income. The basic need for food, to serve minimum dietary energy requirements, is crucial for life. The revenue received for one household should be at least equal to the total expenses on necessary items, including the costs of education and health care, and the key challenge is how to implement New Theory Agriculture practices efficiently and successfully. Therefore, future research should be conducted in this area, in order to simulate these practices in terms of a quantity analysis and thus be able to apply them appropriately and effectively.

REFERENCES

[1] Gustavo Anríquez and Kostas Stamoulis. Rural Development and Poverty Reduction: Is Agriculture Still the Key? Agricultural Development Economics Division, The Food and Agriculture Organization of the United Nations. 2007

[2] John Humphrey, Olga Memedovic. Global Value Chains in the Agrifood Sector. United Nations Industrial Development Organization, Vienna, 2006

[3] Agriculture in the twenty-first century. FAO. 2008

[4] John Dixon and Aidan Gulliver with David Gibbon. Farming Systems and Poverty: Improving Farmers' livehoods in a changing world. FAO and World Bank Rome and Washington D.C. 2001

[5] Vermeulen, S., Woodhill, J., Proctor, F.J. and Delnoye, R. 2008. Chain-wide learning for inclusive agrifood market development: a guide to multi-stakeholder processes for linking small-scale producers with modern markets. International Institute for Environment and Development, London, UK, and Wageningen University andResearch Centre, Wageningen, the Netherlands.

[6] Gustavo Anríquez and Kostas Stamoulis. Rural Development and Poverty Reduction: Is agriculture still the key?. June 2007. FAO.

[7] C. Dowlah The politics and economics of food and famine in Bangladesh in the early 1970s – with special reference to Amartya Sen's interpretation of the 1974 famine. INTERNATIONAL JOURNAL OF SOCIAL WELFARE.*2006:*15: 344–356

[8] J. Edward Taylor, Alberto Zezza, AliArslan Gurkan. Rural Poverty and Markets. FAO 2009

[9] Vermeulen, S., Woodhill, J., Proctor, F.J. and Delnoye, R. 2008. Chain-wide learning for inclusive agrifood market development: a guide to multi-stakeholder processes for linking small-scale producers with modern markets. International Institute for Environment and Development, London, UK, and Wageningen University and Research Centre, Wageningen, the Netherlands.

[10] Brian Ilbery, Damian Maye. Food supply chains and sustainability: evidence from specialist food producers in the Scottish/English borders. Land Use Policy 22 (2005)331–344.

[11] Omar Ahumada, J. Rene Villalobos . Application ofplanning models in the agri-food supply chain: A review.European Journal of Operational Research 195 (2009) 1–20

[12] J.G.A.J. van der Vorst 1 A.J.M. Beulens 2 W. de Wit 3 P.van Beek 4. Supply Chain Management in Food Chains: Improving Performance by Reducing Uncertainty. International Transactions in Operational Research. Volume 5Issue 6, Pages 487 – 499. Published Online: 24 Aug 2006.

[13] Jack G.A.J. van der Vorst, Adrie J.M. Beulens. Identifying sources of uncertainty to generate supply chain redesign strategies. International Journal of Physical Distribution &Logistics Management, Vol. 32 No. 6, 2002, pp. 409-430

[14] *Stephen J. New* . The scope of supply chain management research. Supply Chain Management Volume 2 · Number 1·1997 · pp. 15–22 © MCB University Press · ISSN 1359-8546.

[15] C. Isarangkun., K. Pootrakool. Sustainable Economy Development Through the Sufficiency Economy Philosophy.

[16] Sufficiency Economy. National Economic and Social Development Board. 2006.

[17] Sufficiency Economy. National Economic and Social Development Board. 2006

[18] Somchai Jitsuchon. 'What Is Poverty And How To Measure It?' This article was presented at the 2001 Year-end Conference on Poverty Reduction Strategies, held at the Ambassador City Jomtien, Chon Buri, on November 24th2001.

[19] The 2007 'Household Socio-Economic Survey; Whole Kingdom', National Statistical Office, Ministry of Informationand Communication Technology

Visual Simulation-based Hospital Theatre Management System

Xin Zheng, Hongnian Yu, and Anthony Atkins
Faculty of Computing, Engineering & Technology
Staffordshire University, Stafford, UK ST18 0DG
x.zheng@staffs.ac.uk, h.yu@staffs.ac.uk, a.s.atkins@staffs.ac.uk

***Abstract* -In a competitive environment, hospitals are currently required to reducing waiting lists, improve patient throughput and provide high standards of patient care. Consequently, hospitals should provide fast and efficient services for patients and treat them as soon as possible. A traditional Hospital Information System (HIS) should be able to show the accurate, current situation of a hospital patient through its real-time data collecting modules. However, this system normally lacks a simulation analysis module to help hospital managers review the results under different scenarios, for example changing their policies and procedures of bed management to improve their services. This paper proposes a framework of a simulation-based decision support system of hospitals, and a 3D visual simulation system has been developed for a hospital case study.**

Index Terms- Decision Support System; Hospital Information System; Simulation; Witness

I. INTRODUCTION

In a competitive environment, hospitals are under pressure to reduce waiting lists, increase patient throughput and provide better healthcare. Information systems are being introduced into hospital management to assist in improvements in efficiency and reduction of costs [1]. Consequently, governments are encouraging investments in IT budget allocations for improving management information systems in hospitals [2].

Currently, most Hospital Information Systems (HIS) are providing healthcare information at operational level and this could be enhanced by using simulation techniques to aid decision-making by „what-if" analysis. Simulation models are being developed for most aspects of HIS, such as in emergency departments (ED) [3, 4], income estimation [5], operating rooms, training [6, 7], bed management [8] etc.

In an ED, a discrete-event simulation model was developed to mimic patient treatment flows at the University of Kentucky Chandler Hospital by SIMUL8, and this model emulates workflows with such factors as laboratory tests, doctors and nurses which are based on a bed assignment for only one patient. In this situation, a patient occupies a bed, and a series of events which include testing, doctor and nurse visits will occur according to predetermined work flows [4]. A simulation model integrated with an optimization technique of a Genetic Algorithm (GA) is proposed to reduce patients" waiting time for an ED and it tries to generate a "good enough" nursing schedule using GA [3]. In Japan, a simulation model has been used to calculate hospital incomes using several effect factors such as Average Length of Stay (ALOS) and Bed Occupation Rate (BOR) for a number of patients under different scenarios [5]. Swedish researchers have developed a health simulation model to evaluate the change in numbers of potential patients and the changing costs of a hospital if new surgery legislation is introduced. The model takes medical priority, limited resources of operation rooms, beds, doctors etc. into account for making a surgery schedule and compares the results of different surgical management scenarios [9]. In bed management, a simulation model is designed to find a required number of beds under the current situation by a key factor of LOS ratio of surgical and internal medicine departments [10]. Another simulation model takes LOS ratio, service quality targets and admission levels into account and combines the characteristics of deterministic and stochastic to estimate an optimal number of beds [8]. These successful applications of simulation show simulation is a key method to aid the decision making of a HIS.

In a UK-based case study, a simulation-based decision support system is required for improving efficiency and providing better healthcare services for patients. This simulation system is developed from a proposed framework of the generic simulation-based decision making support system for a small and medium size Supply Chain (SC). This generic framework almost satisfies most of requirements of the hospital and the data collection module will adopt a real-time capturing data mode using a Radio Frequency Identification (RFID) technique to overcome manual mistakes and time delays.

This paper proposes a framework of a hospital decision support system and develops a cell simulator for the theatre management of a UK hospital. The paper is organized as follows: the framework descriptions are demonstrated in Section II. Section III briefly introduces the concept of a simulation

ISBN 978- 974-672-556-9

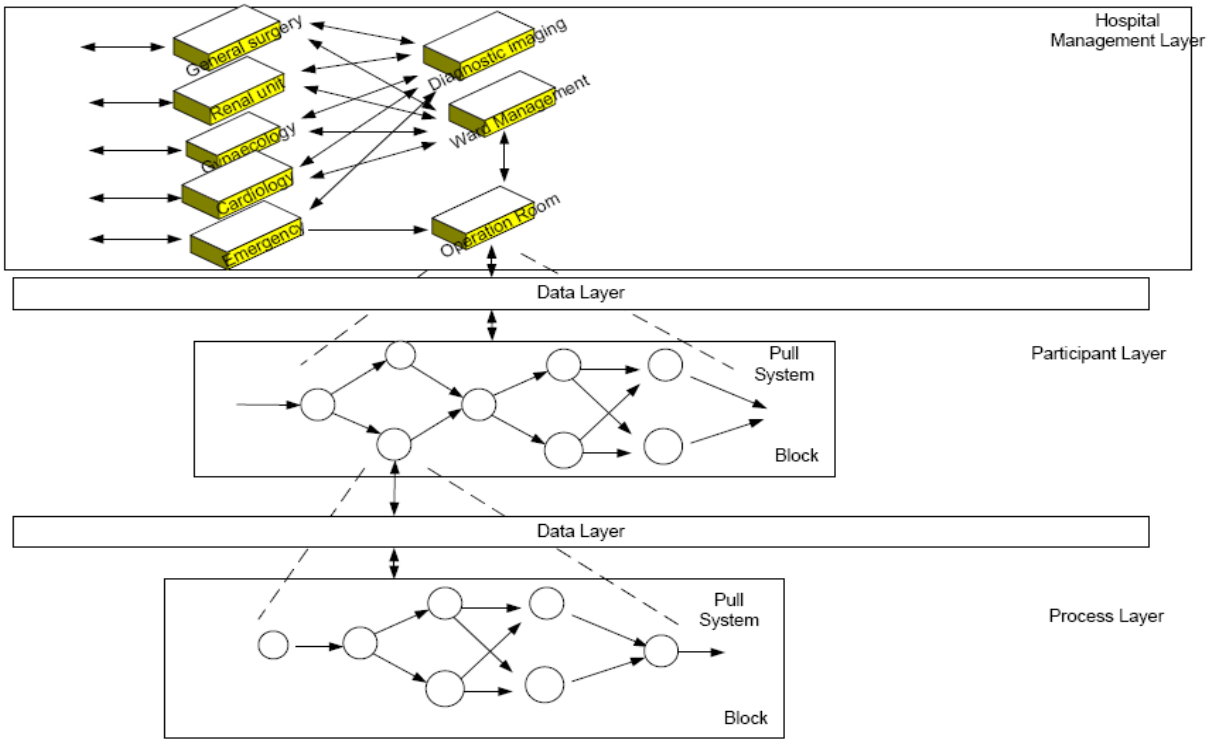


Figure 1. Framework of Hospital Decision Support System

block and shows the structure of the simulation block. Section IV describes a case study and gives patient treatment flows of the theatre room. Section V develops a cell simulator for the case study using a WITNESS simulation software package. Section VI shows the 3D characteristics and performances analysis of the cell simulator. Section VII draws conclusions and outlines future research work.

II. Framework Description

The framework of a hospital decision support system is developed from a proposed generic framework of a decision support system of a supply chain (SC) shown in Figure 1. This framework maintains three layers of the original framework and the top layer is the hospital management level, representing the relationship between different departments or units, such as surgery department, operation room, ward management, diagnostic imaging etc. A department or unit is considered to be an internal individual unit (simulation block) and can communicate with other departments independently. The participant layer expresses the workflows of a department or unit. If a special resource or a series of processes is required to be monitored the process flow can be modelled into the process layer. Different layers communicate with each other through related data layers. This hierarchical structure and an internal independent characteristic are used to simplify a complicated structure, making it easy to assign tasks from the requirement analysis, encode, test, and deploy to developers, as well as decreasing programming risks.

III. Simulation Block

A simulation block represents an internal independent unit and owns its personal workflows. This block can connect other blocks through its data filter and convert layer. In a hospital, a department or unit is able to be modeled as a simulation block.

Figure 2 shows the structure of a simulation block and it adopts a 4P‟s (people, process, product, performance) closed-loop control cycle which can maintain evenness in a dynamic environment and accurate tune to achieve objects.

A simulation block consists of a cell simulator, database, decision maker (i.e. management) or alternatively uses a control program with a global optimisation algorithm. There are inputs which come from the data filter and convert layer and are directly sent to its database. The cell simulator can either take the required data from a database or be inputted manually. The results concerning enterprise performance are provided after running this simulation. Managers (global optimisation algorithms) can check and compare the results and alter the strategic polices or management rules and then obtain the relevant results from the simulation model. The managers (people), input datasets (process), cell simulator (product), and results (performance) comprise a 4P close-loop control cycle. Through this cycle, the performance of an enterprise can be improved step by step.

ISBN 978- 974-672-556-9

IV. CASE STUDY

The case study concerns the theatre room management of a UK hospital and the theatre room is considered to be a simulation block of HIS. Thus, a cell simulator is required to be developed to aid the decision-making aspect.

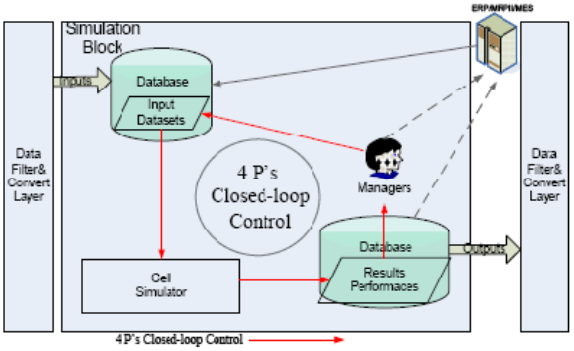


Figure 2. A Structure of a Simulation Block

The patient treatment flow is shown in Figure 3. When a set of anesthetics and theatre rooms is available, the patient is informed and moved from the waiting area via a corridor to the available anesthetics room. In the anesthetics room, a doctor can perform several preparatory works prior to the operation on the patient. The patient is then moved into the theatre room for the operation. After the operation, the patient is moved to the recovery room and occupies a bed. When a doctor has given permission, the patient will be moved back to their ward from the recovery room.

In the hospital, there are only six sets of anesthetic and theatre rooms, and one recovery room is used with six beds. Only one patient can be treated in a set of anesthetic and theatre rooms, and in the recovery room a patient only occupies one bed.

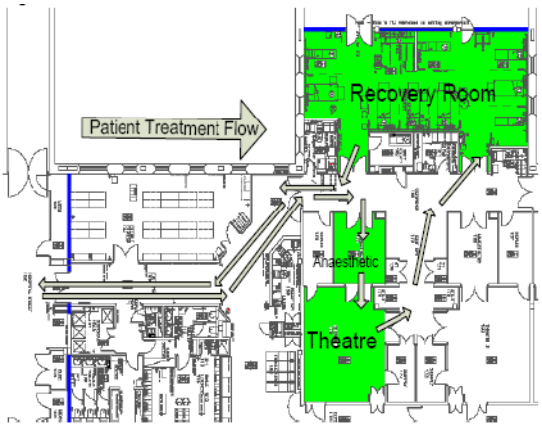


Figure 3 Patient Treatment Flow of Theatre of a Hospital

V. CELL SIMULATOR

A simulation cell model developed by WITNESS simulation software [11，12，13] is used to mimic the patient treatment flow for theatre management as shown in Figure 4. The patient is modeled as a "part", and input into the waiting area (ward management is not considered in the current stage) and set as an active work type. This means the part will be generated according to defined parameters, such as a defined distribution, interval time, or a defined function etc. One buffer, which represents a patient waiting list, is located in the waiting area and accommodates patients from the "part" module. When a combined anesthetic and theatre room set is available after sterilisation, a patient is moved into the available anesthetics room and then to the related theatre room. These two rooms are modeled as two "machines" and both are single types of machines because they only deal with one patient each time. In the output rule of the anesthetics room, a patient is set to move directly into the related operation room without passing through buffers. Doctors and nurses are viewed as "employees" and they go to an anesthetics or theatre room for treating patients according to the labour rules of the related "machine". Because the recovery room with six beds is a limited resource, one buffer is set between the recovery room and theatre rooms. Patients are moved into this buffer after their operation and if one bed is available in the recovery room, a patient is taken from this buffer to occupy the bed. The recovery room is modeled as a "machine group" with six "machines" and every "machine" represents one bed. The length of stay in the bed is set as a cycle time of the related "machine".

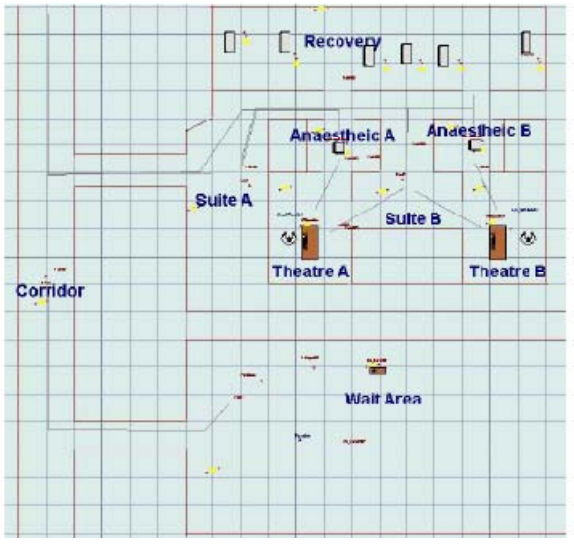


Figure 4. Cell Simulators by WITNESS

VI. PERFORMANCE ANALYSIS

In the cell simulator, there are two queues, and several machines and employees. Two queues are able to adopt Priority Rules (PR), such as First In First Out (FIFO), Shortest Processing Time (SPT), Longest Processing Time (LPT), and Earliest Due Date (EDD) etc., in order to compare their effects on theatre room management. In the "part" model, different distributions or unregulated urgent patients

are used to evaluate the treatment capability of the theatre room and determine its bottleneck.

The proposed cell simulator is able to provide a virtual 3D view by a quick 3D function of WITNESS. Figure 5 shows several different monitoring views of the simulation model. These views enable hospital management to see it from front, slide, and plan camera angles when running the simulation model. Managers are able to choose one employee's or one object's view to review what is going on. These virtual characteristics help mangers to understand the scenario better and bring positive effects to making a decision.

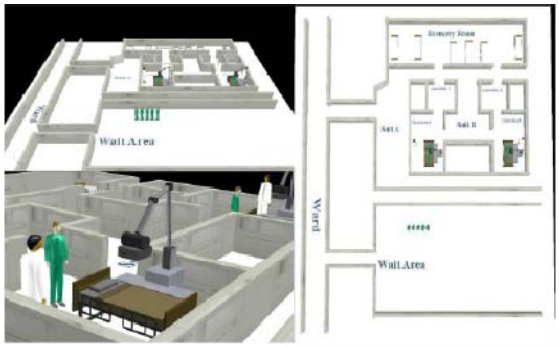


Figure 5. 3D Monitor Views of a Cell Simulator

The theatre cell simulator is able to provide several performance analyses, such as room utilization, bed utilization and waiting list analysis etc.

Figure 6. Room Utilisation

Figure 6 shows different anesthetics and theater rooms' utilisations under an assumed scenario and the theater1 room utilisation is achieved to 85%. Several sets of anesthetic and theatre room utilisations need to be compared to find an optimal dispatch rule to suitably assign patients into sets of anesthetics and theatre rooms.

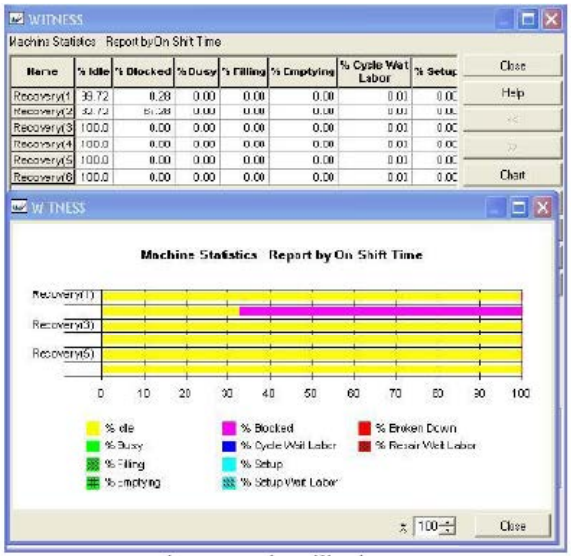


Figure 7. Beds Utilisation

Bed utilization in the recovery room is shown in Figure 7. According to the performance, bed 2 is occupied for a longer time than the other beds. Consequently, there needs to be a management balance for all the beds in the recovery room in order to maintain its efficiency.

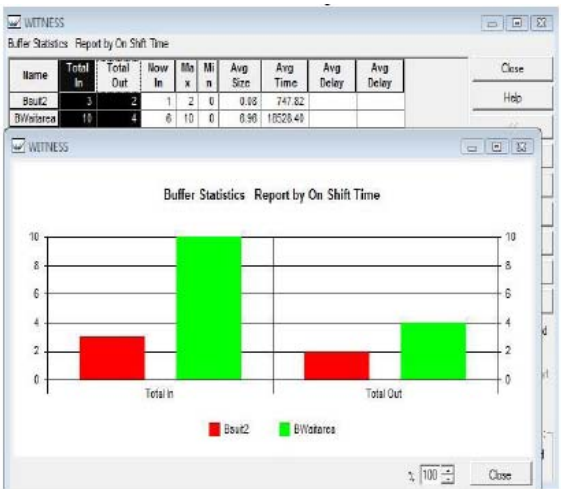


Figure 8. In and Out of Patient Wait List

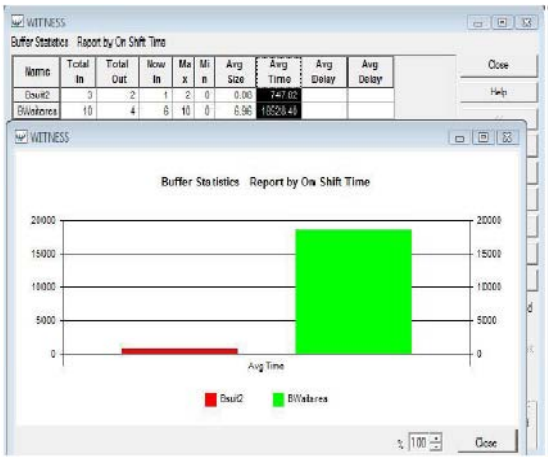


Figure 9. Average Wait Time of Patient Wait List

ISBN 978- 974-672-556-9

Figure 8 and Figure 9 show the performance of patient waiting lists and this helps managers to determine whether there needs to be more space for the list or not. Several PRs, such as FIFO, STP, and LTP etc. will be adopted to evaluate the performance of the theatre room management.

VII. Conclusions and Future Works

This paper proposes the framework of a simulation-based decision support system for a hospital. A cell simulator is developed for theatre room management based on the concept of a simulation block of the proposed framework. The characteristics and performance analyses of the cell simulator are illustrated for further applications.

In future, data will be collected from a HIS of the hospital and classified, mined into the database of the cell simulator. More PRs of the queue system will be applied in the cell simulator to compare their effects and improve the service efficiency of the hospital.

References

[1] P. L. Reichertz, "Hospital information systems--Past, present, future," *International Journal of Medical Informatics*, vol. 75, pp. 282, 2006.

[2] B. Watcharasriroj and J. C. S. Tang, "The effects of size and information technology on hospital efficiency," *The Journal of High Technology Management Research*, vol. 15, pp. 1, 2004.

[3] J.-Y. Yeh and W.-S. Lin, "Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department," *Expert Systems with Applications*, vol. 32, pp. 1073, 2007.

[4]] S. Brenner, Z. Zeng, Y. Liu, J. Wang, J. Li, and P. K. Howard, "Modeling and Analysis of the Emergency Department at University of Kentucky Chandler Hospital Using Simulations," *Journal of Emergency Nursing*, vol. In Press, Corrected Proof, 2009.

[5] T. Abe, S.-i. Toyabe, P. Cao, S. Kurashima, and K. Akazawa, "Development of a simulation program for estimating hospital incomes under the prospective payment system," *Computer Methods and Programs in Biomedicine*, vol. 80, pp. 271, 2005.

[6] J. A. Neil, "Simulation in Nursing Education," *Perioperative Nursing Clinics*, vol. 4, pp. 97, 2009.

[7] P. Blyth, N. S. Stott, and I. A. Anderson, "A simulation-based training system for hip fracture fixation for use within the hospital environment," *Injury*, vol. 38, pp. 1197, 2007.

[8] A. Kokangul, "A combination of deterministic and stochastic approaches to optimize bed capacity in a hospital unit," *Computer Methods and Programs in Biomedicine*, vol. 90, pp. 56, 2008.

[9] M. Persson and J. A. Persson, "Health economic modeling to support surgery management at a Swedish hospital," *Omega*, vol. 37, pp. 853, 2009.

[10] J. M. Nguyen, P. Six, D. Antonioli, P. Glemain, G. Potel, P. Lombrail, and P. Le Beux, "A simple method to optimize hospital beds capacity," *International Journal of Medical Informatics*, vol. 74, pp. 39, 2005.

[11] X. Zheng, H. Yu, and A. S. Atkins, "Inventory Management Simulation using a Novel Replenishment Algorithm," presented at IEEE ICNSC 2007, London, 2007.

[12] X. Zheng, H. Yu, and A. S. Atkins, "A transferring approach – from a Witness model to a Petri net," presented at Proceedings of the 13th International Conference on Automation and Computing, Stafford, England, 2007.

[13] X. Zheng, H. Yu, and A. S. Atkins, "Simulation-based Dispatching Optimisation Algorithm in Batch Scheduling," resented at Proceedings of 2009 International Conference on Software, Knowledge, Information Management and Applications (SKIMA'09), Fes, Morocco, 2009.

Session 6: Security, Software Testing and Engineering

ISBN 978- 974-672-556-9

Review on the Advancements of DNA Cryptography

Beenish Anam, Kazi Sakib, Md. Alamgir Hossain, and Keshav Dahal
School of Computing, The Bradford University
West Yorkshire, UK
b.anam@brad.ac.uk, k.muheymin-ussakib@bradford.ac.uk,
M.A.Hossain1@Bradford.ac.uk,k.p.dahal@Bradford.ac.uk

***Abstract*—Since security is one of the most important issues, the evolvement of cryptography and cryptographic analysis are considered as the fields of on-going research. The latest development in this field is DNA cryptography. It has emerged after the disclosure of computational ability of Deoxyribo Nucleic Acid (DNA). DNA cryptography uses DNA as the computational tool along with several molecular techniques to manipulate it. Due to very high storage capacity of DNA, this field is becoming very promising. Currently it is in the development phase and it requires a lot of work and research to reach a mature stage. By reviewing all the potential and cutting edge technology of current research, this paper shows the directions that need to be addressed further in the field of DNA cryptography.**

Index Terms – cryptography; DNA; DNA cryptography.

I. INTRODUCTION

DNA cryptography, a new branch of cryptography utilizes DNA as an informational and computational carrier with the aid of molecular techniques. It is relatively a new fieldwhich emerged after the disclosure of computational ability of DNA [1]. DNA cryptography gains attention due to the vast storage capacity of DNA, which is the basic computational tool of this field. One gram of DNA is known to store about108 tera-bytes. This surpasses the storage capacity of any electrical, optical or magnetic storage medium [2], [3].

DNA is being proposed to use for many computational purposes. For example, Barish et. al. demonstrated a tile system that takes input and produces output using DNA [4].The method is now also used to solve many NP-complete and other problems. Such as Rothemund et. al. showed that DNA can also be used to compute XOR function which is an essential part of cryptosystems [5]. It is a very potential field of research, as work which has been done in this field suggests that it can put many challenges to the modern cryptosystems [6]. By utilizing DNA cryptography, several methods have been designed to break many modern algorithms like Data Encryption Standard (DES) [7], RSA [8], [9] and Number Theory Research Unit (NTRU) [10], [11].

The research of DNA cryptosystem is still in its early stage. Thus, the scope of doing research on this new field is multi-dimensional. Work needs to be done from theory to realization, as both of the dimensions yet to be matured. Recent development showed that some key technologies in DNA research, such as Polymerase Chain Reaction (PCR),DNA synthesis, and DNA digital coding, have only been developed [12].

Traditional cryptographic systems have long legacy and are built on a strong mathematical and theoretical basis. Traditional security systems like RSA, DES or NTRU are also found in real time operations. So, an important perception needs to be developed that the DNA cryptography is not to negate the tradition, but to create a bridge between existing and new technology. The power of DNA computing will strengthen the existing security system by opening up a new possibility of a hybrid cryptographic system.

This paper gives a simple comparison between traditional and DNA cryptographic methods. It gives an insight to the benefits which can be achieved with the help of DNA cryptography and discusses the techniques which are currently used in this field. It also stresses on the need that both the traditional and DNA cryptographic techniques should be merged in a way that resulting cryptographic systems can enjoy the benefits from both the fields. It also points out some loop holes in this field and discusses that this field needs further research to gain the stage of realization.

The rest of this paper is organized as follows: Section II gives background knowledge which is required to understand DNA cryptography and an insight is given to the field of cryptography, DNA and DNA cryptography. In this section, atypical cryptographic scenario is also explained, which is used in later sections. Section III discusses the techniques which are used in DNA cryptography. What has been done and what still needs to be done will be discussed in Section IV. Finally, conclusion is drawn in Section V.

ISBN 978- 974-672-556-9

II. BACKGROUNDAND RELATED WORK

Modern cryptography uses cross disciplinary interactions between mathematics, computer science, and engineering. Applications of cryptography include online banking, computer authentication, and e-commerce. The discussion starts with the most basic approach for cryptography, and then improvements are demonstrated.

A. Cryptography

Cryptography is a very significant and widely used field, because it is the basis of security of all information. It is a very old field, but in modern times due to the increased use of internet, the significance of this field has been multiplied. The whole manual systems most importantly banking, defense and shopping systems are now being converted into web applications. The precious data which is being transferred over the internet is vulnerable to many security attacks [13] such as IP spoofing, man in the middle attack, teardrop etc. To secure our systems and applications, we rely on the strength of cryptography.

Cryptanalysis runs parallel to the cryptography. The aim of the cryptanalysis is to analyse and try to break the security systems proposed by the discipline of cryptography [14]. So that means, how strong a cryptographic system is, depends on how weak cryptanalysis is possible for that system [15]. A great deal of work has been done on the cryptography and cryptanalysis. And as a result, various systems have been designed (for example, RSA, ECC, etc.) to achieve high level of security.

B. Cryptographic Scenario

The typical scenario in cryptography is that Alice (sender) wants to send some messages secretly to Bob (intended receiver). The message which is to be sent is in the ordinary language understood by all, is called a plaintext. The process of converting plaintext into a form which cannot be understood without having special information is called encryption. This unreadable form is called cipher text and the special knowledge is called encryption key.

The conversion of cipher text again into plaintext with a special knowledge is called decryption, whereas special knowledge for decryption is called decryption key. Only the receiver has this special knowledge and only receiver can decrypt a cipher text with this knowledge called decryption key. In traditional cryptography encryption and decryption is done by algorithms for which currently there is no available solution [6].

C. Types of Cryptography

There are three prominent branches or sub fields of cryptography [6], named as:

1) Modern Cryptography
2) Quantum Cryptography
3) DNA Cryptography.

These three fields depend upon different difficult problems concerning to different disciplines for which there is no known solution until now. The modern cryptography is based upon the difficult mathematical problems such as prime factorization, elliptic curve problem, for which there is no known solution found so far. Quantum cryptography which is also relatively a new field is based upon the Heisenberg uncertainty principle of Physics, and DNA cryptography depends upon the difficult biological processes concerning to the field of DNA technology [6] such as:

- Polymerase Chain Reaction (PCR) for a sequence withoutknowing the correct two primer pairs [6], [12]
- and another is extracting information from the DNA chip without having the knowledge about the sequences present in different spots of DNA chip [16].

A detailed explanation of these techniques is given in Section 3.

D. DNA

Deoxyribo Nucleic Acid (DNA) is the hereditary material of almost entire living organisms ranging from very small viruses to complex human beings [17]. It is an information carrier of all life forms. DNA is a double helical structure with two strands running anti parallel as shown in Figure 1.

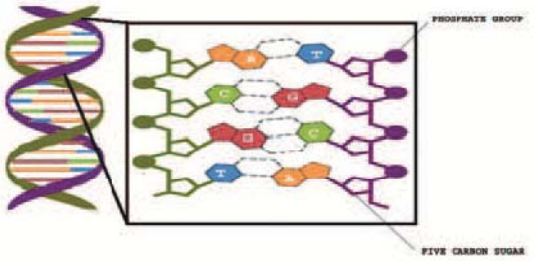


Figure 1. Structure of DNA (A - Adenine, G - Guanine, C - Cytosine and T- Thiamine).

DNA is a long polymer of small units called nucleotides. Each nucleotide consists of three components:

1) a Nitrogenous base
2) a five carbon sugar
3) a Phosphate group

There are four different nucleotides depending upon the type of nitrogenous base they have got. There are four different bases A, C, T, G called Adenine, Cytosine, Thiamine and Guanine respectively [18], [19].

DNA stores all the huge and complex information about an organism with the combination of only these four letters A,C, T and G. These bases form the structure of DNA strands by forming hydrogen bonds with each other to keep the two strands intact. A forms hydrogen bond with T whereas C and G forms

ISBN 978- 974-672-556-9

bonds with one another [20]. It can be seen in Figure 1.

Until 1994, DNA was believed to carry only the biological information but it was Adleman who revealed the computational ability of DNA when he solved NP complete Hamiltonian path problem of seven vertices [1]. After that DNA has been used as a computational tool as well [21].DNA computers deal with the DNA language that consists of four letters A, C, T and G [17], [18]. The computational ability of DNA is now used in cryptography as well. DNA cryptography is a very potential field and if manipulated inappropriate manner can give much harder competition to other fields of cryptography [22].

E. DNA Cryptography

Cryptography is the science that addresses all the aspects for secure communication over an insecure channel, namely privacy, confidentiality, key exchange, authentication, and non-repudiation. As mentioned above, DNA provides an excellent mean to secure data; the technique has been named as DNA cryptography. In such techniques, plaintext message data is encoded in DNA strands by the use of an alphabet of oligonucleotide sequences. Natural DNA obtained from biological sources may be recoded using nonstandard bases, to allow for subsequent processing [1], [17]. Input and output of the DNA data can be moved to conventional binary storage media by DNA chip arrays. Where binary data may be encoded in DNA strands by use of an alphabet of short oligonucleotide sequences.

TABLE I. COMPARISON BETWEEN TRADITIONALAND DNA CRYPTOGRAPHIC METHODS

	Security	Time Complexity	Storage Medium	Storage Capacity	Stability
Traditional Cryptography	One Fold	few seconds	Computer (Silicon) Chips	1 gram of silicon chip carries 16 MB [23]	Dependant on implementation environments
DNA Cryptography	Two Fold	few hours	DNA strands	1 gram of DNA carries 108 TB [22]	Dependant on environmental conditions

In Table I, a simple comparison is shown between traditional and DNA cryptographic techniques by considering attributes including security provided by the technique, time taken to process the technique, storage medium which is used to storethe data in the technique, the storage capacity of the storage medium used and the stability of the results for the particular technique.

By considering security provided by both the techniques, it can be seen that DNA cryptographic techniques provide two fold securities by involving computational difficulties as well as the biological difficulties. In Traditional Cryptography security can be said to be one fold as it relies on only computational difficulties. Time taken by the efficient cryptographic algorithms is few seconds, whereas DNA cryptographic techniques that involve PCR and DNA chip technology can take hours to complete the whole process. Traditional cryptography generally runs on computers over the network, so the storage mediums are silicon chips of the computers, whereas DNA cryptography deals with the DNA strands which are manipulated by biological techniques. If we consider DNA as the storage medium, it has got huge storage capacity as compared to the equivalent amount of the silicon chips. This property of data makes DNA cryptography and DNA computing very tempting and beneficial field of research.

Stability of results of cryptographic techniques refers that encryption and decryption always gives the same results. If the stability of the results of encryption and decryption provided is analysed, it can be seen that traditional cryptographic algorithms depends on the implementation conditions [24].Implementation conditions include the hardware and the language limitations used to encode the algorithm. Whereas the stability of DNA cryptography is very much dependant on the environmental conditions such as temperature, pH. This instability of DNA cryptography is discussed later.

III. DIFFICULT BIOLOGICAL PROBLEMS USED IN DNACRYPTOGRAPHY

DNA cryptography utilizes biological methods for encryption and decryption. Among those, Polymerase Chain Reaction(PCR) [6], [12] and DNA chip technology [16] are the most prominent cryptographic techniques. However, steganography using DNA, is also found in the literature. All these techniques are described below.

A. Polymerase Chain Reaction (PCR)

PCR is an amplification and quantification process of DNA. The purpose of designing PCR is to increase the amount of DNA, as it is very difficult to deal with small amount of DNA strands. The name Polymerase chain reaction comes from the enzyme (biological catalyst) known as polymerase used in the technique and chain represents that this amplification process occurs in many cycles one after another. By performing PCR, short sequences of DNA can be analysed even in samples containing only minute quantities of DNA. PCR can select small strands of DNA and amplifies those. In practice, amplification of DNA involves cloning of segments of interest into vectors for expression. PCR is highly efficient so that untold numbers of copies can be made from small selected DNA. Moreover, PCR uses the same molecules that nature uses for copying DNA. To perform PCR, one should know the sequence of DNA to be amplified to design the right primer for it, where primer is a sequence containing few numbers of nucleotides complimentary to the specific

ISBN 978- 974-672-556-9

sequence of DNA which is to be amplified [6]. In short, we can identify the PCR process into two phases:

• Two "primers", short single-stranded DNA sequences to correspond to the beginning and ending of the DNA stretch.

• Polymerase enzyme that moves along the segment of DNA, reading its code and assembling the copy.

The encryption key, in this case, is compound, consisting of both PCR primers pair and public key. Similarly decryption key consists of complementary primer pairs and private key.

Figure 2. Summary of Cryptographic Technique using PCR

Encryption starts with the exchange of two primers (forward and reverse) between Alice and Bob via a secure channel [6].For Encryption, pre-processing can be done, that the whole algorithm like RSA can be applied first. This step is numbered as 1 in Figure 2. Then cipher text can be converted into DNA sequence by coding scheme represented by number 2in Figure 2. By performing this, entirely different cipher text can be obtained [6]. In literature, cipher DNA refers the term cipher text which is in the form of DNA sequence, and plaintext DNA denotes the plain text which is in the form of DNA.

The prepared cipher DNA is then flanked by the secret primers and mixed with a number of other unknown DNA. Alice sends this DNA mixture to Bob [6]. For decryption, Bob can retrieve the cipher DNA by performing PCR using its secret primer, and reverse the whole process which is done for encryption [6]. Anyone without knowing the two primers cannot retrieve target cipher DNA [12]. Decryption steps are represented as 3 and 4 in Figure 2. In the same figure, number3 denotes that the cipher DNA can be retrieved by using DNA decryption key (secret primers) and is converted in to cipher text by using coding scheme. Finally, in the figure, 4 denotes that the cipher text is decrypted by RSA private key.

The above mentioned PCR technique has numerous different implications, such as Guangzhao Cui et. al. proposed a encryption scheme with the aid of PCR amplification and DNA coding schemes [6]. On the other hand, Tanaka, et al. devised a public-key system using DNA as a one-way function for key distribution using PCR amplification to restore the plain text DNA from other distracter DNA strands [12]. Yamamoto et al. also contributed in cryptography by producing large-scale DNA memory based on the nested PCR [25]. This technique of DNA cryptography provides two fold securities by involving both the molecular techniques and modern algorithms. If one security level is broken some way, the other can keep this technique safe [6]. However, the problem of PCR based techniques lies in the transportation of secret keys between the sender and receiver. This has been discussed in detail in Section IV.

B. Steganography using DNA

Steganography is the technique of hiding information. The goal of cryptography is to make data unreadable by a third party; on the other hand, the goal of steganography is to hide the data from a third party. Formula shown at equation 1 is a very generic description of the steganographic process:

$$cover_{medium} + hidden_{data} + key_{stego} = stego_{medium} \quad (1)$$

In this context, the $cover_{medium}$ is the file in which the data, $hidden_{data}$, is hid. The resultant file may also be encrypted using another key called key_{stego}. Finally, the $stego_{medium}$ is the file that will be transported. Usual, medium of covering file may include the audio or image file. However, due to its massive storage capability, DNA is getting popular to be a steganographic covering medium.

The primitive idea of DNA steganography can be described as follows: For encryption, one or more input DNA strands are taken to be tagged as the plaintext message. One or more randomly constructed secret key strands are appended with the input DNA. Resulting "tagged plaintext" DNA strands are hidden by mixing them within many other additional" distracter" DNA strands which might also be constructed by random assembly. For decryption, given knowledge of the "secret key" strands, resolution of DNA strands can be decrypted by a number of possible known recombinant DNA separation methods. Such as plaintext message strands maybe separated out by hybridization, with the complements of the "secret key" strands might be placed in solid support on magnetic beads or on a prepared surface. These separation steps may combine with amplification steps.

One such method is found in [26]. Viviana I. Risca proposed a steganographic technique using DNA and standard biological protocols. The proposed method encodes the information in the sequence of DNA strand, flanked by two secret primer target regions. The technique uses monoalphabetic encryption key to assign random but unique 3-base DNA codons to 40alpha-numeric and punctuation characters. The key was then used to encode plaintext messages into the base sequence of an artificially synthesized oligonucleotide.

The resultant DNA strand is then hidden into a very large amount of similarly sized garbage DNA strands. To extract the message, one must know the

ISBN 978- 974-672-556-9

primers that will bind the target regions on the message containing DNA strand in order to selectively amplify the required molecule.

Since, this primitive technique is simple and easily breakable, an improvement is proposed by Asish Gehani et al [15].The idea is to distinguishing the probability distribution of plaintext source from that of distracter DNA stands [15].

As it is discussed that DNA based steganography converts plaintext into plain text DNA but do not encrypt it. It only hides the plaintext DNA into bulk of other DNA and person who knows the primer for it can easily locate the plaintext DNA and amplify it. This technique can be very useful as it saves the cost of encryption, but it is vulnerable to statistical analysis. So the PCR based cryptography may be considered to be much safer [6], [22].

C. DNA Chip Technology

A brief introduction to DNA chip technology is given to clear the concept of this technique used in cryptography. Over less than a decade, DNA chips and microarrays have changed the way in which scientists carry out their investigations. DNA chips enable researchers to manipulate the vast amounts of data from genome-sequencing [27]. DNA chip technology is very important for the manipulation of biological data. It is commonly used to find expression of many genes in parallel [11]. These chips like silicon chips can be used to handle and store the data in the form of DNA sequences.

DNA chips consist of large number of spots embedded on a solid surface, most commonly used is a glass slide. Each spot consists of different kind and number of probes, where probes are small nucleotide sequences which are able to bind to the complementary nucleotides. Nucleotides which bind to these probes are fluorescently labelled, whenever any DNA sequence binds to these probes, it is observed under a laser dye and data is calculated electronically depending upon the ratio of the binding of probe with the DNA in each spot [28],[29]. Manufacturers are now developing smaller biochips with better information handling capabilities that will contribute more effectively to numerous fields of research including cryptography.

Technique, considering the typical cryptographic scenario, has following steps:

• Encryption key is a collection of particular probes where decryption key is a collection of corresponding probes having complementary sequence. The decryption key is then sent to the Bob in a secure manner.

• Plaintext is converted into a binary format. This binary format is then embedded into DNA chip as a cipher text (cipher DNA). Without knowing the decryption key one cannot read the plaintext from the DNA chip.

• Bob uses the decryption key and hybridizes the cipher DNA. With the help of computer software he can retrieve plain text [16].

One such cryptosystem is XOR One-time-pad [15]. To construct one-time pad using DNA chip technology, an array of immobilized DNA strands are used, where multiple copies of a single sequence are grouped together in a microscopic pixel. The strands are optically addressable. Distinct DNA sequences at each optically addressable site of the array can be synthesised using various methods. One of the known such technologies is combinatorial synthesis conducted in parallel at thousands of locations. For preparation of oligonucleotides of length L, the 4^L sequences are synthesized in 4n chemical reactions.

To encrypt, each plaintext message has been appended with a unique prefix index of length L_0. A complement of the plaintext message tag is created on one-time-pad DNA sequence by appending unique prefix index tag of the same lengthL_0. Using annealing and ligation each of the corresponding pair of plaintext message and a one-time-pad sequence is concatenated. After that the message is encrypted by bit-wise XOR mechanism. XOR operation is shown in Equation 2, where C is the cipher data or strands, M is the plaintext message and S is a sequence of independently distributed random bits.

$$C_i = M_i \otimes S_i \text{ for } = 1, , n. \quad (2)$$

However, the fragments of the plaintext are converted to cipher strands and plaintext strands are dropped. For decryption commutative property of bitwise XOR operation is used, as shown in Equation 4-5.

$$C_i \otimes S_i = (M_i \otimes S_i) \otimes S_i \quad (3)$$
$$= M_i \otimes (S_i \otimes S_i) \quad (4)$$
$$= M_i. \quad (5)$$

MingXin et. al. used DNA chip technology in a similar manner as described above to design asymmetric key encryption scheme, they referred this encryption scheme as DNA symmetric-key cryptosystem (DNASC) [16]. Recently Lai et al. proposed an asymmetric encryption scheme using DNA chip technology and they also designed signatures using this DNA chip technology so that Alice and Bob can verify each other [30].

The encryption technique using DNA chip technology is not restricted to encrypt only textual data, there are also encryption schemes which are designed to encrypt and decrypt images. One such cryptosystem is designed by Gehani et al. where they produced a substitution one-time-pad system and used it during the process of encryption and decryption of 2D images. It is noteworthy that they also performed PCR to amplify particular DNA strands [15]. Similar technique is also noticed in

Shyam et al.'s work, where they proposed acrypto system and created plaintext and cipher text pairs using DNA chip technology but on images [31].

Although DNA chip provides a wide range of parallel data processing capability, the lack of interoperability of DNA chip and other storage mediums suppressed its potentiality at this moment.

IV. Scope of Research

There are many advantages which seem to be associated to the field of DNA cryptography. The huge storage capacity of DNA makes it a very tempting field for research. Moreover the cryptographic techniques which are designed by involving this field are believed to give very high security level [6].

The research which has been done so far on DNA cryptography shows that several DNA-based methods can be devised in order to break cryptosystems which are currently being used. Many cryptosystems used today are based on RSA public key encryption. RSA public key encryption is based on the intractability of prime factorization as there are no known efficient algorithms to find the prime factors of sufficiently large numbers. As shown in Equation 6.

$$n = p * q \quad (6)$$

where p and q are prime numbers, for a given "n" it is infeasible to find p and q when n is a very large number. If there is any technique that can find how to factor given "n", the whole RSA scheme will be broken. There are techniques which have been devised to break RSA scheme in DNA cryptography. These techniques used self-assembly of DNA tiles to fully break RSA scheme [32], [9]. If these techniques are able to break RSA, RSA will no more remain practical.

Elliptic curve cryptography (ECC) has been applied in key exchange and also in the digital signatures. The security of these cryptosystems is based on the difficulty of solving the elliptic curve discrete logarithm problem [33]. DNA-based methods have been developed to break the cryptosystemsbased on elliptic curves. These methods are accomplished by means of basic biological operations and have developed a parallel multiplier; a parallel divider and a parallel adder for adding points on elliptic curves [34].

The encryption schemes (Ciphers) that uses keys only once are said to be one-time pads (OTP). In theory OTP cipher is absolutely secure. But practically, key distribution and key generation are critical issues to be resolved for the use of OTP ciphers. Key space should be large enough so that keys can only be used once. DNA having huge storage capacity, can be manipulated to generate key space to be used for OTP cipher.

There are also some areas that need to be improved. For example, Time and computational complexity are two of the most important parameters for any kind of cryptographic systems, DNA cryptography dealing with the manipulation of DNA sequences takes a lot of time to deal and work out with DNA sequences as compared to time taken by many very efficient algorithms of traditional cryptography such as, DES, RSA etc. [8], [9].

There is also a risk associated to DNA cryptography that if the cipher DNA gets impured by unwanted DNAs, this process may collapse. This may be handled by taking precautions and maintaining the laboratory environment pure. Still research must be carried on to address the reliability issues.

To summarise the whole discussion, following points are significant.

1) In cryptographic technique involving PCR technology, it can be examined that there are two encryption keys and two decryption keys. One pair of encryption and decryption key is (n, e) pair which was used for RSA public key cryptosystem. The other pair was the forward and reverse primers used for tagging the cipher DNA. These primers were necessary to be securely shared among Alice and Bob as DNA can only be recovered if these primer pairs were known. This means that in addition to the public and private keys, primer pairs are also to be secretly transported. In public key encryption there is only one secret which is shared that is the Bob's private key. However in this technique primer pairs are also to be secretly transported [6]. The transport of primers is not as simple as for the transport of keys of the traditional cryptographic techniques.

   Primer pairs are the sequences of nucleotides; there is a probability that these molecules can be affected by environmental conditions. To take account of this, environmental conditions should be maintained throughout the process of transport.

   Another possibility is that the sequence of primer pair in the form of A,C,T,G can be sent as digital form to the Bob, same as the manner in which keys of traditional cryptographic techniques are transported. Then Bob can synthesize his own primer pairs in the laboratory considering the sequence which is sent to him by Alice. During all this process environmental conditions should remain stable.

ISBN 978- 974-672-556-9

2) DNA steganography system can be the complement of the traditional DNA cryptography, however, potential limitations of DNA steganography methods is its vulnerability. Certain DNA steganography systems can be broken, with some assumptions on information theoretic entropy of plaintext messages. May be the method is brute force, but cannot be ruled out because of the advancement of computing technologies. To make the situation better, plaintext messages can be compressed before encrypting, however, in this case, plaintext messages need to be pre-processed. So the open question is whether DNA steganography systems with natural DNA plaintext input can or cannot be made to be unbreakable. The process of DNA steganography involves the tagging of plaintext DNA with the primers and is hid in other garbage or junk DNA of same length as the plaintext DNA. This is a very simple method to provide confidentiality. Same issue can arise here if the environmental conditions change the binding property of DNA with the primer. There is also a possibility that other garbage DNA can bind to the primer, making it difficult to recover the original plaintext DNA [6], [22].

3) A DNA chip contains arrays of Pico moles of DNA sequences, known as probes. Since an array can contain huge number of probes, a DNA chip can accomplish number of computation in parallel. However, DNA chip data is difficult to exchange between traditional storage medium, due to the lack of standardization in fabrication, protocols, and analysis methods [35]. This problem is defined as interoperability problem in bioinformatics. Moreover by examining the DNA chip technology used in the DNA cryptography, it is seen that encryption and decryption processes are performed on DNA chip. DNA is a biological molecule and its properties are dependent on the environmental conditions. For instance, DNA's property to bind to other nucleotides can be changed by changing environmental conditions. So encryption and decryption process does not remain stable. Due to this instability, encryption and decryption may show different results under different environmental conditions.

In short it can be said that DNA cryptography cannot totally be replaced traditional cryptography which is currently being used. This field requires a lot of research and work to have a position in which it can be implemented and used for practical purposes. There is a need that people from traditional cryptography and DNA technology should exchange knowledge among each other and cryptosystems should be devised in such a way that they can enjoy benefit from both the fields.

V. Conclusion

DNA cryptography is a relatively new cryptographic field of research evolved with the DNA computing. In this particular field, DNA is used as message carrier and the biotechnology such as PCR, is used as the implementation mechanism. The extravagant storage capability and parallel computability of DNA molecules are exploited for encryption, authentication, and authorization. In this paper, the existing DNA cryptographic techniques namely Polymerase Chain Reaction (PCR), Steganography using DNA and DNA chip technology have been discussed. With the summarization of the progress of DNA cryptographic research, the advantages, future trends and several problems have also been identified. All the three kinds of cryptography have their own advantages and disadvantages and can be treated as the complement of each other in future security applications. However, the difficulties that are identified in DNA cryptography are the absence of theoretical basis and practical methodologies which can readily be implemented in the field of security.

The potentiality of DNA in computing can open up further biological molecule based computation methods. Once the DNA cryptography field is developed and analysed, attempts can be made to convert the cipher DNA into cipher proteins, these can give us another level of security. It will be possible only by intensive research and practical work on DNA computing.

Acknowledgment

The Erasmus Mundus project eLINK - east-west Link for Innovation, Networking and Knowledge exchange programme (EM ECW SGA2008-4949) for funding this study.

References

[1] L. Adleman, "Molecular computation of solutions to combinatorial problems," *Science, JSTOR*, vol. 266, pp. 1021–1025, 1994.

[2] G. Cui, Y. Liu, and X. Zhang, "New direction of data storage: DNA molecular storage technology," *Computer Engineering and Application*, vol. 42, no. 26, pp. 29–32, 2006.

[3] J. Chen, "A DNA-based, biomolecular cryptography design," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2003, pp.822–825.

[4] R. Barish, P. Rothemund, and E.Winfree,"Two computational primitives for algorithmic self-assembly: copying and counting," *Nano Letters*, vol. 5, no. 12, pp. 2586–2592, 2005.

[5] P. Rothemund, N. Papadakis, and E. Winfree, "Algorithmic self assembly of DNA sierpinski triangles," *PLoS Biology*, vol. 2, no. 12, pp. 041–2053, 2004.

[6] G. Cui, L. Qin, Y. Wang, and X. Zhang, "An encryption scheme using DNA technology," in *IEEE 3rd International*

conference on Bio-Inspired Computing: Theories and Applications (BICTA08), Adelaid, SA, Australia, 2008, pp. 37–42.

[7] D. Boneh, C. Dunworth, and R. Lipton, "Breaking DES using a molecular computer," in *In Proceedings of DIMACS workshop on DNA computing*, 1995, pp. 37–65.

[8] D. Beaver, "Factoring: The DNA solution," in *4th International Conferences on the Theory and Applications of Cryptology*. Wollongong,Australia: Springer-Verlag, Nov. 1994, pp. 419–423.

[9] Y. Brun, "Arithmetic computation in the tile assembly model: Addition and multiplication," *Theoretical Computer Science, Science Direct, Elsevier*, vol. 378, no. 1, pp. 17–31, 2007.

[10] O. Pelletier and A. Weimerskirch, Algorithmic self-assembly of DNA tiles and its application to cryptanalysis," in *Proceedings of the Genetic and Evolutionary Computation Conference 2002 (GECCO02)*, New York, USA, 2002, pp. 139–146.

[11] X. C. Zhang, "Breaking the NTRU public key cryptosystem using self assembly of DNA tilings," *Chinese Journal of Computers*, vol. 12, pp.2129–2137, 2008.

[12] T. Kazuo, O. Akimitsu, and S. Isao, "Public-key system using DNA as a one-way function for key distribution," *BioSystems, Elsevier Science*, vol. 81, no. 1, pp. 25–29, 2005.

[13] N. Galbreath, *Cryptography for Internet and Database Applications: Developing Secret and Public Key Techniques with Java*. New York, USA: John Wiley and Sons, Inc., 2002.

[14] A. Menezes, P. Oorschot, and S. Vanstone, *Handbook of applied cryptography*. CRC Press, 1996.

[15] A. Gehani, T. LaBean, and J. Reif, *DNA based cryptography*. Germany: Aspects of Molecular Computing, Springer-Verlag., 2004.

[16] L. MingXin, L. XueJia, X. GuoZhen, and Q. Lei, "Symmetric-key cryptosystem with DNA technology," *Science in China Series F:InformationSciences, Springer Verlag, Germany*, vol. 50, no. 3, pp. 324–333, 2007.

[17] L. H. N. C. for Biomedical Communications, *Handbook on Genetic Cells and DNA*. USA: National Library of Medicine, National Institutes of Health, Department of Health and Human Services., 2010.

[18] M. Zhang, L. Sabharwal, and W. Tao, "Interactive DNA sequence and structure design for DNA nanoapplications," *IEEE Transactions on Nanobioscience*, vol. 3, no. 4, pp. 286–292, Dec. 2004.

[19] J. Watson and F. Crick, "A structure for deoxyribose nucleic acid," *Nature*, vol. 171, pp. 737–738, Apr. 1953.

[20] X. Guozhen, L. Mingxin, Q. Lei, and L. Xuejia, "New field of cryptography: DNA cryptography," *Chinese Science Bulletin, Springer Verlag, Germany*, vol. 51, no. 12, pp. 1413–1420, 2006.

[21] A. Leier, C. Richter, W. Banzhaf, and H.Rauhe, "Cryptography with DNA binary strands," *BioSystems, Elsevier Science*, vol. 57, no. 1, pp.13–22, 2000.

[22] G. Cui, L. Cuiling, L. Haobin, and L.Xiaoguang, "DNA computing and its application to information security field," in *IEEE Fifth International Conference on Natural Computation*, Tianjian, China, Aug. 2009.

[23] Farnel. (2010) Silicon storage technology - sst25vf016b-50-4i-s2af - 16m flash memory, spi eeprom, soic8. [Online].Available: http://uk.farnell.com/silicon-storage-technology/sst25vf016b- 50-4i-s2af/16m-flash-memory-spi-eeprom-soic8/dp/1368695

[24] A. A. Jabri, "A statistical decoding algorithm for general linear block codes," in *8th IMA International Conference on Cryptography and Coding*, Cirencester, UK, Dec. 2001, pp. 1–8.

[25] M. Yamamoto, S. Kashiwamura, A. Ohuchi, and M. Furukawa, "Large scale dna memory based on the nested pcr," *Natural Computing, an International Journal*, vol. 7, no. 3, pp. 335–346, 2008.

[26] V. I. Risca, "DNA-based steganography," *Cryptologia, Tylor and Francis*, vol. 25, no. 1, pp. 37–49, 2001.

[27] P. Gwynne and G. Heebner, "Technologies in DNA chips and microarrays: I," *Science*, vol. 4 May, p. 949, 2001.

[28] T. Tsukahara and H. Nagasawa, "Probe-on-carriers for oligonucleotide microarrays (DNA chips)," *Science and Technology of Advanced Materials, Elsevier Science*, vol. 5, pp. 359–362, 2004.

[29] P. Brown and D. Botstein, "Exploring the new world of the genome with DNA microarrays," *Nature Genetics*, vol. 21, pp. 33–37, 1999.

[30] L. XueJia, L. MingXin, Q. Lei, H. JunSong4, and F. XiWen1, "Asymmetric encryption and signature method with dna technology," *SCIENCECHINA Information Sciences*, vol. 53, pp. 506–514, 2010.

[31] V. M. M. Shyam, N. Kiran, "A novel encryption scheme based on dna computing," in *14th IEEE International Conference*, Tia, India, Dec.2007.

[32] Y. Brun, "Nondeterministic polynomial time factoring in the tile assembly model," *Theoretical Computer Science, Science Direct, Elsevier*, vol.395, no. 1, pp. 3–23, Apr. 2008.

[33] V. S. Miller, "Use of elliptic curves in cryptography," in *5th Annual International Cryptology Conference (CRYPTO '85)*, Santa Barbara, Calif, USA, Aug. 1985, pp. 417–426.

[34] K. Li, S. Zou, and J. Xv, "Fast parallel molecular algorithms for DNA based computation: Solving the elliptic curve discrete logarithm problem over gf(2n)," *Journal of Biomedicine and Biotechnology, Hindawi.*, vol.2008, pp. 1–10, Apr. 2008.

[35] J. Fuscoe, W. Branham, C. Melvin, V. Desai, C. Moland, T. Han, L. Shi,W. Tong, A. Scully, and R. Delongchamp, "Technical issues involved in obtaining reliable data from microarray experiments," *Regulatory Research Perspective, National Center for Toxicological Research, USA*, vol. 6, no. 1, pp. 1–22, 2006.

Software Visualization Tools for Software Comprehension

Kamruddin Md. Nur
United International University
kamruddin.nur@gmail.com

Hasan Sarwar
Dept. of Computer Science & Engineering
United International University
hsarwar@cse.uiu.ac.bd

***Abstract*—Softwares are getting ever bigger and complex. Complete software comprehension is important in all streams of software engineering - forward, reverse, round-trip, and reengineering. Software systems are by nature invisible which differs from tangible products. As the size and complexity of software systems increase, a complete comprehensive understanding of software becomes a challenging task. Researchers believe that the key to this problem is better visualization. Better software visualization enables better software comprehension. Although the program comprehension has been emphasized dominantly in research literature in the field of reverse software engineering, software comprehension is less emphasized and should be distinguished from program comprehension, considering software is a collection of programs. Software comprehension should be treated as comprehending the whole software and its underlying program components rather than just segments of code or algorithms. This paper attempts to explore the concepts of software visualization, existing software visualization tools, techniques, methodologies, and finally concentrates on software visualization aspects and means of visualization for software comprehension.**

Index Terms - Software Visualization; Software Comprehension; Software Visualization Tools; Software Engineering.

I. Software Comprehension

Software comprehension is the process of gaining an understanding of how a software system functions. Deimelet al. [1] defines software comprehension as "the process of taking computer source code and understanding it". Burdet al. [2] defines software comprehension as "the activity of understanding existing software systems". Muller [3] defines software comprehension as "the task of building mental models of the underlying software at various abstraction levels, ranging from models of the code itself, to models of the underlying application domain, for maintenance, evolution, and re-engineering purposes". However, the author enhances this definition, defining software comprehension, as "a process whereby a software practitioner understands a software artifact using both knowledge of the domain and/or semantic and syntactic knowledge, to build a mental model of its relation to the situation". Software comprehension should be treated as comprehending the whole software architecture and its underlying program components rather than just segments of code. Understanding a software system essentially involves understanding all subsystems, modules in subsystems, units in modules, code segments in units and their inter-relationships. In general, program comprehension is considered as a reverse engineering activity, while software comprehension embraces all aspects of software engineering - forward, reverse, round-trip and engineering.

II. Software Visualization

The meaning of the word 'visualization' is pretty straightforward as it refers to images that people can see with their eyes. The word visualization contains the root 'visual'. Grant [4] defines, "visual means visible on a computer display and visualization is communicating data with graphics". Price et al. [5] suggests visualization is the conveyance of mental images: "the power or process of forming a mental picture or vision of something not actually present to the sight". Price et al. [5] defines software visualization as, "the use of the crafts of typography, graphic design, animation, and cinematography with modern human-computer interaction technology to facilitate both the human understanding and effective use of computer software". ACM symposium on Software Visualization confines - "Software visualization encompasses the development and evaluation of methods for graphically representing different aspects of software, including its structure, its abstract and concrete execution, and its evolution". Myers taxonomy for program visualization [6] identifies six regions data, code, and algorithm as static or dynamic arranged in a 2x3 matrix (figure. 1.).

ISBN 978- 974-672-556-9

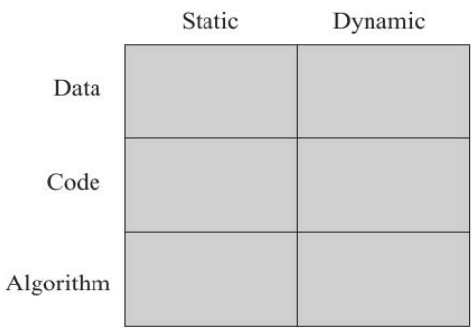


Figure 1. Myer's Taxonomy for program visualization

III. Software Visualization Aspects

Different visualization tools concentrated on different aspects of software systems - code, data, algorithm etc. A valid question is - what are the things exactly we want to see from software visualization? In this section we focus on desired software visualization aspects;

- Software architecture visualization
- Code, method calls, expression evaluation visualization
- Data structure and Algorithm visualization
- Procedural and declarative visual debugging
- Code complexity and dependency visualization
- IDE and language independent visualization
- Standard UML diagram visualization
- Learn ability and Usability

Table I. summarizes the purposes of visualization aspects categorized.

TABLE I. Software Visualization Aspects

	Visualization Aspects	Purpose
Structure	Software Architecture code Function or Method Call Data Structure Code Complexity Code Dependency	To describe the structure of the software
Behavior	Expression Evaluation Algorithm	To describe the behavior of the software
Interaction	Method calls	To describe the interaction of program blocks
Debugging	Procedural Declarative	To enable both procedural and declarative debugging

IV. Visualization Tools

Numerous visualization tools have been proposed, built and utilized in order to visualize a software system. In this section, we explore few of widely referenced existing tools that are being used in different streams of software engineering.

A. Jeliot 3

Jeliot 3[1][7] [8] [9] is the latest version of Jeliot which visualize show a Java program is interpreted as animation. The key feature of Jeliot is the fully or semi-automatic visualization of the data and control flows. Method calls, variables, operation are displayed on a screen as the animation goes on, allowing the users to follow step by step the execution of a program (figure. 5.). Jeliot is based upon the concept of learning-by-doing: designed for pedagogical purposes which help learners construct a mental visual representation of a program [8].

B. BlueJ

BlueJ[2] is an interactive java programming environment designed and implemented by Michael Kolling and John Rosenberg [10] as a pedagogical tool. BlueJ visualizes Java classes in standard UML diagram offering learners a graphical picture of the classes and objects in a system. BlueJ simplifies testing of methods and classes, and removes the necessity for much difficult and confusing Java code [11].

C. jGRASP

jGRASP[3] (Graphical Representation of Algorithm, Structure, and Process) is a lightweight development environment, created specifically to provide automatic generation of software visualizations to improve the comprehensibility of software. jGRASP implemented in Java can produce Control Structure Diagrams (CSDs) for Java, C, C++, Objective-C, Ada, and VHDL; Complexity Profile Graphs (CPGs) for Java and Ada; UML class diagrams for Java.

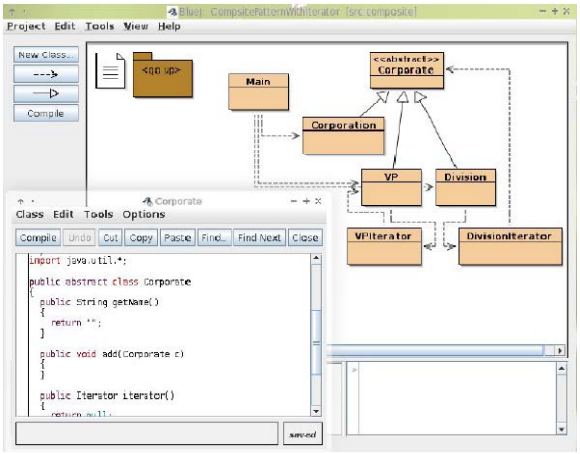


Figure 2. BlueJ class diagram and code pad for editing code

D. X-Ray

X-Ray[4]is open-source software visualization plug-in for the Eclipse framework. X-Ray visualizes

[1]http://cs.joensuu./jeliot/
[2] http://www.bluej.org
[3] http://www.jgrasp.org/
[4]http://xray.inf.usi.ch/xray.php

a software system in polymetric views by utilizing software metrices. A polymetric view [12] is a lightweight software visualization technique enriched with multiple software metrics. X-Ray provides three polymetric views - System Complexity View, Class Dependency View, and Package Dependency View for a given Java project. Polymetric views enable understanding the structure and detect problems of a software system in the initial phases of a reverse engineering process [13].

E. JIVE

JIVE[5] (Java Interactive Visualization Environment) is an interactive execution environment for Eclipse that supports a declarative and visual approach to debugging object-oriented software implemented in Java [14]. JIVE visualizes the runtime state of the software as dynamic object diagram and execution history as time-sequence diagram. JIVE declarative debugging provides extensible set of queries over a program execution history (over individual execution states or entire history of the execution) stored in a relational database [15]. Queries are formulated using the source code or the diagrams, and the results are shown in a tabular format and also as diagram annotations.

V. METHODOLOGY

We first survey the most referenced software visualization tools which are usable and widely referenced for software visualization. Among them, we choose five promising visualization tools for software comprehension. Then we select two ideal software projects as case studies to run on these tools and opted to realize relative, comparative advantages of visualization aspects and means of visualization for each of these selected tools. Then we carry out a qualitative analysis of our experiments and summarize the results.

VI. CASE STUDY

For the purpose of our case study, we have chosen two typical software projects implementing widely used design patterns - composite and strategy design pattern in Java. A design pattern is a general reusable solution to a commonly occurring problem in software design [16] [17]. A composite design pattern composes objects into tree structures to represent part-whole hierarchies and allows a group of objects to be treated in the same way as a single instance of an object. A strategy design pattern defines a family of algorithms, encapsulate each one, and make them interchangeable. Strategy lets the algorithm vary independently from the clients that use it. Our composite design pattern applied software project called 'Employee Management System' (case 1) is a generic employee management project has 7 classes and 29 methods(figure. 3.) and our strategy design pattern applied project 'Racing Game'(case 2) has 13 classes and 25 methods(figure. 4.). The rationale behind choosing software projects coded with software design pattern is for standardization, credibility and familiarity of the solution used worldwide which suffice our need for this research.

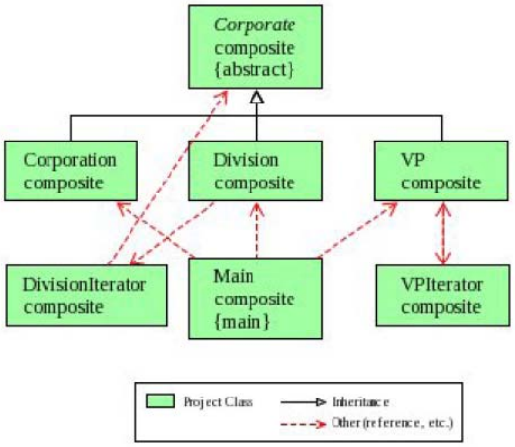


Figure 3. UML class diagram of Composite design pattern project (case-1)

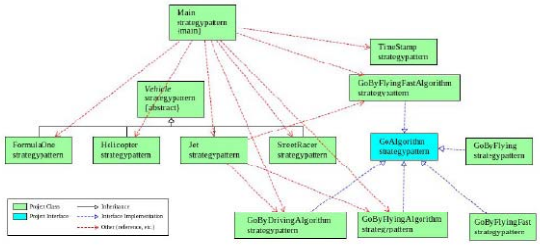


Figure 4. UML class diagram of Strategy design pattern project (case-2)

VII. ANALYSIS

A. Software Architecture Visualization

Software architecture describes the high-level structure of a software system. Bass et al. [18] defines - "software architecture of a program or computing system is the structure or structures of the system, which comprise software elements, the externally visible properties of those elements, and the relationships among them". In general, standard UML diagrams such as package diagram, component diagram, class diagram, and object diagram can be used in software architecture visualization and comprehension [19]. In our case study, among these software visualization tools described above, only BlueJ and jGRASP produces UML class diagram. BlueJ shows classes in package (figure2), however need better support for system level package diagram drawing. JIVE depicts run-time state in dynamic

[5]http://www.cse.buffalo.edu/jive/

object diagram (figure 7), however does not generate any class or package diagram.

B. Data Structure Visualization

Software data structure visualization involves visualizing standard data structures such as - array, list, trees, graphs etc. jGRASP Dynamic viewer [20] for objects and primitives provide visualizations while a program in debug mode or invoke methods for an object on the workbench. When a viewer is opened, a structure identifier [21] attempts to automatically recognize linked lists, binary trees, hash tables, array wrappers (list, stacks, queues, etc.) during debugging or workbench use. jGRASP integrated Java debugger collects run time information necessary to render the dynamic visualizations with the ability to create multiple graphical views of data structures, which are synchronized, automatically generated, and updated as a program executes[22].

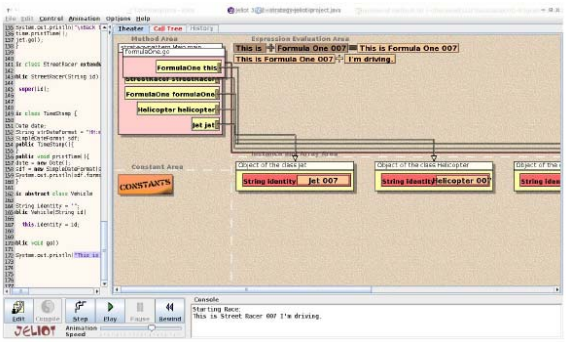


Figure 5. Jeliot 3 Strategy design pattern (case-2) project animation

C. Expression Evaluation Visualization

Expression evaluation visualization involves visualizing instruction, mathematical, or logical expressions to be evaluated in order to proceed to the next instruction. Jeliot 3 has a dedicated expression evaluation area which shows current expression being executed (figure 5). jGRASP has debugger and workbench tab provides text view of expressions being executed. JIVE provides contour model of object calls, debug window and variable window to see expression evaluation step by step. JIVE also provides 'line executed' and 'variable changed' declarative query for finding expression evaluations.

D. Code Visualization

Code visualization involves highlighting code segment being executed or evaluated in the code view and a way of finding code segments to be edited. Jeliot 3 highlights statements, methods and code blocks being executed in the animation. However, in order to edit the code, animation has to be stopped and editing mode needs to be activated. BlueJ does not highlight code segment rather a double click on the UML class diagram pops up editing window of the desired class (figure. 2.). jGRASP generates Control Structure Diagram (CSD) for program block highlighting (figure 6). Many rich IDE today provide pretty printing and program block highlighter which together does the purpose of CSD. X-Ray and JIVE takes the advantage of eclipse code highlighting feature and highlights code as eclipse does for any program.

E. Interaction Visualization

Software component interaction visualization involves visualizing software components interacting with each other. In object oriented (OO) software systems sequence diagrams visualizes sequence of interactions among objects, and in non-OO software systems, by visualizing function calls. Jeliot 3 displays methods and instructions being executed in the method area (figure. 5.). Instances are shown in the instance area. JIVE timed-sequence diagram visualizes sequence of interactions at run time which is a convenient way of finding implementation error by comparing the actual desired interactions (figure. 8.). JIVE provide novel zooming feature which is convenient as opposed to Jeliot 3 scrolling feature.

```
import java.util.Iterator;

public class DivisionIterator implements Iterator
{
  private Corporate[] corporate;
  private int location = 0;

  public DivisionIterator(Corporate[] c)
  {
    corporate = c;
  }

  public Corporate next()
  {
    return corporate[location++];
  }
}
```

Figure 6. jGRASP CSD of Composite pattern DivisionIterator class (case-1)

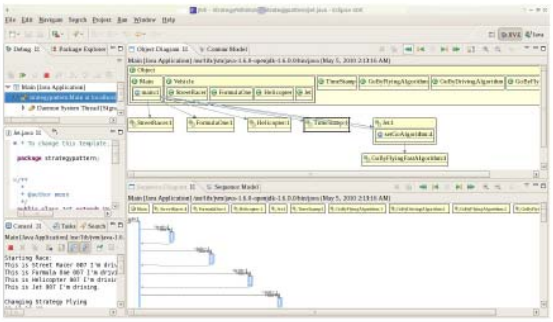


Figure 7. JIVE run-time object diagram (top-pane) and timed-sequencediagram (bottom-pane) (case-2) project

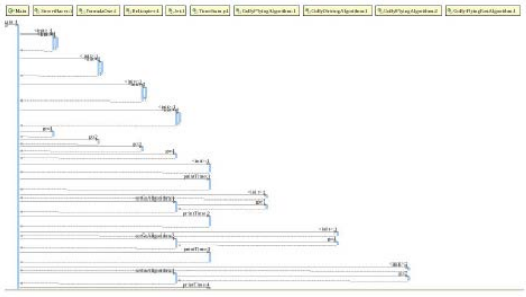


Figure 8. JIVE runtime timed-sequence diagram for strategy design pattern(case-2) project

ISBN 978- 974-672-556-9

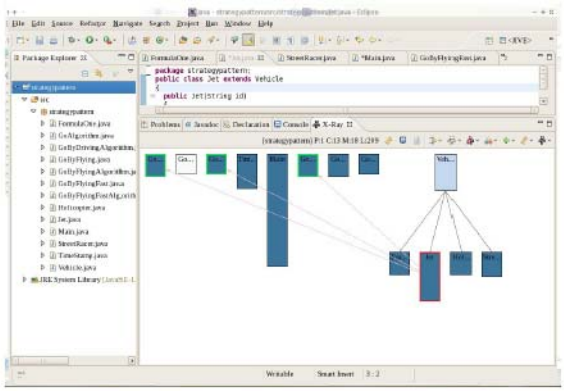


Figure 9. X-Ray System Complexity view of Strategy design pattern(case-2) project

F. Debugging with Visualization

Traditionally, a procedural debugging involves debugging step by step from the beginning to the end of the program. The term declarative focuses on 'what' as opposed to a procedural approach which requires 'what' and 'how'. In a declarative debugging, programmer uses set of graphical queries over program execution [15]. A forward and reverse stepping while debugging greatly enhances software comprehension and software bugs.

Jeliot 3 and BlueJ do not provide any debugger. jGRASP has integrated Java debugger collects run time information necessary to render the dynamic visualizations and procedural debugging. JIVE has both procedural and declarative debugging. JIVE declarative query provides list of available queries and search fields for declarative debugging (figure. 12.) [23][24].

G. Code Complexity and Dependency Visualization

jGRASP CPG does not show the total project complexity graph, it shows only a single file complexity graph (figure11.) at a time which might not be significantly useful for large projects. X-Ray System Complexity Polymetric View (figure 9) shows entire system complexity where different metrics are depicted graphically - width metric, height metric, color metric, and edges represent hierarchy(inheritance) dependency (table. II). The black bordered class for stand-alone class and orange bordered for inner class is a good way of finding design disharmonies [12].

X-Ray Class and Package Dependency Polymetric View is arranged in a bi-dimensional circle (figure. 10.), packages or classes are linked together by dependency links, each of them with a certain weight, highlighting how strength is the dependency between entities [12]. This polymetric view helps finding design defects, providing information about coupling and cohesion.

Both views of X-Ray do not provide any zooming, scaling or filtering and thus suffer from scalability which is a limitation of X-Ray [12].

TABLE II. X-RAY SYSTEM COMPLEXITY POLYMETRIC VIEW GRAPHICALMETRICES

Metrice	***Meaning***
width	number of methods
height	lines of codes
color	Java class type - abstract (light blue), concrete(blue), interface (white), external to the project (green)
edges	hierarchy dependency

H. IDE independent Visualization

Most of the software visualization tools are the outcome of research projects. Most tools are implemented with Java as a standalone application or open-source IDE plugins (e.g. eclipse plugins). As a result, a great deal of effort is needed to export software projects from professional IDEs, importing them to visualization tools and vice versa. IDE dependent visualization hinders interoperability and limits users' choice of using professional IDEs. This problem is more subtle for software project team members working indifferent IDE's for the same project.

I. Learnability and Usability

Learnability is an important aspect of usability [25]. Learnability refers to how easily a system can be learned. Most of the software visualization tools had their own agenda and focus while they were built as a research project implementation. As a result, most of the visualization tools have high learning curve and low usability.

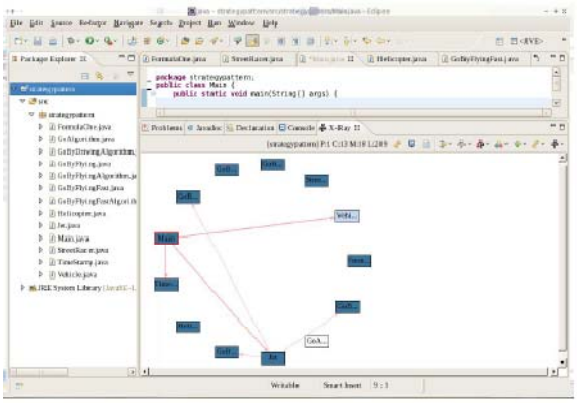


Figure 10. X-Ray Class Dependency view of Strategy pattern (case-2) project

TABLE III. VISUALIZATION ASPECTSAND MEANSOF VISUALIZATION IN SV TOOLS

Aspects	Means of Visualization	Jeliot 3	BlueJ	jGRASP	X-Ray	JIVE
Structure	UML Package diagram		x			
	UML Component diagram					
	UML Class diagram		x	x		
	UML Object diagram					x
	Information mural					
	Graph				x	
Code	Block Highlighting	x		x		x
	Pretty Printing				x	x
	Contour Model					x
Method calls	Block Highlighting	x				
	UML Sequence diagram					x
Data structure	Static graph			x		
	Animation	x				
Expression Evaluation	Code Highlighting	x				
	Animation	x				
Code Complexity	Information Mural					
	Graph				x	
Code Dependency	Information Mural					
	Graph				x	
	Text			x		
	Animation					
Debugging	Procedural	x	x	x	x	x
	Declarative					x
IDE Integration	Multiple visualization view	x		x		x
	Zooming					x
	Standalone*	x	x	x		
	Plugin*				x	x
	Project Import**		x		x	x
	Project Export**				x	x
	IDE Independence**					
	Language Independence**			x		
Execution	Forward	x	x	x	x	x
	Reverse					x
	Runtime visualization					x

* Types of IDEs
** Features of IDE Integration

VIII. RESULTS

BlueJ visualizes brief UML package diagram along with UML class diagram (figure. 2). jGRASP generates UML class diagram of a package, however it does not show UML package diagram (figure. 4.). X-Ray uses graphical rectangular shapes in order to show classes of a software system (figure 9), where graphical size of the rectangle indicates the program code size of the class. JIVE visualizes run-time UML object diagram in association with contour model (figure. 7.), however it does not show UML class diagram (table III, x indicates feature present).

Jeliot 3 highlights code blocks while animation visualizations playing. jGRASP indicates code blocks via automatic CSD generation, however does not indicate which code block is being executed (figure6). X-Ray and JIVE uses eclipse pretty printing, block collapse-expand facilities (as eclipse plugin), however X-Ray does not visualizes code block being analyzed, a double click on the class rectangle just opens the file to be edited. Step by step visualization feature of JIVE provides dynamic highlighted view of code block being executed. Simultaneously, it also shows timed-sequence diagram, dynamic object diagram, and contour model of that corresponding block (figure7).

Method calls visualization achieved by the use of dynamic time-sequence diagram in JIVE (figure8) while Jeliot 3 visualizes method calls and expression evaluation together in its visualization area (figure 5). BlueJ, jGRASP, and X-Ray disregard method calls and expression evaluation visualization.

jGRASP visualizes data structure as text book like graphs. Graphs are drawn in the viewer window from debugger or workbench. Jeliot 3 visualizes array data structure successfully, however does not provide visualization support for other types of data structures (e.g. standard java collections).

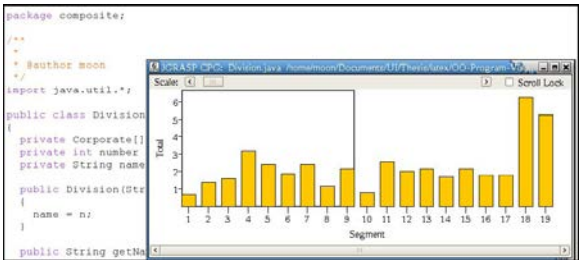


Figure 11. jGRASP CPG for composite pattern (case-1) division class

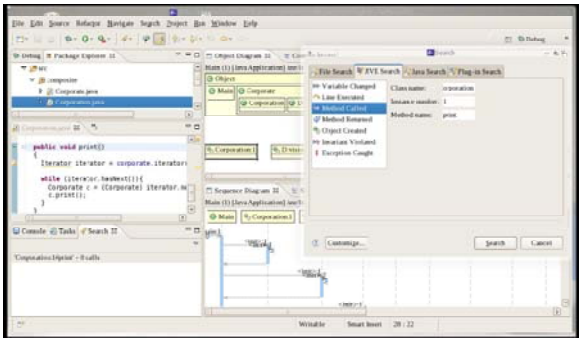


Figure 12. Declarative debugging in JIVE for composite design pattern(case-1) project

X-Ray visualizes code complexity of the entire software system (figure 9.) while jGRASP shows only for a particular class (figure. 11.). X-Ray also

visualizes class dependency as well as package dependency graphs (figure10).

Jeliot 3, BlueJ, jGRASP, and X-Ray provide traditional procedural debugging while JIVE provides both procedural and declarative debugging facilities (figure12).

JIVE and X-Ray both are eclipse plugins and utilizes eclipse export, import features. Jeliot 3, jGRASP, and BlueJhas compiler and debuggers as IDE. Among these tools, only jGRASP is capable of multiple programming language code visualization. As eclipse plugin X-Ray and JIVE has export, import features however visualization is eclipse IDE dependent. BlueJ and Jeliot 3 can export, import projects only between them. JIVE, jGRASP, and Jeliot 3 have multiple visualization views at a time. Only JIVE provides reverse program execution stepping with novel zooming feature at run time.

IX. CONCLUSION

Clearly, none of the visualization tools considered above provides a complete comprehensive software visualization for professional software engineering. In our research, we have selected five widely referenced software visualization tools and discussed their visualization features. Then we conduct a comparative qualitative analysis of these tools for the expected software visualization aspects with our selected standard case study projects. In our research, we have summarized the means of visualization aspects found on these tools and present our findings in the result section. Although there is no single software visualization tool showed complete comprehensive software visualization for software comprehension, these tools showed great promises and contributed towards the aspects and means of software visualization. However, there is indeed, a need of unified process and standardization towards a comprehensive software visualization tool for software comprehension.

REFERENCES

[1] L. Deimel and J. Naveda, "Reading computer programs: Instructors guide and exercises," Technical Report CMU/SEI-90-EM-3, Software Engineering Institute, Carnegie Mellon University, 1990.

[2] L. Burd, M. Munro, and P. Young, "Using Virtual Reality to Achieve Program Comprehension," http://www.year2000.co.uk/munro.html, 2000.

[3] H. Muller, "Understanding software systems using reverse engineering technology," http://www.rigi.csc.uvic.ca, 1994.

[4] C. A. M. Grant, "PhD thesis on Software Visualization in Prolog," http://www.cl.cam.ac.suk/Research/Reports/TR511-camg100-software-visualization-in-prolog.pdf, 1999.

[5] B. A. Price, R. M. Baecker, and I. S. Small, "A principled taxonomy of software visualization," Journal of Visual Languages and Computing, vol. 4, no. 3, pp. 211–266, 1993.

[6] B. A. Myers, "Taxonomies of visual programming and program visualisation," Journal of Visual Languages and Computing, vol. 1, no. 1, pp. 97–123, 1990.

[7] A. Moreno and N. Myller, "Producing an educationally effective and usable tool for learning, the case of the jeliot family, "Proceedings of International Conference on Networked elearning for European Universities, Granada, Spain, 2003.

[8] M. Ben-Ari, M. Myller, E. Sutinen, and J. Tarhio, Perspectives on program animation with Jeliot, ser. Software Visualization, LNCS 2269, S. Diehl, Ed. Berlin Heidelberg: Springer-Verlag, 2002.

[9] N. Myller, "Masters thesis on the fundamental design issues of jeliot 3," http://cs.joensuu.fi/jeliot/pub/theses.php, 2004.

[10] M. Klling and J. Rosenberg, "Blue - a language for teaching object-oriented programming," Proceedings of 27th SIGCSE Technical Symposium on Computer Science Education, ACM, Philadelphia, Pennsylvania, pp. 190–194, March 1996.

[11] K. V. Haaster and D. Hagan, "Teaching and learning with bluej: an evaluation of a pedagogical tool," Information Science and Information Technology Education Joint Conference, Rockhampton, QLD, Australia, June 2004.

[12] J. Malnati, "X-ray - an eclipse plug-in for software visualization,"p. 17, 2007, bachelor Project, Faculty of Informatics, University of Lugano.

[13] M. Lanza and S. Ducasse, "Polymetric views – a lightweight visual approach to reverse engineering," IEEE Transactions on Software Engineering, vol. 29, no. 9, pp. 782–795,September2003.

[14] P. V. Gestwicki and B. Jayaraman, "Methodology and Architectureof JIVE," Department of Computer Science and Engineering, University at Buffalo, Technical Report 2004-20, 2004.

[15] H. Z. Girgis and B. Jayaraman, "JavaDD: a Declarative Debugger for Java," Department of Computer Science and Engineering, University at Buffalo, Technical Report 2006-7,2006.

[16] E. Gamma, R. Helm, R. E. Johnson, and J. Vlissides, Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, 1995.

[17] P. Kuchana, Software Architecture Design Patterns in Java. Boston, MA, USA: Auerbach Publications, 2004.

[18] L. Bass, P. Clements, and R. Kazman, Software Architecture in Practice, 2nd ed. Addison-Wesley Professional, 2003.

[19] C. Hofmeister, R. L. Nord, and D. Soni, "Describing software architecture with uml," in WICSA1: Proceedings of the TC2First Working IFIP Conference on Software Architecture(WICSA1). Deventer, The Netherlands, The Netherlands: Kluwer, B.V., 1999, pp. 145–160.

[20] J. H. C. II, T. D. Hendrix, L. A. Barowski, and J. Jain, "Dynamic object viewers for java," Computer Software and Applications Conference, Annual International, vol. 2, pp.374–375, 2006.

[21] J. H. Cross, II, T. D. Hendrix, J. Jain, and L. A. Barowski, "Dynamic object viewers for data structures," SIGCSEBull., vol. 39, no. 1, pp. 4–8, 2007.

[22] T. D. Hendrix, J. H. Cross, II, and L. A. Barowski, "An extensible framework for providing dynamic data structure visualizations in a lightweight ide," in SIGCSE'04: Proceedings of the 35th SIGCSE technical symposium on Computer science education. New York, USA: ACM, 2004, pp. 387–391.

[23] P. Gestwicki and B. Jayaraman, "Methodology and architecture of jive," in SoftVis '05: Proceedings of the 2005 ACM symposium on Software visualization, New York, USA, 2005, pp. 95–104.

[24] H. Z. Girgis, A. V. Hegde, M. Pushpendran, P. V. Gestwicki, and B. Jayaraman, "Visual Queries for Interactive Execution of Java Programs," 2004.

[25] [25] T. Grossman, G. Fitzmaurice, and R. Attar, "A survey of software learnability: metrics, methodologies and guidelines," in CHI '09: Proceedings of the 27th international conference on Human factors in computing systems. New York, USAACM, 2009, pp. 649

Session 7: Supply Chains Management and Applications

ISBN 978- 974-672-556-9

Sustainable Supply Chain Management: State of the Art

Sarker Rafij Ahmed Ratan[1], Aicha Sekhari[1], Mijanur Rahman[2], Abdel Aziz Bouras[1], and Yacine Ouzrout[1]

[1]University Lumiere Lyon 2, France, [2]University of Dhaka, Bangladesh

***Abstract*— Sustainability is a very vital issue which receives key attention in the media, press, journals, business and academic research arena. Most of the conversation and argument is ambiance the environmental impact that happened on the organizational supply chains. It is mostly regarding how to convert the traditional Supply chain management practices into Sustainable supply chain management (sSCM) system. In this paper we exert our effort to show a state of art regarding sustainable supply chain management. Our research is a systematic literature and analyzed articles based study intended to provide an overview of multifaceted discussion of sSCM by different authors and establish a conceptual basis of this discipline.**

Index Terms - Supply Chain Management, Sustainability, Green Supply chain, Environmental, Economic and Social dimension.

I. Introduction

Sustainable supply chain management has emerged as a proactive approach for improving business processes and activities by social, economic and environmental aspect. All businesses rely on their supply base and customers to survive. Increasingly business peoples are being expected to know that their suppliers are as mindful of their environmental and social responsibilities as they are or should be. This is not always easy to do with local suppliers but it is even harder with suppliers from other countries. However being difficult to do is no excuse because if business body don't find out about their own supply chain, someone else might do it for them. These hard lessons have been learned by many companies and have prompted them to work towards sustainable supply chains.

There is a compelling business issue for companies to work with suppliers to jointly develop products and services which are commercially viable, preserve environmental resources and look after workforce and communities. However sustainable development starts at home and at the top. We need to start implementing it in our own organization before we can start reaching out to suppliers or customers.

The objective of this paper is to find out the state of the art of Sustainable Supply Chain Management and establish a conceptual ground regarding sSCM. It is one of the growing current phenomena to know the details of this field.

In each phase of the sustainable supply chain discipline is nourished by the three key elements like environmental, economic and social aspect.

II. Related Works

Until the early 1970s, most researchers coincide that SCM is both an emerging field of practice and an academic domain [1].

With increase in environmental concerns during the past decade, a consensus is growing that environmental pollution issues accompanying industrial development should be address together with supply chain management, there by contributing to green supply chain management [2] [3].

Environmentally conscious supply chain management refers to the control exerted over all immediate and eventual environmental effects of products and processes associated with converting raw materials into final products [4].

During the 1990s, a fresh posture regarding the relationship among supply chain management, social responsibility and environmental consciousness began to "catch on". Companies embraced the importance of working collaboratively with their supply chain partners to improve social responsibility performance [5].

Sustainable supply chain management is the strategic achievement and integration of an organization's social, environmental, and economic goals through the systematic coordination of key inter-organizational business processes to improve the long-term economic performance of the individual company and its value network [6].

Scope of Sustainable supply chain management is defined from the idea of triple-bottom line "environmental, social, economic dimension". Carter and Rogers also has taken the same things to show a house of sustainable supply chain management [6] [7]. House of Sustainable supply chain is structured as shown in Fig. 1

ISBN 978- 974-672-556-9

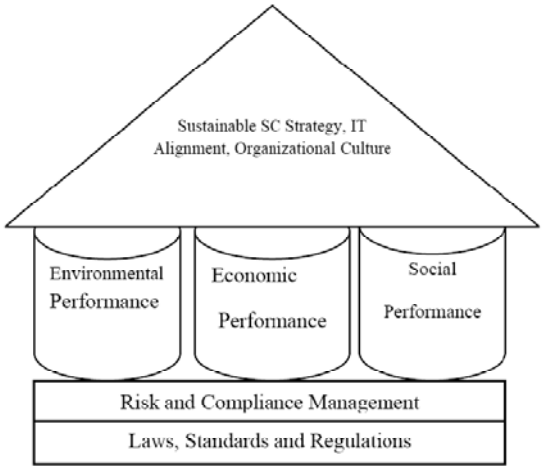


Figure 1. House of sSCM

Researcher Marisa P. et al. [8] Opines that it is still not mainstream practice to systematically take a sustainability approach in tackling supply chain and operations management issues. They conclude through adductive reasoning that the reasons for procrastinating integration of sustainability in supply chain and operations management research are the conflicting nature of the task and the inherent context, which is the focus on operations rather than environmental or social issues.

III. METHODOLOGYOFTHIS WORK

For this paper we follow the Systematic Literature and Analyzed Articles approach. Considering the increasing number of books, journals, conferences and work-shops, the systematic literature review has become an indispensible method [9] [10]. In systematic literature review pertinent work and resent findings are analyzed with regard to a particular research question.

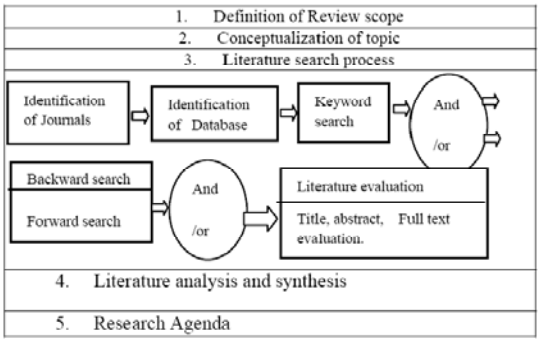


Figure 2. Research Process supported from vom Brocke et al. [11], Webster and Watson [12], Frank Teuteberg and David Wittstruck [9])

The ultimate goal of this study is to find out the state of art sustainable supply chain management. Various sources are supporting this issue by different way. So we have exerted our effort to gather important aspect of sSCM and articulate those aspects for making conceptual understanding. Steps in Fig. 2 are followed for research perspective.

Definition of review and conceptualization of this topic are shown in through 'pros and cons of sustainability' and 'conceptualization of sSCM'. We tried to collect literature in a systematic manner which mentioned in above figure. Sometimes identified the journals and locating database as a first steps. Prominent online sources are used in this regard. Some key words also used here for searching relevant journal papers like "Sustainable Supply chain" , "Sustainability", "sSCM", "Supply chain" , "Corporate Social Responsibility(CSR) and Sustainability", "Green Supply Chain" etc. After that we tried to analysis and synthesis the collecting resources. Following table shows the main theme of sustainability of supply chain emphasized by selected authors in Table I.

Currently Sustainability and sustainable supply chain management is a vital agenda for every business executives for starting their discussion. Sustainability itself creates a new era of business thinking and sources of competitive advantage. Previously SCM activities consider as the behind or back office function to manage somehow. But right now it is one of the major thinking parts of business. Different literature shows the evidence of its importance and emphasizing points also divers as like environmental issue, economic consideration, health footprint, global warming, waste recycling, social perspectives, green awareness of customer, supply chain network and configuration, simultaneous operation and SC strategy etc. Sometimes legislation and legal perspectives also are taken into account for proper sustainability of supply chain management activities. Form the flavor from different papers regarding Sustainable supply chain management, we get a new clue or research agenda that, this field required further study for showing the state- of-the-art of sSCM.

IV. SUSTAINABILITY: PROS AND CONS OF THIS ISSUE

Sustainability is like an iceberg. Most of us know something about it, but there are many aspects hidden beneath the surface. Issues such as acid rain, deforestation, species extinction, third world poverty, and climate change have captured public attention and shown the relationship between social well-being, environmental stewardship and economic activity [26].

Sustainability refers to an integration of social, environmental and economic issues [6]. Shrivastava defines sustainability as "the potential for reducing long-term risks associated with resource depletion, fluctuations in energy costs, product liabilities, and pollution and waste management" [13].

TABLE I. MAIN SUGGESTED THEMESOF SSCM FIELD BY SELECTED AUTHORS

Main suggested theme of sSCM field	*Representative references*
A validated framework of sSCM is developed and proposition of sSCM are outlined. In addition, relevant literature on sSCM is proposed.	Carter C, Rogers DS[6]:International Journal of Physical Distribution and Logistics Management.
A comprehensive view of published literature on aspects and facts of Green SCM.	Srivastava SK [13]:International Journal of Management Reviews.
The objectives of this paper are to present a literature review on sustainable supply chain management and provide a conceptual frame work of research.	Seuring M, Muller S [14]:Business strategy and the Environment
SCM is both an emerging field of practice and an academic domain	Burgess, Singh and Koroglu, [1].
Environmental pollution issues accompanying industrial development should be address together for ensuring green supply chain management.	Sheu et al.[16]; Hsu,C. W.; Hu, A. H [2].
Establishing environmental requirements for purchasing items.	Lamming and Hampson,[15]: British Journal of Management.
More focused on green purchasing for SCM	Yuang and Kielkiewicz-Yuang [16], Rao [17] ,Zhu et al. [18].
Establish an environmental risk management system for Green supply chain management	Bowen et al. [19].
Environmental, social and economic dimension built a house of sustainability of supply chain management.	Elkington [7]; Carter and Rogers [6]
Sustainability is meeting the needs of the present without compromising the ability of future generations to meet their own "needs"	United Nations. Brundtland Commission: [20]
Arraigning main activities of supply chains; exploring the domain of supply chain management.	Borade, A, Bonsad, SV [21]
Environmental ethics for sustainable supply chain management.	Beamon, Benita M.[4]
SC network configuration for Product recovery	Beamon, Benita M.and Farnandes [22]
In this research, an integrated multi-objective supply chain (SC) model is developed for use in simultaneous strategic and operational SC planning.	Sabri, Ehap, H. and Benita M. Beamon [23]
This paper describes the challenges and the opportunities facing the supply chain of the future and describe the various effects these issues have on supply chain design, management, and integration, ultimate goal to establish sSCM.	Beamon, Benita M. [24] *Operations and Supply Chain Management: An International Journal*.
This paper is set out to examine how concern for environmental impact is being incorporating into the strategic and operational management of the supply chains	Simon Croom, et al. [25]: European Operation Management Association (EurOMA) Conference.

Sustainability has become a wide-ranging term that can be applied to almost every facet of life on earth, from local to a global scale and over various time periods. This burning issue is still indicating us how to live our life in this world, how to do our business, how to make our product, how to recycle the used materials. As we have the commitment to the future generation. For all aspect, human being should ensure the issue of sustainable development. The well accepted definition is given by UN Brundtland Commission in 1987, "Meeting the needs of the present without compromising the ability of future generations to meet their own needs" [20]. We can also credit the Brundtland Commission for defining sustainable development which states – "The management of the human use of the biosphere so that it may yield the greatest sustainable benefit to present generations while maintaining the potential to meet the needs and aspirations of future generations"[27].

Sustainability is based on obligations toward future generations and presupposes intergenerational equity. Most ethicists would agree to define sustainability as the following way: Sustainability means that present and future persons have the same right to find, on the average equal opportunities for realizing their concepts of a good human life [28].

The core of mainstream sustainability thinking has become the idea of three dimensions, environmental, social and economic sustainability. These have been drawn in a variety of ways, as 'pillars', as concentric circles, or as interlocking circles [29].

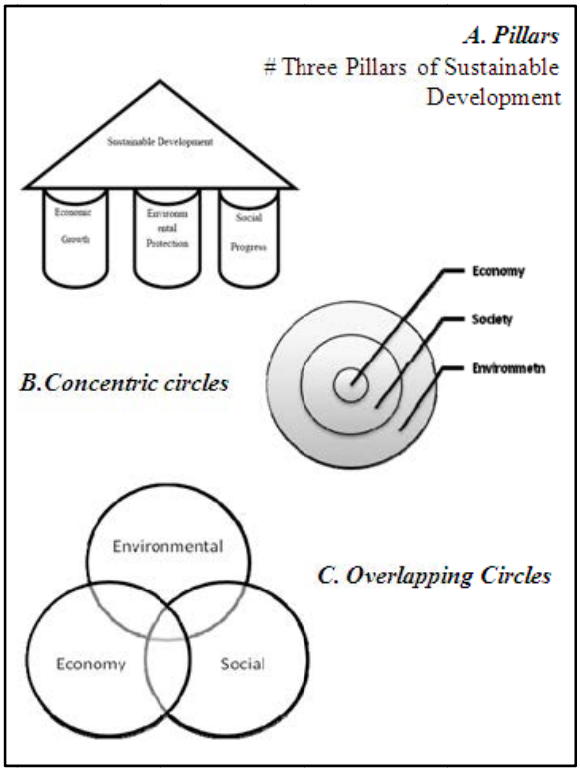


Figure2. Three Visual Representations of Sustainable Development: Pillars, Circles, Interlocking Circles

Sustainability has presented the business community with huge challenges, and therefore,

huge opportunities. This issue simultaneously will generate good and bad effect for the business bodies. Among the business firms, those who are smart enough to manipulate the benefits; will be the pioneer firms for future. Companies which are able to address those challenges properly will exist in future and those are not will extinct soon.

V. Sustainable Supply Chain Management: Conceptualization of this Topic

There are numerous definitions of the terms 'Sustainable' and 'Supply chain'. More simplified and practical oriented definition is added by New Zealand Business council for Sustainable Development, "Management of raw materials and services from suppliers to manufacturer/service provider to customer and back with improvement of the social and environmental impacts explicitly considered." The supply chain considers the interactions between a business and its customers and suppliers [30]. Harland defines supply chain management as "the management of a network of interconnected business involved in the ultimate provision of product and service packages required by end customers" [31].

The concept of supply chain management can be extended by adding sustainability. Sustainable supply chain management integrating environment thinking into supply chain management, including product design, material sourcing and selection, manufacturing processes, delivery of the final product to the consumers, and end-of-life management of the product after its useful life. Besides the environmental issues, there are two other vital part of sSCM are exists, for instances economic aspect and social aspect which are directly involved with the sustaining of organization in the competitive market.

Sometimes green issues are replicated with the Sustainability of supply chain management. Many studies addressed product lifecycle along with supply chain or Green Supply Chain Management, for example, Stonebraker & Liao [32] discussed that the stage of lifecycle variables is associated with the various dimensions of supply chain integration. In each stages of product life cycle, there are certain ups and downs are happened. Sometimes production process emitted toxic things which are deadly harmful to the society. That's why reuse and recycling of different materials are referred. Several papers provided green practices such as Duber-Smith [33]. He suggested some practices for ensuring sustainability and green perspectives of society, including reducing energy consumption, recycle and reuse, using biodegradable and non-toxic materials, minimize harmful emissions, and minimize or eliminate waste.

Sustainable Supply chain is a term that refers to a variety of approaches for the management of natural and human resources from the supplier to the manufacturers/ service provider to the consumer and back.

Sustainable supply chain management has its roots in supply chain management, i.e. it is based on the adoption and extension of its concepts [9].

Paul Hawken's book [34], The Ecology of Commerce, states paradoxically regarding the issue of sustainability, that although businesses of the industrial age contributed to the current state of affairs, business is the only means to implement the changes required in a "restorative economy." Hawken sees sustainability as a must, as we are exceeding the resource capacity of the planet.

A. Potential benefit for sustainable supply chain management

Social responsibility, good environmental and economic performances are not just about the activities that the business firms should carry out within the factory walls. As sustainability demand rose from ultimate consumer, so it is act as a source of different potential benefits and competitive advantages.

There are some potential benefits will rise easily from sustainability of supply chain management field:

- Increased productivity through greater efficiency and higher quality,
- Opportunity for innovation and competitive advantage,
- Risk minimization,
- Protection and enhancement of a company's brand reputation,
- Advantages of considering international agendas and joining sustainability indexes.

From the marketing perspective of sustainability, there are boundless opportunities to include social responsibility and be financially successful. The Canadian Business for Social Responsibility includes bottom-line benefits having [35]:

- Reduced operating costs
- Enhanced brand and image reputation
- Increased sales and customer loyalty
- Increased ability to attract and retain employees
- Publicity and increased public image from good works

VI. Conclusion

In recent years the sustainable supply chain management (sSCM) has received increasing attention and has become a progressively more popular field of academic and research area. Nowadays business bodies must handle manifold

challenges like rapid climate change, growing public interest in green perspective of business (Green Supply Chain Management, Green computing, logistics etc.) and ensuring environmental sustainability. Some legislation also imposed by the concern body like European Union, For instances, EU law for saving environment-Restriction of Hazardous Substance (RoHS),Waste Electrical and Electronic Equipment (WEEE) Eco-design for Energy using products(EuP) etc. It is one of the precise demands of time to make everything on the basis of environmental issue. The sustainable matters are always fit for the environment. So by this paper we tried to address this sustainability issue in the field of supply chain management and showing the way of articulation for different perspectives. And as a secondary goal we also tried to establish a good conceptual basis of this discipline.

ACKNOWLEDGMENT

Authors acknowledge the logistics support and assistance of Laboratory CERRAL, IUT, University Lumiere Lyon 2, and France.

REFERENCES

[1] Burgess, K., P.J. Singh and R. Koroglu (2006).” Supply chain management: a structural literature review and implications for future research.” International Journal of Operations & Production Management.p.703-729.

[2] Sheu, J.B.; Chou, Y. H.; Hu, C.C.,Z(2005): An integrated logistics operational model for green-supply chain management. Trans. Res., 41(4), 287-313.

[3] Hsu, C. W.; Hu, A. H., (2007).Green supply chain management in the electronic industry. Int.J. Environ.Sci.Tec., 5(2), 205-216.

[4] Beamon, Benita M. (2005). “Environmental and Sustainability Ethics in Supply Chain Management”, *Science and Engineering Ethics*, Vol. 11, No. 2, pp. 221-234

[5] Nalebuff and Brandenburger, (1996): Coopetition: a revolutionary mindset that combines competition and cooperation in the market place. New York, NY: Doubleday Press.

[6] Carter C and Rogers DS (2008). A framework of sustainable supply chain management: moving toward new theory. International Journal of Physical Distribution & Logistics Management 38 (5): 360-387.

[7] Elkington J (2004) Enter the triple bottom line. In: Henriques A, Richardson J (Eds) The Tripple Bottom Line: Does It All Add up? Earthscan, London

[8] Marisa P. de Brito and Erwin A. van der Laan (2010): Supply Chain Management and Sustainability: Procrastinating Integration in Mainstream Research. *Sustainability* (*2)*, 859-870, open source: www.mdpi.com/ journal/sustainability

[9] Frank Teuteberg and David Wittstruck, 2010: A Systematic Review of Sustainable Supply Chain Management Research – What is there and what is missing? (MKWI2010- Betriebliches Umwelt – and Nachhaltigkeits management, p 1001-1015. Source: http://webdoc.sub.gwdg.de/univerlag/2010/mkwi/03_anwendungen/betriebliches_umwelt-_und_nachhaltigkeitsmanagement/01_a_systematic_review_of_sscm.pdf. Access date: 01-04-2010.

[10] Fettke P(2006) State-of-the-Art des State-of-the-Art- Eine Untersuchung der Forchungsmethode “Feview” innerhalb der Wirtschaftsinformatik. Wirtschaftsinformatik 48(4):257-266

[11] Vom Brocke J, Simons A, Niehaves B, Riemer K, Plattfaut R, Cleven A (2009) Reconstructing the giant: on the importance of rigour in documenting the literature search process. In: Newell P, Whitley E, Pouloudi N, Wareham J, Mathiassen L (Eds.) Information systems in a globalising world: challenges, ethics, and practices, Proceedings of the 17th European Conference on Information Systems, Verona.

[12] Webster J, Watson R (2002): Analyzing the past to prepare for the future: writing a literature review. Management information system Quarterly 26(2):13-23

[13] Shrivastava SK (2007): Green- supply management: a state-of-art literature review. International Journal of Management Reviews 9 (1): 53-80.

[14] Seuring M, Muller S (2008), Core issue in sustainable supply chain management- a Delphi study. Business strategy and the Environment 17(8): 455-466

[15] Lamming, R and Hampson, J. (1996): The environment as supply chain management issue. British Journal of Management, 7 (special issue 1):45-62

[16] Yuang, A.; Keilkiewicz-Yuang, A., (2001): Sustainable supply network management. Corp. Environ. Manage. 8(3), 260-268.

[17] Rao, P, (2002): Greening the supply chain a new initiative in South East Asia, Int. J. Oper. Prod. Manage. 22(6), 632-655.

[18] Zhu, O.; Sarkis, J.; Geng, Y., (2005): Green supply chain management in china: Pressures, Practices and performance. Int. J. Oper. Prod. Manage. 25(5) 449-468.

[19] Bowen, F. E.; Cousine, P. D.; Lamming, R. C.; Faruk, A. C.,(2001): Explaining the gap between the theory and practice of green supply. Greener Manage. Int.,35, 41-59.

[20] United Nations. 1987: "Report of the World Commission on Environment and Development." General Assembly Resolution 42/187, 11 December 1987. Access date: 06-05-2010.

[21] Borade A, Bonsad SV (2008): The discipline of supply chain management- A systematic literature review. The Icfai Journal of Supply chain Management 5(1): 7-26

[22] Beamon, Benita M. and Fernandez,(2004): Supply-Chain network Configuration for product Recovery. *Product planning and control*, Vol.15, No.3, p-270-281.

[23] Sabri, Ehap H. and Benita M. Beamon (2000). “A Multi-Objective Approach to Simultaneous Strategic and Operational Planning in Supply Chain Design”, *OMEGA*, Vol. 28, No. 5, pp. 581-598.

[24] Beamon, Benita M. (2008). “Sustainability and the Future of Supply Chain Management”, *Operations and Supply Chain Management: An International Journal*, Vol. 1, No. 1, pp. 4-18.

[25] Simon Croom, Sam Barani, Dan Belanger, Tim Lyons, Jaime Murakami (2009): Sustainable supply chain management – an exploration of Current practice. Presented at European Operation Management Association (EurOMA) Conference, June, 2009.

[26] Rohm and Haas industry Report, 2007: “Long term vision for sustainability-2008 and beyond.” Source: http://www.rohmhaas.com/SDreport/long-term-vision-for-sustainability.asp). Access date: 06-04-2010.

[27] Larry Berglund, 2005: Sustainable Supply Chain Management Practices: What In The World Is Going On? Source: [http://larryberglund.wordpress.com/2009/11/07/sustainable-supply-chain-management-practices-what-in-the-world-is-going-on. Access date: 10-04-2010.

[28] Konrad Ott, (1995): Advocated the “The Case for Strong Sustainability.” Source: http://umwethik.botanik.uni-greifswald.de/booklet/8_strong_sustainability.pdf. Access date: 15-02-2010

[29] Adams, W.M. (2006). "The Future of Sustainability: Re-thinking Environment and Development in the Twenty-first Century." Report of the IUCN Renowned Thinkers Meeting, 29–31 January 2006.

[30] New Zealand Business council for sustainable development (2003): Business guide to a Sustainable Supply Chain, p-8, source: http://www.nzbcsd.org.nz/ supplychain . Access date: 08-04-2010

[31] Harland CM (1996). Supply Chain Management: relationships, chains and networks. British Journal of Management 7(1), p. 63-80.

[32] Stonebraker, P. W., & Liao, J. (2006): Supply Chain Integration: Exploring Product and Environmental Coningencies. Supply Chain Management, 11 (1), 34-43.

[33] Duber-Smith, D. C. (August, 2005): The Green Imperative. Soap, Perfumery, and Cosmetics, 78 (8), pp. 24-26.

[34] Paul Hawken (1993): The Ecology of Commerce: A Declaration of Sustainability, New York, Harper Collins Publishers.

[35] Industry Canada: An Overview of Corporate Social Responsibility. Source:http://www.ic.gc.ca/eic/site/csr-rse.nsf/eng/rs00129.html: Access date: 10-04-2010.

A Supply Chain Simulation Based on SCOR Model for the Decision on Entrepreneurial of Plant

N. Reeveerakul[1], Y. Ouzrout[1], N. Chakpitak[2], and A. Bouras[1]
[1]LIESP,Université de Lyon (Lumière Lyon 2), France, CAMT,
[2]Chiang Mai University, Thailand

***Abstract*— The low economic growth, competition among neighboring countries, and strategic used in supply chain collaboration have caused Foreign Direct Investments (FDIs) to relocate their businesses. In order to gain more advantages which help to prevent further business relocation, this paper proposes an integrated framework based on the supply chain to make a decision on plant situations by considering in site location. The Supply Chain Operations Reference (SCOR) model was applied to design process categories and decomposing processes, which implemented in Arena simulation. The simulation describes processes in supply chain collaborations. From this result, the supply chain cost based on basic SCOR processes of: Plan, Source, Make, Deliver, and Return are introduced for selected countries of supply chain partners.**

Index Terms- plant site location; SCOR model; supply chain simulation

I. INTRODUCTION

Since The investment of foreign manufacturing in developing countries result rapidly increasingly on economic growth to the host countries. Such contributions are creating new jobs by foreign companies, increasing the use of multinational distribution networks, or even spending on research and development to support many national projects. These have led to higher productivity through increased capital, which in turn has led to high living standards. Consequently, several developing countries are recognized to attracting foreign investors to invest their manufacturing business that they can gain benefits from them.

However, businesses face many critical challenges with regards to economic crises and the increasing of the intensity of competition. Since 1997, the financial crisis that erupted in Asia has caused severe economic turbulence in the economies of South East Asia. The crisis affected to the global economy for example, the overall investment climate of Thailand seems to have worsened between 2004 and 2007 judging by the opinions of firm managers which have deteriorated significantly [1].

Since the crisis was raised many problems that give Foreign Direct Investments (FDIs) loss profits and increase operational costs. Then they have been recognized on alternative solution, in order to gain more advantages and give them the least problem. One of alternative strategies can be introduced by considering on the effectiveness of supply chain collaboration among the attractive destination of FDI.

For example, since the early 1990s, China as a highly attractive destination for FDI from developed countries [2]. These situation lead Multinational Enterprises (MNEs) relocate their manufacturing facilities and move to China. Besides, from the recent evidence, especially Vietnam and China are country's competitiveness in the areas of labor cost and availability of labor. Besides, when comparing economic conditions, Thailand was stronger in market demand than other competitors, but not as strong as those of India and China [3]. Therefore, several East Asian countries are recognized to improve their capabilities to satisfy the investors better than the competitors. As they can gain benefits from manufacturing FDI in Asia and also foreign manufacturing can gain higher cost advantages.

II. CONTEXTAND PREVIOUS RESEARCHES

From the previous section, the realization on the effectiveness of supply chain collaboration for FDIs was presented. In this section, the review of literature will be explained on the related factors which considered on the decision making on plant entrepreneurial status and how simulation has been recognized to supply chain context.

A. Factors influencing decision making on plant entrepreneurial status

Chan et al.[4], found that the main reasons for business relocations were cost savings and business expansion, regardless of whether the firm was a plant or a headquarters. Other literature from Qi Chun, Jhon C.S. Tang,[5] proposed a genetic algorithm (GA) as an analytical tool to define fitness function as a variable selection algorithm and analysis of inward Foreign Direct Investment (FDI). They found 24 independent variables that affect FDI in China. Among those factors, the most frequently selected determinants of FDI flows into the industries were productivity, research and development investment, labour investment, and education. As faras investment in manufacturing is concerned, profit is a key factor for the investor. Therefore, minimizing

the costs while maximizing revenue is undertaken to maximize profit.

For financial measurement, net present value (NPV) is a widely recognized approach that ascertains the time value of money invested in a business. The approach can be used for optimizing decision making. Moreover, a number of production planning studies have subsequently been carried out at the aim of maximizing the NPV [6]. From the review of literatures, table I synthesizes and classifies the various studies according to the attributes used.

TABLE I. FACTORS ON PLANT RELOCATION AND INVESTMENT DECISION

Factor on plant relocation and investment decision	[7]	[5]	[8]	[9]	[10]	[11]	[12]	Total
1. Supply chain cost	x		x	x				**3**
2. Production quality		x						1
3. Market demand			x	x		x		**3**
4. Logistics cost				x				1
5. Employee skills			x		x			2
6. Expatriate management			x					1
7. Technological support	x							1
8. Loyalty to the employer			x					1
9. Labour productivity			x					1
10. Close to supplier (Lead time)	x			x		x		**3**
11.Economic development policy	x			x		x		**3**
12. In-bounded lead time				x				1
13. Supply chain risk				x		x	x	**3**
14. No. of project in R&D	x						x	**2**
15. Total investment in R&D		x						1
16. Return on asset		x						1
17. Internal culture						x		1
18. Competition	x					x		2
19. Firm size							x	1
20. Capital to labor		x						1
21. Total sales values		x						1
22. Export sales values		x						1
23. No. of high educated employee		x						1
24. Wages of domestic employees		x						1
25. Security					x			1
26. Subcontractors existence					x			1

From a review of literature, there are many factors that have been taken into account for the supply chain. Besides, the majority of researches on plant investment decisions do not agree on a unique set of attributes to typify relations between relocation decisions and the influencing factors. The attributes used in each research are different and the contents of each type of decision remain very general at times, such as the variables affecting FDI based on specific production or transportation costs in the supply chain. Some authors detailed the attributes which influence a decision in each type of situation but their studies were focused on specific orientations, for instance, supplies existence [7],[9],[12], Supply chain risk [9],[11],[12], even the criteria on economic development policy [7],[9],[11], on supply chain cost [7],[8],[9], and research and development support [7],[13] which cannot be generalized. Those factors can be explained as the importance of the supply chain collaboration for the decision on plant relocation and investment. Most researchers emphasize on cost of production, supply chain between supplier and customer or even supply chain risk. This includes the consideration on external factors as an economic development policy. Therefore, the focusing on cost of operational in manufacturing and supply chain among partners are taken into consideration for FDIs.

B. Method used for supply chain analysis.

Many companies have observed that besides their "traditional" risks arising from their business activities, new risks have emerged from sources that are often related to the close collaboration within their supply chain networks [6]. Meanwhile, the supply-Chain Council (SCC) Risk Management Team & Douglas Kent (December,2007) [13] argued that "preservation of shareholder value" is highly important in the supply chain management and is recognized by organizations. The disruption in the supply chain will have an impact on the overall performance of organizations.

Nowadays, companies face many critical challenges with regards to supply chain management. One of the main challenges arises out of trends associated with the supply chain collaboration among supply chain partners. Close co-operation between supply chain partners benefits for performance in the companies. To improve manufacturing industry, the integration of supply chain and risk analysis is necessary. The supply chain analysis is used to identify the impact of changes in the location and/or capacity of facilities on supply, manufacturing and distribution activities. This impact can be measured by determining the required input quantities (materials, labour, capital and energy) and the respective unit prices for alternative locations and capacity level. In addition, factors in the international environment like taxes, barriers to trade and exchange rate, will affect the allocation strategies [8].

However, several authors have developed approaches under various points of view. An example is, N.Viswanadham and Kannan Balaji, 2005, proposed a quantitative model for optimal decision between FDI and outsourcing with multi-stages of supply chain risk caused from the supply chain, inventory and transportation cost. Moreover, to describe supply chain strategy from two sources of uncertainty, F. Cucchiella, M. Gastaldi, 2006, used MatLab programming to cover production capacity

and price fluctuation risks. Besides, to measure risks by considering on operational problems, K. Rojanapibul, J. Pichitlamken, 2005, designed excel spreadsheet integrated with VBA to perform scheduling procedure with ARENA software simulation. From the review of literature, the integration of discrete event simulation and SCOR model on supply chain context have been introduced for several researches as shown in table II.

TABLE II. LITERATURE REVIEW ON METHOD USED IN SUPPLY CHAIN CONTEXT

Authors	Metrics and measures followed SCOR (P,S,M,D)	Real option theory	Matlab programming	A multi-criteria scoring procedure	Discrete event simulation (Arena + Excel spreadsheet)	Combined discrete event and continuous approach	System dynamics
F.Cucchiella, M.Gastaldi, 2006.		√	√				
A.Gunasekaran, C.Patsl, Ronald E. McGaughey	√						
S.Kara, G.Freeman, B.Kayis, T.Ray and H.Abbass, 2005				√			
K.Rojanapibul, J.Pichitlamken, 2005					√	√	
Jennifer V., Kevin P.and Danny J. 2008							
Fredrik Persson, Mirko Araldi, 2009,	√				√	√	
Sameer Kumar and P.Phrommathed, 2006					√		
Y.H.Lee, M.K. Cho, S.J. Kim, Y.B. Kim, 2002	√					√	
G.Bailly, J-C. Vessilier, 1991							√
K.Hafeez, M. Griffiths,J. Griffiths, M.M.Naim, 1996							√

In this regard, we propose an integrated framework to help decisions on plant situations by considering on supply chain cost among site location. The cost analysis in terms of the net present value (NPV) is presented in which corresponding to the model of simulation.

III. PROPOSED MODEL: A FRAMEWORKOF SUPPLY CHAIN NETWORKS BASED ON SUPPLY CHAIN OPERATIONS REFERENCE (SCOR) MODEL

According to the previous analysisof literatures, the proposed framework on supply chain network is introduced. The framework helps the manufacturer to analyse operational cost based on SCOR model characteristic. However, in this paper the collaboration among supply chain partners and cost of supply chain are the main focus. The proposed model is presented in Fig.1.

There are three partners in supply chain network. Those partners are supplier, manufacturer, and customer. The supply chain collaboration describes the relationship of manufacturers, suppliers and customers. There are risks occurred in internal and external manufacturing system. Those risks lead to poor manufacturing performance. So the mitigating of risks will result in more profits and manufacturing effectiveness. From the proposed model, the scope of level 1 processes (Plan, Source, Make, Delivery and Return), process categories, and process element are explained. With the Source, Make, Deliver process elements, a common internal structure has been upon. The proposed model focuses on Make-to-Stock environment.

Then in order to help manufacturers on make a decision for site selection of plant, and improve manufacturing performance the simulation model application based on SCOR (Supply Chain Operations Reference) model analysis is presented.

II. THE SIMULATION MODEL APPLICATION: ARENA SOFTWARE APPLICATION

In order to describe the interaction between relevant parameters and define major factors that affect the system performance, simulation technique using Arena software is implemented. The simulation required the evaluation on performance attributes from SCOR as shown in table III. This evaluation is used to describe their effects on supply chain and system performance.

TABLE III. PERFORMANCE ATTRIBUTESAND ATTRIBUTESAND METRICS USEDFORTHE SIMULATION MODEL.

Performance attribute	*Metric*
Reliability	Perfect order fulfilment
Responsiveness	Order fulfilment lead time
Flexible	Demand chain response time Delivery flexibility
Cost	SC management cost
Asset	Cash to Cash cycle time Inventory days of supply Asset Turns

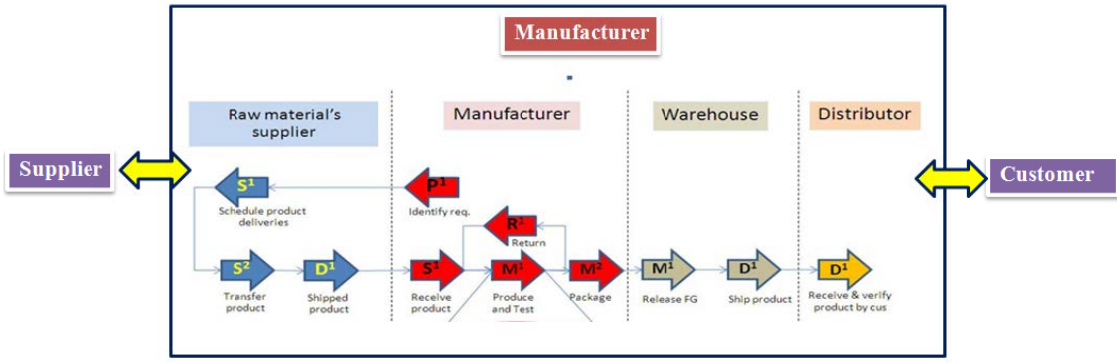


Figure 1. Proposed model of supply chain based on SCOR model.

In this context, quantitative analyses are derived from ARENA simulation. Besides, the simulation based on SCOR-model supported the importance of inbound and outbound supply chain, and the collaboration in chain networking. On this section, the processes simulation helps to describe the proposed model in practical analysis. As shown in following Figure, Fig. 2, 3 and 4, the explanation on process element based on SCOR for manufacturer, supplier and customer are provided.

From Fig.2, the process elements for manufacturer are implemented in ARENA simulation. The three basic processes of SCOR: Source, Make and Deliver are represented. The model focuses on Make-to-Order (MTO) environment. Besides, Figure.3, decompose processes for supplier activities are built. The processes are described on receiving, delaying and sending material to manufacturer. Demands sending to supplier are kept in excel spreadsheet and presents as the input for supplier activities. Finally, schedule product deliveries of customer activities are explained as shown in Figure.4. There are three volumes of customer demand as High, Medium and Low. The demand quantity of high volume is defined as 300 quantities, 50 quantities for medium and 40 quantities for low volume. Time between the arrivals of demand is determined to every 24 operational hours.

V. Results and Analysis

Consequently, the results from the simulation within 6 months are shown on the excel spreadsheet. Those results are average of inventory level, quantities of shipments, raw materials, work in processes, and average time used in each processes. The input data on general information, for example price per unit of product, transportation cost, holding cost, and ordering cost per unit will be provided from decision makers. Then the calculation on supply chain cost will be calculated in terms of total revenue and total expenditure of source, make, deliver and return cost that will be explain as the net present value (NPV).

The simulation was run for 10 replications by the estimation period of 6 months. Then the result from the simulation are shown in the excel spreadsheet as shown in Figure.5. Those results are used for the calculation of supply chain cost.

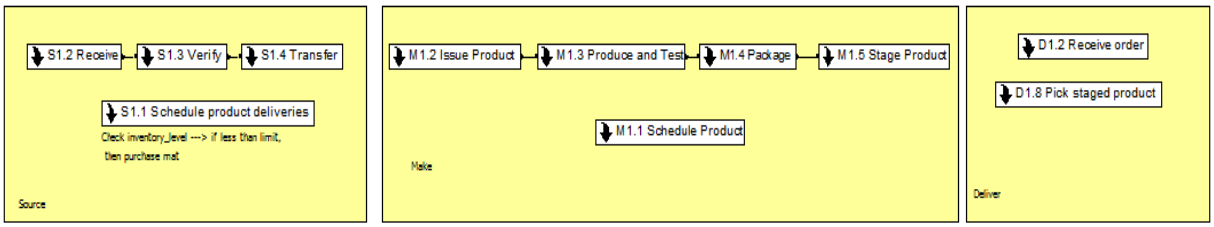


Figure 2. Process element based on SCOR model for manufacturer

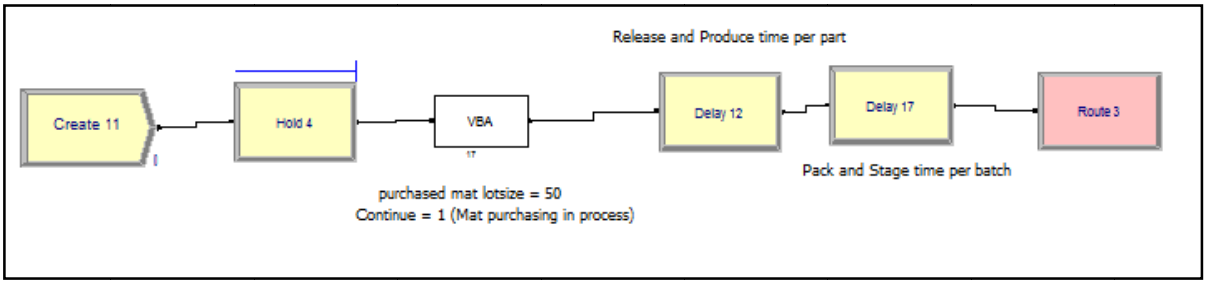


Figure 3. Process element based on SCOR model for supplier

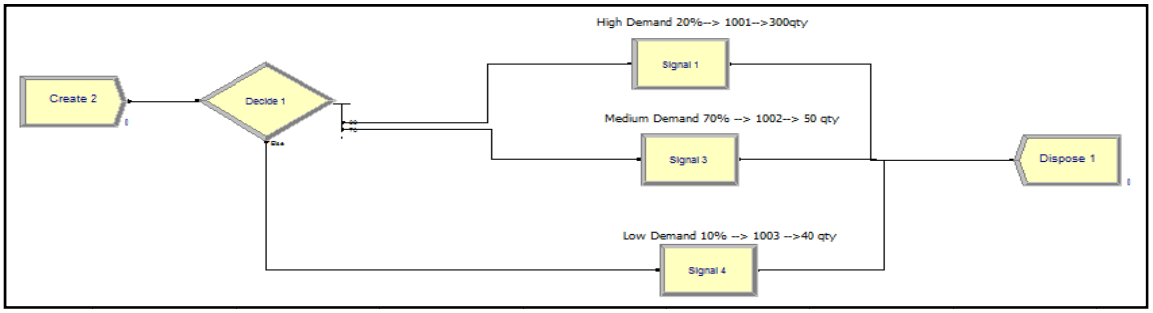


Figure 4. Process element based on SCOR model for customer

	Order_No	QTY_order	Demand_T	Status		Receive_Ti	Verify_Time	Transfer_Time	time_InReceive	time_InVerify	time_InTransfer	IssueTime	ProduceTim	PackTime	StageTime	DelverTime	timeInDeliver		RM_invento
2	1	300	1001	5	2/6/2010 21:49	15	15	15	0.514829525	0.547746941	0.460077433	0	0	0	0.001598	15			199
3	2	50	1002	5	2/6/2010 21:49				0.562800535	0.079716105	0.322658445	0	0	0	0.002212		0.02	2/6/2010 21:49	349
4	3	50	1002	5	2/6/2010 21:49				0.701003886	0.000775253	0.11450217	0	0	0	0.029826		0.02	2/6/2010 21:49	199
5	4	50	1002	5	2/6/2010 21:49				0.685844152	0.380385552	0.332704796	0	0	0	0.01021		0.02	2/6/2010 21:49	199
6	5	50	1002	5	2/6/2010 21:49				0.425299312	0.863489181	1.524614413	0	0	0	0.026237		0.02	2/6/2010 21:49	199
7	6	50	1002	5	2/6/2010 21:49				0.254766019	0.223301116	0.46219574	0	0	0	0.007443		0.02	2/6/2010 21:49	169
8	7	50	1002	5	2/6/2010 21:49				0.002198073	0.003380347	0.117527203	0	0	0	0.004329		0.02	2/6/2010 21:49	219
9	8	50	1002	5	2/6/2010 21:49				0.27910957	0.682819425	0.051881569	0	0	0	0.009034		0.02	2/6/2010 21:49	369
10	9	50	1002	5	2/6/2010 21:49				0.239750315	0.406058678	0.014487132	0	0	0	0.003669		0.02	2/6/2010 21:49	219
11	10	40	1003	5	2/6/2010 21:49				0.541594938	0.083275589	0.046108881	0	0	0	0.005376		0.02	2/6/2010 21:49	369
12	11	40	1003	5	2/6/2010 21:49				0.228795837	0.072509147	0.152228211	0	0	0	0.015431		0.02	2/6/2010 21:49	219
13	12	50	1002	5	2/6/2010 21:49				1.17096255	0.097103139	0.098788725	0	0	0	0.015283		0.02	2/6/2010 21:49	369
14	13	50	1002	5	2/6/2010 21:49				0.019360944	0.182654513	0.127259413	0	0	0	0.000524		0.02	2/6/2010 21:49	219
15	14	300	1001	5	2/6/2010 21:49				0.182424749	0.464003688	0.772458737	0	0	0	0.005091		0.02	2/6/2010 21:49	369
16	15	300	1001	5	2/6/2010 21:49				0.037437808	0.042857538	0.08596294	0	0	0	0.02574		0.02	2/6/2010 21:49	219
17	16	300	1001	5	2/6/2010 21:49				0.091195999	0.206002851	0.081843062	0	0	0	0.001664		0.02	2/6/2010 21:49	369
18	17	300	1001	5	2/6/2010 21:49				0.007212854	0.020481949	0.34085448	0	0	0	0.019243		0.02	2/6/2010 21:49	219
19	18	300	1001	5	2/6/2010 21:49				0.234115268	0.20311617	0.081994761	0	0	0	0.011365		0.02	2/6/2010 21:49	369
20	19	300	1001	5	2/6/2010 21:49				0.136152939	0.406227079	0.247303661	0	0	0	0.012665		0.02	2/6/2010 21:49	219
21	20	300	1001	5	2/6/2010 21:49				0.416855695	0.190987153	0.186564136	0	0	0	0.012286		0.02	2/6/2010 21:49	369
22	21	40	1003	5	2/6/2010 21:49				0.372689879	0.03935305	0.108539698	0	0	0	0.010284		0.02	2/6/2010 21:49	179
23	22	40	1003	5	2/6/2010 21:49				0.052607564	0.09351304	0.331625574	0	0	0	0.022447		0.02	2/6/2010 21:49	189
24	23	50	1002	5	2/6/2010 21:49				0.298282113	0.327477563	0.155287111	0	0	0	0.01365		0.02	2/6/2010 21:49	209
25	24	50	1002	5	2/6/2010 21:49				0.173820384	0.460442169	0.81804738	0	0	0	0.016265		0.02	2/6/2010 21:49	359
26	25	40	1003	5	2/6/2010 21:49				0.11523724	0.012319047	0.363530512	0	0	0	0.00891		0.02	2/6/2010 21:49	159
27	26	40	1003	5	2/6/2010 21:49				0.118719417	0.152486019	0.682047023	0	0	0	0.015787		0.02	2/6/2010 21:49	169
28	27	50	1002	5	2/6/2010 21:49				0.046670624	0.031502754	0.576391435	0	0	0	0.007917		0.02	2/6/2010 21:49	169
29	28	300	1001	5	2/6/2010 21:49				0.018509355	0.036493992	1.023406612	0	0	0	0.018159		0.02	2/6/2010 21:49	169
30	29	50	1002	5	2/6/2010 21:49				0.176218945	0.062417993	0.012828242	0	0	0	0.007558		0.02	2/6/2010 21:49	269
31	30	50	1002	5	2/6/2010 21:49				0.395798852	0.147581471	0.440971462	0	0	0	0.02037		0.02	2/6/2010 21:49	419
32	31	40	1003	5	2/6/2010 21:49				0.008927053	0.678537312	0.013662878	0	0	0	0.05608		0.02	2/6/2010 21:49	269
33	32	50	1002	5	2/6/2010 21:49				0.204059242	1.13195293	0.031106512	0	0	0	0.008149		0.02	2/6/2010 21:49	419
34	33	50	1002	5	2/6/2010 21:49				0.278888346	0.008210822	0.029451119	0	0	0	0.004729		0.02	2/6/2010 21:49	269

Figure 5. Results from Arena

		Year					
	Explanation	**0**	**1**	**2**	**3**	**4**	**5**
Total revenue	Price per unit x Qty_fg_shipped		13,896,195.00	14,591,004.75	14,955,779.87	15,254,895.47	15,559,993.38
% rate change				5%	2.5%	2%	2%
Initial investment cost		10,000,000					
Total expenditure			701,322	736,388	754,798	769,894	785,292
-source cost (ordering - holding+ transporting cost of material)	(Transportation cost per unit x Order qty)+ (Holding cost x Inventory level) + (Ordering cost + No. of ordering)		134,398.00	141,117.90	144,645.85	147,538.76	150,489.54
-Make cost (Direct + indirect + Overhead + additional cost)	(Production cost per unit x Produce amount) + (Unit of elec. Cost x amount of elec. used)		526,737.00	553,073.85	566,900.70	578,238.71	589,803.48
-Delivery cost (Shipped finished good cost)	Transportation cost/pcs x Amount of shipment		36,539.00	38,365.95	39,325.10	40,111.50	40,913.83
-Return cost	(% of non-conform mat. X Order qty x cost of non-conform. Per unit) + (% of non-conform FG x Amount of shipment x cost of non-conform. FG per unit)		3,648	3,830	3,926	4,005	4,085
Net CF			13,194,873	13,854,617	14,200,982	14,485,002	14,774,702
Interest rate (Thailand)		12%	12%	12%	12%	12%	12%

Figure 6. Supply chain cost analysis

From Table 5, the results show total amount of shipment, average inventory level, total amount of customer demand, time between processes, work-in process quantities. Those values take into consider on supply chain cost. Besides, we assume the operating cost with the case in Thailand for example, holding cost per unit, product and material price per unit, cost per unit on transportation,% non-conforming of material and product. Then the sample of supply chain cost corresponding the four basic processes of SCOR: Source, Make, deliver and Return, are presented as shown in Fig. 6. Finally, the NPV for the selected entrepreneurial location of plant is presented as 60,522,973.72 baht.

VI. Conclusion

In order to help manufacturer make a decision on site location for FDIs investment, supply chain collaboration among partners is involved.

Due to the supply chain cost is considered as an important factor used for FDIs investment decision. Then, analysis on operational cost cannot be neglected. The proposed model based on SCOR processes is validated through the application of simulation. From the results of simulation, the Net Present Value (NPV) can be explained in a financial value of money invested for a business. For this reason, NPV can provide useful information on making investment decision by the investor.

Acknowledgment

This work was supported by the European Commission in the **Erasmus Mundus Project: eLINK (e**ast-west **L**ink for **I**nnovation, **N**etworking and **K**nowledge exchange) scholarship**.** This research was conducted in the LIESP (Lumière Lyon 2) laboratory and CAMT (Chiang Mai University). The authors also appreciate the help provided by all partners of the Erasmus Mundus Project

References

[1] BOI, Thailand Executive summary report, 2008
[2] Wu and Pual, 2002.
[3] Reported from 2009 Foreign Investor Confidence survey report.
[4] Chan, S.H., Gua, G.W., Wang, K., 1995. "Stock Market Reaction to Capital Investment Decisions: Evidence from Business Relocation," Journal of Financial and Quantitative Analysis 30, pp.81-100.

[5] QI Chun, Jhon C.S.Tang, "Foreign Direct Investment in Chinese Industries: A Genetics Algorithm Approach," International Conference on Machine Learning and Cybernetics, 2007, 938-944.

[6] Wolfgang Kersten, Philipp Hohrate and Mareike Boger, "An Empirical Approach to Supply Chain Risk Management: Development of a Strategic Framework," Available: http://www.poms.org/conferences/poms2007/CDProgram/Topics/full_length_papers_files/007-0507.pdf (accessed December 2008)

[7] Electronics Industrial Economic Report, Available online at: http://www.bot.or.th/Thai/EconomicConditions/Thai/North/ArticleAndResearch/DocLib_Article/2550_10@ElectronicsIndustryinNorthern.pdf (accessed December 2008).

[8] Bart Vos, Henk Akkermans, Capturing the Dynamics of Facility Allocation. International Journal of Operations & Production Management, 16, 1996, 57-70, 11.

[9] N.Viswanadham and Kannan Balaji, Foreign Direct Investment on Outsourcing: A Supply Chain Decision Model. International Conference on Automation Science and Engineering, 2005, 232-237.

[10] V.M.Mzanda, A.J.Buys, The Ambivalence of technology transfer through Foreign Direct Investments. Portland International Center for Management of Engineering and Technology Conference, 2006, 1827- 1835.

[11] Alan Davis, Eric M.Olson, Critical competitive strategy issue every entrepreneur should consider before going into business. Business Horizon, 2008, 51, 211-221.

[12] R.Sarker, S.Kara, G.Freeman, B.Kayis, T.Ray and H.Abbass, A Multi-agent Simulation Study for Supply Chain Operation. IEEE 2005.

[13] Supply Chain Risk Measurement, Available online at: http://www.drkresearch.org/Contact_Us/Risk_Roundtable/Supply_Chain_Risk_Measurement.pdf (accessed December 2008).

[14] F. Cucchiella, M. Gastaldi, Risk management in SC: A real option approach, Journal of Manufacturing technology management, 2006, Vol.17 No. 6, pp. 700-720.

[15] K.Rojanapibul, J.Pichitlamken, Assessing risk in a job schedule: Integrating a scheduling heuristic and a simulation model to a spreadsheet, 2005

Gap Analysis of Knowledge Supply Chains for E-tourism Curriculum Design

-A Knowledge Management Perspective

Fu Jing
College of Arts, Media & Technology,
Chiang Mai University,
Thailand

Nopasit Chakpitak
College of Arts, Media & Technology,
Chiang Mai University,
Thailand

Matteo Mario Savino
University of Sannio,
Italy

Wided Batat
University of Lyon 2,
France

***Abstract*-With the booming online population throughout the world, e-tourism, as a combination of technology of ICTs and knowledge of tourism, also as an extension of e-commerce, has been changing and revolutionizing conventional tourism industry and tourism education. For most of the developing countries, whose tourism industries are the major national income, are in great need of improving the mechanism and infrastructure for sustainable tourism development. An acute shortage of updated knowledge and skills and qualified working people result from a big gap between what e-tourism industry needs and what e-tourism education provides.**

Index Terms- E-tourism, Gap analysis, Competency Model, Knowledge Supply Chain, ISO 9001:2008, TQM, Knowledge Management

I. INTRODUCTION

In the recent decades, tourism development is the crucial target and main source of national income for most of the developing countries. With abundant natural resources and unique history and culture, these developing countries in Asia, like Thailand and China, are attracting more and more visitors from the western world. Take Chiang Mai as an example, the second largest province in Thailand, in the year of 2008 and 2009, the total number of tourists in Chiang Mai are 4,405,720 and 4,181,878 respectively, the number of foreign visitors are 2,529,420 and 2,598,041.(http://www.cmcity.go.th/aboutus/economic.php). That is to say Chiang Mai annually accommodated foreign travelers nearly 40% of the total visitors. It is found out tourists in Chiang Mai are not "old tourists" any more, who often take mass and conventional tour package with organized tourist business and with lower ability in computer and language, but increasingly belong to the kind of customers who favor online booking and individual traveling and are normally well-educated with more disposable income and buying power. These customers are called as "New Tourist" by Buhalis (2003). He summarized 5 characteristics, named as "5S" for "New Tourist"-namely Segmentation, Specification, Sophistication, Satisfaction and Seduction. Buhalis stated that what new tourists need is gradually transforming from the interests of summer holiday, urban tourism to a mixed demand of technology, edu-tainment (also called as educational entertainment or entertainment education, which is a form of entertainment designed to educate as well as to amuse), eco-environment and multi-culture. Researchers in post-modern tourism, like Lash (1990), Lyon (1994), Hannabuss (1999), stated that tourists in post-modern period travel for experience of the extra-ordinary and de-differentiation. These tourists are more independent and nature lovers who need more individualized, authentic, and enriched travel experiences to meet their educational, cultural, intellectual and sporting interests. ICTs have changed the way of tourism industries in providing tailor-made tourism products and personalized services, enabling tourism companies to efficiently identify the need of their educated, experienced, sophisticated and demanding customers and to formulate personalized value-added chains for individual customers. In 1998, World Tourism Organization (WTO) has pointed out that the key to success lies in using ICT tools for the quick identification of consumer needs and in reaching potential clients with comprehensive, personalized, and up-to-date information. Obviously, ICTs empowered this process and have increased its speed and interactivity. The new demands from these new tourists have increasingly challenged both e-tourism industry and e-tourism education as well.

However, today's tourism industries suffered a lot from a shortage of qualified employees who can make full use of the skills of ICTs and relevant knowledge and experience to meet the change and challenge of new market and to satisfy the requirement of new tourists. As the cradle of the talents at a higher educational level, local universities should take this responsibility to provide competent e-tourism workers with an ability of life-long learning. Then these people will continue to contribute to a real sustainable e-tourism industry.

In 2003, Department of Modern Management and Information Technology (MMIT) was set up in College of Arts, Media & Technology (CAMT) of Chiang Mai University by Chiang Mai Government in Thailand, with an aim to offer a new generation of knowledge workers at an undergraduate level. This

ISBN 978- 974-672-556-9

discipline focuses on using technology and modern management techniques to create new knowledge and to improve efficiency and productivity in their future work. In 2009, CAMT planed a cooperative education program for e-tourism. Students from MMIT are free to choose e-tourism as major electives, known as E-tourism Electives.

A seminar was held among lecturers in e-tourism group and Board of Total Quality Management (TQM) in CAMT to analyze and discuss on job location of students in e-tourism electives according to what they have been taught in CAMT, namely competencies that students acquired. A documental study on overall curriculum has found out knowledge and skills learned from these subjects and programs are not enough to meet the requirement of e-tourism industries.

Since e-tourism has not been set up as a major in any universities in Thailand, even as a practice of a minor has not been proved successful, e-tourism as a major-electives program based on knowledge of management and skills of IT from MMIT majors seems reasonable and feasible. This research is a conceptual framework based on the investigation of the practice of e-tourism electives from MMIT to design competency models of how e-tourism systems are working and what competencies are required, trying to find out the gap between e-tourism industry and e-tourism education. By designing competency models for e-tourism systems, knowledge and skills to achieve required competencies will be directions for matching relevant courses and e-tourism curriculum design. The methodology of the initial finding of the research will be based on a large scale of literature review and cross-disciplinary documental work. A number of discussions of the research will be drawn from the perspective of knowledge management, and further study will be planned from both industrial and educational points of view.

II. Literature Review

A. E-tourism

The concept and domains of e-tourism can be briefly described into three fields-Business (covers management, marketing and finance, Turban, 2002), Tourism (involves transport, travel, hospitality, leisure and heritage, Buhalis, 2003) and IT/IS (consists of information system, information communication, technology and telecommunication, Werthner, 1999)

Buhalis (2003) stated that e-tourism revolutionizes all business processes, the entire value chain as well as the strategic relationship of tourism organizations with all their stakeholders. And ICTs played an irreplaceable role in tourism information systems, especially customer relationship management, by enabling tourism enterprises to react efficiently to environment changes and consumer trends. Today's tourists are called as "new tourists" by Buhalis, however their desires and interests in search of new meaning and dignity of travelling belong to the feature of "post-modern tourists" (Nuryaniti, 1996).

To design a curriculum of e-tourism, the content and domain of e-tourism and trend of customers' needs should be clearly defined, known and considered from industrial, managerial and educational perspectives.

B. The Industrial context: E-tourism and MICROS Systems

ICTs have changed strategic, operational and tactical management of tourism organizations. ICTs play critical role in the production, marketing, financial, operational and distribution functions of tourism sectors. Internet, intranet and extranet are prerequisites to make the whole e-tourism process happen. These technologies revolutionized tourism distribution channels dramatically especially in Computer Reservation Systems (CRSs), Global Distribution Channels (GDSs), and Destination Management Systems (DMSs).

As a representative in e-hospitality systems of e-tourism, MICROS, is one of the best software groups worldwide for hospitality reservation, operation and management. MICROS Systems, Inc. is the world's leading developer of enterprise applications, serving the hospitality and specialty retail industries exclusively. Studying on how MICROS systems work in e-hospitality industry will not only know how hospitality mechanism can work better but reflect on what curriculum needs to be planned to meet the requirement of the systems and to incubate e-hospitality workers.

C. Competency Gap and Competency Model

Spencer (1993) defined competency as an underlying characteristic of an individual that is causally related to criterion-referenced effective and/or superior performance in a job or situation. He put forward five types of competency characteristics, namely Motives, Traits, Self-concept, Knowledge and Skill. He believed that knowledge and skill are the visible competencies which are relatively easy to develop and most cost-effective way to secure abilities. On the other hand, core motive and trait competencies at the base of the personality iceberg are more difficult to assess and develop. Spencer presented *Competency Dictionary* in a generic form, in scales designed to cover behavior in a wide range of jobs, and to be adapted for many applications.

When designing competency models and competency-based curriculum, both visible and invisible competencies should be taken into consideration, meanwhile competency dictionary turns out to be a helpful guideline.

D. Cooperative Education and Total Quality Management

Law (1970) defined cooperative education as a work program associated with a school education in

which student-learners receive supervised payroll experience as part of the school curriculum. John Dewey (1910) wrote that the only adequate training for occupation is training through occupations. He said learning by doing will not only benefit students, but schools, business and industries, even community as well. To effectively plan a cooperative education program in e-tourism, curriculum must be carefully designed, managed and evaluated. In this context, Total Quality Management (TQM) can be considered as a useful tool to better implement cooperative education. As well known TQM focuses on both external factors, like customer satisfaction and internal measures, such as efficiency, calling for full management commitment and participation of all levels. According to TQM concepts, the quality of curriculum design and planning needs multidisciplinary and cross-cultural knowledge and experience from all relevant aspects. The application of TQM can be made through a panel of experts comprising academics and practicing managers who should be involved to the course curriculum design and development process; while non-academic members including representatives from industry, professional bodies of government should also be invited. This collective expertise must work together as a total quality control committee (March, 1991) and total quality learning organization (Garvin, 1993) to communicate and discuss face–to-face, clarify any doubts, predict and resolve any problems and difficulties.

A useful evidence of TQM application can be the certification through the international standard ISO 9001:2008. It is worldwide recognized as a tangible signal of customer's trust of quality assurance and focuses on quality control systems from process of product design to process design, and from production process to service after-sale. The standard provides the basic list of essential requirements for building a quality assurance management system, which can be built via a documented system to establish and maintain an effective operation. This documented procedure usually contains the purpose and scope of an activity and how it shall be controlled and recorded. ISO 9001:2008 can be considered as a precursor and basis of Total Quality Management and has been applied widely into the quality of education, especially in curriculum design (Kie & Palmer, 1999). If we consider both industrial and educational perspective to consider the whole process of how students are fostered, we will see how ISO 9001:2008 can be applied to "Educational Product" design and how it will be improved by Deming's PDCA Wheel (Practice-Do-Check-Act, Deming, 1986). The process can be simplified as the following model (see Figure 1).

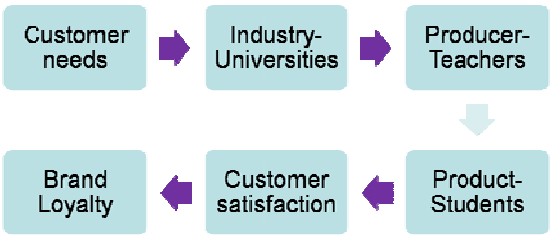


Figure 1. ISO 9001:2008 Applications in Education

It is clear that Universities, which act as educational industries, should be standardized to improve their managerial and executive competency to serve and support educational behaviors according to customers' needs. Teachers, working as producers, should be standardized in their teaching contents and methods to meet both academic and practical educational requirements. Students, as the products of Educational industry, should be produced in a scientific and systematic way to meet the requirement of international standards and satisfaction of their employers. Finally, qualified students will meet the demand the requirement of the industry and win a brand loyalty to their universities. ISO 9001:2008 application and PDCA practice will standardize the whole process with brand-able replication.

E. Knowledge Management in a Managerial Perspective: a competency model

Knowledge is a real experience of knowledge worker especially experts in executing a specific task. Knowledge Management is the making use of working people and useful business information for core business and innovation (Nonaka & Takeuchi, 1995). Cooper (2004) gave a general description and explanation about the relationship between knowledge management and tourism. Peter Senge (1994) put forward a concept of learning organization in his *The Fifth Discipline* by introducing Systems Thinking, Personal Mastery, Mental Model, Shared Vision and Team Learning. Senge's publication on learning organizations opened a new era and aroused more concerns about life- long learning and sustainable development.

All these theories and practice left priceless treasure and challenge for universities to share and create knowledge, to set up appropriate strategies and to take actions. With more qualified and well-educated graduates, both students and universities will be more efficient, competitive and productive. Therefore, e-tourism industries and even the whole society will be brought huge tangible and intangible wealth.

ISBN 978- 974-672-556-9

III. METHODOLOGY

To design a competency-based curriculum for e-tourism, gap identification and analysis are the initial and crucial process. Four steps will be taken as following:

A. Gap between Global E-tourism Systems and General E-tourism Systems

For the first step, the research will take a global glance at the development of e-tourism worldwide. From literature view, all the systems related to e-tourism will be studied and analyzed. As for the major systems, competency models will be designed according to the competency demanded in running the systems. Then a gap analysis will be found between the competency demanded in the global e-tourism industry and general e-tourism industries. This will be clarified in Figure 2.

Global e-tourism systems	Core Competencies	General e-tourism systems	Core competencies
Business systems	Information seeking , Analytical Thinking , Conceptual thinking , Achievement Orientation, Expertise, Concern for Order, Quality and Accuracy, Flexibility , Relationship Building , Organizational Commitment , Teamwork and Cooperation , Customer Service Orientation	Business systems	Flexibility , Relationship Building , Organizational Commitment , Teamwork and Cooperation , Customer Service Orientation , Information seeking
Website systems		Website systems	
E-airline systems		Tourism and Hospitality systems with the aid of IT	
E-hospitality systems			
E-tour operator			
E- travel agency			
E-destination systems			
Gap	Less specific, professional, systematical, competitive Lack of a way of thinking, finding, proper strategies to achieve a higher goal Lack of expert guide, team learning, creativity and innovation		

Figure 2. Gap Analysis between Global E-tourism Systems and General E-tourism Systems

From this figure, we can see e-tourism systems can be classified into business systems, website systems, and tourism systems. Through studying how these systems are operated, we will know the advantages and limits of the systems, and then with the help of an expert panel, we will figure out what competencies may be needed to make these systems work better in a managerial and technical level. Take Management Information System (MIS) as an example; to efficiently make this system work, establishing a database is very crucial. It will provide managers with information and mechanism to support their decision making by pre-planning, designing and specifying the types of information required. With the reference of the Competency Dictionary (Spenser, 1993), Competency of Information Seeking (INFO) matches the competency to establish an interrelated and computerized database. Then, a competency model is designed as Figure 3 shows below.

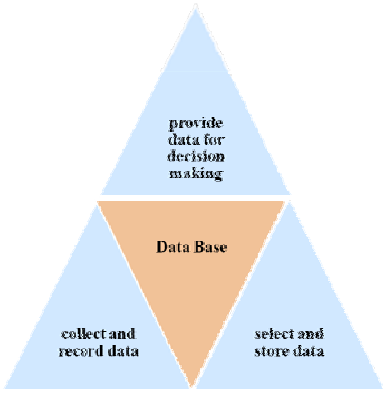


Figure 3. Competency model for Management Information System (MIS)

In order to establish an efficient database, operators must be competent to collect, record, select and store data, and then provide relevant and useful data for decision making. To run the whole process, operators not only need skills to know how to digitalize and store data but more importantly, to analyze and decide what kind of data should be digitalized and used. From the process of running this system and what competencies are needed, relevant courses should be matched and designed in e-tourism curriculum.

Compare the global e-tourism systems with the general tourism systems, a gap is found out that the general systems are less specific, systematical and competitive with a lack of a mechanism of expert guide and team learning.

B. Gap between Present E-hospitality Systems and MICROS Systems

Next, the research will be specified to e-hospitality. To choose e-hospitality systems in Chiang Mai as a case study to find the gap between Present e-hospitality systems and the ultimate solution systems for e-hospitality-Micros is because Chiang Mai has a comparatively mature international tourism environment. Data from Tourism Bureau of Northern Thailand shows, there are 341 hotels and guest houses in Chiang Mai in the year 2008, and the number increased to 418 one year later. These hotels and guest houses have accommodated 4,405,720 and 4,181,878 travelers in 2008 and 2009 respectively both inland and overseas. A lot of these travelers have experience of booking online and received tourism products and services through e-hospitality systems. This tendency calls for more competent hospitality workers. Through a study on the systems, competency models are designed according to the knowledge and skills required in running the systems. Present e-hospitality systems mainly include Customer Reservation Systems (CRS), Customer Information Systems (CIS), Property Management Systems (PMS), and Distribution Systems (DS). However, MICROS systems are more specific which cover and reach far beyond these. MICROS systems pay more attention to the cooperation among departments in the hotel and relationship among

information, operation, revenue management and customer satisfaction. A comparison between present e-hospitality systems and MICROS Systems was made, a gap was found as it is shown in Figure 4:

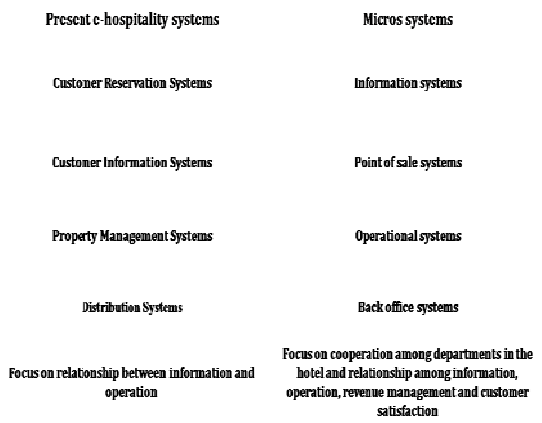


Figure 4. Gap Analysis between Present E-hospitality Systems and MICROS Systems

Normally, MICROS systems are divided into 4 parts, namely Information Systems (IS), Point of Sale Systems (POS), Operational Systems (OS) and Back Office Systems (BOS). Study on how these systems are run will help industries and universities to know the merits and demerits of the systems. Take Central Reservation Systems (CRS) as an example, in order to facilitate marketing research and planning, the staff must have competency to provide and track customers' data and history, to categorize various and frequent individual travelers, and then to design and provide personalized services. Therefore, a Competency of Customer Service Orientation (CSO) is specially needed to meet the requirement and satisfaction of various customers. A competency model is designed as Figure 5 shows.

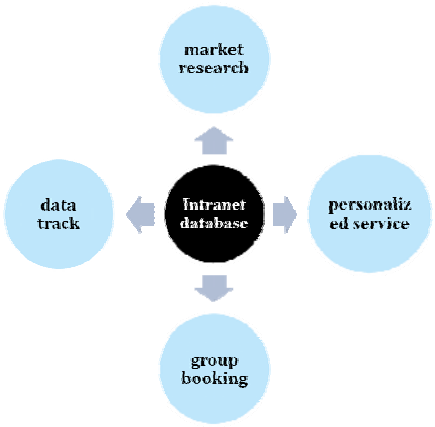


Figure 5. Competency Model for Central Reservation Systems (CRS)

From this model, it is clear to see how CRS is run. During market research, customers' needs will be reflected and collected as digital data. Through tracking data and categorizing booking, the intranet database will store and analyze the need of the customers in order to design specific and tailor-made service to meet the demand of the customers. Competency of Customer Service Orientation (CSO) is crucial to implement the system. According to the process of how the system is run and what competency is needed, then an appropriate training program can be planned.

C. Curriculum Study and Analysis

A documental study on the present tourism curriculum among universities in four parts of Thailand will be the 3rd step. These universities have a comparatively good reputation for their tourism graduates according to the feedback of National Tour Guide Committee of Thailand (2008). These universities have won a brand loyalty because of their excellent graduates for many years. A study on their tourism curriculum found out that each university has his unique features and special focus in accordance with the need of regional economic development, but all of them have not clearly and systematically reflected what e-tourism industries need in the curriculum. See Figure 6 as below:

Present Situation in E-tourism Curriculum (Universities in Thailand)

Universities	Faculties	Location	Core Courses	E-tourism related course
Chiang Mai U	Humanities	Northern T	Tourism management	0
	Business Administration		Hospitality	0
Payap U	Tourism	Northern T	Marketing	0
Khonken U	Humanities	Northeastern T	Business	1 (Computer appliance in business)
Prince Songkla U	International	Southern T	Costal tourism	1(IT for hospitality and tourism)
Mahidol U	International	Central T	English Communication	0
Kasetsart U	Tourism Management	Central T	Management	0
Silpakorn U	Tourism Management	Central T	Eco-Tourism	0
Suan Dusit RU	Tourism Management	Central T	Aviation	1 (Aviation Information)

Figure 6. Present Situation in E-tourism Curriculum in Thailand

Eight universities from 4 parts of Thailand were chosen. They are: Chiang Mai University and Payap University from Northern Thailand; Khonken University from North Eastern Thailand; Prince of Songkla University from Southern Thailand; Mahidol University, Kasetsart University, Silpakorrn University and Suan Dusit Rajaphat University from Central Thailand. Among them, some of them are from top 10 universities in Thailand like Chiang Mai University, Khonken University, Prince of Songkla University, Mahidol University and Kasetsart University. It is worthy to mention that Prince of Songkla University, with a campus in the most famous tourism attraction of Thailand-Phuket Island, has set one course related to e-hospitality called as "IT for Hospitality and Tourism". And Suan Dusit Rajaphat University has set up a course named as "Aviation Information" which is similar to E-Airline. Through investigation on their curriculum, it is found that these universities more or less set up courses to apply computer technology to tourism. All of them are going to improve their students' computer literacy by adding web-marketing and e-commerce

subjects to their curriculum. Even some of them are going to introduce MICROS in the laboratory as the new attempt in hospitality courses. But the concept and mechanism of e-tourism discipline are not systematic and complete, while the capabilities of both teachers and students in applying ICT to tourism still need improvement. A big gap was found between what e-tourism industries need and what e-tourism education provides.

D. Curriculum design, development and assessment

A case study will be done in CAMT in succession in the further research. More junior students from the major of Modern Management and Information Technology, called as MMIT, will be chosen to learn e-tourism as major electives programs from the first semester of 2010 (June-October). All the relevant activities will be involved in curriculum design besides knowledge and skills of e-tourism. From an educational context, lectures to learn knowledge especially IT skills needed for e-hospitality systems and laboratory for e-tourism systems are equally important for students; from a managerial context, cooperative education program with e-tourism industries will be crucial for competency acquisition. Students will gradually be evolved from teacher followers into independent learners then become life-long learners by learning to manage their own knowledge. What to teach and how to teach are the two key points which will lead to the success of e-tourism curriculum. This can be described in the Figure 7 as below:

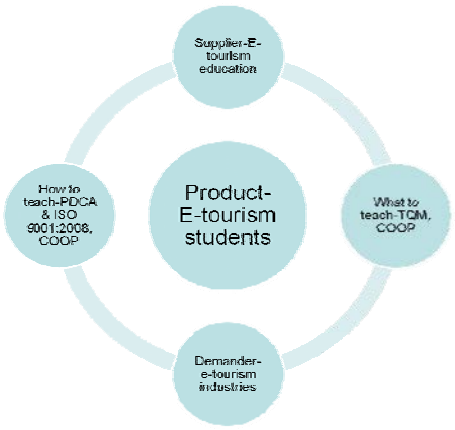


Figure 7. Knowledge Supply Chain of E-Tourism Education

Total Quality Management (TQM) together with Supply Chain Management (SCM) will bring experts from knowledge demander, e-hospitality and other e-tourism industries, to work with lecturers from knowledge supplier, E-tourism Electives Group in CAMT, as an expert panel. Both sides will identify, supervise, check and validate what competencies are needed in running e-hospitality systems, and ISO9001:2008 will ensure the standardization of the whole process with traceable documental work. With the aid of the methods of PDCA (practice-do-check-action) and implementation of Cooperation Education Program, finally curriculum will be checked and evaluated by internal and external standards from both industrial and educational perspectives.

IV. Main Potential Benefitsand Results

The research has potential benefits for both academic and social sectors. For e-tourism industry and education, this research will help them to know not only the knowledge gap between them, but also what kind of qualified e-tourism workers they need to incubate so as to improve effectiveness, efficiency and competitiveness of the whole e-tourism knowledge supply chain by integrating the needs of the e-tourism industry and requirement of post-modern tourists with universities' curriculum design. University and traditional tourism are linked closer to give training and carry out pilot e-tourism projects via cooperative education. During the running of the new curriculum, knowledge will be updated and created. Meanwhile, more qualified lecturers and students will be trained who will bring about competitiveness to e-tourism industry and a brand loyalty to universities. Developing countries in GMS, Africa, Eastern Europe and South America will foster a new generation of e-tourism graduates to develop and operate their e-Tourism infrastructure, finally a more competitive and productive e-tourism market will be formed. For community, this research will bring an innovation to traditional tourism mechanism, so that a more advanced infrastructure of e-tourism will surely bring both spiritual and material benefits to the society.

V. Conclusion

This research is the initial finding of e-tourism curriculum design. Gap identification and gap analysis will be the first step and the direction to fix the disjoint between what e-tourism needs and what e-tourism education provides. The goal of higher education is not to teach students how to survive their jobs, but to train students with competency of long-life learning in their career. Although this research starts from e-hospitality, it will surely be generalized to other fields of e-tourism. Meanwhile, the development and assessment of curriculum take time, so this research needs more work and contribution from academic and practical level in order to ensure that the ideas and outcome are positively coherent. Other universities with tourism discipline or other countries in need of developing e-tourism can compare and contrast with this research. Co-research and collaboration are encouraged to contribute to a worldwide sustainable e-tourism development.

References

[1] D. Buhalis, *E-tourism, Information Technology for Strategic Tourism Management*, London: Pearson, 2003

[2] C. Cooper., "Knowledge Management and Tourism", *Annals of Tourism Research*, Vol.33, 2006

[3] W. E. Deming, *Out of the Crisis*, MIT Center for Advanced Engineering Study, 1986

[4] A. Feigenbaum, *Total Quality Control*, McGraw-Hill, New York, 1986

[5] G. Garvin, "Building a Learning Organization", *Harvard Business Review*, Vol. 71, 1993
[6] K. S Kie, & E. Palmer, "An Empirical Examination of ISO 9000-registered Firms in New Zealand", *Total Quality Management*, Vol.10, 1999
[7] S. Lash, *Sociology of Postmodernism*, Routledge, London, 1990.
[8] G. Law, *Cooperative Education: Handbook for Teacher-Coordinators*, American Technical Society: Chicago, 1970
[9] G. March, "Explorations and Exploitations in Organizational Learning", *Organization Science*, 1991
[10] I. Nonaka & H. Takeuchi, *The Knowledge-Creating Company*, New York: Oxford University Press, 1995.
[11] W. Nuryanti, "Heritage and Postmodern Tourism", *Annals of Tourism Research*, Vol.23, 1996
[12] P. Senge, *The Fifth Discipline Field Book*, Bantam Doubleday Dell Publishing House Group, Inc. 1994
[13] P. Senge, *A Fifth Discipline Resource: Scholl That Learns*, Bantam Doubleday Dell Publishing House Group, Inc., 2000.
[14] L. Spenser, *Competence at Work: Models for Superior Performance*, John Wiley & Sons, Inc., 1993
[15] E. Turban, *Electronic Commerce: A managerial Perspective*, Prentice Hall, New Jersey, 2002
[16] H. Werthner, *Information Technology and Tourism: A Challenging Relationship*, Springer-Verlag, Vienna, 1999.
[17] WTO, "World Tourism Organization, "Hot tourism Trends for 21st Century", *WTO News Letter*, WTO, 1998.
[18] Committee of Tour Guide of Thailand. http://www.pgathaiguide.com/National
[19] Chiang Mai Tourism Bureau http://www.cmcity.go.th/aboutus/economic.php

ABOUT AUTHORS

1) Fu Jing, PhD candidate of Knowledge Management in College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
2) Nopasit Chakpitak, PhD, Dean of College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
3) Matteo Mario Savino, PhD, Professor of University of Sannio, Italy
4) Wided Batat, PhD, Associate Professor of Lyon 2 University, France

Supply Chain of Fruits and Vegetables and Correlated Impact of Managing the Quality

Muhammad Mushtaq Ahmad and Péter Fehér

***Abstract*— Fruits and vegetables processing, trading and marketing require some vital support of integrated supply chain management system. Due to its perishable nature and more consistent application of quality standards, the organizations involved in business of these horticultural commodities have to face colossal quality problems. Therefore, one of the purposes of this research study was to collect information regarding current processing practices and quality attributes applied. For this purpose, an exploratory and qualitative research study was carried out by using the pre-tested questionnaire and to collect information from various organizations dealing in supply chain of these horticultural commodities.**

Index Terms- fruits and vegetable processing; quality management; supply chain management

I. INTRODUCTION

Fruits and vegetables are perishable in nature and cannot be stored for longer periods, which result in very sensitive and complicated trading of these horticultural commodities and exposing big challenges to suppliers, processors and traders. In addition, the inconsistent availability of healthy fruits and vegetables from farm-gate to the consumers, continuous quality assurance push by the traders and supermarkets, competitive global environment, increasing trend of better supply by companies of other competitive countries and also more and more implications of quality standards (FruitVeB, 2008) are also other parallel competitive elements making this business more vulnerable and complex. Fruits and vegetables require certain effective post harvest management practices for better quality and ultimate price. Therefore, to cope with these challenges, the only integrated way to be applied from farm gate collection through washing, grading, packaging, storage and marketing to ultimate consumers is to establish the consistent and sustainable supply system of "Farm-To-Fork Approach", currently transformed as Supply Chain Management (SCM).

Globally, the demand of fruits and vegetables is increasing like anything due to dietetic potentiality, which thereby, augmented the annual production and also enhanced the exports and imports of these horticultural commodities around the globe. According to Statistical Yearbook of FAO (2009), the world production of fruits and vegetables in 1996 was 98.0 million tons which increased to 146.0 million tons in 2007. Similarly, in 1996, the total exports were 20.0 billion US$ which then increased to 44.0 billion US$ in 2007, whereas, imports of these fruits and vegetables were 25.0 and 49.0 billion US$ in the same years, respectively.

In case of food sector, supply chain starts from farm-gate collection of raw materials, then processed and prepared and the finished goods continue downstream to regional and global distribution centers and finally enter retail stores, where they come in contact with the customers for consumption. In general, SCM is one entity, which helps foster to manage and meet customer needs by improving every element of the supply chain, whereas, the role of food safety, quality requirements and their evaluation in international trade is also receiving more and more attention. Among the components of consumer behaviour, motives related to safety are getting stronger and as a result, an increasing attention is given to the interrelations of food safety, quality, supply chain and ethical issues (Joszef et al., 2009).

Presently, food markets are becoming globalised and food trade more and more concentrated and internationalized, but at the same time, quality assurance and traceability requirements against suppliers have also increased significantly (Spriggs, 1999). Increased food safety requirements are now expected to exercise more strongly on fresh products (Unnevehr, 2000) due to the fact that food safety and quality components have an outstanding role with respect to fresh products supplied through marketing channels. Similarly, more safe food demands, traceability and inspection in the developed countries are higher than in countries of medium level and developing one. Risk and uncertainty have been the hallmark of these horticultural products and food business. A long gestation period, seasonal production, significant transportation and logistical costs, the low value/weight ratio, poor infrastructure and the lack of any effective legal system to enforce fair trading, accentuates risk and uncertainty along the supply chain (Trienkens and Van der Vorst, 2007).

The new age consumers are becoming more health conscious in terms of hygiene, source of the food, ingredients of processed food and caloric content. Food safety and quality requirements have an increasing importance around the globe (Kalei, 2008). Dissemination and introduction of the quality assurance systems such as GLOBALGAP, International Food Standards (IFS), Hazard Analysis Critical Control Point (HACCP) etc., as an urgent task, have become integral parts for both production

ISBN 978- 974-672-556-9

and trade (Biacs, 1999). Without the realization of these systems, efficient marketing can be performed neither on the domestic nor on the international market (Tomcsanyi, 1999).

A. Problem of the Study

The thematic approach of SCM in this business is to provide wholesome quality fruits and vegetables to even distant consumers without distorting taste, texture, color also the size of these horticultural commodities. Whereas, an effective application of quality management systems and the flow of efficient and accurate information both upstream and downstream and also an efficient application of SCM system of fruits and vegetables has not been designed to its heart due to the lack of detailed study and development and standardization of particular business process involved in the quality management throughout the whole chain from harvesting to consumer's end.

B. Scope of the Study

Keeping in view the appropriate scope and objectivity and thereby prospective applicability, instill quality and sustainability in supply chain, a research study entitled "Supply chain of fruits and vegetables and correlated impact of managing the quality" was carried out. The main purpose of this study was to chalk out the flaws which are being faced by the various organizations in managing the quality of these perishable horticultural commodities in the supply chain from Farm-To-Fork and also to bring the wholesome fruits and vegetables to the consumers door step in order to satisfy them.

C. Objective of the Study

The objective of this study was: To chalk out the quality-oriented gaps present in the current supply chain of fruits and vegetables in the form of AS-IS and TO-BE Approaches and also to re-design and propose new business processes to develop an impeccable supply chain system.

II. REVIEWOF LITERATURE

Fruits and vegetables are critical source of nutrients and other substances that help protect against chronic diseases, including heart diseases and cancer, stroke and other chronic diseases (Prior and Cao, 2000; Produce for Better Health Foundation, 1999; Quebedeaux and Elisa, 1990; Southon, 2000; Tomas-Barberan and Robins, 1997). Fruits, and vegetables play a significant role in human nutrition, especially as sources of vitamins (C, A, B6, thiamine, niacin, E), minerals, and dietary fiber (Quebedeaux and Bliss, 1988). Their contribution as a group is estimated at 91% of vitamin C, 48% of vitamin A, 30% of folacin, 27% of vitamin B6, 17% of thiamine, and 15% of niacin in the U.S. diet. Fruits and vegetables also supply 16% of magnesium, 19% of iron, and 9% of the calories. Other important nutrients supplied by fruits and vegetables include folacin, riboflavin, zinc, calcium, potassium, and phosphorus. Some components of fruits and vegetables are strong antioxidants and function to modify the metabolic activation and detoxification disposition of carcinogens, or even influence processes that alter the course of the tumor cell (Wargovich, 2000).

Due to their potential goodness for health, the demand of these horticultural commodities is extensively increasing as the consumers are becoming more aware of health risks and health related issues. Food business is the largest business of the world and with the marvelous growth of cash and carries and super markets, as a componential part of this business, the trading of fresh fruits and vegetables is also becoming an essential entity for these business giants due to the rise in customer expectations and time arrests. Fruits and vegetables require impeccable post harvest management practices to manage life and ultimate quality. In this regard, number of techniques are being used through-out the world which include storage at low temperatures, cold and hot water treatments, vapour heat treatment, controlled atmosphere storage, modified atmosphere packaging etc. and due to the stringent quality parameters being applied by most of the countries, the need for cold chain trend is also becoming more intensive.

Hicks (2002) found that modern food packaging technology brings traditional foods into a global arena which increasingly emphasizes their commercial and economic aspects. This means that food and packaging technologists become involved in the entire food supply system. This system ranges from the sea, village farm, plantation, to the markets and consumers in towns and cities, not only in their own country, but also in distant overseas markets. The surplus foods grown in the village have a need to be more carefully harvested, protected from spoilage and damage, packaged, and transported by various means to these markets.

Climatic conditions, especially temperature and light intensity, have a strong effect on the nutritional quality of fruits and vegetables. Soil type, the rootstock used for fruit trees, mulching, irrigation, fertilization, and other cultural practices influence the water and nutrient supply to the plant, which can affect the composition and quality attributes (appearance, texture, taste and aroma) of the harvested plant parts (Goldman et al., 1999). Maturity at harvest and harvesting method influence the commodity's quality and extent of physical injuries. Delays between harvest and consumption or processing can result in losses of flavor and nutritional quality. The magnitude of these losses increases with exposure to temperatures, relative humidity, and/or concentrations of oxygen, carbon dioxide, and ethylene outside the ranges that are

optimum for each commodity during the entire postharvest handling system (Lee and Kader, 2000).

Management intensive application helps to add more value in the supply chain in the form of quality, timeliness, food safety and labour. This has technological, organizational and managerial implications that resound along the chain all the way to upstream and downstream. Value creation potential of an individual firm can be enhanced, in some cases, through a more collaborative relationship with its suppliers and customers, provided that trust and commitment is developed. Herlambang et al. (2009) studied the effectiveness in the supply chain and its relation to the satisfaction level of ultimate customers. Sub-effectiveness could occur if each actor in a supply chain attempts to optimize its own results rather than to integrate its goals and activities with other actors to optimize the results of the entire chain. In this case study on Manalagi mango fruit supply chain in Indonesia, the researcher used price-margin analysis, gap analysis and relationship analysis to explore the effectiveness of the chain. He found that a high level of certain activity cost in the supply chain could indicate the inefficiency built into that activity. The perishable nature of fresh fruit made particular actors in the supply chain can damage all the efforts taken in another stage to deliver more value to customers. Thus, effectiveness is derived from the sum of the contributions of all participants along the chains.

Most of the food we eat derives from crops and livestock (Knight et al., 2002) that are formed specifically for that purpose. With the exception of fish, game and some wild nuts, berries and fruit almost everything else comes initially from crops and livestock. From a wide range of process and production systems, some of which are highly complex and requiring considerable technical input come a vast array of products which are now food on the modern food retailers' shelves. The term "Food Chain" is usually reserved for an understanding of the total supply process from agricultural production, harvest/slaughter, through primary production and/or manufacturing, to storage and distribution to retail sale or use in catering and consumer practice. Many sectors in the food chain now recognize advantages in working partnerships to realize the maximum benefits of a safe, wholesome and efficient food supply business (Adebanjo, 2004).

In today's environment there are a number of reasons why a total integrated food chain approach is necessary including establishing specifications, food safety systems and traceability. It is important that for those in the manufacturing part of the chain, clear specifications are established for raw materials and ingredients such that they meet the requirements for further processing. This relates to both quality attributes and food safety. It is also most important to ensure consistency in developing final products, in terms of keeping and eating quality and traceability (Barrett, 2004). Wysocki (2005) delineated that these are both evolving and challenging times for food distribution and retailing and now the same consumers are behaving in so many different ways due to more shopping experience, awareness, number of varieties and more expectations and quick response from the organizations for marketing.

Hughes and Merton (1996) pointed out that consumers are turning increasingly away from specialist retailers towards supermarkets for their fresh produce requirements, and looks at how supermarket chains are responding. Grimsdell (1996) examined the supply chain of various horticultural commodities which was developed efficiently by British Field Products Ltd, over the year. He outlined the six fundamental requirements for an efficient supply which were scale of operation; strategic alliances, production, flexibility, continuity of supply, quality control and communications and concluded that mutual awareness by all the stakeholders was one of the best ways forward for sustainability. SCM is not a thinking but is a paradigm, characterised by operational effectiveness and efficiency and based on concept of power, value appropriation and robust way of understanding (Cox, 1999).

Fresh-produce consumption has been favorably affected by demographic trends including declining household size, rising income levels, the consumption habits of baby boomers, and growth in the number of consumers. Better quality, increased variety, and year-round availability also have helped boost consumption. Packaging technology adds convenience in the form of precut, ready-to-eat products. In today's system, large retailers have increased their market shares by establishing competitive advantage in areas such as logistics (Hinson, 2005). SCM is a collection of management activities exercised between vertically related firms to improve efficiency, vertical coordination, and overall performance and competitiveness of the participating firms within an industry. Therefore, the research focused on structural change and its impacts in the fruits and vegetables industry has targeted the larger players in the industry, with less attention paid to the impacts of change on mid- and smaller-sized companies or their actions to effectively manage in this environment (Ricks et al., 1999).

Sustainable and wholesome supply of food and food products, availability and fulfilling the implied needs of customers are actually the areas to be focused rationally in global trading. Quality and food safety are the most integrated activities, embedded with each other. For agri-food products, quality may be regarded as a complex characteristic of foods that determines its value and acceptability by consumers.

ISBN 978- 974-672-556-9

Quality is the ongoing process and in case of horticultural commodities, quality cannot be improved after harvest, only maintained; therefore it is important to harvest fruits and vegetables at the proper stage and size and at peak quality. Hygiene and safety and nutritional and organoleptic sustainability accompanied by convenience to use and satisfaction and pleasure are the prime important parameters to be maintained. Supply chain management is the only paramount outlet, help managing the uniform quality and regular availability of safe food everywhere (Sackey et al., 2001).

Different scientists have done good contribution in this regard. For example, Talamini et al. (2005) studied about the growing concern about food safety and indicated that supply chains play predominant role in the productive process. SCM may contribute substantially toward a more uniform process throughout the chain, facilitating the sharing of information and productive practices. Stringer and Hall (2006) studied the generic model of the integrated food supply chain to aid the investigation of food safety breakdowns. He found that increasingly it is becoming important to study and manage various aspects of food production in relation to the total food chain implications and not one stage in isolation. Food safety management systems (CCFRA, 2003) are now built on a detailed understanding of all input variables to the manufacturing process. The internationally recognized approach for assuring food safety is HACCP (Hazard Analysis Critical Control Point), requires an intimate understanding of the interaction between process and product and the identification of critical control points in the manufacturing process, through distribution and storage and including consumer practices.

The progressive liberalization of world trade has created opportunities for Pakistan (Mustafa, 2008) to become integrated into the global trading system and to exploit its national and regional comparative advantages. Pakistan has a potential comparative advantage over developed countries in the production of many agricultural products, such as cotton, rice, fruits, vegetables, flowers etc. However, to maintain or expand its world market share and ability to meet the demands of the world trading system, quality of exportable products and their safety standards and Sanitary and Phytosanitary (SPS) measure (Sheikh et al., 2005).

Warriner (2009) found that food borne illness outbreaks linked to fresh produce are becoming more frequent and widespread. High impact outbreaks, such as that associated with spinach contaminated with Escherichia coli O157:H7 resulted in almost 200 cases of food borne illness across North America. Minami et al. (2009) conducted a study in Thailand (Bangkok and Pathum Thani provinces), from June 2006 to July 2007, in order to assess the prevalence of Listeria monocytogenes, Escherichia coli O157, Salmonella, Shigella and Vibrio parahaemolyticus in foods. Retail raw meats and seafood, including chicken (n = 109), pork (n = 80), beef (n = 108), shrimp (n = 43) and oysters (n = 48), from open markets and supermarkets were analyzed. Salmonella was found in 22 of 61 (36%) open market samples (48% of chicken, none of pork and beef, and 53% of shrimp) and in 12 of 75 (16%) samples from supermarkets (57%, 12%, 24%, 0% respectively). However, a small number of Listeria monocytogenes were isolated, where 6 of 217 (3%) were samples from open markets (6% of chicken and 3% of pork) and 17 of 171 (10%) were from supermarkets (3% of beef, 4% of chicken, and 32% of pork). In both markets, L. monocytogenes was not detected from shrimps, neither from oysters. E. coli O157, Shigella and the-positive V. parahaemolyticus were not isolated in this collection. Several Salmonella and L. monocytogenes isolates were multidrug-resistant (Boer, 1998).

Mergenthaler (2009) studied the implementation of different quality assurance programs (QAPs) and their impacts on international market for horticultural processing firms in developing countries. A representative survey of 50 registered firms processing horticultural produce in Vietnam was conducted. Logistic regression models were used to analyze the effects of QAPs and other firm level factors on participation in different export supply chains. It was found that QAPs influence firms' participation in export supply chains significantly. While the implementation of internationally recognized QAPs improves access to OECD markets, national QAPs are more important for exports to non-OECD countries. Similarly, Murphy et al., (2004) concluded that horticultural processing firms in developing countries can improve their access to export markets when they implement formal QAPs.

In a case of two supply-chain development projects in Thailand: TOPS Thailand and Fresh Partners: related to Introduction of food safety standards for the domestic market and Development of an integrated quality chain for the export market, respectively, Buurma and Saranark (2007) analyzed 50 supermarkets in Bangkok and Chiang Mai. The management decided to introduce a certification system for food safety in order to improve their competitive position and to consolidate their image of a quality supermarket. The introduction resulted in a system of preferred suppliers that had to obtain a certificate for good agricultural practices from the Department of Agriculture. The number of suppliers sharply dropped in the course of the project period. Fresh Partners Thailand is an export company shipping exotic vegetables from Thailand to The Netherlands and surrounding countries in Europe. The management decided to develop an integrated

quality chain in order to comply with the increasing food safety requirements in the European Union and Japan. The investments in quality systems coincide with a growing demand for exotic vegetables in northwestern Europe. Consequently export volumes and numbers of smallholders and laborers are rapidly growing.

III. METHODOLOGY

This research study was carried out in the Department of Information Systems, Corvinus University of Budapest. The main purpose of this study was to examine the current practices of quality assurance being applied in SCM of fruits and vegetables in Hungary and to chalk out the gaps and exploring and recommending the prospective quality-oriented business processes required for an integrated system from farm-gate to consumers end. The Figure 1 depicts the generic elements/ activities of SCM of fruits and vegetables.

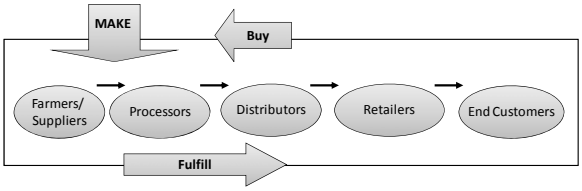


Figure 1. Elements of Fruit and Vegetables SCM (Buy: Purchasing of fresh fruits and vegetables; Make: Processing as finished products or packed as required; Fulfill : Ensuring that customers receive these produce)

In order to conduct this research work and to collect the primary data, this field study was carried out by visiting the organizations mentioned below, which are involved in the current system of supply chain of various fruits and vegetables in this country.

- Hungarian Interprofessional Organization (FruitVeB, Hungary) for Fruit and Vegetables, Budapest, Hungary.
- Garten Ltd. Budapest, Hungary.
- Callidatis Kft. Budapest, Hungary.

For the purpose to collect information, a questionnaire technique was applied by using pre-tested questionnaire and taking interviews from the management of these organizations and exploring the flaws available in current SC of fruits and vegetables. This technique also helped to explore and chalk out the intra and inter business processes gaps throughout the whole chain of these perishables and to measure the differences between AS-IS and TO-BE approaches, of each process involved.

ADONIS Tool is one of the best techniques used to design, develop, link and translate all the old and new processes, activities and sub-activities of SCM into integrated models. It helps to streamline these processes and to restructure the cooperation and supports effort and help reducing the cost. Through its intuitive user interface and as a powerful analyzing mechanism (such as simulation, personnel resource planning, activity-based costing), it emerges as an ideal corporate business process management tool (ADONIS, 2009).

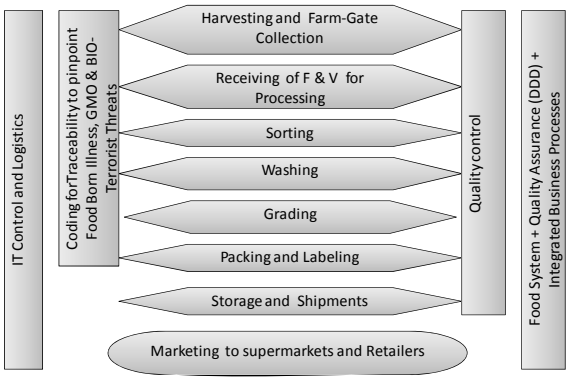


Figure 2. Flow chart – Business processes of SCM of Fruits and vegetables

As the pivotal tracking role of this part of study was to bridge the gaps present in the current quality assurance business process in the form of AS-IS (Field Study) and to devise and suggest the most appropriate and interacted link/ model in the form of TO-BE Approach for improvement and more compatibility.

IV. RESULTSAND DISCUSSIONS

Presently, food markets and food trading are becoming globalised and more and more concentrated and internationalized, but at the same time, quality assurance and traceability requirements against suppliers have also increased significantly. Increased food safety requirements are now expected to exercise more strongly on fresh products due to the fact that food safety and quality components have an outstanding role with respect to fresh products supplied through marketing channels. Similarly, more safe food demands, traceability and inspection in the developed countries are higher than in countries of medium level and developing countries.

Food safety and quality requirements have an increasing importance in international trade. Dissemination and introduction of the quality assurance systems such as GLOBALGAP, International Food Standard (IFS), Hazard Analysis Critical Control Point (HACCP) etc., as an urgent task, have become integral parts for both production and trade. Without the realization of these systems, efficient marketing can be performed neither on the domestic nor on the international market (Tomcsanyi, 1999).

The new age consumers are becoming more health conscious in terms of hygiene, source of the food, ingredients of processed food and caloric content (Loader and Hobbs, 1999). Food safety has become a primary concern across the world (Crespi

and Marette, 2001). The food industry has become vast and diversified, categorized by different segments such as fresh food industry, organic food industry, processed food industry, livestock food industry, etc. All these segments are unique in nature and are interlinked with each other. Each segment needs different supply chain solutions such as procurement and sourcing, inventory management, warehouse management, packaging and labeling system, distribution management etc. (Kalei, 2008).

Therefore, the relation of quality assurance with supply chain management is vital important to develop a food system and sustainable marketing. In this study, application of both physico-chemical analysis (colour, flavor, size or volume or weight, peel thickness, taste, internal pressure of fruits, percent acidity and juice yield of fruits and damage occurred during transportation, presence/ absence of fruit fly eggs etc.) and microbiological quality control for prevalent microorganisms such as Total Plate Count (TPC) for bacteria, Escherichia coli, Salmonella, Listeria monocytogenes, fungi etc. and quality assurance standards such as Hazard Analysis Critical Control Point (HACCP), International Food Standards (IFS) and GLOBALGAP, which are prime important to develop integrated system of supply. Similarly, the information regarding physical treatments of various fruits and vegetables such as Vapour Treatment (VHT), hot water treatment including time and temperature, freezing treatment of citrus fruits for particular time and temperature and quarantine certificate required for export are those phrases or information which would help generating and transforming different questions related to quality and sustainability of this system (EC., 2002).

Quality assurance (QA) is a process to add reliability in the products or services. The Institute of Food Science and Technology (IFST, 1998) defines QA as a process to design and plan, as relevant, raw material specifications, ingredients formulation, processing equipment and environment, processing methods and conditions, intermediate specifications, appropriate packaging and labeling specifications, specifications for quantity.

QA can be a tool to demonstrate regulatory compliance; business efficiency tool to ensure product quality and minimise hygienic risks and communication tool to customers and consumers, globally (Baines and Ryan, 2002) and characterized as the assurance wholesome of a product by means of a system which will manage quality and the product or the prevention of quality problems through planned and systematic activities. Therefore, Manning (2006) delineated that quality Assurance standards are considered to be a proven mechanism for delivering quality of product and service. He analysed critically how effectively this mechanism has been implemented in the integrated food supply chain and concluded that compliance with a set of extrinsic quality standards are pre-requisites in determining the mechanism of supply with particular emphasis on poultry meat.

Quality and health risks are inversely proportional to each other. Stricter the quality, the lesser will be the risks. Owing to perishability, the fruits and vegetables are also inflicted with many microorganisms, if post harvest management practices starting from harvesting or picking, keeping, cooling, loading, transportation, receiving, washing, packaging, storage etc. are not applied appropriately. Mishandling, bruising and poor transport and cold store facilities etc., invite the various pathogenic microorganisms to attack which trigger the conditions for health risk to create in the supply chain and ultimate decline in business.

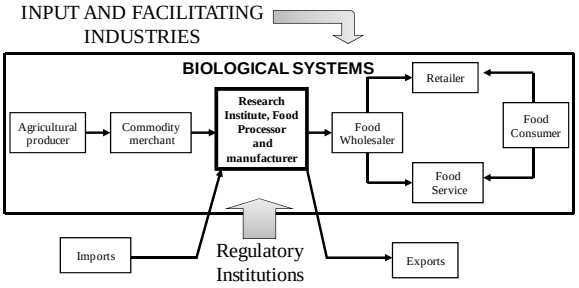


Figure 3. Concept of Food System in SCM

Food system (Figure 3) is an integrated approach which helps to provide food to the consumers table by involving all the stakeholders such as suppliers, research institutions, processors, traders, government agencies and regulations, importers, exporters, wholesalers and customers. When this system is packed up and bounded with quality assurance standards (Due Diligence Defence or 3D) and integrated business processes, then it is named as safe food system or supply chain management (Figure 4). Safety and reliability in supply chain and afterward in the product while purchasing and using, can only be possible by applying the shield of Due Diligence Defence. This 3D approach also helps to develop a brand image of an organization, marketing and ultimate customer satisfaction and good profit.

Quality assurance systems in the food industry have developed from generic standards, such as ISO 9000 and HACCP, to systems specific to the food industry and its subsectors. New national and international rules require increased vigilance on the traceability and quality of inputs processed. Examples are the Dutch QC standard (Quality Control of Feed Ingredients for Animal Feed), the British Retail Consortium (BRC) technical standard for companies supplying retailer branded food products, the Global Food Safety Initiative (GFSI), and the German Quality and Safety (Qualität und Sicherheit GmbH, QS) standard.

These systems build on quality management principles such as the ISO 9000 system and the HACCP concept. They involve quality management principles in processing and distribution and good practices for agriculture. In parallel to the BRC standard, the German retail sector is currently developing an international standard for auditing of private labels products. By outsourcing the auditing according to this standard by third-party certifiers, the German retail industry hopes to reduce the need for and the costs involved in individual audits of private label manufacturers.

In order to restore consumer confidence in meat products, the German meat sector has created the QS label. This label involves firms starting from farm suppliers over individual farms, slaughterhouses, and meat processors up to retailers. Launched after the BSE crisis in 2000/01, the QS system introduces a system of process control and traceability for meat and meat products. The QS system requires documentation and consists of three layers of control: self-auditing, firm audits by independent auditors, and accreditation of auditors. As of March 2003, 869 companies have enrolled in the QS system (QS GmbH, 2003).

However, Critics claim that standards fixed within the QS system do not go far enough beyond legal requirements. The system as such does not lead to a differentiated, high-quality product. The recent case of dioxin contaminated feed in eastern Germany has shown that the QS system leaves gaps in the supply-chain management of quality. While the identified feed producer in Thuringia was not certified under the QS system, he delivered feed ingredients to feed mills participating in QS. Those participating feed mills only rely on self audits and few regulatory controls (Handelsblatt, 2003). Additional problems relate to the fact that the system tries to cover all actors of the meat production chain. As a result, vigilance may not be at its best and moral-hazard problems may reduce care-taking actions of participating firms. Due to its wide scope, the differentiation of products does not seem possible and because of free-rider problems, the investment in reputation may be at risk.

Garetn Ltd, being a quality conscious pack house also uses to put continuous efforts to practice and maintain the spirit of all the related quality assurance systems in their four accredited cold stores in order to develop competitive strength and deterrence. For example, Garten Cold Store, Tiszakecske-tiszabog has been accredited for International Food Standards (IFS) and GLOBALGAP certificates in 2007, Garten Cold Store, Nyirtelek-Gyulatanya for HACCP and GLOBALGAP certificates in the same year and Taszsagi-Mesz Cold Store and Szentesi-Zoldker Cold Store for GLOBALGAP certificates both in 2009. Figure 5 shows the details of all the unit operations involved in quality assurance and quality control during processing and packing of fresh fruits and vegetables. As, quality control is the part of quality assurance, which helps to add excellence in product by analyzing both physic-chemical and microbiological quality during processing. So Garten Ltd. needs to establish these laboratories within their units instead of getting samples analysed from out-side laboratories to get consistent analysis, in all its cold stores as a part of TO-BE unit operation in quality assurance business process.

V. DISCUSSION

Fruits and vegetables processing, trading and marketing require some vital support of integrated supply chain management system. Due to its perishable nature and more consistent application of quality standards, the organizations involved in business of these horticultural commodities have to face colossal quality problems. Therefore, one of the purposes of this research study was to collect information regarding current processing practices and quality attributes applied. For this purpose, an exploratory and qualitative research study was carried out by using the pre-tested questionnaire and to collect information from various organizations dealing in supply chain of these horticultural commodities. On the bases of this study the current situation of supply chain was analyzed which helped to chalk out the gaps present in current (and the and also helped to analyze and measure the gaps between AS-IS and TO-BE approaches).

It was concluded from this study that some of the unit operations of quality assurance in SCM were not applied and practiced completely which result in quality gaps throughout the whole chain. Therefore, in order to develop closely integrated system of supply chain of fruits and vegetables and subsequently, bridging the gaps between current AS-IS and TO-BE approaches of SCM, there is a dire need to implement the quality management system from harvesting through consumer's end in order to develop safe food management system.

VI. RECOMMENDATIONS

Owing to perish ability, increasing trend and global demands of fruits and vegetables, the integrated approach of supply chain from farm-to-fork is the only way out to fulfill the implied needs and expectations of customers. In this study, a new method was developed by using mind mapping technique, in order to collect information, a field study was carried out by using this methodology and then ADONIS tool was applied to chalk out the gaps and also to develop the prospective business processes which were required to help in developing the holistic approach of supply chain management of fruits and vegetables. After completing, the following

recommendations based on the findings of this study were made, which are discussed below:

A. Policies for strategic competitiveness

The growers of fruit and vegetables must be restricted strongly to purchase plant nursery of fruits and seeds of vegetables from the certified nurseries and suppliers.

Effective participation by the government authorities in international standard setting (e.g., participation in international, regional and national codex committees) is a must to develop a base and establish a platform of SCM of fruits and vegetables in Hungary.

A regional level database of cases/ experiences of health risks due to food borne diseases must be maintained by the agricultural statistics department for information which could be used in risk assessment and standard setting process both by concerned government agencies and pack houses.

In order to have an effective SCM, capacity building for compliance of GLOBALGAP, IFS and HACCP is highly important to translate the specifications to last level of an organization. In this regard, Government of Hungary must strengthen the related national and regional institutions, in order to provide training of various standards and bar coding, help to conduct risk analysis and testing and monitoring and enforcement of standards and also to carry out certification.

As an effective SCM triggers with planning from planting-to-produce, best pre-harvest practices such as application of GLOBALGAP must be made mandatory for each grower, for ultimate produce traceability. Otherwise, fresh produce pack houses of competitive countries of EU like Spain, Holland, Italy etc., are intruding more in national market, creating problems for local growers and sweeping out their market benefits due to better pre and post harvest management practices and supply chain system.

B. Harvest management practices

The essence of supply chain management is to reduce the lead time to appropriate level from harvesting/ picking through packing and marketing of these horticultural commodities. Whereas, delayed harvesting, transportation and arrival at various cold stores or pack houses for further processing, affect adversely on the quality of fruits and vegetables due to the presence of field heat for a longer period. So refrigeration transport or efficient field heat removal technique must be applied immediately by pack houses or large growers at farm level after harvest, to prevent any damage to crop.

C. Sustainable availability of risk free produce

Continuous quality control-push by the traders and supermarkets and more implications of quality standards, making this business more vulnerable and complex. So in order to establish deterrence and sustainable supply chain system, the pack houses must be more committed towards application of all required quality standards and instilled quality control laboratories in their premises, for routine physico-chemical analyses including pesticide residues and sensory evaluation, true to spirit, as a part of TO-BE approach in SCM. As an integral component of SCM, this vital activity will also help tracing back any health risk if it could occur in whole supply chain.

Integrated SCM also demands for all the business partners such as suppliers, processors and customers to sit together and join heads twice in a year, to cooperate and make future developments in order to target market needs and promotion of their products and branding.

REFERENCES

[1] Adebanjo, D. (2009). Understanding demand management challenges in intermediary food trading: a case study. Supply Chain Management: Intl. J. 14(3):224–233.

[2] ADONIS (2009). ADONIS, Version 4.0, Method Manual - BPMS Method, BOC, Wien

[3] Barratt, M. (2004). Understanding the Meaning of Collaboration in the Supply Chain. Supply Chain Management: Intl. J. 9(1):30–42.

[4] Biacs, P. (1999). Food quality and safety. Agro-21 Fuzetek, Az agargazdasag jovokepe. 66-70.

[5] Boer, E. (1998). Update on media for isolation of Enterobacteriaceae from foods, Intl. J. Food Microbiology, 45 (1):43-53.

[6] Buurma, J. and J. Ssaranark (2007). Supply-chain development for fresh fruits and vegetables in Thailand # lei, wageningen university and research centre, p.o. Box 29703, 2502 ls, den haag, the Netherlands.

[7] CCFRA (2003). HACCP: A Practical Guide. Guideline No. 42, 3rd Ed., Campden and Chorleywood Food Research Association Group.

[8] Cox, A. (1999). Power, value and supply chain management. Supply Chain Management: Intl. J. 4(4):167–175.

[9] Crespi, J.M. and S. Marette (2001). How should food safety certification be financed?, American J. Agril. Econ. 83(4):852-861.

[10] EC (2002). Laying down the general principals and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Logistics Integration Processes in the Food Industry. Regulation (EC) No. 178/2002 of the European Parliament and of the Council.

[11] FAO (2009). Statistical Yearbook of Food and Agriculture Organization of United Nations.

[12] FruitVeB. (2008). Annual Report of Hungarian Fruit and Vegetable Sector. Hungarian Interprofessional Organization for Fruit and Vegetables, Budapest, Hungary.

[13] Goldman, I.L., A.A. Kader, and C. Heintz (1999). Influence of production, handling, and storage on phytonutrient content of foods. Nutrition Reviews. 57(9):S46-S52.

[14] Grimsdell, K. (1996). The supply chain for fresh vegetables: what it takes to make it work. Supply Chain Management: Intl. J. 1(1):11–14.

[15] Handelsblatt (2003). Dioxin-Skandal deckt Schwachstellen beim QS Prüfsiegel auf. 11 March: 16.

[16] Herlambang, T., P.J. Batt and M.J. McGregor (2009). Developing an Effective Food Chain Management in a

Developing Country: A Case Study on Manalagi Mango Fruit Supply Chain in Indonesia.
[17] Hicks, A. (2002). Minimum Packaging Technology for Processed Foods: Environmental Considerations. AU J.T. 6(2): 89-94. FAO Regional Office for Asia and the Pacific Bangkok, Thailand.
[18] Hughes, D. and I. Merton (1996). Partnership in produce: the J Sainsbury approach to managing the fresh produce supply chain. Supply Chain Management: Intl. J. 1(2):4–6.
[19] IFST (1998). Annual Meeting. Institute of Food Scientists and Technologists. CA. USA
[20] József L., B.I. Csaba, N. Komáromi and Z. Lehota (2009). Development of Traceability in Hungarian fresh vegetable and fruit sector. 4th Aspects and Visions of Applied Economics and Informatics. March 26 - 27. Debrecen, Hungary.
[21] Kalei S.N. (2008). Supply Chain Management in Food Industry. Ifai University Press.
[22] Knight, C., R. Stanley and L. Jones (2002). Agriculture in the food supply chain: an overview Key Topics in Food Science and Technology, No. 5. Campden and Chorleywood Food Research Association Group.
[23] Lee, S.K. and A.A. Kader (2000). Pre-harvest and post harvest factors influencing vitamin C content of horticultural crops. Postharvest Biol. Technol. 20:207-220.
[24] Loader, R. and J.E. Hobbs. 1999. Strategic Responses to Food Safety Legislation. Food Policy. 24(6):685-706.
[25] Manning, L., R.N. Baines and S.A. Chadd (2006). Quality assurance models in the food supply chain. Royal Agricultural College, Cirencester, UK. British Food J. 108(2):91-104.
[26] Mergenthaler, M., K. Weinberger and M. Qaim (2009). Quality assurance programs and access to international markets: the case of horti;cultural processors in Vietnam. Supply Chain Management: Intl. J. 14(5):359–368.
[27] Minami, A., W. Chaicumpa, M.C. Nguan, S. Samosornsuk, S. Monden, K. Takeshi, S. Makino and K. Kawamoto (2009). Prevalence of foodborne pathogens in open markets and supermarkets in Thailand. Available online May 27.
[28] Murphy, R.Y., B.L. Beard, E.M. Martin, A.E. Keener and T. Osaili (2004). Predicting process lethality of Escherichia coli O157:H7, Salmonella, and Listeria monocytogenes in ground, formulated, and formed beef/turkey links cooked in an air impingement oven, Food Microbio. 21(5):493-499.
[29] Mustafa, K. (2008). Barriers Against Agricultural Exports from Pakistan: The Role of WTO Sanitary and Phytosanitary Agreement. Associate Professor/Chairman, Department of Agricultural Marketing, University of Agriculture, Faisalabad. Email: drkmustafa@hotmail.com
[30] Prior, R.L. and G. Cao (2000). Antioxidant phytochemicals in fruits and vegetables; diet and health implications. Hort. Science 35:588-592.
[31] Produce for Better Health Foundation (1999). Dietary guidelines: the case for fruits and vegetables first. Produce for Better Health Foundation, Wilmington, DE (Website: http://www.5aday.com).
[32] QS GmbH. (2003). Participation in the QS System. Fax, 11 March 2003.
[33] Quebedeaux, B. and F.A. Bliss (1988). Horticulture and human health. Contributions of fruits and vegetables. Proc. 1st Intl. Symp. Hort. and Human Health. Prentice Hall, Englewood, USA.
[34] Quebedeaux, B. and H.M. Elisa (1990). Horticulture and human health. Contributions of fruits and vegetables. Proc. 2nd Intl. Symp. Hort. and Human Health. Hort Sci. 25:1473-1532.
[35] Ricks, D, T. Woods and J. Sterns (1999). Supply Chain Management: Improving Vertical Coordination in Fruit Industries. J. Food Distribution Res. 30(3):44–53.
[36] Sackey, B.A., P. Mensah, E. Collison and E.S. Dawson (2001). Campylobacter, Salmonella, Shigella and Escherichia coli in live and dressed poultry from metropolitan Accra. Intl. J. Food Microbiology. 71(1):21-28.
[37] Sheikh, B.A., S.A. Sheikh and G.H. Soomro (2005). Pakistan Agriculture in Global Perspective. Review Paper. Pak. J. Agri., Agril. Engg., Vet. Sc. 21(2): 23-29.
[38] Southon, S. (2000). Increased fruit and vegetable consumption within the EU: potential health benefits. Food Res. Intl. 33:211-217.
[39] Spriggs, J. (1999). What is in the world is going on in food safety. Paper presented at the Food Safety and International Competitiveness Conference, April. Alberta.
[40] Stringer M. F. and M. N. Hall (2006). The Breakdowns in Food Safety. Food Technology Division, Campden and Chorleywood Food Research Association, Chipping Campden, Gloucestershire, GL55 6LD, UK.
[41] Talamini, E., E. A. Pedrozo and A. L. Silva (2005). Supply chain management and food safety: exploratory research into Brazil's pork export supply chain. Gest. Prod. 12(1):107-120.
[42] Tomas, B.F.A. and R.J. Robins (1997). Phytochemistry of fruits and vegetables. Oxford Science Publ., Oxford, UK : 375.
[43] Tomcsanyi, P. (1999). Building market and product quality. Agro-21 Fuzetek, Az agargazdasag jovokepe.:72-78.
[44] Trienekens, J.H., J.M. Hagen, A.J.M. Beulensc and S.W.F. Omta (2003). Innovation Through (International) Food Supply Chain Development: A Research Agenda. Intl. Food Agribusiness Management Review. 6(1):1-6.
[45] Unnevehr, L.J. (2000). Food safety issues and fresh food products export from LDCs. Agril. Econ., 23:231-240.
[46] Van der Vorst J. G. A. J. (2004). Performance levels in food traceability and the impact on chain design: results of an international benchmark study. Dynamics in Chains and Networks. Wageningen Academics Publishers, Wageningen.
[47] Wargovich, M.J. (2000). Anticancer properties of fruits and vegetables. Hort. Sci. 35:573-575.
[48] Warriner, K., A. Huber, A. Namvar, W. Fan and K. Dunfield (2009). Recent Advances in the Microbial Safety of Fresh Fruits and Vegetables. Advances Food Nutrition Res. 57:155-208.
[49] Wysocki, A.F. (2005). A Frictionless Marketplace Operating in a World of Extremes. The Magazine of Food, Farm and Resources. 4 (20):45-47.

Supply Chain in Teak Furniture Business

Sakda Phuatngam
College of Arts, Media and Technology, Chiangmai University, Thailand
sakda1234@yahoo.com

.

***Abstract* — Teak wood Furniture SMEs Firm has the competition at most. All strategy had applied in the organization for the most profit, including the supply chain management and innovation management system had applied in the organization. Problem: Knowledge management, Supply Chain management, Innovation management Organizations increasingly find that they must rely on effective supply chains, or networks, to successfully compete in the global market and networked economy. In Peter Drucker's (1998) new management paradigms, this concept of business relationships extends beyond traditional enterprise boundaries and seeks to organize entire business processes throughout a value chain of multiple companies. During the past decades, globalization, outsourcing and information technology have enabled many organizations, such as Dell and Hewlett Packard, to successfully operate solid collaborative supply networks in which each specialized business partner focuses on only a few key strategic activities (Scott, 1993). This inter-organizational supply network can be acknowledged as a new form of organization. However, with the complicated interactions among the players, the network structure fits neither "market" nor "hierarchy" categories (Powell, 1990). It is not clear what kind of performance impacts different supply network structures could have on firms, and little is known about the coordination conditions and trade-offs that may exist among the players. From a systems perspective, a complex network structure can be decomposed into individual component firms (Zhang and Dilts, 2004). Traditionally, companies in a supply network concentrate on the inputs and outputs of the processes, with little concern for the internal management working of other individual players. Therefore, the choice of an internal management control structure is known to impact local firm performance (Mintzberg, 1979).**

In the 21st century, changes in the business environment have contributed to the development of supply chain networks. First, as an outcome of globalization and the proliferation of multinational companies, joint ventures, strategic alliances and business partnerships, there were found to be significant success factors, following the earlier "Just-In-Time", "Lean Manufacturing" and "Agile Manufacturing" practices. Second, technological changes, particularly the dramatic fall in information communication costs, which are a significant component of transaction costs, have led to changes in coordination among the members of the supply chain network (Coase,1998).

Many researchers have recognized these kinds of supply network structures as a new organization form, using terms such as "Keiretsu", "Extended Enterprise", "Virtual Corporation", "Global Production Network", and "Next Generation Manufacturing System" In general, such as tructure can be defined as "a group of semi-independent organizations, each with their capabilities, which collaborate in ever-changing constellations to serve one or more markets in order to achieve some business goal specific to that collaboration" (Akkermans, 2001). If Knowledge management, Supply Chain management and Innovation management applied in furniture SMEs firm. Good managements can solve a life products cycle problem and New Competitive Products became productivity.

Index Terms — Innovation Management; Supply chain Management; Teak Furniture

I. INTRODUCTION

In 1937 government support to grow teak forests. Now teakwood from grow teak forests more than ten millions trees able to cut down for teaks furniture SMEs firms. Solving problem a lot of teak wood from grow teak forests already to cut down. Teak wood furniture for export, The way to solving problem and support the teak wood furniture SMEs firms. Teak wood stands out from other kinds of wood with its distinctive natural patterns and golden brown color. Teak wood is a hardwood so it is durable and free from termites and weevils, because it contains a natural substance call "o-cresyl methyl ether" that was discovered by the Thai Department of forestry to have an anti-insect and anti-fungus effect. it is, however, very easy to cut into shapes, so teak wood is one of the most ideal raw materials for furniture production. (Thailand International Furniture Fair,2008).

Now Teak wood over 25 years of age are cut down and replaced by new trees more than ten millions trees. Effective supply chain management (SCM) has become an emergent strategy for Teak furniture SMEs firms to secure competitive advantage, add value and improve organizational performance in today's highly volatile global market. Supply chain management (SCM) is an important issue facing any organizations worldwide. SCM is an important area that helps maximize competitiveness and profitability for the company as well as other supply chain members which integrate and coordinate across their whole extended network (Lambert and Cooper,1998). Managing the supply chain has become a way of improving competitiveness by reducing uncertainty and enhancing customer service.

Not all Teak SMEs firms are successful in achieving a higher level of performance even if they have implemented the SCM concepts. In order for

firms to succeed, survive, and sustain their competitive positions, certain characteristics of SCM have to exist. This research tries to identify the characteristics that bring or create strategic advantages to Thai manufacturers as well as help them sustain their competitive advantages and maximize performance. In the long run, these SCM characteristics add value to the Teak SMEs firms.

II. LITERATURE REVIEW

Value chain describes the full range of activities which are required to bring a product or service from conception or design, through different phases of production, delivery to final consumers, and final disposal after use. Value chain provides a systemic understanding of the production process of a particular good. With the growing division of labor and the global dispersion of the production of components, systemic competitiveness has become increasingly important.(Kaplinsky and Morris,2000).

Value chain analysis is actor oriented. It is very effective in tracing product flows, showing the value adding stages, identifying key actors and the relationship with other actors in the chain. Often, however, these actors operate within certain rules that are set by others. (Schmitz, 2005).

Participation in global markets is not just governed by trade policies in final market countries. It also reflects the strategic decision of the lead firms in the value chain. Gareffi (2003) constructed the theory of value chain governance based on(1) Complexity of transactions; (2) Ability to codify transactions; and (3) Capability of the supply-based. If these factors are valued by ‘high’ and ‘low’ then there are eight possible combinations, of which five are actually found, i.e., market, modular, relational, captive and hierarchy. The emergence of an impressive variety of voluntary regulation systems is a key element of the globalization process. These systems are normally meant to set quality, social, or environmental standards, and typically involve a larger degree of coordination, traceability, and monitoring along different agents of the commodity chain (Muradian and Pelupessy, 2005).

Knowledge is action, focused innovation, pooled expertise, special relationships and alliances. Knowledge is value-added behavior and activities. For knowledge to be of value it musts be focused, current, tested and shared (Newman, 1991).

Popular Visualization

The value chain, also known as value chain analysis, is a concept from business management that was first described and popularized by Michael Porter in his 1985 best-seller, *Competitive Advantage: Creating and Sustaining Superior Performance*.

A value chain is a chain of activities. Products pass through all activities of the chain in order and at each activity the product gains some value. The chain of activities gives the products more added value than the sum of added values of all activities. It is important not to mix the concept of the value chain with the costs occurring throughout the activities. A diamond cutter can be used as an example of the difference. The cutting activity may have a low cost, but the activity adds much of the value to the end product, since a rough diamond is significantly less valuable than a cut diamond. Typically, the described value chain and the documentation of processes, assessment and auditing of adherence to the process routines are at the core of the quality certification of the business, e.g. SO 9001.

The value chain categorizes the generic value-adding activities of an organization. The "primary activities" include: inbound logistics, operations (production), outbound logistics, marketing and sales (demand), and services (maintenance).The "support activities" include: administrative infrastructure management, human resource management, technology(R&D), and procurement. The costs and value drivers are identified for each value activity. The value chain framework quickly made its way to the forefront of management thought as a powerful analysis tool for strategic planning. The simpler concept of value streams, a cross-functional process which was developed over the next decade, had some success in the early 1990s.

The value-chain concept has been extended beyond individual organizations. It can apply to whole supply chains and distribution networks. The delivery of a mix of products and services to the end customer will mobilize different economic factors, each managing its own value chain. The industry wide synchronized interactions of those local value chains create an extended value chain, sometimes global in extent. Porter terms this larger interconnected system of value chains the "value system." A value system includes the value chains of a firm’s supplier (and their suppliers all the way back), the firm itself, the firm distribution channels, and the firm's buyers (and presumably extended to the buyers of their products, and soon).

Capturing the value generated along the chain is the new approach taken by many management strategists. For example, a manufacturer might require its parts suppliers to be located nearby its assembly plant to minimize the cost of transportation. By exploiting the upstream and downstream information flowing along the value chain, the firms may try to bypass the intermediaries creating new business models, or in other ways create improvements in its value system.

Value chain analysis has also been employed in the development sector as a means of identifying

poverty reduction strategies by upgrading along the value chain. Although commonly associated with export-oriented trade, development practitioners have begun to highlight the importance of developing national and intra-regional chains in addition to international ones.

The Supply-Chain Council, a global trade consortium in operation with over 700 member companies, governmental, academic, and consulting groups participating in the last 10years, manages the Supply-Chain Operations Reference(SCOR), the *de facto* universal reference model for Supply Chain including Planning, Procurement, Manufacturing, Order Management, Logistics, Returns, and Retail; Product and Service Design including Design Planning, Research, Prototyping, Integration, Launch and Revision, and Sales including CRM, Service Support, Sales, and Contract Management which are congruent to the Porter framework. The *SCOR* framework has been adopted by hundreds of companies as well as national entities as a standard for business excellence, and the US DOD has adopted the newly launched Design-Chain Operations Reference (DCOR) framework for product design as a standard to use for managing their development processes. In addition to process elements, these reference frameworks also maintain a vast database of standard process metrics aligned to the Porter model, as well as a large and constantly researched database of prescriptive universal best practices for process execution.

Value Reference Model

A Value Reference Model (VRM) developed by the global not for profit Value Chain Group offers an open source semantic dictionary for value chain management encompassing one unified reference framework representing the process domains of product development, customer relations and supply networks. The integrated process framework guides the modeling, design, and measurement of business performance by uniquely encompassing the plan, govern and execute requirements for the design, product, and customer aspects of business.

The Value Chain Group claims VRM to be next generation Business Process Management that enables value reference modeling of all business processes and provides product excellence, operations excellence, and customer excellence.

Six business functions of the Value Chain:

- Research and Development
- Design of Products, Services, or Processes
- Production
- Marketing & Sales
- Distribution
- Customer Service

The rapidly changing environments, coupled with the move of competition from “among organizations” to “between supply chains”, have resulted in a great challenge in managing an effective SCM practice in the hope of reducing supply chain costs, securing competitive advantage and improving organizational performance. Kuei, Madu, andLin (2001) present an empirical study of the relationship between supply chain management and operational performance. Li et al. (2006) supported the belief that SCM practices have direct impacts on competitive advantage and organizational performance. Their findings indicated that SCM practices produce competitive advantage to the organization which will, in turn, lead to improved organizational performance. Chu (2006) found that communication is one of the factors that directly and positively affects trust in supply chain partnership. Simatupang, Wright, and Sridharan (2002) found that information technology.

Knowledge is the basis for, and the driver of, our post-industrial economy. Knowledge is the result of learning which provides the only sustainable competitive advantage. Knowledge is the next paradigm shift in computing following data processing 1945-1965 and information management 1966-1995. Knowledge is action, focused innovation, pooled expertise, special relationships and alliances. Knowledge is value-added behavior and activities. For knowledge to be of value it musts be focused, current, tested and shared (Newman, 1991).

During the past many years, Moreover, significant recognizing of most organizations that ‘Knowledge’ is a powerful strategic that gives them sustainable competitive advantage (Drucker, 1993). Thus, now they are finding out the best ways to manage knowledge with suitable in their organization. KM is an effective way in organization to encourage systematic creation, sharing and using knowledge.

In the context of knowledge management which is a business process through which a firms create and use its collective knowledge involving the identification and analysis of available and required knowledge, and turns such knowledge into planning and control of actions to fulfill organizational objectives (Civi,2000). A knowledge management system therefore plays an important role in setting up business strategy of a firms. This is supported by Amaravadi *et al.*(1995) who introduced the concept of intelligent marketing information systems (IMkIS) as a decision-making aid for managers. An IMkIS addresses the need to share marketing expertise in a competitive marketing environment. It is an extension of Kotler’s marketing decision support system that it uses techniques of knowledge representation from the field of artificial intelligence. The distinguishing feature of IMkIS is its knowledge base, which contains knowledge about variousaspects of marketing. In IMkIS concept, there are eight major groups within an organization which impact on marketing decisions: online databases, domestic and international planning, product/price management,

promotion, public information, distribution and cost management. With these components, an IMkIS can assist in the successful assimilation of knowledge among the different segments of an organization, resulting in knowledge transfer from successful segments of the company to the newer segments. The knowledge-based component of the IMkIS enables marketing expertise to be accessed and shared easily.

The organizational knowledge creation model of Nonaka &Takeuchi (1995) underlies the principles of Knowledge Management. According to their "Socialization -Externalization - Combination - Internalization" (SECI)model, the continuous and dynamic interaction between tacit and explicit knowledge that happens at the level of the individual, the group, the organization, and between organizations can be significant for the *sustainable development of knowledge*. Socialization is the process of creating tacit knowledge from tacit knowledge, whereas externalization is that of articulating tacit knowledge into explicit concepts. Combination involves the process of systemizing concepts into an explicit knowledge system. Internalization is the process of embodying explicit knowledge into tacit knowledge. Nonaka et al (2003) also suggest that at the foundation of this modeling lies *ba*, the context in which knowledge needs to exist, in which it is shared, created, and utilized.

III. Research Methodologyand Program

The aim and objectives of the research are:

- To maintain and protect existing knowledge for Teakwood furniture production.
- To raise awareness for the value of Teak wood product.
- To create best-practice knowledge management model for Teak wood SMEs firms.

At the exploration phase, the objective is to explore and survey purposive knowledge about Teak wood furniture production; Cutting, Turn off edge, Puncture, Grooving, Carvings from primary source, secondary source and interview. The knowledge of social value, belief and attitude of Thai people for teak wood furniture, development of their personality using traditional Thai culture based, temple based, and local wisdom will be investigated. The key knowledge will be identified, observe, and capture from key sampling with experts The data collection will be also from expert people for made teak furniture , book, and internet. The process of data collection will apply knowledge management SECI model. The semi-structure interview questions will explore both verbal and nonverbal signs. Data interpretation will be analyzed and validated.

The model development phase: Knowledge management systems for personality enhancement model and training programs will be developed. Proper communication channels will be implemented to disseminate the knowledge findings.

In this research study, both qualitative and quantitative methods will be applied.

Justification and motivation of methodology (Novelty): Knowledge Management Methodology and personality theory will be applied for examples:

Knowledge Management Model

SECI Model

The SECI model and COP theory will be employed from capture, analysis, validation, modeling, sharing, and learning the new knowledge creation for personality enhancement model for SMEs firm. The knowledge of Thai traditional and local wisdom has many important special tacit characteristics, which requires exploring and involving both explicit and tacit natures. According to Nonaka (1998, 2000), the transformation of knowledge, SECI Model, has four phases and the process will repeat like a spiral. Some tacit can be transferred straightforwardly to other people as tacit knowledge. Tacit knowledge can also be expressed and externalized into explicit knowledge, which can be combined and therefore new explicit knowledge can be produced. Explicit knowledge should be internalized and become one's own tacit knowledge before it can be applied and generated. All these transfer and transformation of knowledge creation produce within appropriate *Ba*, which are physical and virtual places, tools and environment facilitating the learning processes.

Knowledge Model

Knowledge Model tool using CommonKADS methodology and knowledge mapping to capture, analyze, model and store best practice of domain knowledge from experts. The session may be semi-structured interviews, observations of the expert's personality both verbal and nonverbal characters, and follow-up interview. All processes are audio taped and videotaped for transcription and in-depth analysis. This type of knowledge acquisition is useful for detailed knowledge to be encoded in the knowledge-based system (KBS) such as best-practice and role-model expert systems.

Types of knowledge and stored.

- Declarative knowledge
  - Text files, elements of diagram
- Procedural knowledge
  - Sequence, text diagram, procedure program
- Semantic knowledge
  - Articulated diagram
- Episodic knowledge
  - Animation, Simulation and Video

Learning in Action

Learning in action (Garvin, 2000) is an important process for adaptive training design. Human learning requires inquisitiveness and openness to tackle conventional wisdom and routines as accepted for "the way people do things."People's characteristics and personality are both explicit and implicit. They are invoked without conscious thought. For effective personality training model, acquiring information at multiple points in multiple ways and encouraging conversation to maximize learning will be approached to foster effective learning. In this research study, learning in action will be categorized into 2 types, which are must do list and intelligent learning. The researcher will first explore, observe and set the effective requirement criteria for personality enhancement programs as "must do lists." Then, the best-practice from role models will be searched and identified.

The research study will be continuing for three-year divided into 5 different phases as follows:

The first stage, proposal stage, the research problems are explored and identified. Literature review will be searched for. Interview questions will be established for pre-test of problem statement and ready to be surveyed for the data collection phase. The final proposal and qualifying examination needed to be done.

The second stage will be data collection phase. The appointment will be scheduled for the key target interview. During this stage face-to-face interview will be conduct. Participation in seminar and meetings for additional relevant information search will be explored.

The third stage is the analyzing and interpretation phase. Knowledge validation will be done at the end of this stage.

The fourth stage is the model development phase from knowledge identification, knowledge creation and acquisition, knowledge codification and refinement to knowledge employment and verification.

The fifth stage is the knowledge access and sharing through conference paper for presentation and final thesis.

IV. ANALYSIS

1. Demand of raw material find now teak wood from grow teak forests more than ten million units able to use for teak furniture SMEs firms.(Administration of Phrae province)

2. In this case study find Special of golden teak wood raw material value-added because Teak wood stands out from other kinds of wood with its distinctive natural patterns and golden brown color. Teak wood is a hardwood so it is durable and free from termites and weevils, because it contains a natural substance call "o-cresyl methyl ether" that was discovered by the Thai Department of forestry to have an anti-insect and anti-fungus effect. It is, however, very easy to cut into shapes, so teakwood is one of the most ideal raw materials for furniture production. (Thailand International Furniture Fair, 2008).

3. Characteristic of design. This paper synthesized four researches concerning the knowledge management of wood carving local wisdom in the upper-northern area of Thailand: 1) Knowledge Management of Wood Carving Local Wisdom in Baan Tam Pah Tong, Muang, Chiang Rai; 2) Knowledge Management of Wood Carving Local Wisdom in Baan Kiw Lae Noi, San Pah Tong, Chaing Mai; 3) Knowledge Management of Wood Carving Local Wisdom in Baan Look, Mae Tah, Lampang; 4)Knowledge Management of Wood Carving Local Wisdom in Baan Klong Toey, Klong Laan, Kam Paeng Pet. The result of the synthesis could be divided into 2 parts: the context of the communities, and the development and potential of the woodcarving groups.

3.1. The context of the communities. There are 4aspects from which the context of the communities could be seen. 1) Local resources and environments. In all 4 areas, the lands are fertile and have reservoirs for agriculture. The raw materials for wood-carving are abundant. 2) Community social capitals or cultural heritages. The studies found cultural and racial diversities in the areas. People live in peace and with simple life style. 3) The social change allows skilled woodcarvers to create their works and make it a career.

3.2. The development and the potential of the woodcarvings groups. The studies found that wood-carving in all four areas was influenced by the nature-loving life style of local people, the environment, and experiencing the splendor of woodcarvings. The major raw material for woodcarving is teak wood, which is available in local area. The tools and processes of woodcarving are similar. The studies also found that the beauty and the neatness can add value to the products. The leadership of people also helps create the opportunity to develop the local wisdom. The common problems for all 4 areas are the raw material and product design. Regarding the knowledge transferring process, personal relation makes the learning process easier. There are 2 forms of knowledge transfer:

1) Demonstration,

2) Demonstration and practice.

The experts teach the apprentices verbally and informally. The contents of learning are wood selection, tools handling, and woodcarving techniques. The apprentices are taught at the woodcarver's home. The knowledge of woodcarving

is mainly drawn from the work experience, and their imagination. From the knowledge transfer assessment, it was found that most woodcarvers would follow-up, probe, and observe the progress of the apprentices closely, so the knowledge would be preserved in the communities. The major problem with the knowledge transfer is that woodcarving is a tedious and complicate job. The people in all 4 areas are proud that their communities have the local wisdom of woodcarving.

The development of the woodcarving local wisdom can be done in2 ways:

1) to follow the conventional life style of people,
2) to develop and adapt the knowledge to match the outside world.

The woodcarvers and people in their communities are proud of their village to possess their own local wisdom, and to earn more family income. There are efforts to preserve the local wisdom of woodcarving with the community. They setup career groups to support their local occupations and increase the income of the group members, so they would be self reliant in their own community, and to maintain the local wisdom which is their cultural capital. However, the groups still rely on the supports from the outside, such as the government agencies, especially, in terms of investment and marketing.

4. Industries, Trade and Marketing. Thailand was once classified as a major timber-producing country with rich forest resources. The net effect of illegal logging, ineffective forest management, expansion of agriculture and livestock breeding, and shifting cultivation, accompanied by the continuously increasing demand from wood-based industries, has resulted in a timber scarcity.

Currently, Thailand has to import logs, sawn wood, and other resources from neighbouring countries to meet the domestic demand, which has turned Thailand from a timber exporting to an importing country. After the Government banned all logging operations in over 300 forest concessions inJanuary1989, the supply of domestic timber ceased. As a direct impact of this decision, numerous sawmills had to close down because of the log shortage. This also affected the plywood and veneer industries, who now have to rely mainly on imported logs and have therefore cut mill capacity to as low as 50-60%.

Thai teak had been continuously exploited from natural forests since before the establishment of the Royal Forest Department(18 September 1896) and it is recorded that export earnings were second only to rice for many decades. After heavy exploitation for many years, production has been substantially reduced. However the imposition of a logging ban in 1989prohibiting teak harvesting from natural forests, except from areas with land titles, as well as from dams and reservoir construction, has sharply reduced the teak harvest. Records indicate that the amount of teak harvested during 1989-93 was 23,857; 10,835; 1,890; 350; and 2,967 m3 respectively (Anon, 1993).Exports of teak logs and lumber have not been allowed since 1991, except in the form of processed products or veneers.

In a great reversal, teak log and lumber were imported amounted to more than 200,000 m3/year as raw material for export products, as well as for domestic consumption (Moonrasarn,1992). Statistics of the Royal Forest Department indicate that Thailand imported 169,472 m3 of teak logs and 19,531 m3 of sawn timber in 1993.

V.CONCLUSION

Current golden teak furniture is widely popular in both domestic and international. Although gold production of teak furniture are not new issues, and produced continuously; however, the attractiveness of this type of furniture is still available directly. "Difference" that we can create (design) was how to ... from the collection of information. Found that “Creative Design" is at the heart of the golden teak furniture manufacturer. This must have something special or different than Design. General (Standard) to add value to golden teak furniture (Something Difference, Something Special).Both shape and look to meet market demand. Should be designed using cultures of ancient Greek or Roman Theme, which must come one way or Colonial Decor second stencil(Ginger Bread) or engraving (Carvings) This is a classic line. Popular ever. Obsolete and not bring a product vertical like Modern. Consistent with the text of Kaplinsky and Morris as follows: Value chain describes the full range of activities which are required to or service from conception or design, through different phases of production, delivery to final consumers, and final disposal after use. Value chain provides a systemic understanding of the production process of a particular good. With the growing division of labor and the global dispersion of the production of components, systemic competitiveness has become increasingly important.(Kaplinsky and Morris,2000).

This research happened from the motivation in many idea such as production problem, the quality of product, the requirement of the market, new product, administration and Technical skill which from the situation that appear this, This research has the objectives for capture knowledge of expert form teak furniture and transfer to the others teak furniture business SMEs firms. The other is to create new product for export.

This research will focus on how KM tool – new technique, which is combined with KM theories – new technique that has never been offered offers before, could manage the problems. These problems

will be characterised by dynamic interactions. So the new technique will be proposed as a new strand of knowledge to create process. It may operate as a dynamic nature of knowledge creation and manage such a process effectively.

As mentioned, the KM tool will be based on the SECI process and the four types of *ba*, therefore the combination of the theories will be mobilised together with the SECI process and *ba*, in which knowledge assets have to interact originally and dynamically. So the knowledge will be mobilised and shared in each mode/*ba* where tacit knowledge may be held by individuals, converted and amplified by the spiral of knowledge through socialisation, externalization, combination, and internalisation.

I think the new technique is key capable of solving these problems which will be sustainable and competitive advantage in the long run.

REFERENCES

[1] Anon. 1993. Forestry Statistics of Thailand 1993.Data Centre, Information Office, Royal Forest Department.

[2] Drucker, P. F. (1999), "Management Challenges for the21th Century, Harper Business, New York.

[3] Kaplinsky, R., and Morris, M. 2000. A Handbook for Value Chain Research. Prepared for IDRC.

[4] Kaosa-ard, Apichart. 1980. Teak (Tectona grandis Linn.f.), I. Natural Distribution and Related Factors.

[5] Technical Paper No. 9, Teak Seed Centre, Ngao, Lampang. Mimeograph. 1 Mahaphol, S. 1954. Teak in Thailand. Paper No. R.16. Royal Forest Department Ministry of Agriculture. 30 pp9 pp.

[6] Ministry of Industry (MoI) of Republic of Indonesia.2006. The development of small and medium furniture industry through cluster approach. Final report. Jakarta.

[7] Muradian, R., and Pelupessy, W. 2005. Governing the coffee chain: the role of voluntary regulatory systems. World Development 33:2029–2044.

[8] Nonaka, I. et al.: A Theory of Organizational Knowledge Creation: Understanding of the Dynamic Process of Creating Knowledge. In: Dierkes, et al. (eds.): Handbook of Organizational Learning and Knowledge. Oxford university Press, Oxford (2003) 491-517

[9] Pandey, D., and Brown, C. 2000. Teak: A global overview. Unasylva 201:3-13.

[10] Schmitz, H. 2005. Value Chain Analysis for Policy-Makers and Practitioners. Institute of Development Studies, University of Sussex, UK.

SCOR Model Based on RFID Enable Supply Chain Management Module on ERP

Matteo Savino
Department of Engineering, University of Sannio, Piazza Roma, 21 Benevento 82100, ITALYmatteo.savino@unisannio.it

Pongsak Holimchayachotikul
College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, 50200 THAILAND

Alessandro Brun
Politecnico di Milano, Department of Management, Economics and Industrial Engineering, Via Lambruschini 4, Milano ITALY

***Abstract* — Well-organized Supply Chain Management (SCM) suitably delivers exactly packaged product, quantity and product information to the specified customer at the right time. Recently, many companies have applied ERP system to manage their supply chain. However, this system consumed much more time and resource to collect the required data in its supply chain module. Moreover, companies are very frequently gathering a plethora of indicators, without developing a unifying framework. SCOR (Supply Chain Operations Reference) model has widely used to support redesign and re-engineer of the supply chain in many industrial sectors. The present research is to propose a framework integrating the SCOR Model and the supply chain module in ERP system, for a Supply Chain based on RFID (Radio Frequency Identification). The methodology will be presented through a case study. In particular, we will first of all introduce a diagram depicting the AS-IS situation, to explain the supply chain operation in each unit and to define information, which could be gathered through RFID. The lean thinking paradigm has also been applied to define the non value added activity. The RFID enabled Supply Chain based on SCOR framework resulted in shorter total lead-time and increased traceability of the products, as real time accurate information were shared among partners. Each process of the SCOR Model is analyzed and customized with respect to the RFID application, with a deep analysis of the improvements that can be obtained thanks to this technology. This customized and lean configuration of the SCOR model allows a targeted representation of the main processes at each level. The Key Performance Indicators (KPIs) proposed by the SCOR model are then used to monitor these processes. The adoption of the SCOR Model based on RFID, along with the definition of standardized KPIs, allowed a fast implementation of the ERP software. The new postal stamp distribution on the Italian territory in 2003 was used as a case study. The case study produced also some best practices such as stamps tracking, tracing and a specific warehouse layout configuration which led to improved operations management performances.**

Index Terms- RFID, SCOR, ERP, SCM

I. INTRODUCTION

Recently, Supply Chain Management (SCM) has played a crucial role of business operation strategy management in order to achieve customer satisfaction in highly competitive business environments. Many companies are investing in a number of technologies to facilitate and increase their performances. The dramatic advance of logistic and IT tools have stimulated supply chain unit improvement such as real time information tracking and automatic data collection. On turn, it has also been possible to identify wastes and non-value-added activities. With this regards, RFID is a cutting edge technology, which offered the solution to this problem. RFID is a seemingly simple technique. Data is stored in RFID tags which are attached to objects or located in smart cards, and this data can be read using radio signals and presented on a display by using a suitable reader. The data can then be transmitted automatically to an Information Technology (IT) system for further processing. Lately, several large-scale enterprises (such as Wal-Mart, Hewlett-Packard and many others) have also applied this technology to manage their delivery tracking, inbound and outbound quality control warehouse, and inventory replenishment. These resulted in reducing labor costs and improving inventory control [1]. The US Department of Defense has used RFID to improve the control over military supplies [2]. In addition, software vendors such as Microsoft, Sun, IBM, and others have focused on the RFID technology to improve customer response in Europe, Japan and other countries [3]. To further demonstrate the relevance of this tool, a lot of SCM researchers have concentrated on the study of how RFID can improve SCM in many aspects [4].

With regards to the logistics management issues of the new postal stamp distribution on the Italian territory, existing supply chain units employed traditional methods to manage and retrieve the data and information. Since RFID and its application on SCM is really new, we noticed that, more often than not, RFID implementations in a SC are characterized by a “standard” implementation of the ERP system, with a configuration derived from the SCOR methodology yet not taking into account idiosyncrasies of RFID systems. For this reason the present paper is aimed to explore the possibility of integration of RFID technology with the well known

SCOR model in its ERP system. To turn the ERP implementation weakness in term of the resource usages among complicated data collection, RFID has been considerably recognized as data collection facilitator. In particular a case study will be discussed, in which the Supply Chain related to the distribution of the Italian value charts is configured according to the SCOR and, for certain phases, we tested the RFID implementation for pallets tracing in order to get a better identification along the chain.

The introduction and supportive literature review are shown in the present section. In section II, the background of SCOR model and RFID technology are discussed. The research methodology is presented in Section III. The case study is then illustrated in section IV. The possibility of integration of RFID technology with the well known SCOR model in SCM module of ERP is demonstrated in section V. Finally the conclusion and future work are summarized in section VI.

II. BACKGROUND

Supply chain management and SCOR model

Supply chain management (SCM) is defined as design, maintenance and operation of supply chain processes for satisfaction of end user needs and also described as a set of organizations directly linked by one or more of the upstream and downstream flows of products, services, finances, and information from a source to a customer [7]. In SCM, a key role is played by the most recent Information and Communication Technologies (ICT) that can support better information system integration and a quicker response to customer demand. ICT make it possible to process more information, more accurately, more frequently, from more sources.

SCOR [5] model was developed in 1996-1997 by Supply-Chain Council Inc. (an independent not-for profit corporation), as the cross industry standard for supply chain.

TABLE 1. FIVE MAIN MANAGEMENT PROCESSES

FIVE MAIN MANAGEMENT PROCESSES
PLAN: it is the typical integrated planning which involves all the actors of the chain; the included sub processes serve in order to balance the demand, the supplies and to develop a conduct line that satisfies the rules established from the markets.
SOURCE: it is the supplying phase of the material and informative resources; in this phase, the included sub-processes serve to obtain the goods and the services in order to meet the forecasted demand.
MAKE: this phase directly regards the realization of the product or service; the processes have the task to transform raw materials and components in the finished product required to satisfy the needs of the customers.
DELIVER: it is the delivery and the distribution of the product or service; typically, it includes order management, transportation and warehousing activities.
RETURN: Return is used in order to maintain trace of the processes involved in the returns of raw materials to the supplier and in the return of finished products coming from the customer.

SCOR is recognized as an international standard and a process diagnosis tool, which addresses processes, as well as performance measures and best practices.

SCOR allows companies to describe a supply chain configuration using a standard language, to identify what areas have to be examined, which are the best indicators to evaluate a benchmarking and how financial costs and returns on specific investments have to be determined.

The SCOR Model, which has been developed to describe all of the activities leading to customer satisfaction, describes a generic supply chain in five main integrated management processes summarized in Table 1.

Radio frequency identification

Radio frequency identification (RFID) is a seemingly simple technique. Data is stored in RFID tags, which are attached to objects or located in smart cards, and this data can be read using radio signals and presented on a display by using a suitable reader. The data can then be transmitted automatically to an information technology (IT) system for further processing.

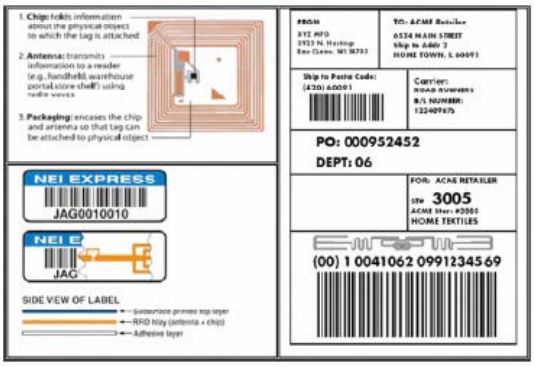


Figure 1. Structure of RFID tag.

To identify product using RFID technology, we can precisely monitor the locations of goods in different locations or racks. In that case the locations or racks will be fitted with RFID antennas. Fig 1 shows the different parts of an RFlD tag including a packing and chip [6]. The tags send their ID numbers to the antennas. Unlike the situation with barcodes, this occurs automatically and as often as desired, and it can also occur after a defined event such as closing the rack door. A sensor controller monitors the antennas in the rack and communicates the read out data to the central system management programs; RFID readers can generate useful information by collecting data for parameters such as temperature and events such as opening and closing rack doors. Opening a door can signal an unauthorized access, so it should trigger an alarm (perhaps a silent alarm). Even without an alarm, removal of a component can be recognized by triggering an automatic scan after the door is closed and comparing the result with the

existing inventory list. In any case, the assets database is updated, and it presents an accurate picture of the rack history and the components installed in the rack.

An RFID system component is typical made of an RFlD tag on the item itself, an antenna included to "interrogate" the tag and a reader. The software is vital to organize the captured data and link it back main database. These RFID implementations have provided benefits in terms of data synchronization and information flows within their supply chain.

III. METHODOLOGY

The methodology of the SCOR model based on RFID is shown in Fig. 2. Firstly, the new postal stamp supply chain operation on the Italian territory was collected as the primary data. Secondly, SCOR model attributes were used to match with the each operation units, while the lean thinking paradigm was also applied to define non value added activities and list the information unit. In addition, the AS-IS diagram describes the supply chain operation in each unit and defines the information that can be potentially collected through RFID.

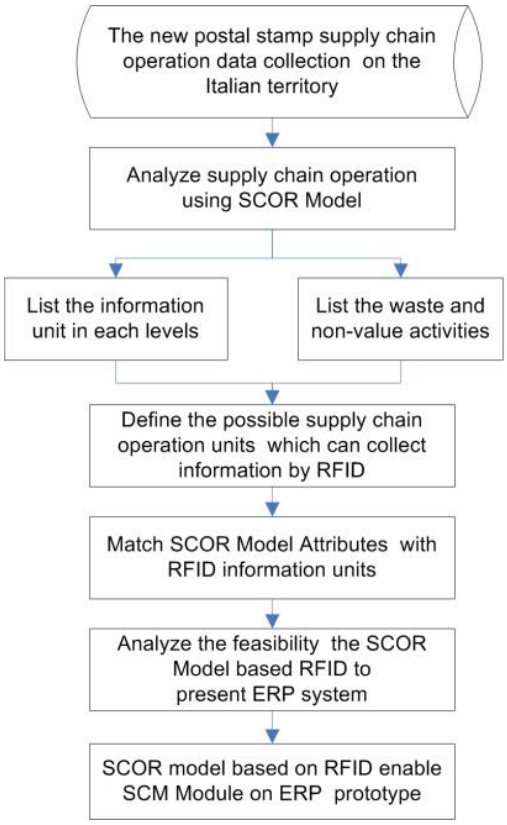


Figure 2. The SCOR model based on RFID enable SCM Module on ERP

The RFID enabled Supply Chain based on SCOR framework resulted in shorter total lead-time and increased traceability of the products, as real time accurate information were shared among partners. Each process of the SCOR Model is analyzed and customized with respect to the RFID application, with a detailed analysis of the improvements that can be obtained thanks this technology. The case study allowed showing how the adoption of the SCOR Model based on RFID, along with the definition of standardized KPIs, resulted in a fast implementation of the ERP software.

IV. THE CASE STUDY RESULT AND DISCUSSION

In 2004 the Italian postal delivery service, Poste Italiane – PI had in charge the distribution of the postal value charts on the whole Italian territory and decided to re-organize its logistic processes. The Value charts had to be distributed from the central supplier (Italian Poligraphic Institute and Mint – IPIM). Types of Value charts to be distributed were: regular postal value Chart (stamp), Philatelic value charts, phone cards, stamped values, checks and money orders. The supply chain is a typical multi-stage supply chain structured as shown in Fig. 3.

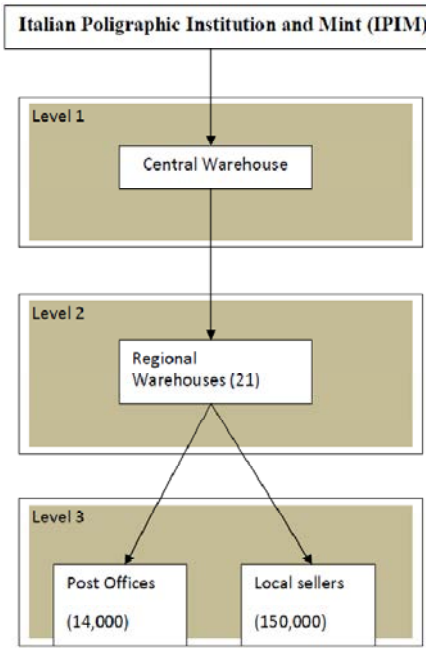


Figure 3. The Value Chart supply chain

The data related to this supply chain can be summarized as follows: 14,000 post offices, 150,000 local sellers, a total of 5,700,000 pieces to be distributed (average monthly basis), 75,000 pallets

The supply chain is organized in a mixed Push/Pull way, i.e. the final distribution points ask for inventory replenishment with respect to their specific needs, while IPIM produces the value charts according to Material Requirement Planning criteria, with a Master Production Schedule made on yearly basis with the following inputs: requests of the previous year, inventory status, sell trends.

With real time communication, enabled by RFID technology, the system prints documents automatically, when the related pallet is effectively

loaded on the truck. In this way, shipping errors (i.e. pallet not loaded on the vehicle but the Document of Transport printed) are avoided.

V. POSSIBILITYOF INTEGRATIONOF SCOR BASED RFID

The supply chain processes were analyzed and matched to elements of SCOR model in each level. Moreover, we also analyzed how RFID could have been applied in the various processes of the analysed Supply Chain. A brainstorming session with the operator, officer and manger in each supply chain operation unit was conducted to identify the feasibility of RFID adoption to reduce wastes and to facilitate operations. Central warehouse, regional warehouse, post offices and local sellers were analyzed resulting in the following.

Level 1 (central warehouse)

The Logistics Execution System (LES), a major component of the supply chain management (SCM) application of SAP, allows an administrator to manage information and processes involved in all stages of the supply chain, from raw materials procurement to finished products distribution. LES connects SCM processes involved in procurement, order processing, production, storage, inventory management, shipping, and sales. The LES module receives the Transfer Requests related to the specific quantities and types of materials and, upon availability, transform these requests in Transfer Orders. ERP creates the Material Transfer List in which, for each lot, the regional deposit of final destination, the shipping number and the expedition time are specified. At level 1, the central deposit is divided into three areas as shown in Table II. At this level, we proposed to use the RFID when, for each Transfer List, one Handling Unit (HU) is created. The Handling Unit is related to each single lot (made by one type of value chart) to be sent to the regional warehouses and is called HU_L. For the HU_L, the associated RFID tag carries the information shown in Table III.

TABLE II. AREAS OF LEVEL 1.

AREAS
Material receiving, in which material coming from IPIM are received and controlled in terms of quality and conformity.
Stock area, where materials wait to be sent to level 2 (regional warehouses).
Shipment area, in which operations related to shipment are set-up and materials are put on the means of transport.

TABLE III: INFORMATION CONTAINEDIN RFID TAG FOR HU_L.

INFORMATION
Content of the HU (type of value charts)
Value of the content
Weight
Regional warehouse of destination
HU_L Tracking number
Number of the Transport Document (i.e. the document given to the transporter and that has to travel together with him the mean of transport used for the expedition)

1.1) SOURCE:

Before arriving to level 1, RFID tags should be attached to pallets and boxes at IPIM with necessary information. RFID technology is used in source point of level 1 to swiftly count the quantity of boxes/pallets, to match materials with the documents sent with vehicles and to send information to ERP instantly with almost zero paperwork, less human contact and high accuracy.

With a RFID reader the tags on the pallet are scanned. After this also the single boxes inside the pallet are opened and the labels on these boxes are scanned as well. In this way the data related to both each pallet and each single box are sent to the ERP database.

Regarding the Source phase, for this level S1.2 Receive Product, S1.3 Verify Product, S1.4 Transfer Product processes are configured. The S1.1 process is not considered since it will be replaced by another process in the other levels. Also the Process S1.5 is not present because in this phase there are no payments to be authorized and settled. Conformity verification of the supplied materials is done. If conformity is verified, there is an entry confirmation; otherwise the material is placed into the non conformities area.

1.2) MAKE:

The Make phase is not present because materials are not processed and, after the incoming controls, materials are transferred into the stock area where they will be picked up for the subsequent shipment.

1.3) DELIVER:

The Deliver phase is characterized by the set-up of the expedition lots with respect to the order received by the regional deposits. Here D1.9 Pick Product, D1.10 Load Vehicle, Generate Ship Docs, Ship Product, D1.11 Receive & Verify Product at Customer processes are configured. The list of excluded elements is given in Table 4 with the related motivation such as “No prices to be quoted” and “Orders are based on standard parameters and there is no need to monitor them”, so D1.1 and D1.2 were not assigned. In the Reserve Inventory & Determine Delivery Date (D1.3) did not also

determine because it was managed by the Source and Make Process respectively in Stage 1 and 2. No routes have to be determined because the locations are ever the same and the routes are pre-defined and carriers are defined and no shipment rate is necessary, thus D1.4-D1.7 were also not considered. The inspection of incoming products is made by S1.3; therefore, D1.8 was not allocated. D1.12 and D1.13 were not considered because "Products do not need installation" and "No payment is done for the products delivered", respectively.

1.3.1) Inventory:

After conformity tests, the conforming pallets/boxes will be scanned by RFID reader at the entrance of warehouse and send the information automatically to ERP software.

Product Identification: by using RFID technology we can precisely monitor the locations of goods in different locations or racks. In that case the locations or racks will be fitted with RFID antennas. The tags send their ID numbers to the antennas. Unlike the situation with barcodes, this occurs automatically and as often as desired, and it can also occur after a defined event such as closing the rack door. A sensor controller monitors the antennas in the rack and communicates the read out data to the central system management programs. RFID readers can generate useful information by collecting data for parameters such as temperature and events such as opening and closing rack doors. Opening a door can signal an unauthorized access, so it should trigger an alarm (perhaps a silent alarm). Even without an alarm, removal of a component can be recognized by triggering an automatic scan after the door is closed and comparing the result with the existing inventory list. In any case, the assets database is updated, so it presents an accurate picture of the rack history and the components installed in the rack.

1.3.2) Advance Information about shipment:

When each HU_L exits the warehouse, the transponder sends the signal to the ERP repository with related information. In real time the regional warehouses are informed that the shipping has been done with related arriving times. Another advantage of the RFID application at this level is the immediate printing of the transport document when the pallet is put on the vehicle. This document is very important from a fiscal point of view. Usage of RFID technology in shipment phase is shown in Fig 4. With real time communication given by RFID technology the system prints this document automatically only when the related pallet is actually loaded on the truck. In this way the shipping errors (i.e. pallet not put on the mean of transport but Document of transport printed) are impossible.

1.3.3) Pallet management:

We can attach RFID tags to the pallets so that they can be identified with RFID readers at desirable points. The following data could be acquired:

- Trolley number, driver, delivery customer (recipient), and delivery date and time.
- Trolley loading, unloading and transfer times.

Regional warehouses are organized in the following areas:

1.4) RETURN:

When customer returns the product, a rewritable RFID tag will be attached to the pallet or box and the cause of returning will be written on the tag and read through a RFID reader, information send to SAP system and the product will be send back to previous level destination.

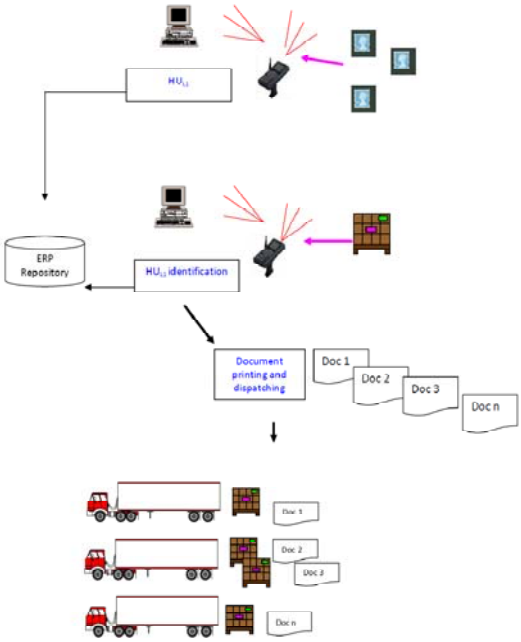


Figure 4. Usage of RFID technology in shipment phase

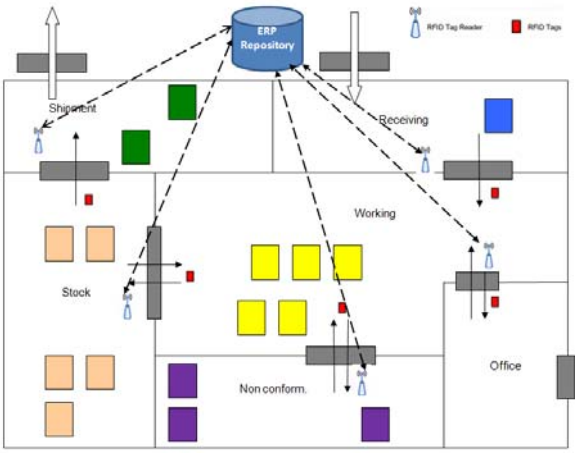


Figure 5. RFID technology usage of regional warehouses.

Level 2 (regional warehouses)

In this phase, operations on materials are performed because the expedition lots sent by the central warehouse are de-aggregated in several small lots (which sometimes can contain as small as ten pieces or less) with respect to the order of the Post Offices and local sellers. The RFID usage in regional warehouses is shown in Fig 5. The list of excluded elements is given in Table IV. Level 2 is the only level where a Make phase is performed.

TABLE IV. AREASAND THEIR CORRESPONDING RESPONSIBILITIESOF REGIONAL WAREHOUSES

AREAS
Receiving: shipments from the central warehouse are received and checked in term of quantities.
Working; in this area pallets coming from Receiving are de-packed and the shipment lots are formed according to the orders received from Post Offices.
Stock; where worked materials are placed before shipping.
Non-conformities; in which are stored those materials that must be shipped back to the central deposit due to a wrong shipment or to wrong (exceeding) quantities provided.
Shipment; where the pallets are loaded on the vehicle.

TABLE V. ACTIVITIES RELATEDTO WORKING AREA OF A REGIONAL WAREHOUSE.

ACTIVITIES
The incoming shipment (i.e. coming from the Receiving buffer).
The different mix of expedition lots are formed and included into the shipment.
For each expedition lot a RFID Tag which contains new information is associated.
Another RFTAG is also associated to the entire handling unit expedited.
The handling units formed are transferred into the Stock area, ready to be shipped.

2.1) SOURCE:

RFID technology is used here to swiftly count the quantity of boxes/pallets, to match materials with the documents sent with vehicles and to send information to ERP instantly with almost zero paperwork, less human contact and high accuracy.

Labels on the pallet are scanned by means of a RF reader. After this, the single boxes inside the pallet are opened and the labels on these boxes are scanned as well. In this way the data related to the pallet and to each single box are acquired into the ERP database and there is the tracing between the handling unit of the entire pallet and the handling unit of each single box. In addition, S1.2 Receive Product, S1.3 Verify Product, S1.4 Transfer Product processes are configured. S1.1 and S1.5 is not present because it will be replaced by another process in the other levels.

2.2) MAKE:

In this area pallets coming from central warehouse are de-packed and the shipment lots are formed according to the orders received from Post Offices. Tags with new information are attached to the pallets/boxes. The activities performed in this Working area are listed in Table V.

2.3) DELIVER:

The Deliver phase is characterized by the set-up of the expedition lots with respect to the order received by the regional deposits. Here D1.9 Pick Product, D1.10 Load Vehicle, Generate Ship Docs, Ship Product, D1.11 Receive & Verify Product at Customer processes are configured. The list of excluded elements is due to the same reasons already mentioned in deliver in level 1.

2.3.1) Inventory: Pallets/boxes will be scanned by RFID reader at the entrance of warehouse and information is automatically sent to ERP software.

2.3.3) Advance Information about shipment: The activities performed in this Shipment area are shown in Table VI. The LES receives the Transfer Requests related to the specific quantities and types of materials and, upon availability, transform these requests in Transfer Orders. The ERP creates the Material Transfer List in which, for each lot, post offices / local distributors of final destination, the shipping number and the expedition time are indicated.

TABLE VI. ACTIVITIES RELATEDTO SHIPMENT AREAOF A REGIONAL WAREHOUSE

ACTIVITIES
The transfer requests are collected and sent to the working area, where the handling unit is set-up according to the activities described in the previous section.
Once the shipment is prepared, the transfer request is transformed into an order request. The pallets are set into specified subareas with predefined criteria related to the time of picking-up, quantities, etc.
Two information objects are created: ▪ The first is *Supply* in charge to advice the post office. It contains arrival dates of the supplied material, specific content and so on. ▪ The second one is the *Delivering* object that is in charge to engage the material into the stock area (so that it will not be wrongly used for another shipment) and to activate the shipment process.
When the Delivering object is created in the ERP, the exact day and time of delivering to the post offices of final destination is also defined. In fact into the ERP the delivery time for each post office of each warehouse has been stored. The system then calculates the arrival date of shipment. This way, the post office is able to see the expedition lot and its content and can consider it as *available* from the date arrival indicated. This allows the post office to sell values in advance, giving to the customer the arrival date of the material.

At this level the RFID is used when, for each Transfer List, two types Handling Unit (HU) are created as follows the first type is related to each single lot ordered by the regional warehouses and is called HU_L. The second type of HU is related to the

pallet that is formed and is called HU_P. For the HU_L the associated RFID transponder carries the information synthesized in Table VII. When each HUp exits the regional warehouse, the transponder sends the signal to the ERP repository with related information. In real time, the regional warehouses are informed that the shipping has been done with related arriving times. Another advantage related to the RFID application at this level is the immediate printing of the transport document when the pallet is put on the mean of transport. This document is very important by a fiscal point of view. With real time communication given by RFID technology the system print this document automatically only when the related pallet is effectively put onboard of the truck. In this way the shipping errors (i.e. the printed pallet not put on the transportation) are impossible.

2.4) RETURN:

When customer returns the product, a rewritable RFID tag will be attached to the pallet or box and the reason of return will be written on the tag and red through a RFID reader, information sent to SAP system and the product will be sent back to previous level destination.

TABLE VII. INFORMATION TYPES, CONTAINED BY THE TAGSOF HU_L.

INFORMATION	
- Content of the HU	- Value of the content
- Regional warehouse of destination	- HU_L Tracking number
Number of the Transport Document (i.e. the document given to the transporter and that has to travel together with him the mean of transport used for the expedition)	

TABLE VIII. INFORMATION TYPES, CONTAINED BY THE TAGSOF HU_P.

INFORMATION	
- Number of HU_L included	- Value
- Weight	- HU_P Tracking number
Post offices/local sellers of destination	

Level 3 (post offices & local seller)

At this level the director of the post office takes in charge the shipment arrived and controls what has been delivered with the transport document provided with the shipment and the printing of the Delivering sheet given by the ERP. Once correspondence between these documents and the received shipping is verified, the material is accepted and the procedure is closed. In this step, signals of arriving date and times, as well as the correct quantities received are stored into the ERP to be able to calculate the related KPIs. In case of non conformity, shipment is sent back to the regional warehouse, a NC procedure is open with motivation of rejection and the delivering object is cancelled to discharge the material from the post office

3.1) SOURCE:

RFID technology used in this point to swiftly count the quantity of boxes/pallets very fast, to match materials with documents and to send information to SAP instantly with almost zero paperwork, less human contact and high accuracy. With a RF reader the labels on the pallet are scanned. After this, single boxes inside the pallet are opened and labels on these boxes are scanned as well. In this way the data related to the pallet and for each single box is acquired into the ERP database and there is the tracing between the handling unit of the entire pallet and the handling unit of each single box. In addition, the source components of this phase have also defined as the same configuration of the previous phase.

3.2)MAKE:

The Make phase is not present because materials are not processed and, after the incoming controls, materials are temporary stored from where they will be to the end users/customers.

3.3) DELIVER:

Inventory: Pallets/boxes will be scanned by RFID reader at the entrance of warehouse and the information is automatically sent to ERP software.

3.4) RETURN:

When customer returns the product, a rewritable RFID tag will be attached to the pallet or box and the cause of return will be written on the tag and read through a RFID reader, information send to SAP system and the product will be send back to previous level destination.

After SCOR Model attribute definition based on RFID, the KPIs have also defined from this prototype results to measures this integration of the ERP software.

VI. CONCLUSIONS

The present paperdescribed the integrated application between SCOR Model based on RFID and supply chain module in ERP system. After the proposed methodology implementation, the result is extended to the RFID installation guidelines to increase the ERP performance of this case study supply chain. The SCOR integrated RFID powered supply chain allowed to correct real-time information sharing between the partners, to reduce the total lead-time, also resulting in better traceability and lower operation costs. In all the phases where RFID technology is suggested to be used, the supply chain enjoyed the following advantages:

- Data can be read very fast minimizing handling time.
- No paperwork and less people involved.
- Less manual error. Data accuracy is very high.
- Automatic inventory update and location update.
- Correct sequence always ensured.

- Immediate printing of documents.
- Data acquisition costs are reduced.
- New data can be added in the same tag and tags can be reused erasing the previous data.

As a further development of this research project, we are now developing a performance analysis between the present ERP system without SCOR Model based on RFID and the proposed prototype in term of quantitative economic and financial investment analysis. Moreover, we are planning to develop a simulation tool to analyse a SCOR Model based on RFID. Our goal is to run simulation experiments in order to get quantitative measures of the company's performance, depending on the attributes proposed in the framework.

ACKNOWLEDGMENT

The financial is supported by Erasmus Mundus, External Cooperative Window Project (**eLINK*e***ast-***w***est ***L***ink for ***I***nnovation, ***N***etworking and ***K***nowledge exchange) and College of Arts, Media and Technology Chiang Mai University, THAILAND

REFERENCES

[1] Pradhan S. el aI, "RFID and Sensing in the Supply Chain: Challenges and Opportunities", *liP Laboratories Palo Alto lIPL,* 2005 , February 9, pp. I-I I.

[2] Khan A. et aI, "Exploring the Potential Benefits of RFID: A Literature Based Study", *Department of Information Systems, University of Melbourne,* 2006, pp. 1-12.

[3] RFID Journal (2004), "IBM to guide DoD RFID rollout," available at: www.rfidjournal.com/article/articleview// 836/1/1

[4] eWeek (2004), "Suppliers, competitors balk at WalMart's RFID lead", available at: www.eweek.com/article2/0,1759, 1643386,00.asp

[5] Supply Chain Council, 2004 "SCOR version 8.0 Overview"

[6] Maloni M. et aI, "Understanding Radio Frequency Identification (RFID) and Its Impact on the Supply Chain", While Paper, 2006, July 26, pp. 144.

[7] Mentzer, J.T. et al. (2001): Defining Supply Chain Management, in: Journal of Business Logistics, Vol. 22,No. 2, 2001.

Session 8: Network Management and Collaboration

ISBN 978- 974-672-556-9

Network Workload and Security Evaluation in the Distributed Measurement System

D.L. Carnì, D. Grimaldi, and F. Lamonaca
Dept. of Electronics, Computer and System Sciences, University of Calabria, Rende –CS, Italy
dlcarni@deis.unical.it, grimaldi@deis.unical.it, f.lamonaca@deis.unical.it

***Abstract*—The Distributed Measurement Laboratory (DML) is an heterogeneous system and the network workload evaluation and the security system measurement are become important research topics. The workload evaluation is important aspect of the performance management strategies and is taken into examination in the case the specific requirements must to be respected. The security system measurement is an important information to contrast the attacks from viruses, worms and hackers. The paper describes the methods (i) to evaluate the workload of the cooperating hardware equipments of the DML, and (ii) to measure the network security system. In this manner useful indexes are achievable to monitor the functioning conditions of the whole DML. The methods are based on four different tools. In particular, they concern with (i) the synchronization among the clocks of the different co-operating PCs, (ii) the One-Way Delay parameter measurements of the path from the source to the destination, (iii) the workload measurement of the CPU of each PC, and (iv) the measurement of the Intrusion Tolerant System operational security. The results of experimental tests show the performance and the accuracy of the four methods. Based on the analysis of these tests, practical criteria are inferred to point out design criteria and performance management strategies.**

Index terms - Distributed Measurement Laboratory; DML; network workload evaluation; security system measurement

I. INTRODUCTION

The Distributed Measurement Laboratory (DML) [1]-[4] operates on physical environments distributed in local area and/or geographic wide area. It consists of a collection of distinct Distributed Measurement Systems (DMSs) [5]-[8] devoted to specialized measurement (sub) tasks, co-operating one with each other. Each DMS is responsible for local measurements and includes various hardware platforms to perform different measurement tasks. The hardware includes: server of local network, automatic measurement stations, and Local Area Networks (LANs). The automatic measurement station is equipped by Hardware Interface to the network node, usually constituted by Personal Computer (PC), standard interfaces (RS232, RS485, IEEE488 and wireless), and Measurement Instruments. The connection among different DMSs can be obtained through the network bridge and/or the services of standard communications networks, including internet in the case of geographic area network. Therefore, the DML is an heterogeneous system and the performance management strategies and the security system evaluation to contrast the attacks from viruses, worms and hackers are become important research topics. They can be very complex in the case specific requirements must be respected.

Several important challenging requirements to the management have emerged in the contest of the expanding DMLs [5]-[10]. Indeed, there are trends towards (i) more efficient access to measurement information, (ii) measurement services that are customized according to individual needs, (iii) powerful tools for real time monitoring and measuring, and (iv) strong minimization of the communication network overhead. Examples of the actual requirements can be summarized as follows:

1) the correlation in the time domain of the measurements given by independent DMSs is imposed, the requirement is the DMS synchronization or the detection of the time delay between the clock of different DMSs;
2) the Mobile Agent technology is adopted for measurement tasks, the requirement is the selection of the path from the source to the destination characterized by the shorter delay time of the packets. This parameter, with the available bandwidth and packet loss rate, are important variables characterizing the behavior of the network traffic flow;
3) the management of the queue connection to the DML server is imposed for monitoring and measurement services, the requirement is the client selection by detecting the path characterized by shorter delay time of the packets;
4) the distributed processing of the measurement results is adopted, the requirement is the selection of the processing resource characterized by shorter workload, among all available in the DML, to permit speed processing.

ISBN 978- 974-672-556-9

Several important requirements to measure the security of the network have emerged in the contest of different strategies to point out the security mechanism. The network security is not simple to realize and a priori to evaluate because the DML is complex system, and the relations among the variables that describe their behavior are not simply enumerable or determinable [11]. In order to evaluate the security of the network, the fundamental strategies to point out the security mechanism must be taken into consideration. The network security can be managed with two different strategies. In the first, the concept is to point out mechanisms to prevent the attack. In the second, the concept is to limit the damages of the attack, once the first strategy faults.

In order to implement the first strategy, the firewall is used [12], [13]. It consists on a set of hardware and software components developed to manage and control the data traffic, among the network links, and in different TCP/IP stack levels.

In order to implement the second strategy, the Intrusion Tolerant System (ITS) is used. It must detect and stop the anomaly. The input/output data are managed by the device that implements the ITS to permit at the system to operate in continuity when it is not in the good state. To reach this goal, the mechanisms included into the ITS are (i) intrusion detection, (ii) fragmentation, (iii) replication, (iv) migration, (v) masking, (vi) isolation, (vii) containment, and (viii) recovery. One of the most important components of the ITS is the Intrusion Detection System (IDS), devoted to detect the attack and to generate the intrusion triggers. Once detected the intrusion, the successive step is the minimization of the damage caused by the attack. Therefore, the security evaluation and measurement is associated to the dependability concept. This is the capacity of the system to deliver service that can be justifiably considered trusted [14]. In the following, the operational security of the ITS is evaluated.

In the paper the methods able to give adequate answer to all these requirements concerning with network performance estimation and the security evaluation are proposed. The general architecture of the DML taking into account is shown in Fig.1. The basic idea is to distinguish in four different groups the tools to be used.

The first one concerns with the synchronization among the clocks of the different co-operating PCs, as an example PC#1of DMS#1 and PC#m of DMS#n. In this manner the requirement 1) can be satisfied.

The second one is correlated to the previous one and concerns with the One-Way Delay (OWD) parameter measurements of the path from the source to the destination. As an example, in Fig.1 the path can be constituted by (i) the connecting network#1, (ii) the connecting network #b, and (iii) one of the connections (α, β) into internet. In this manner the requirements 2) and 3) can be satisfied.

The third one concerns with the workload measurement of each PC, as an example the PC#1 or the PC server#1 of the DMS#1. This is performed by measuring the CPU occupancy parameter in the fixed and well determined conditions, not influenced by the different clock frequencies of the CPUs. In this manner the requirements 4) can be satisfied.

The fourth one concerns with the evaluation and the measurement of the ITS operational security, as an example the ITS defending the DMS#1. This is performed by considering in the unified view all the security strategies and mechanisms to realize the defense system.

The clock synchronization, the OWD measurements and the CPU occupancy are based on three independent procedures known and tested in literature [11]-[19]. Both the OWD measurement and the CPU occupancy parameter are useful indexes of the workload evaluation of the DML.

The method to measure the ITS operational security meets the idea proposed in [20] and [21] concerning with the relationship between security analysis and mechanism of dependability analysis. It is based on the analysis proposed in [22], but introduces as novelty the use of the semi-Markov chain to emulate the variability of the fault rate.

In order to clearly describe the four methods, initially the description of the activities to evaluate the workload of the DML is given. In particular, (i) the activities that must be executed by both the client and the server of the DML are highlighted, (ii) the factors motivating the choice of the procedures for the clock synchronization, the OWD measurements, and the CPU occupancy estimation are discussed, (iii) and the modifications justifying their adaptation to be used into the DML are highlighted. Successively, the basic aspects to evaluate the operational security of the Intrusion Tolerant System (ITS) are given. In particular, (i) the mapping of the concept of system security into that of the system dependability is established, (ii) the dynamic behaviour analysis of ITS state transition is discussed, and (iii) the measurement procedure of the ITS operational security is presented.

At the end, the results of experimental tests showing the performance and the accuracy of the proposed methods are described. Based on the analysis of the results of these tests, some criteria are highlighted to be advantageous used in the design and management strategies of the DML.

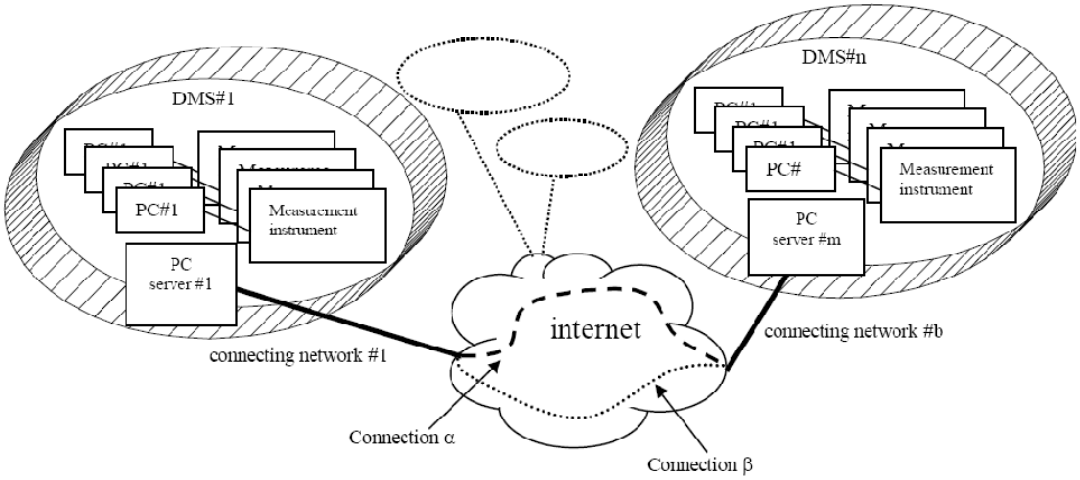


Figure 1. General architecture of the DML.

II. WORKLOAD EVALUATION OF THE DML

The workload of the DML is determined by the workload of the cooperating hardware equipments [23], as: the connecting networks, the PCs configured as network server, and the PCs configured to perform measurement tasks. According to the used method, all the PCs involved in the workload evaluations are considered as server, without differentiation on the basis of their functionality into the DML. Therefore, the new scenario to be taken into account in the workload evaluation is constituted by: the PC server, the connecting network, and the PC client. Both the PC server and the connecting network are cooperating hardware interested to the workload evaluation. The client PC is dynamically defined into the DML and executes the activities scheduled by the method. The monitoring activities are shown in Fig. 2a). The PC synchronization is performed if the OWD measurement needs. The other activities are performed in cooperation between the client and the server. According to Fig. 2c), the PC client generates the probe packets, monitors the received packets, evaluates the mean delay of the communication, determines the CPU occupancy, and computes the loss packets. According to Fig. 2b), the PC server receives the User Datagram Protocol (UDP) packets, attaches the time stamps according to delay evaluation of the communication network, executes the procedures according to the CPU workload evaluation, and send the UDP packets to the client.

Software tools are used to perform the clock synchronization, the OWD measurement, and the CPU occupancy estimation.

A. Clock synchronization

The clock synchronization is used to measure the delay of the communication between two PCs, as an example PC#1 and PC#m of Fig.1. Among the various procedures, the Network Time Protocol (NTP), now established as an Internet standard protocol [15], is used to organize and maintain the

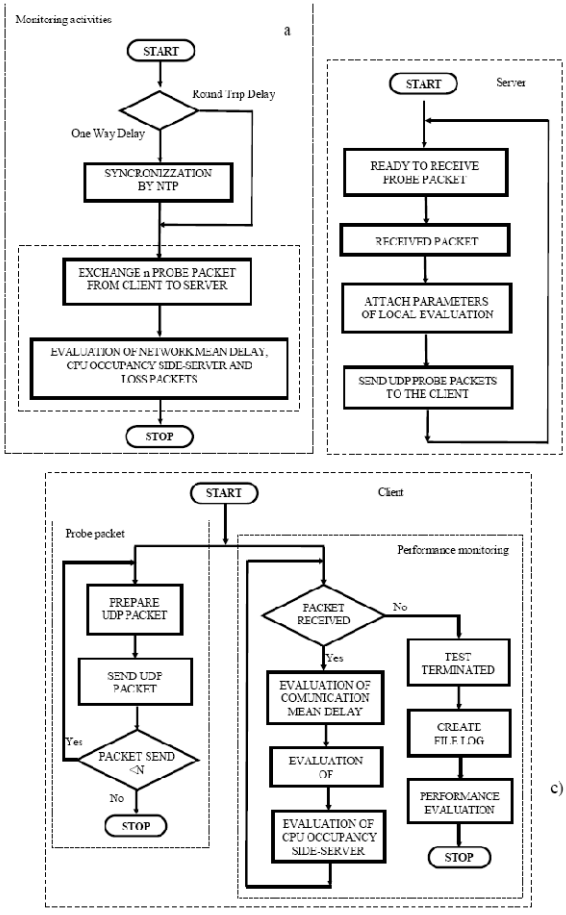


Figure 2. Flow diagram implementing the propose method.

clock synchronization of the PC to the national time computer service. NTP is built on the Internet Protocol and UDP [16], which provide a connectionless transport mechanism. Fundamental advantage of the NTP on DML is the fact that its protocol includes procedure to compensate for the effects of statistical delay variations encountered in Wide Area Network (WAN), and it is suitable for accurate and high resolution synchronization

throughout internet. Numerous experimental tests show that, as a consequence of the clock drift of the PC, the synchronization interval is an influencing factor of both the accuracy and the stability of the synchronization. Fig. 3 shows the trend of the synchronization offset between two PCs, in the case the time interval is equal to ten minutes and one minute, respectively. In this last case the synchronization is more accurate.

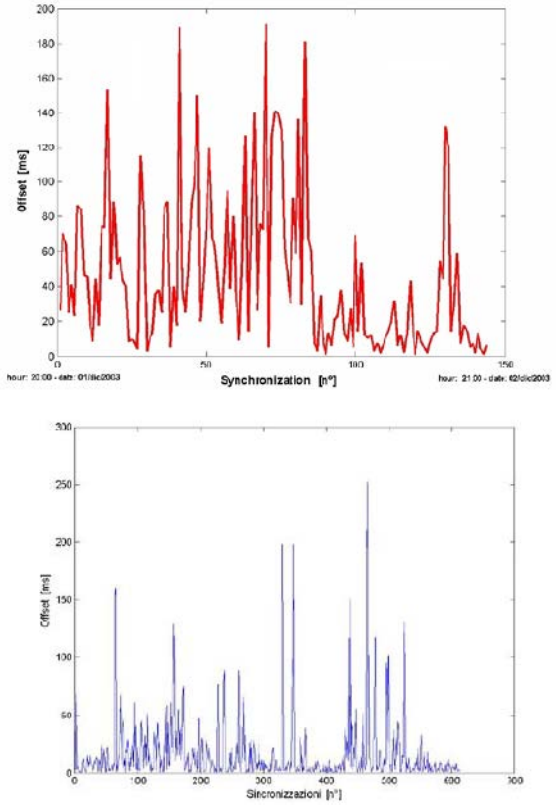


Figure 3. PC synchronization in presence of random perturbation a) every ten minutes, and b) every one minute.

B. One-way delay measurement

The OWD measurement is used as easy parameter to evaluate the workload of the network connection under examination. The OWD can be evaluated once both the PCs, source and destination, are synchronized by NTP. This coarse parameter furnishes useful information for all type of network connection, internet included [17]. The measurements are realized through "active probing" [18] which injects packets in the path to measure and by observing the series of parameters of which the packets are spoiled. Each packet is characterized by timestamp at the sender and at the receiver. The protocol used for the connection is the UDP. Owing to the reduced control steps if compared to the TCP protocol, the UDP shows minimum lateness of the packets. The programming language for this purpose is JAVA [24]. The log file containing the timestamps of the packets is created when the measurements are terminated. The OWD parameter is evaluated by means of the occurrences of the delays detected by means of the timestamps. In similar way the Round Trip Delay (RTD) parameter can be evaluated. As an example, in Fig.1 the path can be constituted by the connecting network#1, the connecting network#b, and the connections α into internet for the sent packet and the connection β for the received one. In this last case does not need the clock synchronization. Fig.4 shows the sequence of the operations performed on the travelling packet at both the client and server side, to evaluate the RTD. Fig.5 shows the structure of the travelling packet. It is constituted by 60 bytes divided as it follows:

- bytes 0-7 hold the identification key of the packet (ID packet),
- bytes 8-20 hold the start time instant at the sender of the packet from the client to the server (Timestamp0),
- bytes 21-33 hold the receiver time instant of the packet at the server (Timestamp1),
- bytes 34-46 hold the start time instant at the sender of the packet from the server to the client (Timestamp2),
- bytes 47-59 hold the receiver time instant of the packet at the client (Timestamp3).

Only the bytes 0-33 are used for the OWD measurement.

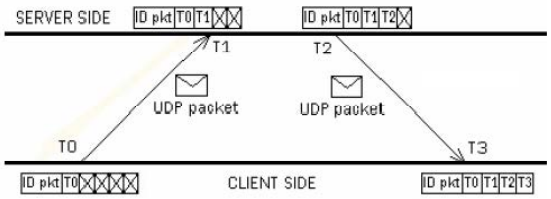


Figure 4. Sequence of the operations performed on the traveling packet at both the client and server side in order to evaluate the RTD.

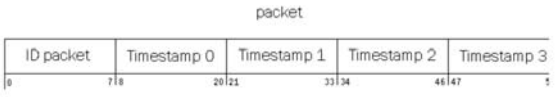


Figure 5. Structure of the travelling packet.

C. CPU occupancy estimation

The PC workload can be evaluated by means of the CPU occupancy parameter. It is defined on the basis of the delay time occurring to the CPU to answer at the system call. Obviously, higher the occupancy of the system CPU, longer is the delay time of the answer. Low CPU occupancy parameter means that the PC can support other measurement tasks. The estimation of the CPU time response is performed by using system library tied to the operating system. The technique is based on the intelligent probe packet sent by the PC client. As an example, in Fig.1 the PC#1 can be the object of the

CPU occupancy evaluation and it must be configured as server. Consequently, another PC, as an example the PC#m, must be configured as client. Each probe packet operates according to the following steps:

- n consecutive request to the system clock are sent,
- the first and the n-th answer containing the timestamps of the clock are taken into account and stored,
- the CPU occupancy parameter is evaluated as difference between the two timestamps stored.

In order to make the CPU occupancy parameter independent from the clock frequency of the system CPU, the final value is furnished as the percentage ratio between the time interval defined from the two timestamps stored and the maximum time delay of the CPU response. The reliable solution is accomplished by stepping outside Java and writing C code lines able to be used in the different operating systems [26] (Windows 2000/XP and Linux), and integrated with Java application [19], [24] via Java Native Interface.

III. Operational Security Evaluation of Intrusion Tolerant System

The analysis of system security can be related to the mechanism of system dependability analysis. Strong connection is between the concept of system security and that of system dependability [26]. Concepts defined in the system dependability analysis can be mapped into the system security analysis. Example of the mapping is the following:

- *input space*, defined in the dependability analysis as the totality of the system input, can be mapped into the security field as the totality of the admissible input devoted to the normal functionality of the system or to malicious intent;

- *usage environment*, defined in the dependability analysis as the mechanism selecting the inputs from the input space, can be mapped in the security field as the set of permitted or malicious user request.

- *system failure*, used to quantify the dependability of the system, can be mapped in the security field as system breach, that represents the evaluation of the probability of operative block toward the failure.

- *load*, defined in the dependability analysis as the stress in time unit of the system, can be mapped in the security field as the attacker effort to reach his goal.

On the basis of the relation among the system dependability concepts and the security system concepts, it is possible to use the mechanisms devoted to the dependability estimation to define and quantify the parameters to measure the system security.

The ITS can be described by the direct state transition diagram (Fig. 6). The model proposed in [27] is modified by including different states to take into consideration the modality of the attack and the system redundancy.

Initially the system is in the Good state G, and no vulnerabilities are found. If the attack founds the vulnerability, the system state changes into the Vulnerability state V. During this time interval the attacker tries to compromise the system security. It is possible that the system administrator detects and resolves the attack, than the state of the system comes back to G. Otherwise the system state changes in attacked state A. In this state the IDS can detect the anomaly by changing the system state in TRiage state TR, or, if the anomaly is not detected, the state is the Un-Detected state UD. In the TR state there are others two possibilities of evolution: (i) the state changes in Service Degradation state GD, and (ii) if the redundancy component is present, the state can change in Masquerade state MC. If the detection system doesn't recognize the attack typology, the ITS change the TR system state in the Fail-Secure state FS.

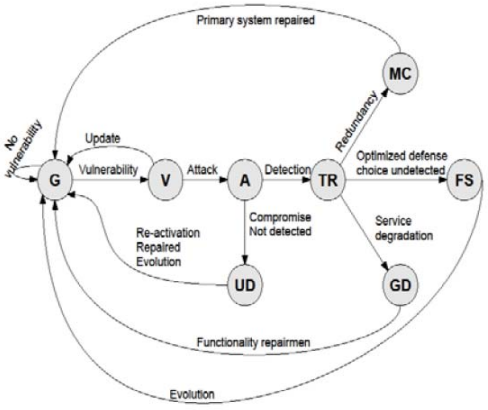


Figure 6. State transition diagram describing the behavior of the intrusion tolerant system.

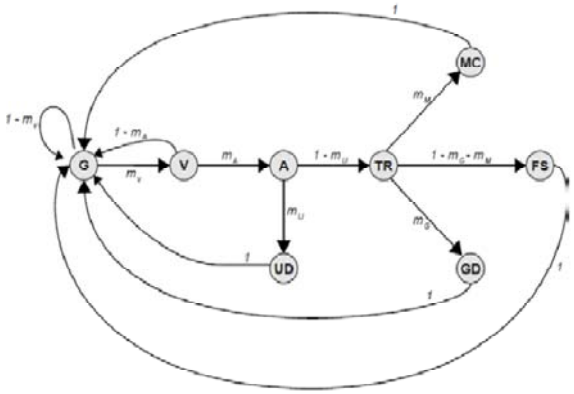


Figure 7. Directed graph describing the Discrete Time Markov Chain.

A. Measurement procedure

The direct graph shown in Fig.7 is used to evaluate the security system parameters of availability, integrity, confidentiality, and the security index Mean Time To Security Failure (MTTSF). It is taken into consideration that the failure rate has Weibull distribution. This distribution allows to represent the scenarios where the failure rate is increasing, decreasing, and constant [28].

The ITS evaluates the presence of illegal activity with the hypothesis of exponential distribution of the detection rate. If an anomaly is detected, the apparatus runs the defense procedure with the goal to came back in correct functionality state. With the assumption of non exponential distribution of the state transition, the stochastic process can be described by the continuous time semi-Markov chain.

By considering S the state set, and Ω the probability space, it is possible to define the following random variables:

X_n: Ω→$S(1)$

T_n : Ω→N

where X_n represents the occupied state of the transition n, and T_n the time instant in witch occurs the event.
Defined m_{ij} the transition probability between the state i-th to j-th, in the steady state condition, it is [29]:

$$m_{ij} = \lim_{t\to\infty} q_{ij} \quad (2)$$

The transition matrix M = [m_{ij}] describes the Embedded Discrete Time Markov Chain (EDTMC). The elements of this matrix are represented by the labeled directed graph shown in Fig. 7. The meaning associated to these elements is denoted by the parameters described in the Table1.

TABLE 1 - PARAMETER DESCRIPTION OF EMBEDDED DISCRETE TIME MARKOV CHAIN.

Parameter	Description
m_v	Existing vulnerability
m_A	Exploit start
m_U	Attack not detected
m_G	Service degradation
m_M	Redundancy service active.

The chain of Fig. 7 is ergotic, and only one probability vector (π_G, π_V, π_A, π_{UD}, π_{TR}, π_{GD}, π_{MC},π_{FS},) is available [30]. In particular, the probability of each state of the EDTMC in steady state is according to [31].

According to the transition state diagram of Fig.7, the parameter evaluating the availability (A) of the system in the full service is:

$$A = 1 - (\pi UD + \pi FS) \quad (3)$$

Without information about the characteristics of the system, the parameter evaluating the integrity Itr can be assumed coincident with that evaluating the confidentiality C. The two attributes A and Itr can't be used with absolute certainty to characterise the security level if the typology of the system is unknown.

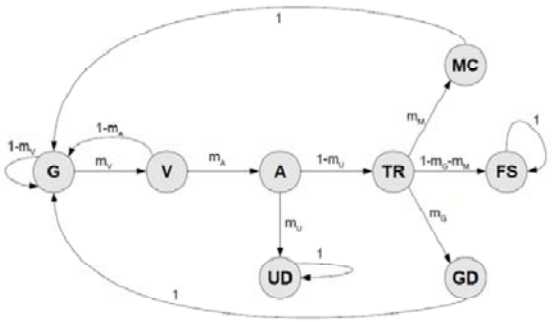


Figure 8. Labeled directed graph describing the Embedded Discrete Time Markov Chain.

By considering the absorbing Markov chain, Fig.8, the matrix M can be renumbered considering that the transient state came first, followed by the absorbing states. By taking into consideration this renumbering, the canonical form of the M matrix is:

$$M = \begin{bmatrix} T & R \\ O & I \end{bmatrix} \quad (5)$$

where, I is the identity matrix, T includes the transient state transition, and R includes the absorbing state transition [30]. The dimension of T is (NtrxNtr), with Ntr number of transient states, the dimension of R is (NtrxNabs), with Nabs number of absorbing states. The dimension of I is (NabsxNabs). Therefore, the dimension of the matrix M is (Ntr+Nabs)x(Ntr+Nabs). From the matrix M is computed the fundamental matrix F that give the expected number of times that the process is in a transient state, starting from another transient state [30]. F is defined as:

$$F = (I - T)^{-1} \quad (6)$$

In this paper, the start state taken into consideration is G, and the absorbing states are UD and FS. The other states are transient. With the calculation of F is possible to estimate the resilient time f_j into transient state before the state changes into the absorbing state.

Finally, on the basis of the strong connection between dependability and security, previous discussed, the first row of the matrix F multiplied by the elements of the vector :

$$(h_G, h_V, h_A, h_{UD}, h_{TR}, h_{GD}, h_{MC}, h_{FS}), \quad (7)$$

where the elements represent the mean residence time in the state (G, V, A, UD, TR, GD, MC, FS),

respectively, furnishes the relation for the evaluation of the parameter MTTSF [31]. It is:

$$MTTSF = \sum_{j \epsilon S_{TR}} f_j h_j = \frac{h_G (m_A m_V)^{-1} + h_V m_A^{-1} + h_A + \tilde{m}_U (h_{TR} + m_G h_{GD} + m_M h_{MC})}{1 - \tilde{m}_U (m_G + m_M)} \quad (8)$$

IV. Experimental Tests

Experimental tests were carried out in order to assess the reliability and the effectiveness of the four methods. Different operating conditions of both the connecting network and the PCs are examined.

A. OWD and RTD measurement

The first test concerns with the OWD and the RTD measurements on two different WANs, in the following denoted by WAN1 and WAN2. Both the two WANs have the same client node and different server node. Each WAN includes optical fiber three, switched communication lines, routers, switches, and modem connection. After the clock synchronization between the client and server by means of the NTP procedure, both the OWD and the RTD measurements are executed by sending 1000 probe packets. The network schemes to be taken into account in these measurements correspond to the two different paths that can be detected in the Fig.1. The first refers to the OWD measurement and is constituted by the connecting network#1, the connecting network #b, and the connections α into internet. The second refers to the RTD measurement and includes the first one with added the connections β for the return back of the packets. Fig. 9 a), and c), show the OWD and RTD measurement executed on the WAN1 for each probe packet, Fig.9 b), and d) the corresponding occurrences of the time delays. Fig. 10 a), and c), show the OWD and RTD measurement executed on the WAN2 for each probe packet, and Fig. 10 b), and d) the corresponding occurrences of the time delays. It can be noted that the occurrence distributions for the WAN1 have the Gaussian shape, in the contrary for the WAN2 the distribution is far from the previous one. This is a consequence of the fact that the working conditions of the WAN2 are congested, if compared to the working conditions of the WAN1.

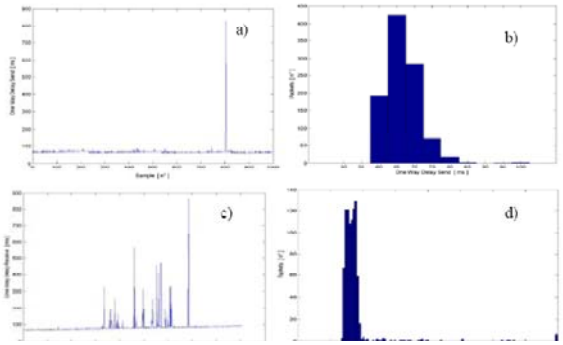


Figure 9 a), c), OWD and RTD measurement executed on the WAN1 for each probe packet, and b), d), the corresponding occurrences of the time delays.

B. CPU occupancy

The second ones is the CPU occupancy measurement on the PC configured as server. This measurement is not dependently from the connecting networks. The experimental tests are executed by sending 1000 probe packets at the time interval of 1s from the client to the server. Both the server and the client are equipped by platform Windows 2000. Fig. 11 shows the trend of the CPU occupancy parameter versus the successive probe packets in the working condition characterized by about 50% of occupancy.

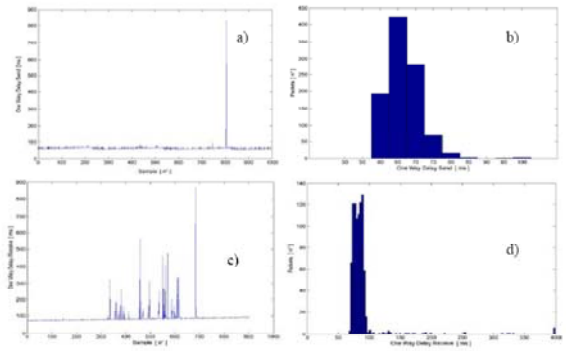


Figure 10 a), c), OWD and RTD measurement executed on the WAN2 for each probe packet, and b), d), the corresponding occurrences of the time delays.

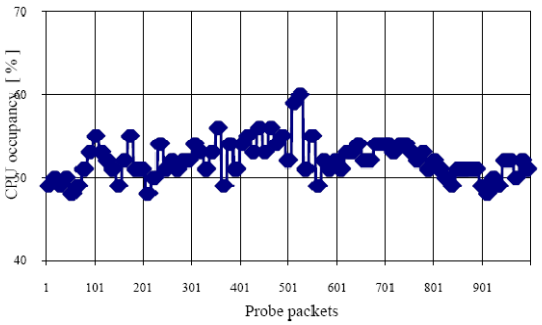


Figure11. CPU occupancy measurement.

C. System security measurement

The attack to the system security can be classified versus different aspects: (i) target, (ii) technique, and (iii) attacker typology. The Information Assurance

Program (IAP) [32] gives the guidelines to classify the different threat toward the system security and provides the base to identify the different case studies. In order to validate the proposed approach and the measurement procedure, three different case studies are chosen: (i) E1, high protection level, (ii) E2, medium protection level, and (iii) E3, low protection level. Tab.2 and Tab.3 show the correspondent values assumed for the transition probability, and the mean residence time, for each security test.

From (3) can be evaluated that the system can correctly operate for 98% of the time in the case E1, 89% of the time in the case E2, and 78% of the time in the case E3. From (4) can be evaluated that the integrity (I) of sensible data is 6% in the case (E1), 17% in the case (E2), and 26% in the case (E3).

According to the values of transition probability (Tab.2) and the mean residence time (Tab.3), the values obtained for MTTSF are shown in Tab.4. As expected, the value of MTTSF decreases as decreases the security system level. By taking into consideration the second scenario E2, Fig. 12 shows the trend of the MTTSF versus the mean residence time h_G evaluated in relative time unit and the attack probability m_A. As the attack probability increases, the MTTSF decreases. Moreover, as the residence time decreases, the MTTSF decreases.

Table 2 - Transition probability.

	m_V	m_A	m_U	m_G	m_M
E1	0.5	0.4	0.2	0.3	0.3
E2	0.7	0.6	0.5	0.2	0.2
E3	0.9	0.8	0.7	0.1	0.1

Table 3 - Mean residence time in relative time unit.

	h_G	h_V	h_A	h_{UD}	h_{TR}	h_{GD}	h_{MC}	h_{FS}
E1	60	40	30	20	50	40	40	20
E2	30	60	50	40	70	60	60	40
E3	10	90	80	70	90	80	80	70

Table 4 - MTTSF.

	MTTSF
E1	941
E2	335
E3	253

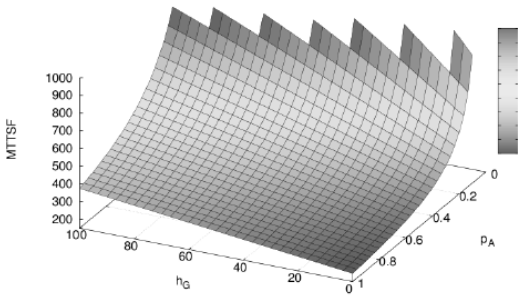


Figure 12. MTTSF versus h_G and m_A.

V. CONCLUDING REMARKS

In the paper the methods are presented for evaluating the workload of the connecting networks and the PCs of the DMLs, and for measuring the system security.

The proposed methods are based on four different tools known and tested in literature. The first one concerns with the synchronization among the clocks of the different co-operating PCs. The second one concerns with the One-Way Delay parameter measurement of the path from the source to the destination. The third one concerns with the workload measurement of the CPU of each PC. The fourth one concerns with the evaluation and the measurement of the ITS operational security, It is based on the practical assumption that the system security analysis has strong connection with the system dependability analysis.

The experimental results show the validity of the proposed methods and make its useful in order to monitor the functioning conditions of the whole DML.

Moreover, practical criteria can be drawn from the analysis of these experimental results to point out performance management strategies and criteria for the security system design. In particular:

- the lower bound of the correlation in the time domain of the measurements into the DMS can be established a priori by basing on the fact that the effectiveness of the synchronization increases as decreases the time interval to repeat the procedure;
- the reliable range of the delay time of the communication into the DMS can be evaluated a priori by considering the paths characterized by the Gaussian shape of the occurrences of the delay time of the probe packets. Indeed, in this path the reliable range of the delay time can be estimated by the standard deviation of the mean in the Gaussian distribution;
- the queue can be managed by selecting the connecting paths into the DMS characterized by the shorter value of both the mean and the standard deviation of the mean;
- the security system set up can be a priori established on the basis of the established protection level required and the value of the parameter Mean Time To Security Failure.

REFERENCES

[1] Grimaldi D., Nigro L., Pupo F., "Development issues of distributed measurement systems", Proc. of Imeko TC-4, Budapest, pp.111-114, September 1996.

[2] Arpaia P., Cennamo F., Daponte P., Savastano M., "A distributed laboratory based on object oriented systems", Proc. of IMTC/96, Brussels, 4-6 June, pp. 27-32, 1996.

[3] Arpaia P., Baccigaluppi A., Cennamo F., Daponte P., "A distributed measurement laboratory on geographic

network", Proc. of IMEKO TC-4, Budapest, 10-12 September, pp.294-297, 1996.

[4] Arpaia P., Baccigaluppi A., Cennamo F., Daponte P,. "A distributed measurement laboratory on geographic network for remote test experiments", Proc. of IMTC/98, St. Paul, Minnesota (USA), May 18-21, pp.206-206A, 1998.

[5] Daponte P., Grimaldi D., Nigro L., Pupo F., "Distributed Measurement Systems: an Object-Oriented Architecture and a Case Study", Computer Standards and Interfaces, vol.18, No.5, pp. 383-395, June 1997.

[6] Grimaldi D. Nigro L., Pupo F., "Java Based Distributed Measurement Systems" IEEE Trans. on Instrum. and Measur, vol. 47, No. 1, pp.100-103, February 1998.

[7] Fortino G., Grimaldi D., Nigro L., "Multicast control of mobile measurement systems", IEEE Trans. on Instrum. and Measur, vol. 47, No. 5, pp.1149-1154, October 1998.

[8] Grimaldi D., Marinov M., "Distributed measurement systems", Measurement, 30, pp.279-287, 2001.

[9] Daponte P., Grimaldi D., Marinov M., "Real-time measurement and control of an industrial system over the Internet: implementation of a prototype for educational purposes**",** IEEE Trans. on Instrumentation and Measurement, August 2002.

[10] Fortino G., Grimaldi D., Nigro L., "An Agent based measurement laboratory over internet", Proc. of IEEE Autotestcon'99, S. Antonio, Texas, USA, Aug. 30–Sept. 2, 1999.

[11] Jun Lu, Yu Wang, Zhongwang Wu, Yu Lu, "Network behaviour description and behaviour base modeling method", Multiconference on Computational Engineering in Systems Appl., IMACS, pp. 1353 – 1356, February 2007.

[12] Chiong R., Dhakal S., "On the insecurity of Personal Firewall", Intern. Symposium on Information Technology, ITSim 2008, vol.4, pp. 1 - 10.

[13] Al-Shaer E., Hamed H., Boutaba R., Hasan M., "Conflict classification and analysis of distributed firewall policies", IEEE journal on selected areas in communications, vol. 23, N. 10, October 2005.

[14] Avižienis A., Laprie J.C., Randell B., "Fundamental concepts of dependability", LAAS Report n. 01-145, 2001.

[15] Mills D. L.: "Internet Time synchronization: the Network Time Protocol", Electrical Engineering Department University of Delaware..

[16] RFC 768 - User Datagram Protocol - J. Postel ISI 28 August 1980.

[17] Gurewitz O., Sidi M., "Estimating One-Way Delays from cyclic-path delay measurements", IEEE INFOCOM, pp.1038-1044, 2001.

[18] Pàsztor A., Darryl Veitch: "High Precision Active Probing for Internet Measurement"–http://www.emulab.ee.mu.oz.au/~attila/pub/inet2001.pdf.

[19] Iarantsev A. "Performance meter and memory leaks detector", Dec 2001, http://www.codeproject.com/debug/compinfo.asp.

[20] Nicol D. M., Sanders W. H.,. Trivedi K. S, "Model-based evaluation: From dependability to security", IEEE Transactions on Dependable and Secure Computing, vol. 1, pp. 48–65, January-March 2004.

[21] Jonsson E., Stromberg L., Lindskog S., "On the functional relation between security and dependability impairments", Proceedings of the New Security Paradigms Workshop 1999, Sep. 22–24 1999.

[22] Ortalo R., Deswarte Y., Kaaniche M., "Experimenting with quantitative evaluation tools for monitoring operational security", IEEE Transactions on Software Engineering, vol. 25, no. 5, pp. 633–650, Sept/Oct 1999.

[23] Grimaldi D., "Workload measurement method into a distributed measurement laboratory", Measurement, vol.39, N.7, 2006, pp.655-663.

[24] Prentice Hall - Sun's Core Java 2, Volume II Advanced Features, 5th Edition – Chapter 11 Native Methods.

[25] Bradford E.G.: "RunTime: High-performance programming techniques on Linux and Windows 2000"-Senior Programmer, IBM - 1 April 2001

[26] Littlewood B., Brocklehurst S., Fenton N., Mellor P., Page S., Wright D., "Towards operational measures of computer security", Journal of Computer Security, vol. 2, pp.211–229, 1993.

[27] Wang D., Madan B., Trivedi K. S., "Security analysis of SITAR intrusion tolerance system", Conf. on Computer and Comm.s Security, 2003.

[28] DeLong T. A., Smith D. T., Johnson B. W., "Dependability metrics to assess safety-critical systems", IEEE Trans. on Rel., vol. 54, n.3, pp. 498 – 505, Sept. 2005.

[29] Haver Kort B.R., Niemegeers I.G., "Performability modelling tools and techniques", Perf. Ev., vol. 25, pp. 17-40, 1996.

[30] Trivedi K. S., "Probability and statistics with reliability, queuing, and computer science applications", John Wiley&Sons, 2001.

[31] Grinstead C. M., Snell J. L., "Introduction to probability", AMS, 1997.

[32] Lowry J., "An initial foray into understanding adversary planning and courses of action. DARPA", Information Survivability Conference and Exposition (DISCEX II), vol. 1, pp. 123–133, 2001.

Vehicular Ad Hoc Network Paradigm and Evaluation of IEEE 802.11a and IEEE 802.11p Using Realistic Vehicular Traces

Joarder Mohammad Mustafa Kamal, Mohammad Shahidul Hasan,
Alison L. Carrington, and Hongnian Yu

***Abstract*- In recent years, the vehicular communication has become an innovative and sophisticated way to provide necessary information to the rural or urban travelers in roads and motorways. Through the idea of multi-hop ad hoc networking, it is possible to efficiently disseminate traffic related information to the drivers and utilise the information collected from on-board sensors from neighbouring vehicles to provide more safe travel to the passengers. This paper presents an evaluation of two proposed wireless standards for vehicular network communication – IEEE 802.11a and IEEE 802.11p using TCP and UDP data transmission. This paper also investigates various vehicular mobility models and traffic generators for simulations and several well-known routing protocols for inter-vehicular communication. The simulation uses AODV and DSR routing protocols in a realistic vehicular environment using a real-world topological map extracted from TIGER data set. VanetMobiSim is used to generate realistic mobility model and the wireless network is simulated using the dominant network simulator ns-2. From the simulations it is found that IEEE 802.11p performs better than IEEE 802.11a in case of TCP transmission while performs almost similarly in UDP transmission.**

Index Terms – network evaluation; network paradigm; vehicular network communication; wireless standard evaluation

I. INTRODUCTION

Nowadays travel, traffic and transportations have become an integral part of our daily life. Real-time traffic data collection and dissemination is able to provide backend support to the end users' applications and services e.g. active navigation. Again with the increasing number of vehicles on roads government organisations and vehicle manufacturers need to provide sufficient measures in both planning and development on traffic management and ensuring public safety. The main concern is the traffic data dissemination in a more appropriate and precise way which can be used for real-time decision making. Intelligent Transportation Systems (ITS) addresses the challenges faced in traffic information collection and dissemination, advanced highway signaling, real-time traffic monitoring and surveillance, mobility data mining and knowledge discovery and a large number of internet-based applications providing entertainment and multimedia services. All of these ITS technologies depend on the efficiency of the communication techniques between vehicles and roadside infrastructures. Vehicular Ad hoc Network (VANET) is nowadays in a more focused stage through real-life implementations and academic researches. Although the primary reason of interest behind VANET research only emphasises the traffic and road safety but it has opened new windows for internet access, distributed computing, delay-tolerant networking, e-commerce etc. Although many promising applications e.g. congestion avoidance, emergency road maintenance notifications etc. are seen today to use the single-hop point-to-point VANET but it is still a challenge to implement real-life applications that will utilise multi-hop ad hoc networking technique. This paper investigates the latest vehicular communication technology paradigm, mobility models and mobility generators, routing protocols, simulation tools and major performance criteria. The rest of the paper is organised as follows: section II gives the background insight of VANET architecture. Section III presents mobility models and vehicular traffic simulators which are widely used to simulate and measure the performance of VANET. Section IV discusses different VANET routing protocols and their characteristics. Finally section V presents a simulation work which uses two well-known VANET simulation tools VanetMobiSim [1] and ns-2 [2] to evaluate the performance of IEEE 802.11a and 802.11p using two popular MANET routing protocols (Ad hoc On-demand Distance Vector (AODV) [3] and Dynamic Source Routing (DSR) [4]) with realistic vehicular traffic traces. Section VI includes some conclusions, challenges and future work directions.

II. BACKGROUND

The IEEE 802.11p draft amendment to the popular IEEE 802.11 standard focuses on the enhancements of physical medium and medium access techniques to ensure inter-vehicular and roadside communications. It includes the 5.9 GHz

ISBN 978- 974-672-556-9

licensed ITS band and enables Dedicated Short Rage Communications (DSRC) channels which is specially designed for one-way or two-way vehicular communications [45]. Recently multi-hop ad hoc networking opened a new era in inter-vehicular communications (IVC), vehicle-to-vehicle (V2V) and infrastructure-to-vehicle (I2V) communications. VANET architecture can be described in three different categories: pure WLAN/Cellular, pure ad hoc and hybrid. In the pure WLAN/Cellular architecture, access points or base stations are able to provide connectivity to the vehicles.

In pure ad hoc network all vehicles are responsible to construct and maintain the network without any network infrastructure. In hybrid architecture, vehicle which has both WLAN and Cellular networking capabilities act as the gateways or routers for other vehicular nodes. As vehicles can move at high speed, it is much harder to construct and maintain the communication network. Thus vehicular nodes frequently experience node disconnection, lost route and re-discovery problems. Delay tolerant and opportunistic routing therefore seems to be a better choice for VANET. Energy and power management which is a major concern for many ad hoc network types is not a challenging issue for VANET because of the onboard vehicle battery power supply. Similarly processing and storage capability obstacles can also be handled if vehicles are equipped with on-board computing devices. The availability of the computing devices also ensures on-demand, multimedia and roadside business applications. Another important characteristic of VANET is that it can be used for vehicular traffic mobility prediction. As in most of the big cities and highways, the vehicle mobility patterns are fixed through lane separation, traffic lights, speed cameras etc. it is convenient to use mobility data mining techniques to provide improved traffic management services. Mobility data mining could provide valuable knowledge about predictive movement, future positions based on daily, weekly and monthly movement patterns of vehicles.

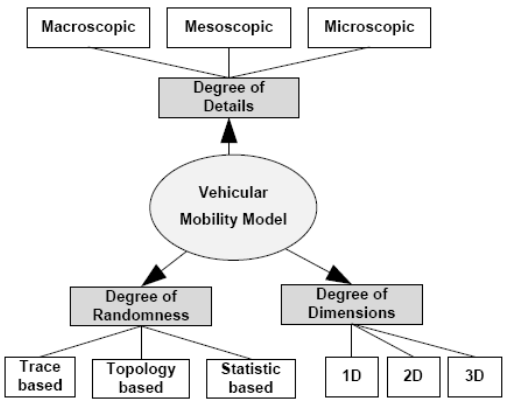


Figure 1. Different aspects of vehicular mobility model generation

III. MOBILITY MODELSFOR VEHICULAR ADHOC NETWORK

To design, model and simulate any VANET architecture it is necessary to have realistic network model and traffic data or mobility pattern taken from real scenarios in daily life. As conducting real-life experiments is not always possible due to proper environment, safety, setup cost, equipments etc., simulation is the only feasible way to test and evaluate network protocols for VANET. Fig. 1 shows different aspects of a mobility model. Macro and micro mobility features are the two main categories for vehicular mobility attributes. The *macro-mobility* features include road topology, road structure, lane formation, speed limits, restrictions, traffic signs etc. It also considers the effects of points of interests which exhibit particular mobility patterns for vehicles. The *micro-mobility* features include individual vehicle, driver behaviours based on sex, age and mental conditions; driver's interactions with other drivers, with the traffic signs and various driving conditions; vehicle acceleration, deceleration, overtaking criteria etc. *mesoscopic-mobility* feature [9] describes the traffic flows from an intermediate level between the macro and microscopic features. Fig.1 shows a breakdown of various degrees and levels of categorisation for vehicular mobility model generation. A comprehensive discussion on VANET mobility models can be found in [5]-[8].

The multilayer description of vehicular mobility patterns consist of trip modeling, path modeling and flow modeling [9] and based on these criteria the authors categorise the mobility models as random models, flow models, traffic models, behavioural models and trace-based models. A concept map is presented in [5] which states two primary building blocks – Motion Constraints and Traffic Generator which are linked together with time patterns. Motion Constraint also employs Topological Maps which also includes Speed Constraints, Attraction Points and Obstacles. On the other hand, Traffic Generator is further decomposed into Car Generation Engine and Driver Behaviour Engine. These decompositions also include car's type and particulars, centres of interest, social habits, mobility predictions and driver's danger assessments.

Random models e.g. Random Waypoint (RWP) model, Random Walk model (RWalk), Reference Point Group model (RPGM), node following model and Gauss-Markov model are popular choices of many research works for both (Mobile Ad-hoc Network) MANET and VANET. Although these random models are widely used within the research community, these are not able to generate realistic traffic data for vehicular network simulations [5]. The few first attempts to make a realistic mobility model are through the introduction of Simple Freeway model and Manhattan (or Grid mobility) model. Simple Freeway model restricts vehicle's

ISBN 978- 974-672-556-9

movement into several bi-directional multi-lane freeways while the Manhattan Grid mobility model restricts the movement on urban grids [5]. But these models do not consider the macro and micro mobility features [5]. Many recent traffic generation tools are capable of generating realistic traffic and mobility data for vehicles.

IMPORTANT [10] and Java based BonnMotion [11] tools implement several variations of random mobility models while considering only the macro-mobility features. IMPORTANT only features the Car Following model which features car-to-car inter-distance control a specific type of micro-mobility attribute. The generated scenarios can also be exported into several well-known network simulators like ns-2 [2], GloMoSim [12], QualNet [13] etc. Mobility Model Generator of Vehicular Networks (MOVE) [14] adds the TIGER/Line (Topologically Integrated Geographic Encoding and Referencing system) [15] map (available from U.S. Census Bureau) extraction capability as well as random and manual mobility traces generation. This map parsing and mobility trace generation schemes also add improved capability into SUMO [16] vehicular mobility simulator. Both the Street Random Waypoint (STRAW) [17] tool and GrooveSim [18] are capable of parsing TIGER data files. STRAW implements an intersection management scheme using traffic signs and traffic lights. GrooveSim mainly introduces non-uniform distribution of vehicles speed on the roads considering motion constraints and speed limitations. Therefore, vehicles are not able to maintain the initial velocity set by the model.

MobiREAL [19] mainly focuses on pedestrian mobility showing a guideline and direction for future vehicular mobility model design. UDel model [20] is a set of tools for generating urban mobility along with the calculation of radio propagation. It works based on the statistical data obtained from the U.S. Department of Labour and is capable of parsing Geographical Information System (GIS) data which makes it more realistic while producing radio signal propagation information. SHIFT [21] traffic simulator, developed within the PATH [30] project, is a complete microscopic tool which generates mobility traces according to validated mobility models. Voronoi Model [22] is based on the voronoi graphs which utilises voronoi channels to represent roads and other spatial area based on Voronoi Tessellation algorithm [22]. It introduces global moving direction and local direction patterns for vehicular mobility thus it mainly improves the motion constraints mentioned previously. The obstacle mobility model [23] utilises random building corners and voronoi tessellations in order to identify the movement path between buildings. It also includes a radio propagation model; wireless communication and movements are restricted using the paths identified by the voronoi graph which is based on the presence of individual obstacles.

The CanuMobiSim [24] is a Java-based flexible user mobility modelling tool which is able to generate mobility traces for ns-2, QualNet, GloMoSim. While most of the mobility trace generation tools only consider macro-mobility attributes CanuMobiSim considers micro-mobility attributes which enables it to generate more realistic mobility traces. It implements several car-to-car interaction models like Fluid Traffic model [24], Intelligent Driver Model (IDM) [24] etc. The tool also includes a complex traffic generator that can utilise source-destination based path calculation using Dijkstra's shortest path algorithms or can also model trips between attraction points. It also identifies a separate class of users or drivers and their individual motion patterns. Extractions of spatial information from Geographical Data Files (GDF) or TIGER data sets are also possible.

CanuMobiSim is primarily focused on general purpose MANET. To extend its capability towards the VANET VanetMobiSim [25] is introduced as an extension of it. The model is the pioneer to consider the mobility patterns of a vehicle through a driver's point of view. To define road topology, VanetMobiSim introduces user-defined graph and spatial data extraction from GDF map, TIGER map and clustered voronoi map. Road topology is characterised by introducing multiple lanes in both directions, physical separation of traffic flows for opposite directions, traffic signs or traffic lights and speed limits. Trip generation is based on either random motion or activities sequencing which consists of multiple sets of ―start‖ and ―start and stop‖ points. VanetMobiSim introduces three categories of micro-mobility models like considering mobility behaviour in a deterministic way or a function of nearby vehicles in either a single lane or multiple traffic flows. The Graph-Based Mobility Model (GBMM), the Constant Speed Motion (CSM) and the Smooth Motion Model (SMM) originally introduced by CanuMobiSim fall under the deterministic categorisation. The Fluid Traffic Model (FTM) and Intelligent Driver Model (IDM) falls under the single lane or multiple traffic flows categorisation. To model realistic vehicular mobility patterns, a tight relationship is maintained between the traffic generation mechanism and topological map. Therefore, while a driver approaches towards a traffic signal it slows down and acts as per the traffic light's indication. Again a close relationship is maintained according to the traffic signs and state of the traffic lights and other neighbouring vehicles activities. To model intersection management schemes VanetMobiSim introduces Advance Intelligent Driver Model (AIDM) which introduces acceleration and deceleration mechanism in the road intersection points. The Intelligent Driver Model with

Intersection Management (IDM-IM) and Intelligent Driver Model with Lane Changes (IDM-LC) are the inherent models from AIDM. Furthermore, the IDM-LC model is actually extends the IDM-IM model through introducing lane changing model. Minimising Overall Breaking deceleration Induced by Lane changes (MOBIL) model, which is another interesting feature in this tool to model lane changes and maintain compatibility with AIDM also. More details on VanetMobiSim and its features and a comprehensive validation can be found in [7].

IV. Routing Protocolsfor Vehicular Ad Hoc Network

Unicast routing protocols for vehicular ad hoc networks could be categorised broadly into two categories; topological and geographical as shown in Fig. 2. Proactive and reactive protocols maintain link state routing tables or discover route on-demand. Geographic routing mechanisms use location of the source node and its neighbouring nodes to make routing decisions by utilising neighbourhood discovery process. In [26] the authors classified geographic routing protocols into three categories; non-Delay Tolerant Network (non-DTN), Delay Tolerant Network (DTN) and hybrid. The non-DTN does not consider network partitioning or disconnectivity while the DTN type routing protocols consider this issue and act accordingly. Hybrid protocols use both kinds of measures to handle partial network connectivity and temporally disconnectivity.

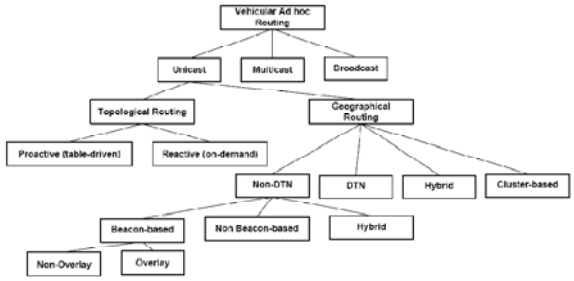


Figure 2. Classification of Vehicular Ad hoc Networks Routing Protocols

Greedy Perimeter Stateless Routing (GPSR) [27] is one of the widely discussed geographic routing protocols in literature. A node greedily forwards network packets to a neighbouring node that is geographically close to it. A local maxima is reached when there is no direct communication path between the source and forwarding nodes due to the presence of obstacles e.g. buildings and trees or there is no other neighbour node which is closer to the destination node. If such a situation is encountered, GPSR uses face routing and right hand rule to go around and again trying to resume in greedy mode. As GPSR uses planer graph to build the network, therefore routing loops can occur. Again, mobility introduces a great performance impact on GPSR because of frequent network partitions and disconnectivity. GPSR with Advanced Greedy Forwarding (AGF) [28] is an improvement over GPSR which solves two problems. By increasing the frequency of beaconing it solves the problem of having outdated information at each node. And by introducing speed, direction and total time to travel into the beacon packet, every node computes the deviation of the destination node's estimated current position from its previous position.

Greedy Perimeter Coordinating Routing (GPCR) [29], a beacon-based overlay routing protocol, utilises nodes at the junctions or intersections of roads which follow a natural planer graph. It represents the planer graph using underlying roads and nodes using both greedy and perimeter routing along the edge. Upon reaching the junction (J) a coordinating node guides the packets to the next edge of the planer graph using the right hand rule. In a realistic network simulation in ns-2, it has been shown that GPCR performs better than GPSR with high packet delivery rate [29]. Geographic Source Routing (GSR) [29] assumes to have static city map information which will provide global topological knowledge about the total network. Source node can determine the junction nodes using the map information and directly forwards packets to them in the road intersections. Therefore, GSR performs more accurately in the city areas and has better packet delivery rate with low bandwidth consumption comparing to two well-known topological routing protocols AODV and DSR [30]. Similarly like GSR, Anchor-based Street and Traffic Aware Routing (A-STAR) [31] also utilises city street map information to compute a series of junction points in advance. But A-STAR selects the anchor points based on the traffic flow along the street. It chooses two types of paths – one is along the bus routes which indicate the static path for traffic and another one is dynamically rated path with latest traffic information. When nodes fall into local maxima it computes another anchor path immediately while marking that region as "out-of-service" for other network packets. It remains in "operational" state after a timeout period.

Street Topology Based Routing (STBR) [32] is a beacon-based overlay routing scheme which computes the road connectivity at the junction nodes selecting a junction node as a master. The master nodes exchange information with each other. Thus these are able to sense either the next junction node for a packet delivery is up or down. Unlike GSR or A-STAR, STBR computes its route based on geographic distance. Another overlay routing mechanism is Greedy Traffic Aware Routing (GyTAR) [33] where a junction node receives a packet and will decide which will be the best junction node to forward to it. It also assumes the number of cars on a road from the roadside units and thus

determines the connectivity, traffic density and physical distances from each other. Based on this information GyTAR marks the neighbouring junction nodes with weights and in time of making a packet forwarding decision it utilise these weight values accordingly.

Landmark Overlays for Urban Routing Environments (LOUVRE) [34] is a geo-proactive beacon-based overlay routing technique which determines the sequence of overlaid junction nodes in advance. Considering a threshold of vehicular density on a road it chooses a connected link thus not taking the spatial information of the road into account. Thus it decreases the delay of computing overlay routes and increases the global route optimality while fails to scale as much as is expected. Topology-assist Geo-Opportunistic Routing (TO-GO) [35] is non-DTN hybrid routing which acquires 2-hop neighbour information to select the best target node based on the greedy mechanism to forward network packets and introduces opportunistic forwarding. The protocol always chooses the target node instead of the destination node. It is unlikely that the destination is in the same street as the forwarding node or the source node. As the packet is expected to travel through several junctions, network packets are opportunistically forwarded to the target nodes and those are constantly making progress towards the destination node.

Based on the predictable vehicular mobility, Vehicle Assisted Data Delivery (VADD) [36] employs opportunistic delay tolerant networking. At a junction point each node makes its decision to forward packets to the next node based on the smallest packet delivery delay. This delay is computed based on road density and distance, vehicle velocity etc. Several variations are also available which choose the forwarding node after the forwarding path is selected. Location First Probe (L-VADD), Direction First Probe (D-VADD), Multi-path Direction First Probe (MD-VADD) and Hybrid Probe (H-VADD) are some of the mentionable variations. Geographical Opportunistic Routing (GeOpps) [37] is another similar type of routing protocol that utilises vehicle's onboard navigation systems to greedily find out the next forwarding node which is close to the destination. It computes the shortest distance from the packet's destination to the nearest point (*NP*) of vehicle's moving path and the estimated arrival time to the destination. If another neighbour vehicle is found which has lower arrival time for packets towards the destination the packets are forwarded to that vehicle and repeats until it reaches the destination. GeoDTN+Nav [38] is a hybrid approach which includes both greedy, perimeter and DTN mode of operations. The greedy and perimeter mode operation is same as previously mentioned for others routing mechanism. Based on the network connectivity (by measuring the number of hops the packet has passed so far), neighbour node's packet delivery quality and neighbours moving direction, a vehicular node can determine possible delay tolerance capabilities on different paths. It can also switch back and forth between the DTN and Non-DTM mode to utilise both its delay tolerant and greedy routing capability.

Cluster-based routing protocols can be also used for VANET. A cluster is consisting of a cluster-head and other members. A cluster-head is responsible for maintaining the member information dissemination and inter-cluster communication. The main drawback is the instability of the lifetime of a cluster-head. Clustering for Open IVC Networks (COIN) [39] is a cluster-based VANET routing protocol where a cluster-head is elected estimating the vehicular dynamics and driver intentions rather than traditional cluster-head election procedure. Thus it provides a more stable virtual cluster infrastructure, increases the lifetime of a cluster-head and decreases the frequency of cluster membership changes.

Broadcast routing is essential for vehicular network communication in the case of dissemination of traffic data, emergency, congestion, weather forecast, roadside business promotion and advertisements etc. Although simple flooding technique works well in simple stable network topology, in the case of a highly dynamic VANET, more efficient broadcast mechanisms are required. BROADCOMM [41] is stated as an emergency broadcast routing protocol which utilises hierarchical network topology considering a group of vehicles as a virtual cell. These types of small virtual cells follow the moving direction of the vehicles consisting of them. There are also cell-reflectors which are the nodes located close to the geographical boundary of the virtual cells. Cell reflectors behave as a cluster-head or base station for a certain period of time. It also handles emergency messages within its own cell and the neighbouring cells and act as intermediate routers for neighbouring cells. These simple mechanisms work well in motorways. Urban Multi-hop Broadcast (UMB) [41] is another broadcast routing protocol which handles issues related to interference, collision, hidden node problem etc. during multi-hop message dissemination. The mechanism always selects the furthest nodes in the broadcast direction for forwarding and acknowledging packets without any prior information.

Geocast routing is basically location-based multicast routing. Thus in geocast routing information is delivered to a group of network nodes identified by their geographical locations and service region. Therefore, simplified multicasting techniques can be used by defining multicast groups for specific service regions. In [42] a simple geocast technique is described for inter-vehicular communications which uses selective rebroadcast technique with waiting

time to see whether it receives the same message from any other nodes or not. If it receives the same information from other neighbouring nodes before the expiry of the waiting time it will not rebroadcast the packet. A similar type of idea is also used in Inter-Vehicles Geocast (IVG) [43] protocol. In [44] another specialised geocast routing mechanism is described which will broadcast packets to the nodes which stay in the geocast service region for a certain period of time within its lifetime. It is more like a client-server based communication technique used for service oriented applications like location based services (LBS), advertising and publish-and-subscribe. This idea of periodic retransmission of geocast message is called abiding or stored geocast. Table I shows a summary of the routing protocols.

TABLE I SUMMARY OF VEHICULAR ROUTING PROTOCOLS

Routing Protocol	Type	Unique Properties
GPSR	Unicast	Non-DTN, Beacon-based, Non-Overlay, greedy forwarding, perimeter and face routing
GPSR with AGF	Unicast	Non-DTN, Beacon-based, Non-Overlay, solves the problem of routing loops in GPSR
GPCR	Unicast	Non-DTN, Beacon-based, Overlay, Junction Points
GSR	Unicast	Non-DTN, Beacon-based, Overlay, Static City Maps
A-STAR	Unicast	Non-DTN, Beacon-based, Overlay, Static City Maps with dynamic traffic rating
STBR	Unicast	Non-DTN, Beacon-based, Overlay, Junction Points as Master Nodes
GyTAR	Unicast	Non-DTN, Beacon-based, Overlay, consider traffic density, connectivity and distance
LOUVRE	Unicast	Non-DTN, Beacon-based, Overlay, assumes the sequence of overlaid junction nodes in advance
TO-GO	Unicast	Non-DTN, Hybrid, greedy forwarding and opportunistic routing
VAAD	Unicast	DTN, smallest packet delivery delay based on road density and delay
GeOpps	Unicast	DTN, calculates the Nearest Points in the moving path from the destination point
GeoDTN +Nav	Unicast	Hybrid, combining greedy, perimeter mode and DTN routing mechanism
COIN	Cluster-based	Clustered, elect cluster head based on vehicular dynamics and driver intention(Unicast)
BROAD COMM	Broadcast	Hierarchical topology, virtual cells of vehicles
UMB	Broadcast	Consider interference, packet collision and hidden node problem
Geocast	Multicast	Selective rebroadcast(Simple)
IVG	Multicast	Selective rebroadcast
Geocast	Multicast	Consider location based services(Specialised)

V. EVALUATIONOF IEEE 802.11A AND IEEE 802.11P STANDARDS

IEEE 802.11p draft standard and IEEE 1609 WAVE (Wireless Access for Vehicular Environment) standards are an emerging technology for vehicular communications and ad hoc networking operating in the 5.9 GHz frequency bands. It uses OFDM-based physical layer construction with recommended 3 Mbps data rate. While IEEE 802.11a standard which is amended in IEEE 802.11-2007 standards also use OFDM-based technology in the 5 GHz frequency band with 54 Mbps data rate. Both of these technologies utilise short range communication facility having a larger operational frequency than other popular wireless technologies in 2.4 GHz range e.g. IEEE 802.11b and IEEE 802.11n.

DSRC radio technology also fits well in IEEE 802.11a based mechanism. Therefore, an effort has been carried out since 2004 to include DSRC into IEEE 802.11a standard which resulting IEEE 802.11p draft standard [46]. Again, IEEE 802.11p includes the enhancement stated in IEEE802.11e Enhanced Distributed Channel Access (EDCA) mechanism therefore it is designed to support multiple channel access and message prioritisation using Access Categories (AC). So, IEEE 802.11p can be used for reliable data transmission e.g. TCP more efficiently [48]. Although IEEE 802.11a is a legacy standard but due to its efficiency in short range communication it is useful to make performance comparisons with IEEE 802.11p. Therefore, for short range vehicular communication both of these specifications are quite suitable in nature.

In this paper, a vehicular ad hoc network simulation has been designed using the popular network simulator ns-2 and realistic mobility traffic generator VanetMobiSim. Intelligent Driver Model (IDM) with Lane Changes is applied as the driver model and a real life topological map of the District of Columbia; USA extracted from the TIGER/Line 2006 Edition data sets is used. A 2000 x 2000 m^2 area is chosen for the simulation. Fig. 4(a) and 4(b) exhibit the VanetMobiSim screenshot and corresponding Google Map, respectively to show the similarity of the simulated area with real-life. In Table II, a list of key parameters is given which have been used in VanetMobiSim to generate realistic traffic data and produce mobility traces for ns-2. These traces are applied on two well known topological routing protocols AODV and DSR. IEEE 802.11a and IEEE 802.11p are used respectively using TCP and UDP data transmission between a source and destination node for 100 seconds. Note that, the simulation results shown in this paper are only based on IEEE 802.11a and 802.11p PHY and MAC layer enhancement available in ns-2. The IEEE 1609/WAVE specifications are not considered here. Table III shows the list of related parameters that are used in ns-2. In all cases node 0 is designated as the packet source and node 1 as the destination. For TCP data transmission, CBR traffic over TCP is used while FTP over UDP is used in another case.

TABLE II. VANETMOBISIMPARAMETERSFOR VEHICULAR MOBILITY DATA GENERATION

Parameters	Values
Region	District of Columbia, Washington, USA
Data Set	TIGER/Line files 2006 Second Edition
Dimension of Area	2000 X 2000 m2
Maximum Traffic Lights	Maximum 5 in an intersection
Number of Lanes	Maximum 4, minimum 2
Trip Generator	Random trip generator with minimum stayof 5 sec and maximum stay of 30 sec
Path Selection	Dijkstra's shortest path algorithm -default
Position Generator	Random initial position generation every time
Mobility Model	Intelligent Driver Model with Lane Changes (IDM-LC)
Lane Changing Model	Minimizing Overall Breaking deceleration Induced by Lane changes (MOBIL)
Vehicle Movement Speed	Minimum 8.33 m/s (30 km/h) and maximum 13.89 m/s (50 km/h)
Vehicle's Length (l)	5 m (Default)
Maximal Acceleration (a)	0.5 m/sec$_2$ [0.6 m/sec$_2$(Default)]
Comfortable Deceleration (b)	0.5 m/sec$_2$ [0.9 m/sec$_2$(Default)]
Jam Distance (s0)	1 m (Min.distance to a standing vehicle)
Recalculation time of movement parameters	0.5 s (step)
Maximum Safe Deceleration	4 m/sec$_2$(Default)
Driver's Politeness Factor (p)	0.7 [0.5 (Default)]
Threshold acceleration (athr)	0.2 m/sec$_2$(Default)
Number of Vehicles	20
Simulation Time	100 sec

TABLE III. NS-2 PARAMETERSFOR NETWORK SIMULATION

Parameters	Values	
ns-2 Version	2.34	
PHY and MAC	IEEE 802.11a and 802.11p with 80211PHYEXT	
Propagation	Shadowing model with path loss	
	IEEE 802.11a	***IEEE 802.11p***
Sensitivity	-82 dBm	-85 dBm
Frequency	5.18 GHz	5.9 GHz
Bandwidth	-96 dBm for 10 MHz bandwidth	-99 dBm for 10 MHzbandwidth
Power Monitor Sensitivity	-99 dBm	-102 dBm
Header Duration	20 µs	40 µs
Antenna Type	Omni-directional	
Transmission	150 m	
Packet Size	1000 with packet sending interval 0.005 sec	
TCP Packet Size	1460 with window size 32	
Routing	AODV, DSR	
Time	100 sec	

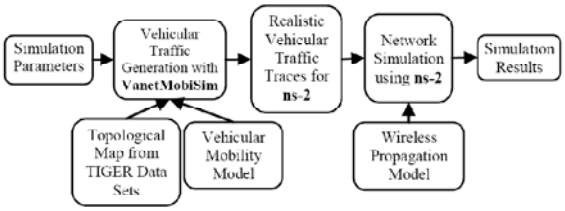


Figure 3. Overall simulation procedure using VanetMobiSim and ns-2

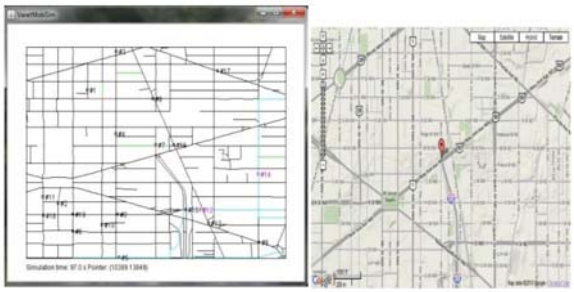


Figure 4. Street Map of District of Columbia, Washington, USA in (a). VanetMobiSim (based on TIGER/Line 2006 Second Edition Data Set) (b). Google Map Screenshot (Year 2010)

A block diagram is shown in Fig. 3 to describe the overall simulation procedure using VanetMobiSim and ns-2. Fig. 5 shows the performance throughput graphs of AODV routing protocol for both IEEE 802.11a and IEEE 802.11p standards. From Fig. 5(a) it is seen that for TCP traffic though there is a late start for both protocols, at later stage IEEE 802.11p shows more stable throughput than IEEE 802.11a. Again Fig. 5(b) shows that for UDP, traffic IEEE 802.11a produces better throughput when comparing to IEEE 802.11p which carries out with a slow start but maintains a reasonable throughput performance for the entire simulation. Fig. 6 shows the throughput performance of DSR protocol. Fig. 6(a) shows that for TCP traffic IEEE 802.11p performs better than IEEE 802.11a. However, from Fig. 6(b) it is hard to compare the performances of the two standards for UDP traffic as both of them show a constant throughput performance throughout the total simulation period.

For TCP transmission, AODV routing protocol suffers from frequent packet drops and low packet delivery ratio but DSR routing shows more stable throughput. In case of UDP transmission, DSR also performs better than AODV routing in terms of throughput measurement. IEEE 802.11p uses OFDM-based radio technology in the physical layer while FDMA/TDMA based techniques in MAC layer for link bandwidth management. It also uses dedicated control channels to exchange network control packets. On the other hand, IEEE 802.11a uses back-off periods to avoid collision detected in the MAC layer which delays the estimated packet transmission time [47].

ISBN 978- 974-672-556-9

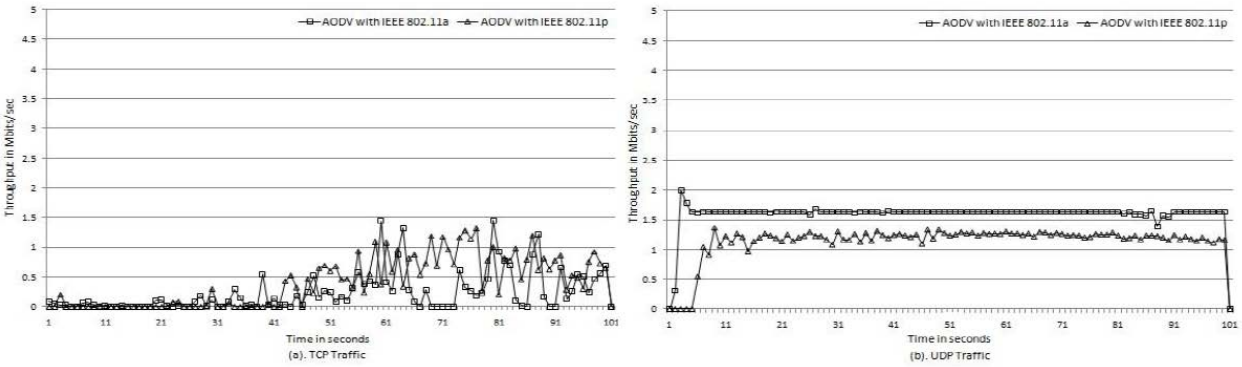


Figure 5. AODV throughput using IEEE 802.11a and IEEE 802.11p

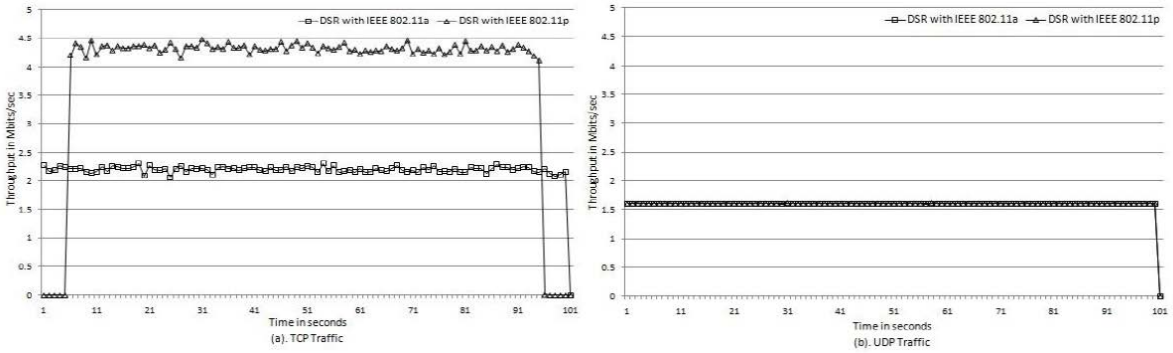


Figure 6. DSR throughput using IEEE 802.11a and IEEE 802.11p

According to the WAVE standards, the communicating nodes need to perform the authentication and association process therefore can start actual data transmission right away. These features make IEEE 802.11p draft standard suitable for vehicular communications rather than IEEE 802.11a standard. In this simulation, it is seen that TCP traffic suffers less while using IEEE 802.11p specifications. As TCP produces higher control packets and ensures reliable end-to-end communication therefore IEEE 802.11a adds extra overhead over the data transmission phase. But in unreliable UDP transmission both of these wireless standards perform similarly in AODV and DSR routings. DSR performs better due to its reactive nature of utilising source routing rather than proactive routing table management in AODV. Therefore, it might be beneficial to choose IEEE 802.11p as the wireless standard for vehicular ad hoc networking and WAVE systems as in today's Internet world a big portion of data traffic is carried out by TCP. A more deep investigation is needed to clearly understand TCP and UDP performance behaviour over ad hoc networks.

VI. Conclusion

This paper presents an ns-2 simulation using a realistic vehicular traffic generator VanetMobiSim to observe the performance of two proposed IEEE standards IEEE 802.11a 802.11p. It is found that DSR routing performs better than AODV and IEEE 802.11p draft standard shows a more stabilised nature with both AODV and DSR protocols in TCP and UDP transmissions. It is seen that the enhanced quality of service features in IEEE 802.11p makes it more suitable for reliable data transmission like TCP although a simulation with full WAVE architecture can able to produce more clear aspects. Large scale simulations are needed to observe the scalability of other topological and geographical routing protocols within these two IEEE standards. A comprehensive discussion has also made on vehicular mobility models and recognized routing protocols. Our future work will extend towards a more large scale network simulation with more emphasis on evaluating the performance of reliable data transportation using IEEE 802.11p/IEEE 1609 WAVE network architecture.

Acknowledgement

This research is supported by eLINK Project, EU Erasmus Mundus External Cooperation Window – Asia Regional Call (EM ECW – ref. 149674-EM-1-2008-1-UK-ERAMUNDUS).

References

[1] vanet.info - a site about research on vehicular ad hoc networks (VANETs) [Online]. Available:http://vanet.eurecom.fr/

[2] The Network Simulator – ns-2 [Online]. Available: http://www.isi.edu/nsnam/ns/

[3] C. Perkins, E. Belding-Royer, S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," Internet Specification, IETF, July, 2003, RFC 3561.

[4] D. Johnson, Y. Hu, D. Maltz, "The Dynamic Source Routing Protocol (DSR)," Internet Specification, IETF, February, 2007, RFC 4728.

[5] Jerome Harri, Fethi Filali and Christian Bonnet (2006). *Mobility Models for Vehicular Ad Hoc Networks: A Survey and Taxonomy*. [Online]. Available: http://www.eurecom.fr/util/publidownload.fr.htm?id=1951

[6] F. L. Yu Wang, "Routing in Vehicular Ad Hoc Networks: A Survey," IEEE Vehicular Technology Magazine, June 2007.

[7] M. Fiore, J. Harri, F. Filali, C. Bonnet, "Vehicular Mobility Simulation for VANETs," 40th Annual Simulation Symposium, ANSS '07, pp. 301-309, 2007.

[8] W. Wang; F. Xie; M. Chatterjee, "An Integrated Study on Mobility Models and Scalable Routing Protocols in VANETs," Mobile Networking for Vehicular Environments, pp. 97-102. 2007.

[9] H. Hartenstein, K. Laberteaux (Ed.), "*VANET Vehicular Applications and Inter-Networking Technologies*," Wiley, 2009.

[10] F. Bai, N. Sadagopan, A. Helmy, "The IMPORTANT Framework for Analyzing the Impact of Mobility on Performance of Routing for Ad Hoc Networks," AdHoc Networks Journal - Elsevier Science, vol. 1, issue 4, pp. 383-403, November 2003.

[11] BonnMotion [Online]. Available: http://net.cs.uni-bonn.de/wg/cs/applications/bonnmotion/

[12] GloMoSim – Global Mobile Information Systems Simulation Library [Online]. Available: http://pcl.cs.ucla.edu/projects/glomosim/

[13] QualNet [Online]. Available: http://www.scalable-networks.com/

[14] Rapid Generation of Realistic Simulation for VANET [Online]. Available: http://www.csie.ncku.edu.tw/~klan/move/index.html

[15] TIGER/Line Data Set [Online]. Available: http://www.census.gov/geo/www/tiger/

[16] SUMO Project [Online]. Available: http://sourceforge.net/apps/mediawiki/sumo/?title=Projects

[17] D. Choffnes, F. Bustamante, "An Integrated Mobility and Traffic Model for Vehicular Wireless Networks," 2nd ACM Workshop on Vehicular Ad Hoc Networks (VANET 2005), September 2005.

[18] R. Mangharam, "GrooveSim: a topography-accurate simulator for geographic routing in vehicular networks," 2nd ACM Workshop on Vehicular Ad Hoc Networks (VANET 2005), September 2005.

[19] A. Uchiyama, "Mobile Ad-hoc Network Simulator based on Realistic Behavior Model," 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2005), Urbana-Champaign, IL, USA, May 2005.

[20] UDel Models For Simulation of Urban Mobile Wireless Networks [Online]. Available: http://udelmodels.eecis.udel.edu/

[21] The SHIFT Traffic Simulator [Online]. Available: http://path.berkeley.edu./

[22] H. M. Zimmermann, I. Gruber, "A Voronoi-based Mobility Model for Urban Environments," In Procediongs of the European Wireless 2005 (EW'05), April 2005.

[23] The Obstacle Mobility Model [Online]. Available: http://moment.cs.ucsb.edu/mobility/

[24] CANU Mobility Simulation Environment (CanuMobiSim) [Online]. Available: http://canu.informatik.uni-stuttgart.de/mobisim/index.html

[25] VanetMobiSim [Online]. Available: http://vanet.eurecom.fr/

[26] K. C. Lee, U. Lee, M. Gerla, "Survey of Routing Protocols in Vehicular Ad Hoc Networks, "Advances in Vehicular Ad-Hoc Networks: Developments and Challenges," IGI Global, Oct, 2009 (Book Chapter).

[27] B. Karp, and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," Mobile Computing and Networking, pages 243-254, 2000.

[28] V. Naumov, R. Baumann, T. Gross, "An evaluation of Inter-Vehicle Ad Hoc Networks Based on Realistic Vehicular Traces," Proceedings of ACM MobiHoc'06 Conference, May, 2006.

[29] C. Lochert, M. Mauve, H. Fussler, and H. Hartenstein, "Geographic routing in city scenarios," SIGMOBILE Mob. Comput. Commun. Rev., vol. 9, no. 1, pp. 69–72, 2005.

[30] C. Lochert, H. Hartenstein, J. Tian, D. Herrmann, H. Füßler, and M. Mauve , "A routing strategy for vehicular ad hoc networks in city environments," In Proceedings of IEEE Intelligent Vehicles Symposium (IV2003), pp. 156–161, June 2003.

[31] B.-C. Seet, G. Liu, B.-S. Lee, C. H. Foh, K. J. Wong, K.-K. Lee, "A-STAR: A Mobile Ad Hoc Routing Strategy for Metropolis Vehicular Communications," NETWORKING, pp. 989-999, 2004.

[32] D. Forderer, "Street-Topology Based Routing," May, 2005 (Master's thesis, University of Mannheim).

[33] M. Jerbi, S.-M. Senouci, R. Meraihi, and Y. D. Ghamri, "An improved vehicular ad hoc routing protocol for city environments," Communications, 2007. ICC '07. IEEE International Conference, pp. 3972–3979, 24-28 June, 2007.

[34] K.C. Lee, M. Le, J. Haerri and M. Gerla, "Louvre: Landmark overlays for urban vehicular routing environments," Proceedings of IEEE WiVeC, 2008.

[35] K.C. Lee, U. Lee, M. Gerla, "TO-GO: TOpology-assist geo-opportunistic routing in urban vehicular grids," Wireless On-Demand Network Systems and Services, 2009. WONS 2009. Sixth International Conference on, pp.11-18, 2-4 Feb, 2009.

[36] J. Zhao, G. Cao, "VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks," INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings, pp.1-12, April 2006.

[37] I. Leontiadis, C. Mascolo, "GeOpps: Geographical Opportunistic Routing for Vehicular Networks," World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium. pp.1-6, 18-21 June 2007.

[38] P.-C. Cheng, J.-T. Weng, Tung, L.-C., K.C. Lee, M. Gerla and J. Haerri, "GeoDTN+NAV: A Hybrid Geographic and DTN Routing with Navigation Assistance in Urban Vehicular Networks," In Proc. of the 1st International Symposium on Vehicular Computing Systems (ISVCS'08), Dublin, Ireland, July 2008.

[39] J. Blum, A. Eskandarian, and L. Hoffman, "Mobility management in IVC networks," In Proc. of IEEE Intelligent Vehicles Symposium, 2003.

[40] M. Durresi, A. Durresi, and L. Barolli, "Emergency broadcast protocol for intervehicle communications," In ICPADS '05: Proceedings of the 11th International Conference on Parallel and Distributed Systems—Workshops (ICPADS'05), 2005.

[41] G. Korkmaz, E. Ekici, F. Özgüner, and Ü. Özgüner, "Urban multi-hop broadcast protocol for inter-vehicle communication systems," In ACM International Workshop on Vehicular Ad Hoc Networks, pp. 76–85, 2004.

[42] L. Briesemeister, L. Schäfers, and G. Hommel, "Disseminating messages among highly mobile hosts based on inter-vehicle communication," In Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 522–527, 2000.

[43] B. Karp and H.T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," In Proc. of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom), 2000.

[44] C. Maihöfer, T. Leinmüller, and E. Schoch, "Abiding geocast: time-stable geocast for ad hoc networks," In Proc. of the 2nd ACM international workshop on Vehicular ad hoc networks (VANET '05), pp. 20–29, 2005.

[45] Federal Communications Comission (FCC), USA, Wireless Services: Dedicated Short Range Communication (DSRC) Service [Online]. Available: http://wireless.fcc.gov/services/index.htm?job=service_home&id=dedicated_src

[46] D. Jiang, L. Delgrossi, "IEEE 802.11p: Towards an International Standard for Wireless Access in Vehicular Environments," In Proc. Of IEEE Vehicular Technology Conference, 2008. pp. 2036-2040, VTC Spring 2008.

[47] T. Murray, M. Cojocari, F. Huirong, "Measuring the performance of IEEE 802.11p using ns-2 simulator for

vehicular networks," IEEE International Conference on Information Technology, EIT 2008, pp. 498 - 503, 2008.

[48] S.-Y. Wang, H.-L. Chao; K.-C. Liu; T.-W. He, C.-C. Lin, C.-L. Chou, "Evaluating and improving the TCP/UDP performances of IEEE 802.11(p)/1609 networks," IEEE Symposium on Computers and Communications, ISCC 2008, pp. 163-168, 2008.

Session 9: Software Agent and Soft Computing Applications

ISBN 978- 974-672-556-9

Flow Forecasting for a Large River Basin of Nepal Using ANN

K. N. Dulal[1, 2], K. P. Dahal[2], M. A. Hossain[2] and R. Rijal[1]
[1]Kantipur Engineering College, Nepal
[2]University of Bradford, UK

***Abstract*— This paper investigates the applicability of Artificial Neural Network (ANN) technique for forecasting of flows for a large river basin of Nepal, called Narayani River basin. It is a snow-fed perennial river, which is poorly gauged or ungauged in the upper part and has lower gauge density in the lower part. The possibility of using freely available Tropical Rainfall Measuring Mission (TRMM) satellite rainfall data as input to the model is explored. Daily rain gauge data as well as TRMM data is used for forecasting flow one day ahead. The result of the model shows that the model performance with TRMM data is almost similar to that of rain gauge data. As satellite data is not the actual measurement of rainfall on the ground, it is subjected to error. The result confirms that the error in input data is absorbed by ANN during training, which makes ANN applicable for flow forecasting using satellite data.**

Index Terms- Artificial Neural Network; flow forecasting; Narayani River basin; Nepal

I. INTRODUCTION

Hydrological processes include precipitation, evaporation, infiltration, ground water flow and surface runoff. The activating signal of the hydrological process is precipitation whereas river discharge is the output signal. Hydrological processes are highly non-linear and exhibit a high degree of spatial and temporal variability. Hydrological model represents the complex hydrological processes in a simplified way. It is an effective tool for the planning, management and development of water resources system. Three types of models are used in hydrology [2]: Data-driven (black box/empirical models), conceptual models and physically based models. Data-driven models are based on the input and output data of the system. Conceptual models represent the various hydrological processes in a conceptual way by representing processes as storages and fluxes with semi-empirical form of equations. Physically based models are based on the physics of the system.

Nepal has vast water resources, from which the country can uplift the economy. However, due to the technical and financial constraints as a developing country, the resources have not been utilized fully. Nepal is often affected by the natural disaster such as floods. Steep mountain topography, fragile geological condition, narrow river valleys and rising bed levels in plain areas, and obstruction of flows near the border with India often exacerbates the frequency of flooding. Although Nepal is contributing the least for the rising trend of global warming, its Himalayan mountains and glaciers will have tremendous impact in the future due to this phenomenon, which may also result in higher frequency of floods and droughts.

Flood forecasting and warning system has not been established in a systematic way in Nepal. Only a few hydro-meteorological stations are equipped with real-time data acquisition system. Real time data of higher resolution, which is required for flood forecasting, is not available for most of the stations for developing operational forecasting model [16]. The large river basins of Nepal are snow fed and perennial. Upper part of the basin is practically inaccessible due to difficult mountainous topography in the Himalayan region. Few hydro-meteorological stations are installed in the lower part, which are not adequate to provide the proper spatial representation. Therefore, these rivers are poorly gauged or ungauged in the upper part and have lower gauge density in the lower part. Hence, there is a need to look at alternative data sources rather than only depending on the ground-observed data source for developing and applying models.

Now-a-day a number of satellite rainfall data are available at different temporal and spatial resolution covering the entire world. In case of developing country like Nepal, which needs considerable time for developing infrastructures, there is a need to investigate whether we can benefit from freely available satellite data in managing our water resources.

As the hydrology of large river basins of Nepal is very complex, it would very difficult to implement conceptual and physical-based models. Therefore, data-driven models, which are developed purely from the input and output data and does not require to look into the complex physical processes in detail, can be an alternative for such basins.

Among many data-driven techniques, Artificial Neural Network (ANN) is the most popular in Hydrology. Its application to hydrological modelling can be found in several papers, such [4]-[6], [8], [10], [12], [14] and [15]. Few papers, such as [11] and [13] are found on the use of satellite rainfall data for

ISBN 978- 974-672-556-9

modelling Nepalese river basins.

The objective of this paper is to evaluate the possibility of using satellite rainfall data for flow forecasting of one of the large river basins of Nepal. The observed rainfall data is used as a basis for comparing the model performance. Among different data-driven models, Artificial Neural Network (ANN) will be used to map the relationship between input and output data of the river basin.

II. Study Area and Dataset

Narayani River basin (Fig.1), also known as Gandaki, is a major river basin of Nepal. It is a snow fed river. It originates from the Himalayan range, crosses Mahabharat range and flows to the plains of Terai. The major tributaries are the Trishuli, Budhi Gandaki, Marsyangdi, Seti, Kali Gandaki and East Rapti. Among these rivers, some parts of the Kali Gandaki River, the Budhi Gandaki River and the Trishuli River lie in Tibetan territory and flow down through the Himalayan range to Nepal. The climate of the basin varies from subtropical zone in the South to Tundra in Higher Himalayan region. Almost 80% of rainfall occurs in Monsoon period, which spans from June-September. For the rest of the period, there is no or very little rainfall. The spatial variation of annual rainfall is also very high in the basin. The Narayanghat hydrological gauging site, an outlet of the mountainous watershed, is taken as outlet station for this study. The drainage area of the Narayani is 31100 Km^2 at this gauging station.

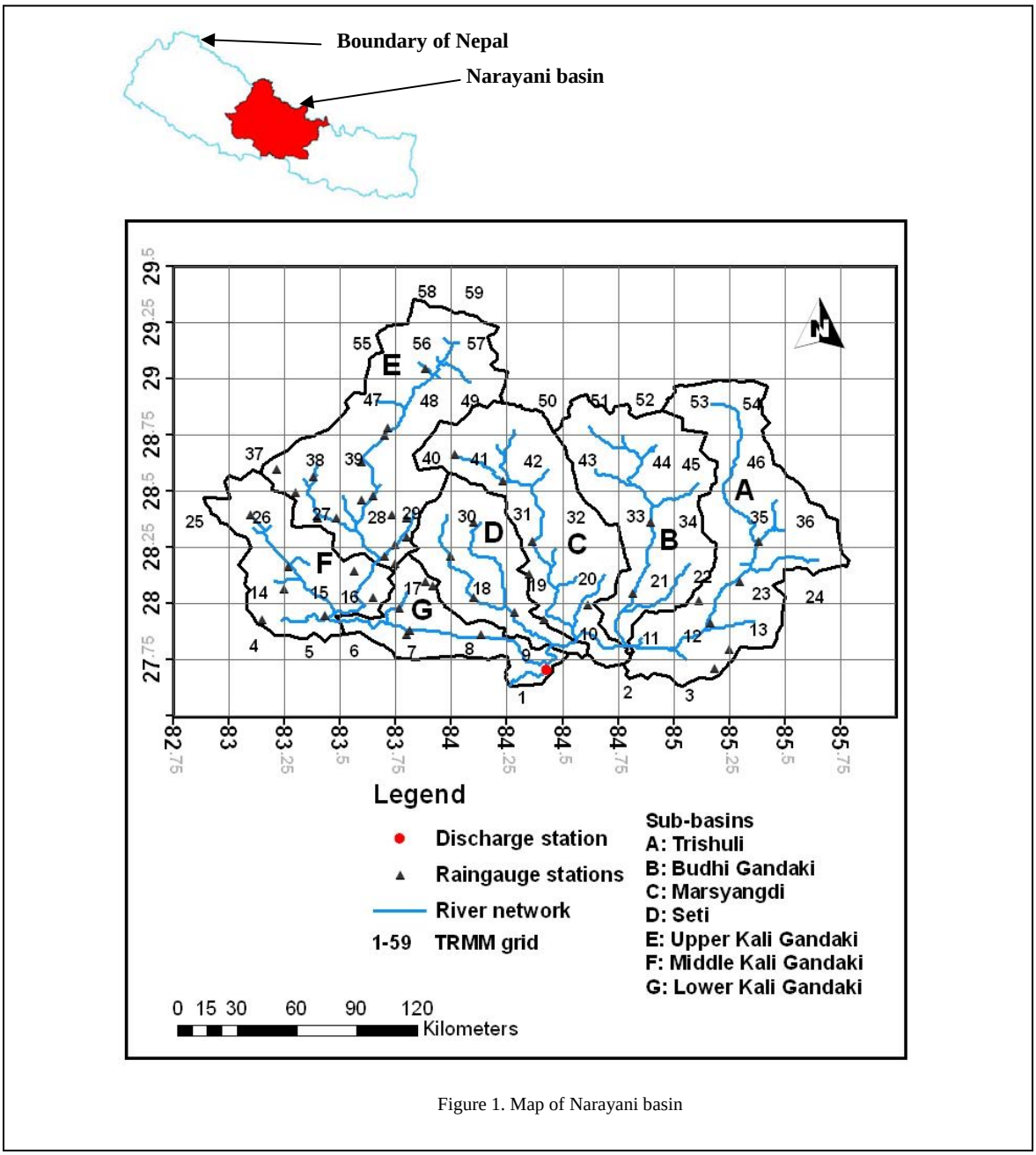


Figure 1. Map of Narayani basin

ISBN 978- 974-672-556-9

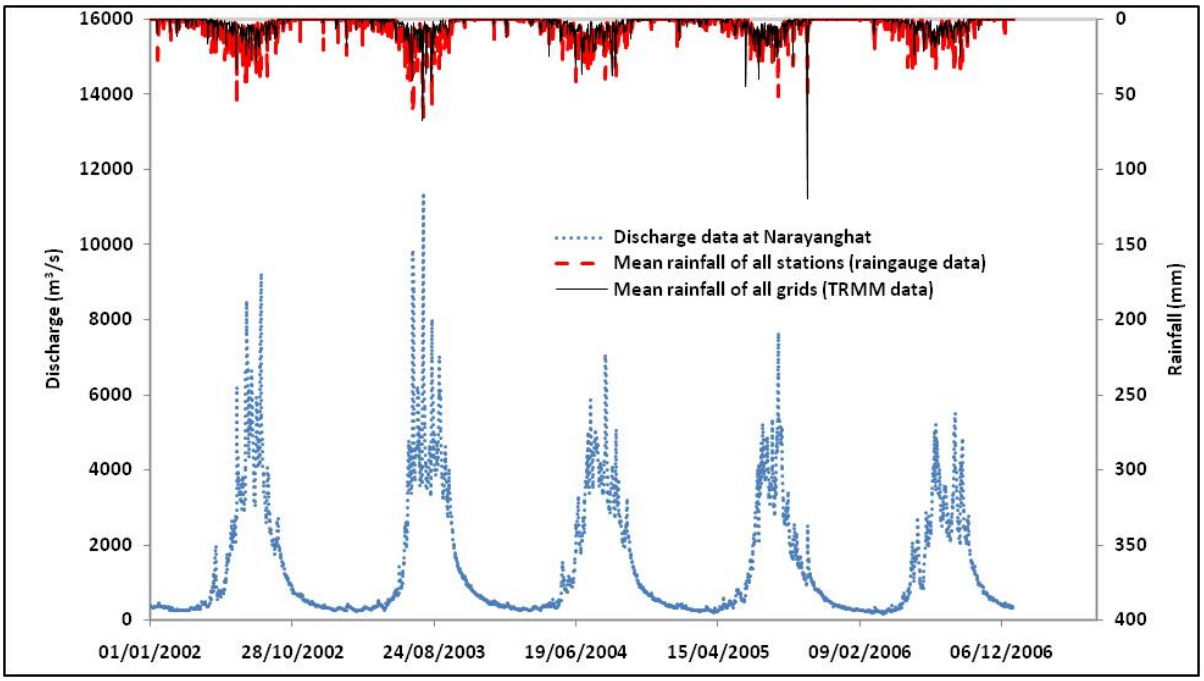


Figure. 2. Plot of time-series data

In this study, the part of the Narayani basin with Narayanghat station as outlet (shown in the lower part in Fig. 1) is considered due to the availability of hydrological data at this station. The river enters the plain area after crossing this station and finally drains to India.

The following data sets are available for the study:

- Observed daily rainfall data (mm) of 48 stations for the period of 2002-2006 [16]
- Observed daily discharge (m^3/s) at Narayanghat discharge station for the period of 2002-2006 [16]
- Daily rainfall data (mm) from Tropical Rainfall Measuring Mission (TRMM) satellite downloaded for 2002-2006 for $0.25^0 \times 0.25^0$ grid resolution. The number of grids covering the basin is 59. [17]

Fig. 2 shows the plot of mean daily discharge data at Narayanghat together with mean rainfall of 48 stations and 59 TRMM grids. The mean of discharge data of 2002-2006 is 1412 m^3/s while the minimum is 163 m^3/s and maximum is 11300 m^3/s. The mean annual rainfall (based on 2002-2006 data) from rain gauge data is 2009 mm and from TRMM Satellite data is 1345 mm. According to the records of discharge data from 1963 to 2006, the maximum instantaneous discharge during this period is 15400 m^3/s on 5 Aug, 1974 and the minimum instantaneous discharge is 146 m^3/s on 30 Mar, 2006.

III. METHODOLOGY

A. ANN with Back-propagation algorithm

ANN is a technique in artificial intelligence, which is loosely inspired by the working principle of human brain [7]. It is a universal function approximator. It consists of nodes (neurons), layers and connections between nodes (weights). The layers between input and output layer are called hidden layers. ANN learns the pattern between input and output through the set of examples (instances/data), which is known as training.

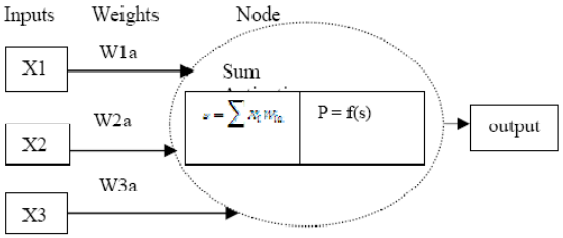


Figure 3. Processing in a node

ANN is a method of non-linear regression. Each node in hidden or output layer does two tasks (Fig. 3): summing up the product of connected input and weights and transforming the summation to output by activation function. The output from a node will subsequently be input to the next layer in case of multi-layered network.

In this study, back-propagation algorithm (gradient descent) [3] is applied for the training of ANN. The training is the optimization of weight vectors. In this algorithm, steps are made in the direction opposite to the direction of error gradient to

obtain optimized weights.

In back-propagation algorithm, weights are initialized in the beginning. Inputs are presented to the network and the output is computed. The computed output is compared with the observed (target) output and error function (E) is computed. Then E is propagated backwards to update the weights. The same procedure is repeated with the updated weights until the weights are optimized.

Weight update equation is written as

$$W_{n+1} = W_n - \eta \frac{\partial E}{\partial w} \quad (1)$$

Where E = error function (usually sum of squared error)

η = learning rate

W = Weight

The derivative is computed at the output and hidden layer using the chain rule, hence propagating errors from output to hidden layers. The limitation of equation (1) is that the optimization might get trapped in local minima. To overcome this problem, gradient descent with momentum learning is applied in this study. The equation in this case is

$$W_{n+1} = W_n + dW \quad (2)$$

Where dW = Change in weight, which is given by

$$dW = \mu\,(dw)_{n-1} - \eta(1-\mu)\,\partial E/\partial W \quad (3)$$

Where μ = momentum coefficient

B. Steps for the implementation of ANN technique for f low forecasting

- Input data set preparation: As the basin is large, it is divided into seven sub-basins as shown in lower part of Fig. 1. For each sub-basin, mean areal precipitation (MAP) is computed for both rain gauge and satellite data. The number of nodes at input layer is fixed based on the hydrological knowledge of the basin and statistical analysis.
- Prepare the target data to forecast.
- Separate data set for training and testing.
- Fix the number of hidden layers
- Fix the activation functions, training parameters, number of epochs etc.
- Optimize the number of nodes at each hidden layer by trial and error procedure.
- Train ANN until satisfactory performance is obtained
- Test the performance of trained ANN with new data set.

C. Performance measure

Nash-Sutcliffe Coefficient of efficiency (NSE) [1] is used for evaluating the performance of the model.

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left((X_{observed})_i - (X_{predicted})_i\right)^2/n}{\sum_{i=1}^{n}((X_{observed})_i - \overline{X_{observed}})^2/n} \quad (4)$$

Where

NSE = Nash-Sutcliffe Coefficient of efficiency

n = number of samples

$X_{observed}$ = Observed variable (observed discharge)

$X_{predicted}$ = Predicted/forecasted variable (forecasted discharge)

$\overline{X_{observed}}$ = Mean of observed variable

NSE is a better measure of overall performance, which is the most widely, used one in Hydrology [9]. It is a form of Normalized means squared error (NMSE). Equation (4) is further reduced to

$$NSE = 1 - \frac{MSE}{Var_{observeddata}} = 1 - NMSE \quad (5)$$

Where

MSE = means squared error

$Var_{observed\ data}$ = Variance of observed data

The closer the value of NSE to 1, the closer the NMSE to zero and the better the model performance.

D. Software

The software to be used for ANN modelling is MATLAB NNtoolbox. A code for implementing the multilayered perception neural network with Back-propagation algorithm is written, which uses the various commands available in the NNtoolbox.

IV. Results and Discussions

A. Comparison of satellite rainfall data with gauge rainfall data

Table 1 presents the result of correlation analysis and the mean annual rainfall. The result shows low correlation between rain gauge data and satellite data, ranging from 0.2 to 0.6. Mean annual rainfall estimated by satellite rainfall is lower for most of the grids. From Fig. 2, it is found that the mean of satellite rainfall of all grids is also lower than that of gauge data of all stations. Figs. 4 (a) to 4(d) present sample plots of gauge and satellite rainfall for comparison. For the grid in high altitude area, such as grid 47, TRMM rainfall is higher than gauge rainfall. In the lower part of the basin, TRMM rainfall is lower than gauge rainfall except for grids 5 and 10.

ISBN 978- 974-672-556-9

TABLE 1. STATISTICAL INDICATORS BETWEEN GAUGEAND SATELLITE DATA

Grid No.	Index No. of Rain gauge	Correla -tion	Mean annual rainfall (mm)	
			TRMM	Gauge
4	715	0.3	1291	1429
5	701	0.3	1328	1088
7	726, 810, 826	0.4	1432	1570
8	827	0.3	1567	1704
9	808, 817	0.3	1701	1630
10	809	0.3	1669	1498
12	1004, 1007, 1038	0.4	1579	2101
14	725	0.3	1150	1596
15	722	0.3	1282	2312
16	613, 614, 620, 622	0.6	1431	2269
17	804, 805, 832	0.4	1659	3360
18	815	0.3	1632	2304
19	807	0.3	1577	2798
21	1002	0.3	1603	2212
22	1057	0.4	1587	2959
23	1055	0.3	1303	1899
26	615	0.3	1181	2353
27	621, 627	0.3	1223	1878
28	606, 619, 626, 829	0.5	1269	2403
29	814, 821	0.5	1757	4340
30	824	0.4	1849	3701
31	802	0.3	1452	3398
33	801	0.3	1093	1588
35	1001	0.2	757	940
37	616	0.4	1261	1777
38	628, 629	0.4	1301	2621
39	607	0.2	1057	1377
41	816, 820	0.3	1255	695
47	601, 604	0.3	710	344
56	610	0.2	405	124

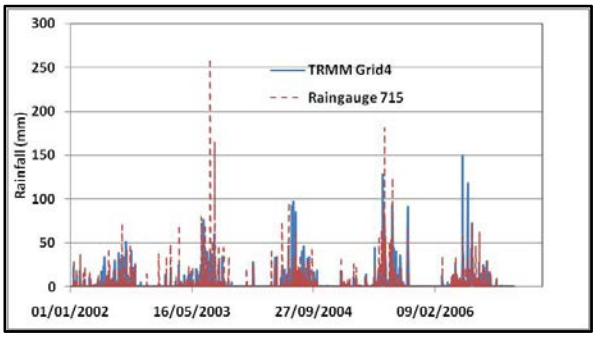


Figure 4(a). Plot of satellite rainfall data for grid 4 with corresponding rain gauge data

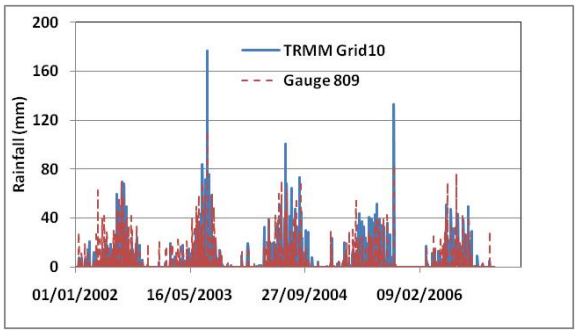


Figure 4(b). Plot of satellite rainfall data for grid 10 with corresponding rain gauge data

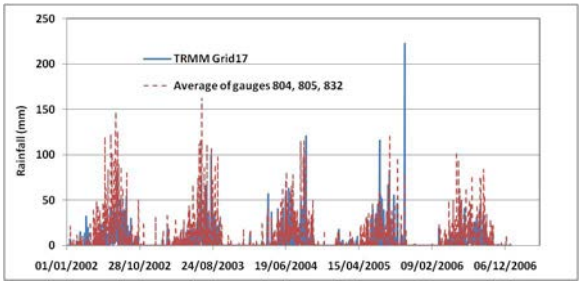


Figure 4(c). Plot of satellite rainfall data for grid 17 with corresponding rain gauge data

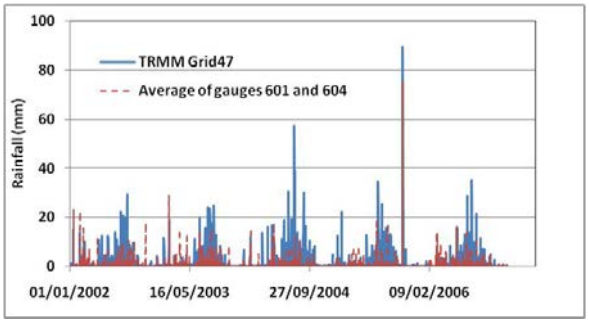


Figure 4(d). Plot of satellite rainfall data for grid 47 with corresponding rain gauge data

B. Data analysis for input data set preparation

As precipitation falling at a certain location takes time to reach the outlet as a stream flow/discharge, we have to consider the response time of rainfalls for preparing input data set. Cross-correlation analysis between rainfall and discharge is performed to obtain the time window, i.e. lag time of rainfall data for each sub-basin.

Besides rainfall, past discharge data is also used as input data set. The purpose of past discharge is to make ANN learn the pattern during no-rain period and also to tell ANN the response for past rainfalls.

Cross-correlation analysis between mean areal precipitation (MAP) of each sub-basin and the discharge at the outlet is shown in Table 2. From the Table, it is found that MAP at time t has highest correlation with discharge at time t for Trishuli,

Budhi Gandaki, Seti and Lower Kali. For the remaining sub-basins, the highest correlation exists between MAP at time t-1 and discharge at time t. Based on this analysis, MAP at time t is taken as input for four basins and MAP at time t-1 and t is taken as rainfall input for remaining three basins. The autocorrelation analysis of discharge shows that discharge at time t-1 has the highest correlation of 0.96, followed by the discharge at time t-2 with correlation of 0.93.

TABLE 2. CORRELATION ANALYSIS

Cross- correlation between rainfalls and discharge at time t					
Sub-basin	R(t)	R(t-1)	R(t-2)	R(t-3)	R(t-4)
Trishuli	0.72	0.71	0.69	0.68	.67
Budhi Gandaki	0.58	0.57	0.55	0.55	.53
Marsyangdi	0.61	0.62	0.58	0.55	.55
Seti	0.598	0.596	0.56	0.53	.51
Upper Kali	0.75	0.77	0.75	0.74	.73
Middle Kali	0.63	0.65	0.64	0.61	.61
Lower Kali	0.55	0.54	0.50	0.47	.45
Autocorrelation of discharge at t with past discharges					
Autocorrelation	Q(t-1)	Q(t-2)	Q(t-3)	Q(t-4)	Q(t-5)
Q(t)	0.96	0.93	0.91	0.90	0.88

In summary, the variables chosen for input layer are as follows:

R1(t), R2(t), R3(t), R3(t-1), R4(t), R5(t), R5(t-1), R6(t), R6(t-1), R7, Q(t), Q(t-1)

where
R1 = MAP of Trishuli sub-basin
R2 = MAP of Budhi Gandaki sub-basin
R3 = MAP of Marsyangdi sub-basin
R4 = MAP of Seti sub-basin
R5 = MAP of Upper Kali Gandaki sub-basin
R6 = MAP of Mid-Kali Gandaki sub-basin
R7 = MAP of Lower Kali Gandaki sub-basin
Q = Discharge at Narayanghat

Two past discharges are also given as input and the variable at output layer is Q(t+1). Hence, the rainfall data and discharge available up to time t is used to forecast discharge data at time t+1, i.e. one day ahead. In functional form, we can write

Q(t+1) = f(R1(t), R2(t), R3(t), R3(t-1), R4(t), R5(t), R5(t-1), R6(t), R6(t-1), R7, Q(t), Q(t-1)) (6)

C. ANN modeling

Three years data (2002-2004) is used for training ANN, while remaining two years data is used for testing the performance of trained ANN. Three layered feed-forward network (I-H-O) is used for the modelling. The input layer consists of 12 nodes, output layer consists of 1 node and the nodes in hidden layer are optimized by hit and trial approach. Activation functions applied are: hyperbolic tangent (tansig) at hidden layer and linear (purelin) at output layer. The network is trained using gradient descent momentum algorithm. The learning parameter and momentum coefficient are also optimized by hit and trial approach. The training stops when the maximum number of epochs is reached or the error falls below the threshold value. The maximum number of epochs is set to 30000 and the threshold value of mean squared error is set to 0.00001. From several trials with both gauge rainfall and satellite rainfall, it was found that 6 hidden nodes produces relatively better performance. Hence the feed-forward network adopted in this study is 12-6-1.

Several trials were performed for 12-6-1 network with both gauge rainfall data and satellite rainfall data. The result of the trial giving the best performance was selected.

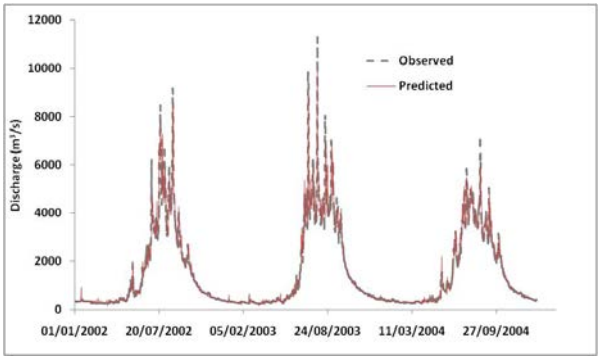


Figure 5. Hydrograph of training with gauge rainfall data

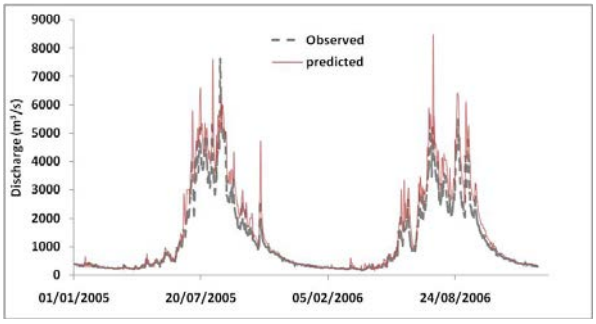


Figure 6. Hydrograph of testing with gauge rainfall data

The result of training with gauge rainfall is presented in Fig. 5. The result shows that the ANN model reproduces the low flow very well and the high flows are also captured well. The overall pattern of the hydrograph is reproduced. The plot of the result for test data set in this case (Fig. 6) shows slight overestimation of peaks with very good fit for low flows. The NSE value of 0.96 is obtained for training and 0.83 is obtained for test data. This statistics show that the overall performance of the model is good.

ISBN 978- 974-672-556-9

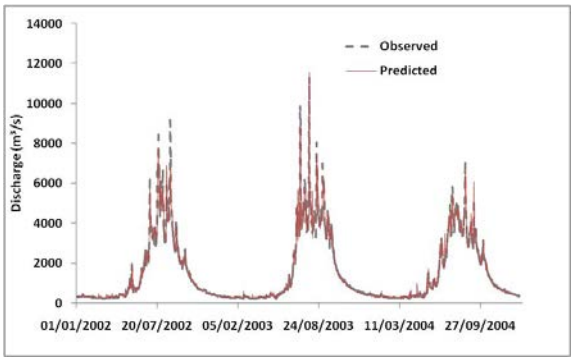


Figure 7. Hydrograph of training with TRMM rainfall

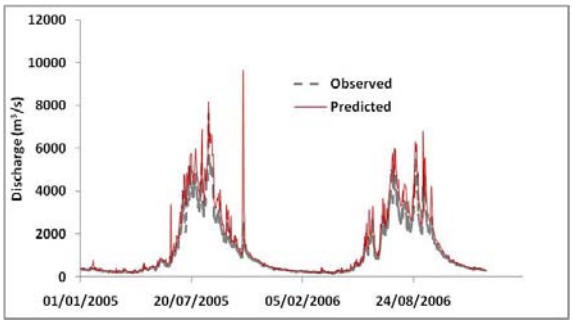


Figure 8. Hydrograph of testing with TRMM rainfall

In the next experiment, satellite rainfall data was applied for ANN modelling. The result of the training in this case (Fig. 7) shows almost similar performance with the gauge rainfall case. The plot of result for testing is depicted in Fig. 8, which shows good fit for most of the parts. The peaks are reproduced well except one peak of small magnitude, which is highly overestimated. As the rainfall immediately preceding this peak is high, the simulated value is also high. The statistical performance indicator, NSE, is 0.97 for training and 0.81 for test data.

The good fit for low flows and slight underestimation for some peaks in case of training can be explained on the basis of availability of enough data to learn the pattern by ANN. As peak flows do not occur very frequently, there is less data set in this region to learn the pattern, whereas there is adequate data in the low flow region.

Comparing the model performance with two sets of rainfall data, it is found from this study that the ANN result for both sets is good and almost similar statistically. Even though satellite rainfall data is lower than the gauge rainfall data for most of the grids, the model performance does not deviate significantly. Gauge rainfall data is the actual rainfall measured at a location. However, rain gauge stations in the higher mountain area are sparse and the representation of spatial variability by a few rainfall stations is not perfect. On the other hand, satellite rainfall data are available for the whole basin, which gives better picture of spatial representation. Input data set is subjected to different sources of errors. But the training process in ANN absorbs certain error in data as it is actually a data fitting procedure based on the pattern. This is the reason for the similar type of performance although satellite rainfall data underestimates the observed gauge rainfall.

V. CONCLUSIONS

The satellite rainfall data underestimates gauge precipitation for most of the parts of the basin, but the performance of ANN model for forecasting of flow one day ahead for Narayani basin with satellite rainfall data as input data is comparatively similar to that with gauge rainfall. This concludes that satellite rainfall data will be useful for forecasting flows. As the peaks are also reproduced well, this data can be useful for flood forecasting.

Satellite rainfall data are available in near-real time which can be used for forecasting purpose, but rain gauge data in Nepal are not available in real-time except for few stations. However, the limitation of satellite data is that satellite does not directly measure precipitation on the ground and the precipitation data is not free of errors. That means the accuracy of satellite data is still in question. Therefore, validation and necessary correction of satellite rainfall data is necessary before applying for modeling.

REFERENCES

[1] J. E. Nash and J. V. Sutcliffe, “River flow forecasting through conceptual models. Part I: a discussion of pricnciples,” *Journal of Hydrology*, vol. 10, pp. 282-290, 1970.

[2] J. C. Refsgaard, “Terminology, modelling protocol and classification of hydrological models,” in *Distirbuted Hydrological Modelling*, M. B. Abbott and J. C. Refsgaard, Ed., Kluwer Academic Publishers, 1996, pp. 17-39.

[3] D. E. Rumelhart, G. E. Hinton and R. J. Williams, “Learning internal representation by error propagation,” in *Parallel Distributed Processing. Vol. I: Foundations*, D. E. Rumelhart and J. L. McClelland, Ed., MIT press, Cambridge, 1986.

[4] A. W. Minns and M. J. Hall, “Artificial neural networks as rainfall-runoff models,” *Hydrological Sciences Journal*, vol. 41, pp. 399-417, 1996 .

[5] K. Thirumalaiah and M. C. Deo, “Real-time flood forecasting using neural networks,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 13, pp. 101-111, 1998.

[6] M. Campolo, P. Andreussi and A. Soldati, “River flood forecasting with a neural network model,” *Water Resources Research*, vol. 35, pp. 1191-1197, 1999.

[7] S. Haykin, *Neural Networks, a comprehensive foundation*, Prentice hall, New Jersey, 1999.

[8] M. P. Rajurkar and U. C. Kothyari, “Artificial neural networks for daily rainfall-runoff modelling,” *Hydrological Sciences Journal*, vol. 47, pp. 865-877, 2002 .

[9] R. H. McCuen, Z. Knight and G. Cutter, “Evaluation of the Nash-Sutcliffe Efficieny Index,” *Journal of Hydrologic Engineering*, vol. 11, pp. 597-602, 2006.

[10] M. A. Antar, I. Elassiouti and M. N. Allam, “Rainfall-runoff modelling using artificial neural networks technique: a Blue Nile catchment case study,” *Hydrological Processes*, vol. 20, pp. 1201–1216, 2006.

[11] M. K. Hazarika, T. P. Kafle, R. Sharma, S. Karki, R. M. Shrestha and L. Samarkoon, “Statistical approach to

discharge prediction for flood forecasts using TRMM data," in *Proc. of the 5th Annual Mekong Flood Forum*, Ho Chi Minh City, Vietnam, 2007.

[12] I. Pulido-Calvo and M. M. Portela, "Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds," *Journal of Hydrology*, vol. 332, pp. 1-15, 2007.

[13] M. S. Shrestha, G. A. Artan, S. R. Bajracharya and R. R. Sharma, "Using satellite-based rainfall estimates for stream-flow modelling: Bagmati Basin," *Journal of Flood Risk Management*, vol. 1, pp. 89-99, 2008.

[14] M. K. Akhtar, G. A. Corzo, S. J. van Andel, and A. Jonoski, "River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: case study of the Ganges River basin," *Hydrology and Earth System Sciences*, vol. 13, pp. 1607–1618, 2009.

[15] A. S. Islam, "Improving flood forecasting in Bangladesh using an artificial neural network," *Journal of Hydroinformatics* (in press), 2009.

[16] Precipitation and streamflow data, Department of Hydrology and Meteorology, Kathmandu, Nepal.

[17] TRMM Satellite data, URL: http://disc.sci.gsfc.nasa.gov/precipitation

Disjunctive Naïve Bayesian Classifier to Enhance Accuracy for Dynamic Prediction

Md. Faisal Kabir[1], Hossain Ashik Mahmud Chowdury[1], Keshav Dahal[1], Alamgir Hossain[1] and Chowdhury Mofizur Rahman[2]

[1] School of Informatics, University of Bradford, Bradford BD7 1DP, UK

[2] Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh.

Abstract— A novel supervised learning algorithm named disjunctive naïve Bayesian classifier is presented in this paper. In conventional naïve Bayesian classifier only one set of class conditional probabilities are calculated from the given data. However, in our proposed approach we divide the data using k-means clustering –and save the center of each cluster. We then train these k clusters in naïve Bayesian classifier. For a new case to classify we compute similarity with the previously obtained cluster centers and based on the best match, we use the appropriate cluster set conditional probability to predict the class. We tested our proposed model on a number of benchmark data and attained higher classification accuracy rates than conventional naïve Bayesian classifier.

Index Terms - disjunctive naïve Bayesian classifier; clustering; dynamic prediction accuracy

I. INTRODUCTION

Classification [1] is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. Classification has many applications in customer segmentation, business modeling, marketing, credit analysis, and biomedical and drug response modeling [2-3] [1]. There are several algorithms for classification such as decision tree, naïve Bayesian, generalized linear models, support vector machine etc [1]. The naïve Bayesian classifier (NB) is one of the most popular data mining techniques for classifying the large dataset. It has been successfully applied to the different problem domains of classification task such as weather forecasting, intrusion detection, image and pattern recognition, medical diagnosis, loan approval and bioinformatics etc [2-4]. Naive Bayesian classifier also efficiently applied in feature selection [5], exudates pixel classification [5] and web classification [6].The classification task is to map the set of attributes of sample data onto a set of class labels, and naïve Bayesian classifier particularly suitable as proven universal approximates. The NB classifier is a probabilistic approach for performing supervised learning that provides an optimal way to predict the class of an unknown example [7-8]. In NB class conditional probabilities for each attribute values are calculated from the given training data, and then these probabilities are used to classify the known or unknown examples. Clustering is a process of partitioning a set of data or objects in a set of meaningful sub-classes, called clusters. It assists users to understand the natural grouping or structure in a data set. It has wide applications in pattern recognition, spatial data analysis, economic science, market research, document classification and so forth [9]. Widely used clustering methods are K-means clustering and hierarchical clustering: agglomerative and divisive respectively. In K means clustering algorithm, the selection of K and the initial center are crucial. The performance of K-means clustering depends on initial guess of the partition [10]. Selecting inappropriate initial center (randomly) may converge in local maxima which can be solved using genetic algorithm [11].

In this paper, we propose a new classifier named disjunctive naïve Bayesian classifier. Like naïve Bayesian classifier it calculates the prior and conditional probabilities from a given training data and classifies the training examples using these probabilities. However, before applying naïve Bayesian classifier first we divide the training data using k-means clustering algorithm and save the clusters center and calculate the conditional probabilities for each cluster. For new data to classify, we apply similarity measure with the previously obtained clusters center to select the most similar cluster. Based on the appropriate cluster set and using naïve Bayesian classifier for that cluster set it predicts the class for unknown data. We tested our proposed model on a number of benchmark data and achieved higher classification accuracy rates than conventional naïve Bayesian classifier.

In section 2 of the paper, we briefly describe classification and clustering. Our proposed model, disjunctive naïve Bayesian classifier which gives better accuracy than naïve Bayesian classifier is

ISBN 978- 974-672-556-9

described in section 3. Section 4 illustrates the experimental results and analysis with an example and section 5 concludes the paper with future research direction.

II. Classification and Clustering

Classification

Classification [1] is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks. A classification task begins with a data set in which the class assignments are known. For example, a classification model that predicts credit risk could be developed based on observed data for many loan applicants over a period of time. The simplest type of classification problem is binary classification. In binary classification, the target attribute has only two possible values: for example, high credit rating or low credit rating. Multiclass targets have more than two values: for example, low, medium, high, or unknown credit rating. In the model building (training) process, a classification algorithm finds relationships between the values of the predictors and the values of the target. Different classification algorithms use different techniques for finding relationships. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown. Classification models are tested by comparing the predicted values to known target values in a set of test data. The historical data for a classification project is typically divided into two data sets: one for building the model; the other for testing the model.

Naïve Bayesian (NB) Classifier

Naïve Bayesian (NB) classifier is a simple probabilistic classifier based on probability model, which can be trained very efficiently in a supervised learning framework [12]. The NB classifier is given as input a set of training examples each of which is described by attributes A_1 through A_k and an associated class, *C*. The objective is to classify an unseen example whose class value is unknown but values for attributes A_1 through A_k are known and they are *a1, a2,……, ak* respectively. The optimal prediction of the unseen example is the class value *c* such that *P(C=ci|A1=a1…Ak=ak)* is maximum. By Bayes rule this probability equals to [13]:

$$\text{argmax}_{c_i \in C} \frac{P(A_1 = a_1, \ldots A_k = a_k \mid C = c_i)}{P(A_1 = a_1, \ldots A_k = a_k)} P(C = c_i) \qquad (i)$$

Where, $P(C=c_i)$ is the prior probability of class c_i, $P(A_1=a_1,\ldots A_k=a_k)$ is the probability of occurrence of the description of a particular example, and $P(A_1=a_1,\ldots A_k=a_k|C=c_i)$ is the class conditional probability of the description of a particular example c_i of class *C*. The prior probability of a class can be estimated from training data. The probability of occurrence of the description of particular examples is irrelevant for decision making since it is the same for each class value *c*. Learning is therefore reduced to the problem of estimating the class conditional probability of all possible description of examples from training data. The class conditional probability can be written in expanded form as follows:

$$\begin{aligned} & P(A_1=a_1,\ldots A_k=a_k \mid C=c_i) \\ & = P(A_1=a_1 \mid A_2=a_2 \wedge \ldots A_k=a_k \wedge C=c_i) \\ & * P(A_2=a_2 \mid A_3=a_3 \wedge \ldots A_k=a_k \wedge C=c_i) \\ & * P(A_3=a_3 \mid A_4=a_4 \wedge \ldots A_k=a_k \wedge C=c_i) \\ & * P(A_4=a_4 \wedge \ldots A_k=a_k \wedge C=c_i) \end{aligned} \qquad \text{(ii)}$$

In NB, it is assumed that outcome of attribute *Ai* is independent of the outcome of all other attributes *Aj*, given class *c*. Thus class conditional probabilities become:

$$P(A_1=a_1,\ldots A_k=a_k|C=c_i) = \prod_{i=1}^{k} P(A_i = a_i \mid C = c_i)$$

If the above value is inserted in equation "(i)" it becomes:

$$\equiv \text{argmax}_{c_i \in C} P(C=c) \prod_{i=1}^{k} P(A_i = a_i \mid C = c_i) \qquad \text{(iii)}$$

In Naïve Bayesian classifier, the probability values of above equation are estimated from the given training data. These estimated values are then used to classify unknown examples.

Clustering

Clustering is a process of partitioning a set of data or objects in a set of meaningful sub-classes, called clusters. It assists users to understand the natural grouping or structure in a data set. It has wide applications in pattern recognition, spatial data analysis, economic science, market research, document classification and so forth [9]. Widely used clustering methods are K-means clustering and hierarchical clustering.

K-means clustering [14] partitions objects into groups that have little variability within clusters and large variability across clusters. The user is required to specify the number k of clusters a priori. Estimation is iterative, starting with a random allocation of objects to clusters, re-allocating to minimize distance to the estimated "centroids" of the clusters, and stopping when no improvements can be made. The centroid is the point whose attributes take the mean expression level of the objects in the clusters. Hierarchical clustering is used to partition objects into a series

ISBN 978- 974-672-556-9

of nested clusters [15]. There are two kinds of hierarchical clustering approaches: agglomerative and divisive. The agglomerative approach begins by assuming that each object belongs to its own separate cluster. This process continues, where the best cluster partition is chosen at each step until all objects are in their own clusters. Details of hierarchical clustering can be found in [16].

III. Proposed Model

Our proposed model works in two phases. First of all, from a given training data, our proposed model clusters or divides the data into k number of groups by using simple K-means algorithm and saves the similarity measures - centroid of each cluster. We have selected the value of K in error and trail basis. If we get better accuracy rate than non-clustered data set then we stopped changing the value of K. Furthermore, it trains each cluster by using Naïve Bayesian classifier and save the training results. Secondly, for new or testing data, we first identify which cluster it belongs to on the basis of similarity measure – centroid. Afterward, to make class prediction for new data it uses the prior and conditional probabilities generated from the training data of selected or appropriate cluster only.

Training phase

First of all, we select the value of K and initial center randomly. We have selected the value of K in error and trail basis. After that we split the training data into K number of groups using conventional K-means algorithm where each observation belongs to the cluster with nearest mean. We then save the cluster centers for each cluster so that it can be used further to find the appropriate group for new data. Finally, we train the examples of each cluster by using Naïve Bayesian classifier and save the training results, i.e., save the class conditional probabilities for each group. The correctly classified instances of training data are calculated by adding the classified instances of each cluster. The whole simulation is performed using WEKA 3.6.2 which is well known data mining software. The flow chart of training phase is shown in figure 1.

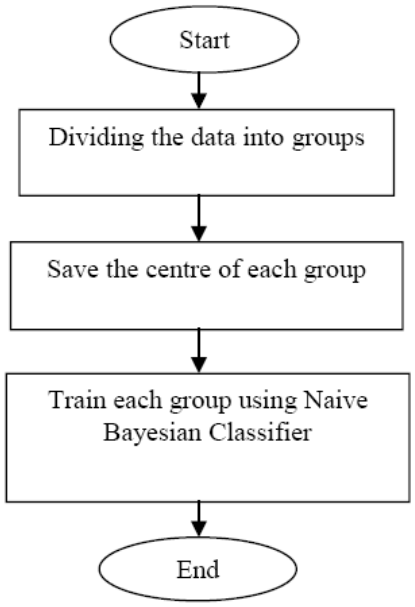


Figure1. Flow chart for training phase

Testing phase

In this stage at first we calculate the distance between the test data and the center of each cluster obtained in the training phase. This distance is calculated by the following Euclidian distance formula:

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

After that we select the appropriate group for the new data on the basis of least distance. Finally, we classify the new example using trained data of selected cluster. The correctly classified instances of testing data are calculated by adding the classified instances of each cluster. The flow chart of testing phase is shown in figure 2.

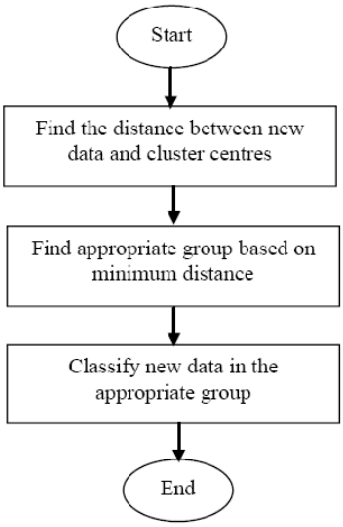


Figure 2. Flow chart for testing phase

ISBN 978- 974-672-556-9

IV. Experimental Results and Analysis

1. *Application Example*

We performed experiments on four data sets obtained from UCI repository [17]. For simplicity, we will show how our proposed model works for iris dataset [17]. There are 150 and 68 instances for training data and testing data respectively. The testing data are randomly obtained from training data as testing data for iris dataset are not available in the UCI repository [17]. For the simplicity of calculations we replaced class attributes Iris_Setosa, Iris_Versicolour and Iris_Virginica by 1, 2 and 3 respectively. The training data are divided into three clusters using WEKA 3.6.2. The parts of results and the graphical clustered assignments are shown in table I and figure 3 respectively.

TABLE I. Training Data After Clustered

A	B	C	D	Class	Cluster
5.1	3.5	1.4	0.2	1	cluster2
4.9	3	1.4	0.2	1	cluster2
.	.	.	.	.	.
.	.	.	.	.	.
5.3	3.7	1.5	0.2	1	cluster2
5	3.3	1.4	0.2	1	cluster2
7	3.2	4.7	1.4	2	cluster1
6.4	3.2	4.5	1.5	2	cluster1
.	.	.	.	.	.
.	.	.	.	.	.
5.1	2.5	3	1.1	2	cluster1
5.7	2.8	4.1	1.3	2	cluster1
6.3	3.3	6	2.5	3	cluster3
5.8	2.7	5.1	1.9	3	cluster3
.	.	.	.	.	.
.	.	.	.	.	.
6.2	3.4	5.4	2.3	3	cluster3
5.9	3	5.1	1.8	3	cluster3

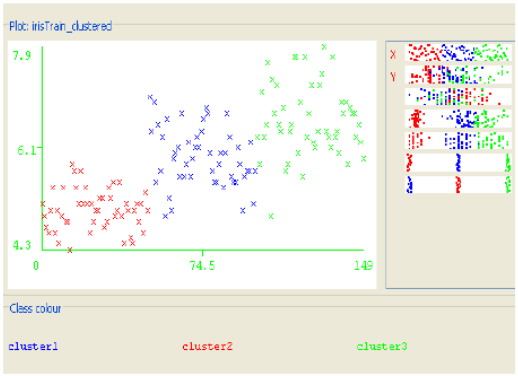


Figure 3. clustered assignments

Center of Clusters

We then calculated center of each cluster. It has been done by taking mean of corresponding attributes within a cluster. The results are shown in table II. For example, center of cluster 1 is (5.936, 2.77, 4.26, 1.326, 2).

TABLE II.CentersOf Each Cluster

Cluster	A	B	C	D	class
1	5.936	2.77	4.26	1.326	2
2	5.006	3.418	1.464	0.244	1
3	6.588	2.974	5.552	2.026	3

In testing phase we clustered the testing data. First we calculated the distance between instances and each center of cluster. Here, D1, D2 and D3 are the distance between instance and the center of cluster1, Cluster2 and cluster3 respectively. And CLUSTER for particular instance is selected on the basis of least distance. Table III contains the partial results of calculation.

TABLE III. Testing Datawith Their Assigned Cluster

A	B	C	D	Class	D1	D2	D3	CLUSTER
5.1	3.5	1.4	0.2	1	3.4	0.1	5.2	2
4.9	3	1.4	0.2	1	3.4	0.4	5.2	2
.								
.	.							
5.3	3.7	1.5	0.2	1	3.3	0.4	5.1	2
5	3.3	1.4	0.2	1	3.4	0.1	5.2	2
7	3.2	4.7	1.4	2	1.2	4.1	1.5	1
6.4	3.2	4.5	1.5	2	0.7	3.7	1.6	1
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
5.1	2.5	3	1.1	2	1.5	2.2	3.3	1
5.7	2.8	4.1	1.3	2	0.3	3.1	2.1	1
6.3	3.3	6	2.5	3	2.4	5.6	0.8	3
7.2	3.2	6	1.8	3	2.5	5.6	0.8	3
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
6.9	3.1	5.1	2.3	3	1.9	5.01	0.6	3
5.8	2.7	5.1	1.9	3	1.4	4.6	0.9	3

We classified each testing cluster on the basis of corresponding training cluster. Here, we used naive Bayesian classifier in WEKA 3.6.2. Our proposed model can correctly classified 67 instances out of 68 in contrast with conventional naive Bayesian classifier which can correctly classified 65 instances. The following table IV shows the classification for testing data.

TABLE IV. Correctly Classified Testing Instances in Our Proposed Model

Training Cluster	Testing Cluster	Number of Instance at Testing Cluster	Correctly Classified Instance
Cluster 1	Cluster 1	30	29
Cluster 2	Cluster 2	18	18
Cluster 3	Cluster 3	20	20
	TOTAL	**68**	**67**

2. *Experiments with Benchmark data*

To estimate the performance of our approach with the naïve Bayesian classifier on different problem domains, we performed experiments on

four data sets obtained from UCI repository [17] that is shown in table I. The experimental results in table 5 demonstrate that our proposed approach attained better classification/accuracy rates than naïve Bayesian classifier. This is because to predict the class for an unknown data we use probability of the appropriate cluster only whereas in conventional naïve Bayesian classifier probabilities are calculated from the whole data set.

TABLE V. CLASSIFICATION/ACCURACY RATES(%)FOR NB CLASSIFIER AND PROPOSED CLASSIFIER

Dataset	No. of class	No. of cluster	No. of attributes	No. of Training data	No. of test data	NB classifier %	Proposed method %
Adult	2	5	14	32561	16281	83.13	87.38
Soybean (large)	19	5	35	307	376	88.56	89.63
Annealing	6	4	38	798	100	74.00	95.00
Iris	3	3	4	150	68	96	98.53

V. CONCLUSION

Classification is one of the significant functions of data mining which accurately predicts the target class for each case in the data. In our proposed model we have focused on the improvement of the classification rates for naïve Bayesian classifier. We tested our proposed model on a number of benchmark data and achieved higher classification accuracy rates than conventional naïve Bayesian classifier. This is because to predict the class for an unknown data we use probability of the appropriate cluster only whereas in conventional naïve Bayesian classifier probabilities are calculated from the whole data. In our future works, we plan to apply this concept in other classification methods. In addition, we plan to apply this approach in real world problem domains.

ACKNOWLEDGMENT

This research has been supported by the EU Erasmus Mundus Project - eLINK (east-west Link for Innovation, Networking and Knowledge exchange) under External Cooperation Window – Asia Regional Call (EM ECW – ref. 149674-EM-1-2008-1-UK-ERAMUNDUS).

REFERENCES

[1] http://united-states.bangkok.ws/documentation/oracle-/database/11.1/datamine.111/b28129/classify (accessed on, 21th April, 2010)

[2] R. O. Duda et al, Pattern Classification, *2nd ed. Chichester, U.K.: Wiley-Interscience*, 2000

[3] K. M. A. Chai et al, Bayesian online classifiers for text classification and filtering, *Proceedings of SIGIR 2002*, Tampere, Finland, pp. 97-104.

[4] T. Hastie et al, The Elements of Statistical Learning: Data Mining, Inference and Prediction, *Heidelberg, Germany: Springer-Verlag, 2001*.

[5] Xhemali, Daniela and J. Hinde, Christopher and G. Stone, Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages, IJCSI International Journal of Computer Science Issues, Vol. 4, No. 1, 2009

[6] Sopharak, K. Thet Nwe, Y. Aye Moe, Matthew N. Dailey and B. Uyyanonvara, "Automatic Exudates Detection with a Naive Bayes Classifier", *Proceedings of the International Conference on Embedded Systems and Intelligent Technology*, pp. 139-142, February 27-29, 2008.

[7] Kononenko I, 1990, Comparison of inductive and naïve Bayesian learning approaches to automatic knowledge acquisition, *Wieling, B. (Ed), Current trend in knowledge acquisition*, Amsterdam, IOS press.

[8] Langely, P., An analysis of Bayesian classifier, *Proceedings of the 10th national Conference on Artificial Intelligence* (San Matro, CA: AAAI press), pp. 223-228, 1992.

[9] http://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690-/slides-/Chapter8 (accessed on, 15th April, 2010)

[10] Ting Su, Jennifer Dy, "A Deterministic Method for Initializing K-Means Clustering," ictai, pp.784-786, 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'04), 2004

[11] Bashar Al-Shboul, and Sung-Hyon Myaeng, "Initializing K-Means using Genetic Algorithms", World Academy of Science, Engineering and Technology 54 2009

[12] Margaret H. Dunham, *Data Mining: Introductory and Advanced Topics*. Pearson Education Pte. Ltd., Singapore, 2005.

[13] Dewan Md. Farid, Jerome Darmont, Nouria Harbi, and Chowdhury Mofizur Rahman, "A New Supervised Learning Algorithm using Naïve Bayesian Classifier," In Proc. of the IADIS International Conference Information Systems 2010, 18-20 March, 2010, Porto, Portugal, pp. 78-84.

[14] Hartigan, JA, Wong MA. A k-means clustering algorithm. *Applied Statistics* 28; 100– 108, 1979.

[15] http://www.comp.brad.ac.uk/intranet/modules/AIWA/2005 /AIWA12Data%20Mining-clustering.pdf (accessed on,-25th April, 2010)

[16] Kaufmann L, Rousseeuw PJ. *Finding Groups in Data: An introduction to Cluster Analysis*. New York: Wiley 1990.

[17] http://archive.ics.uci.edu/ml/, Machine learning Repository, Center for Machine Learning and Intelligent system, university of California, Irvine(accessed on,-15th April ,2010).

Comparing the Performance of Wearing Helmet Behavior Model While Driving Motorcycle by Binary Logistic Regression Analysis Method and Learning Vector Quantization of Artificial Neural Network

Vatanavongs Ratanavaraha, Phattarasuda Witchayaphong,
Napat Lakhawattana, and Khedsadaporn Chaiwong

Abstract: **This study aimed to compare between efficiencies of motorcyclist helmet-wearing behavior models using binary logistic regression and learning vector quantization artificial neural network. The binary logistic regression yielded mathematical model that could forecast future motorcyclist helmet-wearing behavior. The other method, learning vector quantization (LVQ) would feed past data to the system and train it to learn the relationship among the data. This training enabled the artificial neural network to understand and memorize the pattern to predict other sets of data. This study employed a sample group of 8,635 motorcyclists from all provinces in Thailand. The data were analyzed in a descriptive statistical approach to describe motorcyclist helmet-wearing behavior using a statistical SPSS package. A motorcyclist helmet-wearing model was developed using binary logistic regression by a computer program, and LVQ artificial neural network by a Matlab program. When comparing efficiencies of the two models in term of percent accuracy, the study team found that binary logistic model yielded 71.35 percent while the LVQ offered 71.24 percent accuracy. Therefore binary logistic regression model is more accurate when used to forecast motorcyclist helmet-wearing behavior.**

Index Terms- artificial neural network; Binary logistic regression; helmet wearing behavior; learning vector quantization; motorcyclist.

I. Introduction

Rapid growths in technology, economy, and population have led to increase in travel demand. Accelerating development of the country resulted in development of transportation facilities which in turn raised a great amount of traffic accidents. These traffic accidents have been one of the causes of great loss to society and economy.

A Department of Highway's study, Value of Accident Damages in Thailand, conducted by a group of experts, found that Thailand suffered economic loss of 232,200 million baht per year or equivalent to 2.8 percent of gross domestic product. This study reflected that not only did accident reduction save lives, but it also prevented unnecessary economic loss.

Effective accident prevention and control planning depended on basic detailed data and behaviors of risk population group. This was to gain true appreciation on course of accident which would lead to the most on-target and efficient accident prevention and control plan.

The study objective is to compare motorcyclist helmet-wearing behavior using binary logistic model and LVQ artificial neural network. Each individual's decision to wear helmet depended on a number of reasons and factors such as hearing, experience on public relations or advertisement on accident prevention, sex, age, type of agency, type of perceived media, attitude and acceptance of traffic laws, and knowledge of traffic laws. The main objectives of applying logistic regression analysis method are to study relationship between dependent and independent variables and use the regression model to estimate or predict occurrence and opportunity of the dependent variables. The LVQ artificial neural network application fed past information to the system so it could learn data relationship. Training focused on memorizing patterns and used them for predicting other sets of data. The effectiveness was measured from errors. Small errors reflected that the artificial neural network could learn and memorize data relationship effectively.

II. Relevant Theories

A. Binary Logistic Regression

This is used when a dependent variable Y is dichotomous such as

$$Y = \begin{cases} 0 \text{ when sometimes wear helmet,} \\ 1 \text{ when always wear helmet} \end{cases}$$

and the binary logistic regression equation is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon \quad (1)$$

Where Y = dependent variable
X_1- X_n = independent Variable
β_0 = y –axis intercept
β_1-β_n = regression factor i to n
ε= error term

Coefficient of determination, R_2, shows degree that independent variables X described dependent variable Y. R_2 may be determined from

$$R^2 = 1-\frac{SSE}{SST} \quad (2)$$

Adjusted$R^2$is another goodness of fitmeasurement but it takes into account degrees of freedom of SSE and SST.

$$adjustR^2 = 1-\frac{SSE/(n-k-1)}{SST/(n-1)} = 1+\frac{(n-1)}{(n-k-1)}(R^2-1) \quad (3)$$

Where
adjustR^2=adjusted coefficient ofdetermination
n = number of sample
k = independent variable

B. Artificial Neural Network

Learning Vector Quantization (LVQ) is expansion of Kohonen network in a trained environment. LVQ network classifies input vectors by designating attributes for classes. Classes are predefined, some of which are used for classification and others are for training. LVQ network structure imitated self-organizing network. The difference is that each neuron output is labeled a class as seen in Figure 1. Normally, a class consists of more than one output neuron. Similar to Kohonen network, weighed neuron vectors will communicate with corresponding input vectors. When input x enters the network, a neuron output that is most similar to x will be a winner neuron. The network will adjust a weighed neuron vector of the winner neuron and other neurons according to the following relationship:

$$W_{ij}^{new} = \begin{cases} W_{ij}^{old} + \alpha(x_i - W_{ij}^{old}) & \text{if classification is correct} \\ W_{ij}^{old} - \alpha(x_i - W_{ij}^{old}) & \text{if classification is correct} \end{cases} \quad (4)$$

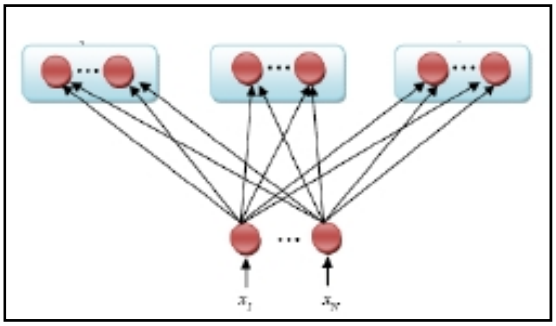


Figure 1. LVQ Network Structure

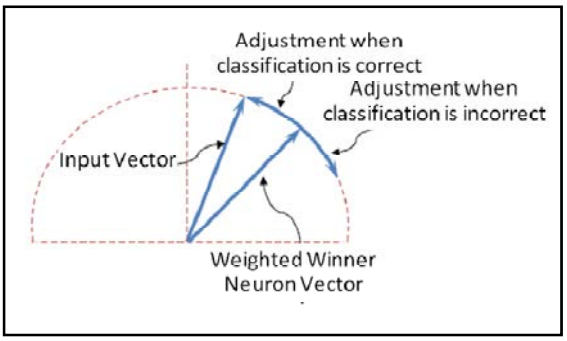


Figure 2. LVQ Network Learning

The LVQ learning rule rewards a winner neuron which matches the class of the correct input vector by adjusting the weighted neuron vector so it approach the input vector. On the other hand, if the winner neuron does not correctly classify the data, it will be punished by adjusting weighted neuron vector further from the input vector as shown in Figure 2.

C. Comparison of Model Effectiveness

The two models were developed from different grounds and their result accuracies were compared. The comparison of motorcyclist helmet-wearing behavior model was defined by percentage correct which could be calculated as equation (10)

$$\%Correct = \sum_{n=1}^{N} W_n / N \quad (5)$$

where

$$Y = \begin{cases} 1 \text{ when } n_{th} \text{ sample select class i or } P_n(i) > 0.5 \\ 0 \text{ otherwise} \end{cases}$$

Mean absolute percentage error (MAPE) is a measurement tool for statistical time series error which can be expressed as

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (6)$$

where A_t = actual value
F_t = forecasted value
n = number of tested samples

III. Research Methodology

Survey was accommodated by Department of Disaster Prevention and Mitigation including questionnaires, traffic accident campaign awareness, and knowledge, attitude and acceptance of traffic laws of sample groups from 26 provinces throughout the country. These provinces consisted of Bangkok, Kalasin, Kamphaengphet, Chanthaburi, Chumphon, Tak, Nakhon Nayok, Nakhon Pathom, Nakhon Sri Thammarat, Buriram, Prachuap Khirikhan, Phayao, Phang-gna, Phitsanulok, Phrae, Mahasarakham, Yasothon, Roi Et, Ranong, Satun, Samut Prakan, Sa Kaew,

ISBN 978- 974-672-556-9

Saraburi, Nongkhai, Ang Thong and Uttaradit. A set of 8,635 data points were divided into 6,045 data points (70%) for training, and 2,590 data points (30%) for testing. The percentages between the two were maintained the same across all sampled provinces.

This study involved effectiveness comparison between the two models. The analysis was divided into two sections as follows:

3.1. A development of motorcyclist helmet-wearing behavior model using binary logistic regression to determine the relationship between independent and dependent variables. Data were collected from samples in 26 provinces of Thailand for a total of 2,590 data points. The variables considered in this model include:

TABLE 1. ANALYSIS VARIABLES IN THE MOTORCYCLIST HELMET-WEARING BEHAVIOR MODEL

Variables	Collected Data
Y	1 motorcyclist wears helmet every time when riding a motorcycle 0 Otherwise
NEWS	1 Has never seen traffic accident campaign advertisement. 2 Sometimes see traffic accident campaign advertisement. 3 See traffic accident campaign advertisement almost every day. 4 See traffic accident campaign advertisements every day.
RULES	1 Follow traffic laws but not very strictly. 2 Strictly follow traffic laws.
SEX	1 Male. 0 Otherwise.
AGE	1 Age between 15-25. 0 Otherwise.
KNOWLEDGE	1 Possess standard knowledge on traffic laws. 0 Otherwise.

The model could be expressed in a mathematical form as follows:

$$Y = \beta_0+\beta_1(NEWS)+\beta_2(RULES)+\beta_3(SEX)+\beta_4(AGE)+\beta_5(KNOWLEDGES) \quad (7)$$

where: Y = Motorcyclist helmet-wearing behavior (sometimes/always)
NEWS = Awareness of traffic accident campaigns
RULES = Acceptance of traffic laws
SEX = Sex
AGE = Age
KNOWLEDGE = knowledge of traffic laws

3.2. A development of motorcyclist helmet-wearing behavior model using LVQ neural network. The network designated 1,000 epochs as the highest value. 70 percent of the data (6,045 data points) were used for training and 30 percent of data (2,590 data points) were used for testing. The test data were sampled from each province with the same portion.

This model development assigned dummy variables (0, 1) for all data input. The development of neural network designated motorcyclist helmet-wearing behavior as a target variable which included two results, sometimes and always wear helmet. Then the outcomes were compared to the real data to verify how accurate artificial network could predict the results.

$$Y = \begin{cases} 0 \text{ if sometimes wear helmet} \\ 1 \text{ if always wear helmet}. \end{cases}$$

IV. DATA ANALYSIS

A. *Application of Binary Logistic Regression Model.*

The model's influence factors were analyzed by SPSS package. Constants, factors and accuracy of the prediction were determined to develop an accurate and significantly acceptable model as shown in Tables 3 and 4.

TABLE 2. DEPENDENT VARIABLE ENCODING BY SPSS

Original Value	Internal Value
Sometimes	0
All Times	1

TABLE 3. GOODNESS-OF-FIT OF THE MODELS BY SPSS

Step	-2 Log Likelihood	Cox & Snell R Square	Nagelkerke R Square
1	3029.229	.190	.254

TABLE 4. PREDICTION ACCURACY ANALYSIS

Observed		Predicted		
		Y		Percentage Correct
		Sometimes	Always	
Step 1 Y	Sometimes	986	408	70.7
	Always	334	862	72.1
Overall Percentage				71.4

TABLE 5. ESTIMATES OF FACTORS FOR INFLUENCE VARIABLES

		B	S.E.	Wald	df	Sig.	Exp (B)
Step 1(a)	X1	-.108	.053	4.156	1	.041	.898
	X2	1.964	.095	426.930	1	.000	7.126
	X3	-.404	.088	20.810	1	.000	.668
	X4	.245	.093	6.942	1	.008	1.278
	X5	.431	.090	22.796	1	.000	1.539
	Constant	-1.036	.182	32.434	1	.000	.355

* Sig at 0.05

From Table 5 factors in the regression equation could be determined as shown in Equation (13) :

$$Y = -1.036 - 1.080X_1 + 1.964X_2 - 0.404X_3 + 0.245X_4 + 0.431X_5 \quad (8)$$

where
Y = Motorcyclist helmet-wearing behavior (sometimes/always)
X_1 = Awareness of traffic accident campaigns
X_2 = Acceptance of traffic laws]
X_3 = Sex
X_4 = Age
X_5 = Level of knowledge of traffic laws

B. Application of LVQ Artificial Neural Network

Let motorcyclist helmet-wearing behavior be a target variable which could be in two forms, sometimes and always wear helmet.

$$Y=\begin{cases} 0 \text{ if sometimes wear helmet} \\ 1 \text{ if always wear helmet}. \end{cases}$$

Input variables were selected from a suitable model. The number of variables equaled to the number of data points used in the development of binary logistic regression model. These variables were all converted to dummy variables and tested in LVQ artificial neural network which could be expressed as follows.

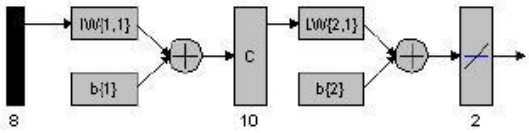


Figure 3. LVQ Artificial Neural Network Architecture for Development of Helmet-wearing Behavior Model

The number of neurons varied between 10 and 20. This training aimed to instruct the artificial neural network to understand and memorize relationship between input and output fed into the system and to predict other data sets. The effectiveness of this training would be measured by mean square error (MSE). Small MSE showed that the artificial neural network could effectively learn and memorize relationship between input and output.

TABLE 6. TRAINING AND TESTING RESULTS FROM DEVELOPMENT OF HELMET-WEARING BEHAVIOR MODEL USING LVQ ARTIFICIAL NEURAL NETWORK, MEASURED BY MEAN SQUARE ERROR

Network Architecture	Epochs	Training Results	Test Results
8-10-2	20	0.287345	0.2919
8-10-2	50	0.286849	0.2876
8-10-2	70	0.287345	0.2876
8-10-2	100	0.286849	0.2876
8-10-2	200	0.286849	0.2876
8-20-2	20	0.287014	0.2876
8-20-2	50	0.286352	0.2876
8-20-2	70	0.286849	0.2876
8-20-2	100	0.286849	0.2876
8-20-2	200	0.286849	0.2876

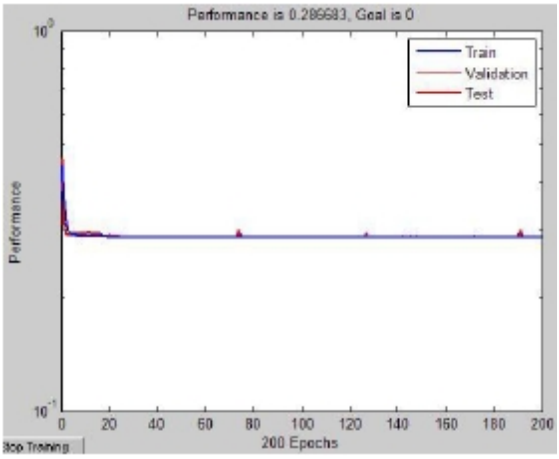


Figure 4. Graph showing Result of Effectiveness Test from LVQ Artificial Neural Network

The test result from the model using LVQ artificial neural network was shown in Table 7.

TABLE 7. RESULTS OF ACCURACY ESTIMATION ON FACTORS INFLUENCING MODEL USING ARTIFICIAL NEURAL NETWORK METHOD

Network Architecture	Accuracy Percentage
8-10-2	70.81%
8-10-2	71.24%
8-10-2	71.24%
8-10-2	71.24%
8-10-2	71.24%
8-20-2	71.24%
8-20-2	71.24%
8-20-2	71.24%
8-20-2	71.24%
8-20-2	71.24%

V. CONCLUSION

An analysis of motorcyclist helmet-wearing behavior using binary logistic regression yield a ρ_2 of 0.254 which could describe the behavior to an acceptable level of accuracy with an equation:

$$Y = -1.036 - 1.080x_1 + 1.964x_2 - 0.404x_3 + 0.245x_4 + 0.431x_5$$

It was found that variables affecting motorcyclist helmet-wearing behavior included hearing/seeing traffic accident campaign advertisement, sex, age, newspaper and magazine media exposure and knowledge on traffic laws.

The LVQ artificial neural network architecture that produced the best result yielded MSE 0.286849 after training. This indicated that the network's learning effectiveness was within an acceptable level. Testing result yielded MSE 0.28760 which was higher than that from training. This was because the test samples were of smaller size and the LVQ network could not establish as good relationship between input and output.

ISBN 978- 974-672-556-9

Comparison between the two models revealed that the binary regression analysis model produced 71.36 percent accuracy while the LVQ artificial neural network produced 71.24 percent accuracy. It could be concluded that the binary regression analysis model yielded 0.11 percent better results than the LVQ artificial neural network.

The binary logistic model generated a utility function that could describe factors affecting alternative selection. The results were merely prediction on each individual's helmet-wearing behavior. The LVQ artificial neural network, on the contrary, was a black box that clearly classified each individual's decision but could not explain significance of each governing factor.

ACKNOWLEDGEMENT

We would like to thank a number of relevant government agencies which provided secondary data which was crucial in success of this study.

REFERENCES

[1] Srikaew, A., Computational Intelligence, Bangkok, 2009, Charan Sanitwong Publishing.
[2] Wanichbancha, K., Advance Statistical Analysis Using SPSS Package, Department of Statistics, School of Commerce and Accounting, Chulalongkorn University.
[3] Amornnimit, U., Data Analysis Using Logistic Regression Method, Department of Applied Statistics, School of Sciences, University of Thai Chamber of Commerce.
[4] Whelan, G., "Modeling Car Ownership in Great Britain", Transportation Research Part A: Policy and Practice, Vol. 41, Issue 3, pp. 205-219.
[5] Kanitpong, K., Boontob, N., and Thanaboriboon, Y., Transportation Research Record: Journal of the Transportation Research Board, Helmet Use and Effectiveness in Reducing the Severity of Head Injuries in Thailand.
[6] Jomnonkwao, S., Development of Total Travel Distance and Mode Choice Model in Nakhon Ratchasima Municipality, Master Thesis (Transportation Engineering), Suranaree University of Technology.
[7] Suksawat, S. Factors Affecting Behavior of Injured Motorcyclists, Master Thesis (Nursing), Burapha University, 2000.
[8] Na Chiangmai, T., Attitude and Behavior on Using Safety Helmet of Employees in Sankamphaeng, Chiang Mai, Master Thesis, Chiang Mai University, 2

Session 10: Education and Applications

ISBN 978- 974-672-556-9

Evaluation of Student Performance with Data Mining: An Application of ID3 and Cart Algorithms

Manawin Songkroh
College of Arts, Media and Technology,
Chiang Mai University, Chiang Mai, Thailand

Andrea Kő
Corvinus University of Budapest, Hungary

Abstract: This paper has exploited RapidMiner- the open-source software- in classify students by their academic performances. The process is complied with CRISP-DATA MINING to ensure its completeness and accuracy. CART and ID3 algorithms are used in this paper to serve the purpose. There is slightly difference in model evaluation- using cross validation and model testing with 20% of evaluation data set- by both algorithms but both shows the high accuracy ratio which can be deployed to the advising process for the college.

Index Terms- algorithm application; CART algorithm; data mining; ID3 Algorithm; RapidMiner; student performance evaluation

I. INTRODUCTION

Evaluating students' performance is a complex issue, which can't be restricted for the grading. Reasons of good or bad performances belong to the main interests of teachers, because they can plan and customize their teaching program, based on the feedback. Data mining is one of the approaches, which can provide an effective assistance in revealing complex relationships behind the grades. This paper gives an overview about the application of data mining for students' performance evaluation. CRISP methodology was utilized as a framework in our data mining approach and RapidMiner was our implementation environment.

Data mining is the process of extracting patterns from data with various names, i.e., knowledge discovery in database (KDD), information discovery, knowledge extraction .and data pattern processing. [1] differentiates the most two popular terms, i.e., data mining and KDD which is "KDD refers to overall process of discovering useful knowledge from data and data mining refers to a particular step in this process (p.36)."

Data mining has gained its popularity about 20 years ago in many fields, for example, banking industry for fraud detection [2], NASA for detecting new stars [3], bioinformatics Industry [4]. Its popularity has increased horizontally and vertically, namely, more number of people are using or trying to use it and so do the other fields.

This study is one example from educational institution that would like to apply data mining, based on the data collected, to classify the students, hoping to assist them in their academic success.

A. Business Understanding

Define Business Objectives of the Data Mining Model.

The college of this study- CAMPUS- has been founded in 2004. It offers major in Software Engineering (SE), Modern Management and Information Technology (MMIT), Animation (ANI) and Knowledge Management (KM). It is the newest faculty of CMU which is experiencing with many opportunities. In this study, the focus is on MMIT department. The study aims to classify students' academic performance according to subjects taken with the goals of answering these questions:

What kind of classes, associations and rules can we determine for students' performance applying data mining? Can we take use of the classification from the algorithm? Can advisors apply the model to advise to their students more efficiently?

Consequently, the model would retain more students in the program or even ensure study to graduate under the appropriate time frame, i.e., within four years.

Assess the Business Environment for Data Mining

The data contains CAMPUS students' record starting from 2006-2009, comprising of personal information, grade in each subject, and GPA (students' grade point average) in dbf file formats. The data is acquired from Registration Office of CMU with the approval from the head of the department. The confidentiality terms has been agreed upon, i.e., the research team has no intention of disclosure of individual students' record and the data is used for the purpose of this study only.

The open-sourced RapidMiner 4.4 is used to analyze data. The Rapid Miner 4.4 User Guide contains operator reference and developer tutorial which assist in using the RapidMiner efficiently.

Formulate the Data Mining Goals and Objectives

Data Mining Goals. Building a suitable database: As data contains of many students' record, it is difficult to utilize all of those data. One of the data mining goals in this study is to prepare the data to be ready for analyses. Data quality is a key issue in data

mining project, if data has poor quality the result is useless.

Objectives, tasks and subtasks for each goal are well explained in CRISP-DATA MINING model. The data mining goals should be "deploy a good predictive model to meet the business objective (p.37)" [5].

B. Data Understanding

1. Data Acquisition

Data was received in January 2010 from the Registration Office of CMU in 5 dbf files. Data for data mining must be in the flat-file format. The original data were ranked by ascending rows so they have been re-arranged in the proper format and combining all the five files together in one flat-line, big sheet by manually arranging it. The correctness has been checked to ensure the accuracy of the data.

2. Data Description (Variables, Cases, Descriptive Statistics, report)

Data contains of students' grade point average (GPA) in each semester starting from semester 1/2006 to semester 1/2009, including provinces and gender.

From Table 1, 2 and 3, there are 235 male and 272 female students. From this amount, 346 students are in regular program and the rest of 161 students are in special program. Students from Chiang Mai province are the majority reflecting of 305 students (60.16%). The second biggest is students from Lamphun (60 students reflecting 11.83%). The third biggest are students from Chiang Rai province in the amount of 47 students (9.27%). The rest are distributed to Lampang (35 students), Payao (19), Prae(13), Nan (5) and other (19).

TABLE 1. FREQUENCY DISTRIBUTIONOF MMIT STUDENTSBYGENDER.

Gender	%
Male	46
Female	54
Total	100

TABLE 2. FREQUENCY DISTRIBUTIONOF MMIT STUDENTSBY PROGRAM

Program	%
Regular	68
Special	32
Total	100

TABLE 3. FREQUENCY DISTRIBUTIONOF MMIT STUDENTSBYPROVINCES

Province	%
Chiang Mai	60.16
Lamphun	11.83
Chiang Rai	9.27
Lampang	6.90
Payao	3.75
Prae	2.56
Nan	0.99
Bangkok	0.99
Other	3.55
Total	100

TABLE 4 TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOF STUDENTS WITH GRADE A

Subject	Amount
Students Quality Dev to thru Act I (C262)	88
Rapid App Development (C242)	53
IT Literacy (C238)	50
Usage of Thai Language (C035)	49
Introduction to STAT Concept I	46

Table 4-14 represents the top 5th frequency of each grade, i.e., A, B+,.....I. The most frequent A grades are in C262 (955100: Student Quality Development thru Act I) (88), C242 (954243: Rapid App Development) (53), C238(50), C035 (050100: Usage of Thai Language) (49), and C101 (208100, Intro to Stat Concept) (46). As this study concerns about students with academic performance, we will take a look at the frequency table with grade lower than C (or 2 point). Table 10 shows that there are 112 students got D+ in C143 (951101: Art and Animation History) and 55 students in C029 (013103: General Psychology) and C008 (001104: Foundation English II). Table 11 shows that there are 84 students with D grade in C073 (177321: Labor Law) and 58 students with D grade in C008(001104:English Foundation II). Table 12 shows the frequency of F grade students. There are 32 students with F grade in C008 and C003. W stands for "Withdrawn" subjects. Table 13 shows that C101 (208100: Introduction to STAT Concept) has the highest withdrawn frequency, in the amount of 75 and the second highest is in C115 (701100: Elementary Accounting) (50 students).

TABLE 6. TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOF STUDENTSWITH GRADE B+

Subject	Amount
IT Literacy (C238)	85
Rapid App Development (C242)	53
Information and COMM Technology (C239)	51
Usage of Thai Language (C035)	50
English for ACAD Purpose(C010)	41

TABLE 7. TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOF STUDENTSWITH GRADE B

Subject	Amount
Usage of Thai Language (C035)	166
IT Literacy (C238)	135
Information and COMM Technology (C239)	73
English for ACAD Purpose (C010)	70
Integrated Sciences (C075)	70

TABLE 8. TOP-5 SUBJECT FREQUENCY DISTRIBUTION OF STUDENTSWITH GRADE C+

Subject	Amount
Usage of Thai Language (C035)	157
IT Literacy (C238)	120
Integrated Sciences (C075)	94
Art and Animation History (C143)	91
English for ACCAD Purpose (C013)	75

TABLE 9 TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOF STUDENTSWITH GRADE C

Subject	Amount
Survey of Mathematics (C080)	118
Fundamental in Socio & Anthro (C065)	97
Art and Animation History (C143)	95
Integrated Science (C075)	91
ORG Safety (C106)	74

TABLE 10. TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOF STUDENTSWITH GRADE D+

Subject	Amount
Art and Animation History (C143)	112
General Psychology (C029)	55
Foundation English II (C008)	55
IND ORG & PROD MGT (C107)	51
Foundation English I (C003)	51

TABLE 11. TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOFSTUDENTSWITH GRADE D

Subject	Amount
Labor Law (C073)	84
Foundation English II (C008)	58
Art & Animation History (C143)	50
Fundamental in Socio & Anthro (C065)	44
Foundation English I (C003)	37

TABLE 12. TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOF STUDENTSWITH GRADE F

Subject	Amount
Foundation English II (C008)	32
Foundation English I (C003)	32
Labor Law (C073)	25
Principles of Management (C121)	13
Introduction to Stat Concepts (C101)	13

TABLE 13. TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOF STUDENTSWITH GRADE W

Subject	Amount
Introduction to STAT Concept (C101)	75
Elementary Accounting (C115)	50
Art and Animation History (C143)	33
Survey of Mathematics (C080)	27
General Psychology (C029)	27

ISBN 978- 974-672-556-9

TABLE 14. TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOF STUDENTSWITH GRADE I

Subject	Amount
KM System (C244)	51
Student Quality Development Thru Act I (C262)	36
Introduction to Manufacturing Information System (C257)	6
Art and Animation History (C143)	3
Rapid App Development (C242)	3

TABLE 15. TOP-5 SUBJECT FREQUENCY DISTRIBUTIONOF STUDENTSWITH GRADE N

Subject	Amount
Integrated Science (C075)	183
Foundation English I (C003)	183
Usage of Thai Language (C035)	33
Art and Animation History (C143)	31
IT Literacy (C238)	28

Table 15 shows the frequency of "N". C238 has 28 students with N and C143 (Art & Animation History) has 31 students. There are 33 students in C035 and there are 183 students in C003 (Foundation English I) and C075 (Integrated Science).

3. Data Quality Assessment (missing values, outliers, report)

Some records contain incomplete data (no information at all) that have been deleted from this analysis. In some records, provinces are missing; thus, maximum likelihood imputation has been used to fill those data. In this case, there are 10 missing province data so those records have been imputed to "Chiang Mai."

The following points were noted during this stage:

There are no outliers in the data.

Data Transformation was done by replacing more or equal to 2.5 GPA to "Good" (performance) and less than 2.5 GPA to "Bad" (performance).

Data should be divided into two sets, i.e., training set and evaluation in the proportion of 80:20 [5]. The division has done by random sampling.

As the data contains of students' records in different years, there are many subjects that students has taken the course yet. Thus, Year 1 students has been deleted for the analysis as it shows many "N" (not available) data which will affect the analysis,

Variables are also deleted due to lack of "valid attributes" as some subjects are not required to take. Only required variables (required subjects) for the first two years are used in the analysis to over the mentioned incompleteness of data and it is enough to use only those required subjects because students who has academic problems would be expelled (due to GPA lower than 2.0) in the third academic year (the 7th semester).

No data derivation is required.

Feature selection is beneficial when data is comprised of many attributes/variables. As [6] have put it "Reducing the dimensionality of the data reduces the size of the hypothesis space and allows algorithms to operate faster and more effectively. In some cases accuracy on future classification can be improved; in others, the result is a more compact, easily interpreted representation of the target concept." However, in this study feature selection has been done manually based on the applicability of resulted models. Attributes contains of subjects from the whole 4 years of academic study plan. As the purpose of this study is to detect the students with poor academic performance (GPA < 2.5) in the early period of the study; namely, in the first and the second year, variables has been reduced from 45 subjects to 13 subjects. In these 13 subjects are the mandatory subjects for every student for the first two years of the study plan. The subjects in the 3rd and 4th year have been taken out due to it applicability.

II. MODELING

The Following models were selected from the arsenal of data mining techniques:

ID3 Classification

Standard Classification Trees with Deployment

Classification algorithms are used to classify students' GPA so that proper advise can be provided to students. These two algorithms are ID3 Classification and CART.

CART can be described as "the operator leans decision trees from both nominal and numerical data. Decision trees are powerful classification methods which often can also easily be understood. The actual type of the tree is determined by the criterion, e.g., using gain ratio or Gini for CART([7],p.165)."

ID3 can be described as "The operator learns decision trees without pruning using nominal attributes only. Decision trees are powerful classification methods which often can also easily be understood" ([7],p.178). To put in the other words, ID3 learns an unpruned decision tree from nominal attributes only.

ID3 Classification

ID3 Classification Trees are powerful tools for classification and prediction. The attractiveness of decision tress is due to the fact that, in contrast to neural networks, decision trees represents rule that are easy to interpret.

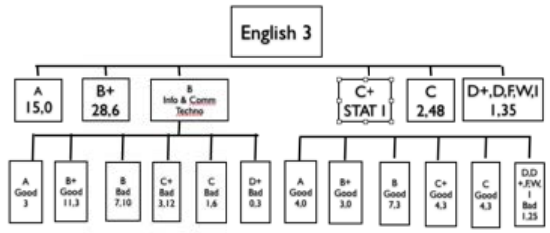


Figure. 1 Classification from ID3 Algorithm

Each box in the tree in Fig. 1 represents a node. In each node contains information about the number of instances at the node, and the distribution of dependent variables (academic results, i.e., good or bad). The instances in the root node contain 262 instances from the training set. Below the root note (parent) is the first split that splits the data into 11 nodes (children) based on the predictor: Grade in C010 (English for Academic Purpose).

The left most node resulting from this split contains 15 instances associated with good GPA, and this is the case of pure node and will not be split further as most of the stances have the same value of the dependent variables. The second node from the left is further split based the grade of Information and Communication Technology Class. The first child node of Node Information and Communication Technology contains 3 instances with good GPA (also the pure node). The second child node of C239 contains of 7 instances with good GPA and 10 instances with bad GPA.

The third child node of Info and Comm Technology (B+) contains of 7 good and 3 bad GPA. The interpretation of this ID3 Classification can be done similarly to the other child nodes. The last node (the bottom nodes) is called "leaf" which is helpful for prediction purpose. In summary, the leaves may be either *"pure"* or "un-pure" (containing two dependent variables in one node). There are pure nodes in two extreme scenarios, that is, for students with Grade A in C010 and with Grade D, F, I, NA, and W.

The rules for the leaf nodes in Figure X are generated by following path down the branches until a leaf node is encountered. For instance,

IF Grade in English 3 = B
AND Grade in Info and Comm Technology = A
Then Result= Good.

Model Assessment (Classification Matrix: ID3 Model)

The classification matrix shows the number of cases that were correctly classified and those that were miss classified as the other category.

TABLE 16. CLASSIFICATION MATRIXOF ID3MODEL

	True Good	True Bad	Class Precision
Predict. Good	77	41	65.25%
Predict. Bad	18	125	87.41%
Class Recall	81.05%	75.30%	

* Classification Error 22.63%

In this case, the overall model could correctly predict whether the students' academic performance (GPA) was good or bad with 77.37% (100-22.63%). Our primary good for this study is to find students' with bad academic performances. The percent of correct predictions for bad category is 75.30%; namely, if there are 100 (academically) bad students, our model will correctly classify approximately 75 as bad which would assist the college to provide them with proper academic assistance.

Assessment of ID3 Model

To assess the model, the performance evaluation was utilized; Figure 2 shows steps in model evaluation.

The data sources have been divided into two sets which are the training set (the upper node and the training set (the third node). As mentioned earlier, there is no rule of thumbs in terms of the number of training and evaluation set. [5] suggests that 80:20 of training : evaluation is appropriate.

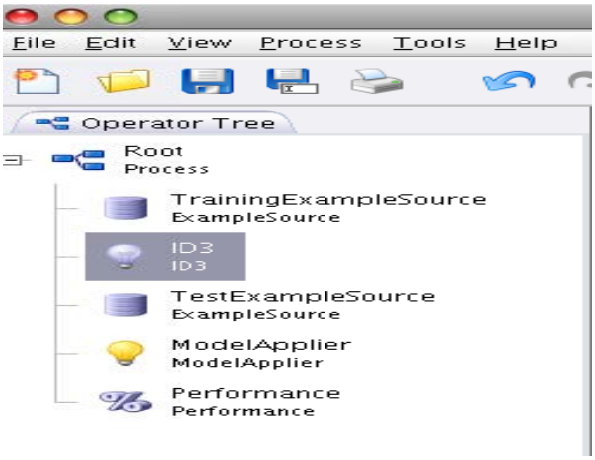


Figure 2. Model Evaluation Step

TABLE 17. MODEL EVALUATION MATRIXON ID3

	True Good	True Bad	
Predict. Good	13	7	65.00%
Predict. Bad	6	38	86.36%
	68.42%	84.44%	

*Hit Ratio (Accuracy)= 79.69%

Hit ratio of 79.69% indicates that model can predict students' performance quite well for both good and bad students. The percentage of correct predictions for the bad category is 84.44% which is higher than that of good category by 16.02% (84.44-68.42%). In other words, if there are 100 bad students, the model can correctly detect them in the number of 84 students and if there are 100 good students, the model can correctly detect 68 of students.

Standard Classification Trees (C&RT or CART)

[5] mentions that CART is "classification and regression tree algorithm developed by Breiman et al. (1984) [7]" After running CART algorithms, the result is shown in Figure 3

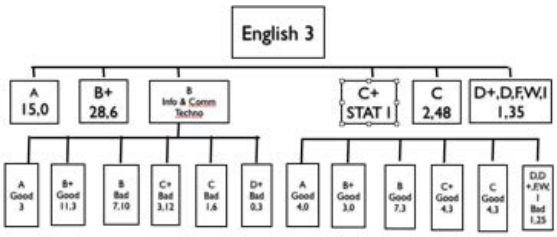


Figure 3. Classification from CART Algorithm

Fig. 3 shows CART splits nodes to 26 leaves. From the most left, the first node contains of 15 cases (students with Grade A from English 3). To the last 6 leaves on the right, all nodes shows "bad" results for students with Grade D, D+, F, I, NA, and W.

The rules for the leaf nodes in Figure 3 are generated by following the path down the branches until a leaf node is encountered. For example,

IF English 3 Grade = C+
AND Stat I Grade = A
Then Performance = Good.

It is interesting to see whether if there is any difference between these two models in evaluation and assessment step.

Model Assessment (Classification Matrix: CART)

TABLE 18 CLASSIFICATION MATRIX OF CART MODEL

	True Good	True Bad	Class Precision
Predict. Good	62	28	68.89%
Predict. Bad	33	138	80.70%
Class Recall	65.26	83.15%	

*Error 23.36%

Table 18 shows the overall model can correctly predict whether the students' performance was good or bad with 76.64% accuracy. The percent predictions for the bad category are 83.15% and for the good category is 65.26%; namely, the model can predict bad category better than good category in the number of 17.89%. In other words, if there are 100 bad students, the model will correctly classify 83 as bad and on the other hand, if there are 100 good students, the model will correctly classify 65 as good.

CART Evaluation

The other way to evaluate the model is by applying the model in the evaluation data set.

TABLE 19. MODEL EVALUATION MATRIXON CART

	True Good	True Bad	Class Precision
Predict. Good	12	9	57.14%
Predict. Bad	7	36	83.72%
Class Recall	63.16%	80.00%	

*Accuracy = 75%

Table 19 shows the overall model could correctly predict whether the students' performance was good or bad with 75% accuracy. The percentages of correct predictions for the bad and good category are 63.16% and 80.00% respectively. In other words, based on the evaluation data set, if there are 100 bad students, the model can detect 63 of them as bad and if there are 100 good students, the model can detect 80 of them as good.

Comparison of ID3 and CART Model

The comparison of both models would provide us some useful information on whether which is the best model, under the circumstance, to use. Table 20 shows the comparison of model performances from cross validation and model testing.

TABLE 20. COMPARISON OF MODEL EVALUATION

Evaluator	Overall Accuracy	Accuracy for Good	Accuracy for Bad
ID3-Cross Validation	77.37%	81.05%	75.30%
ID3-Test Set	**79.69%**	**68.42%**	**84.44%**
CART-Cross Validation	76.64%	65.26%	83.15%
CART-Test Set	**75.00%**	**63.16%**	**85.00%**

Table 20 shows that for ID3 cross validation, the overall accuracy is 77.37% which is higher than that of CART by 2.37%. Apart from that, when tested with the evaluation data set, ID3 model shows the higher overall accuracy of 79.69% compared to 75.00%. The prediction of good and bad category accuracy of ID3 model is 68.42% and 84.44% respectively. When compared to those of CART model, the prediction of good category is higher by 5.26%; however, the rate is lower for the prediction of bad category by 0.06%.

Deployment

A. Plan Model deployment (create deployment plan.) The model will be introduced to MMIT students' advisor and encourage them use them with the first year student of Year 2010. The important step is that advisors should be agreed upon the model and use it, apart from the other relevant concerns. Focus group meeting can be arranged to accomplish this objective.

B. Plan model monitoring and maintenance. At the end of the first year (after the students receive their GPA), the accuracy and predictability of model

will be evaluated by plugging new grades to the algorithm. The ratio of good/bad performance will be recorded to check the applicability of the model by comparing the ratio from the previous year.

III. Conclusions and Future Work

Data preparation has consumed the majority of the time in this study and this is very common in conducting data mining project. The results provide the information on which mandatory subjects are essential in determining the success of students. These two classifiers were used to differentiate the students. ID3 algorithm is better than CART algorithm when tested for model performance with cross validation and tested on evaluation data set (20% number of data, randomly selected). The model can predict "bad" or "poor" students than that of "good" students (85% compared to 63.16%).

The classification model will *not only* predict students' performance *but also* served as the protective model for advisors in providing additional and appropriate classes. For example, "English for Academic Purpose" is the subject that determines which class the students belong to (good or bad). The second determiners for the splits are "Introduction to STAT concept" and "Information and Communication Technology." Based on the new information from the algorithms, students should be informed so that more effort can be put into these subjects.

The future work of data mining, related to this topic, will be done in different majors, that is, in Software Engineering and in Animation. We expect to have more input data and it is good to try different algorithms like Neural Networks.

Acknowledgement

This work was the result of exchange program by Erasmus Mundus Program with additional support from Corvinus University of Budapest-EU partner, especially from Dr. Andrea Kő for data mining procedure and Dr. Gabor Andras for arranging the program, including Dr. Nopasit Chakpitak for suggestion on the topic.

References

[1] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth & R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining.* The MIT Press, 1996. Print.

[2] K. Chan, F. Philip, F. Wei, Prodromidis, Andrea, Stolfo, & J. Salvatore. *Distributed Data Mining in Credit Card Fraud Detection.* Ebook. 21 March 2010.

[3] U. M. Fayyad, S.G. Djorgovski, & N. Weir. "From Digitized Images to Online Catalog Data Mining a Sky Survey." *AI Magazine* 17.2 (2010)Print.

[4] L. Y. Koh, M. L. Lee, A. M. Khon, P.T.I. Tan & V. Brusic, ed. *Duplicate Detection in Biological Data using Associating Rule Mining.* Proceeding of the 2nd Europe Workshop on Data Mining and Text Mining in Bioinformatics. n.d. Print.

[5] R. Nisbet, J. Elder & G. Miner. *Handbook of Statistical Analysis and Data Mining Applications.* Har/DVD ed. Boston: Academic Press, 2009. Print.

[6] M.A. Hall, & G. Holmes, "Benchmarking Attribute Selection Techniques for Discrete Class Data Mining." *IEEE Transactions on Knowledge and Data Engineering* N.A..N.A. (2002): 1-16. Print.

[7] Mierswa, M. Wurst, R. Klinkenberg, M. Scholz, & T. Euler, ed. *Rapkd Prototyping for Complex Data Mining Tasks.* . Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2006. Dortmund, Germany: Rapid-I. 2001. Print.

The Application of a Situated Sign Language and Total Communication with an Animation Dictionary for Improving English Vocabulary Learning of Thai Primary School Students with Hearing Impairment

S.Wicha, N. Chakpitak and K. Meksamoot
College of Arts, Media and Technology
Chiang Mai University 239 Huay Kaew Road, Muang District, Chiang Mai, Thailand
santichai@mfu.ac.th,nopasit@camt.info, and komsak@ieee.org

***Abstract*- From the first phase of the study the researcher proposed a way for teaching and how to learn the English vocabulary by using the 'Total Communication with Animation Dictionary' (TCAD). This is a continuation from the previous work that has been extended to illustrate a new way for teaching hearing impaired students to improve their learning abilities by using a supplementary educational technology tool. Since the TCAD had already been implemented we found that the tool supported vocabulary long-term memory retention as well as discovering how to improve the vocabulary acquisition in a more effective method.**

The aim of this study is to enhance the vocabulary cognition learning for the hearing impaired through the use of situated sign language bundle with TCAD for improving the vocabulary comprehension. This quantitative research selected a target group of 70 primary students, from grades one to three, at the Anusarnsoontorn School for the Deaf in Chiang Mai, Thailand. The population group used total communication with a dictionary that was composed of animated characters combined with situated sign language that classified the vocabularies in TCAD by the category of location. This study revealed that total communication with animation dictionary plus situated sign language is essential supplementary vocabulary retention and a comprehensive instruction tool for the hearing impaired.

Index Terms- Hearing Impaired; Language Immersion;3D Animation; Total Communication; Vocabulary Language Acquisition; Web-based Learning

I. INTRODUCTION

Many studies on hearing loss indicate that the hearing impaired has difficulties in speech understanding but they have a greater visual perception capacity. In terms of learning, when individuals with normal hearing abilities are compared with individuals that have a hearing impairment, the acquirement procedure of knowledge is absolutely different. Individuals with a hearing impairment learn from a sign based technique. This technique, however, has a limitation on vocabulary learning. Individuals that have a hearing impairment are denied the opportunities to learn from speech-based technique, basic language acquisition, and techniques which are text-based. Without learning text-based techniques and speech-based techniques not only is there a limit for acquiring vocabularies but these students also cannot socially function in a normal environment; especially communicating with ordinary people.

II. LITERATURE REVIEW

In this research, there were two major areas involved. First, cognitive learning is a theory that discusses about learning, memorizing, recalling, and recognizing. Second, disability learning describes the learning process that is focused on auditory processing order.

According to Schuman's research [1] learning theory can be classified into behaviorism, cognitivism, and constructivism. Behaviorism focuses on a new behavioral pattern being repeated until it becomes automatic. Based on the thought process behind the behavior, cognitivism is used as indicators to describe what is happening inside the learner's mind. Constructivism focuses on preparing the learner to solve problems that are in ambiguous situations.

A. Cognitive Learning

Cognitive learning is the result of listening, watching, touching or experiencing. This is described in the following figure (Cognitive Model of Learning by Sharon Derry's review of cognitive learning theory)

Based on the Cognitive Model of Learning by Sharon Derry [2], learning and remembering meaningful information for a person is set in a pattern. First, new information is obtained through human contact, reading literature, media, sound, etc. In addition, prior knowledge comes from experience or learning. Therefore, people have already gone through life absorbing useful information that provides benefit for their life and a way for living. From here, new information and prior knowledge has

ISBN 978- 974-672-556-9

been fused together to form comprehension or a working memory. Next, elaboration is formed when there is a connection between new information and prior knowledge, thus increasing their relationship. After elaboration, learning starts to resume as new information has been integrated with the knowledge network. If utilized effectively knowledge then becomes meaningful and useful. Finally, learning revolves around prior knowledge. This enables the person to retrieve knowledge that was specifically learned. However, the construction of knowledge was never specifically learned. The conclusion derived from the knowledge network. Based on the literature of cognitive learning, the study provides a good start to set the concept of how a person learns. Although the work may seem to be biased against the hearing impaired, the cognitive learning model by Derry can be used as the critical link that maps out a structured plan in creating a device for the hearing impaired.

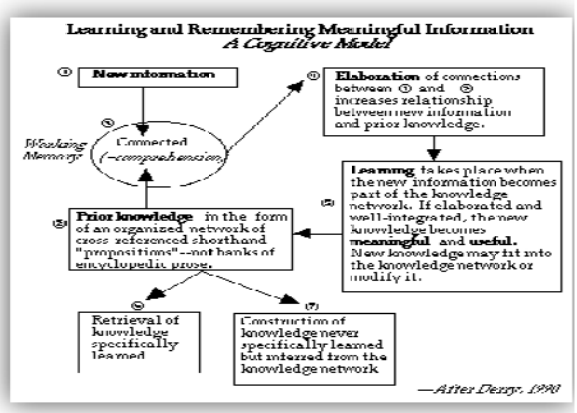


Figure 1. A Cognitive Model of Learning from Derry 1990

B. Memory Retention

"Learning Strategies Development" [3] outlined the details of the things that make people forget. The assertion for "forgetting and remembering" are as followed

The interference or the confusion factor (e.g. mental overcrowding, multi-tasking) is one of the primary reasons that can make the person likely to forget things. Other reasons for forgetting can be identified by the negative attitude or self concept of the individual. Sometimes these individuals do not learn well enough and easily forget things. The individual can sometimes be faced in a situation with not having the right cue. For example the person has studied in one way but the test question is presented in another way. Also, these individuals can experience a lack of attention or concentration on the material at hand. Things are remembered by the process that starts from thinking to encoding to rehearsing and to retrieving. From here, things start to be committed into memory that is based in pictures.

Memory is developed by organizing, funneling information, associating or connecting with prior knowledge, applying emotion, and grouping the sequence of information together. Long term memory is developed by repeating, reciting, rehearsing, elaborating, connecting, and teaching someone.

C. Cone of Learning

The cone of learning [5] illustrates the learning level of normal people after being taught for two weeks. From the figure above there are two types of learning: passive learning and active learning. Passive learning is found in a traditional classroom: instructors lecture and verbalize information to passive note-taking students. The instructor is a 'verbal' textbook that reads the definitions to the class. Students are considered as 'empty' vessels being filled with knowledge or human 'tape recorders'. Passive learning handles only 10% to 50% of the knowledge retention when time passes to two weeks.

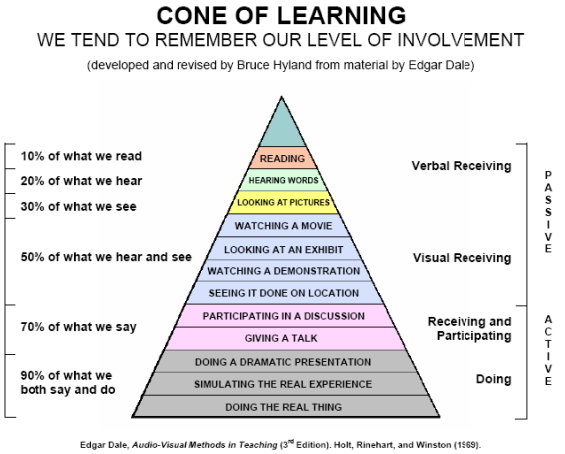


Figure 2. Edgar Dale, Audio-Visual Methods in Teaching (3rd Edition). Holt, Rinehart, and Winston [4]

In the case of active learning, the instructors created 'a learning environment' in which students can learn to restructure the new information and their prior knowledge into new knowledge about the content and for application. According to this cone, activities that stimulate verbal reception, visual reception, participation, and performing are brought to class. While conducting active learning, the instructors create 'a learning environment' in which students can learn to restructure the new information and their prior knowledge into new knowledge about the content and for putting it into practice. According to this cone, activities that stimulate verbal receiving, visual receiving, participating, and doing are brought to the classroom. Active learning handles about 51% to 90% of the knowledge retention when two weeks time passes by.

The cone of learning model is a classic example that paints a clear picture on how a person should learn. Since the participation of activities reinforces learning the device created will be able to act as a

supporting learning tool to give the students with a hearing impairment a rapid means to learn effectively. The device will link the learning behavior, environment, curriculum agendas, as well as instructors to produce a positive outcome for these students' learning abilities.

D. Vocabulary Language Acquisition

Vocabulary Language Acquisition (VLS) constitute a subclass of language learning strategies, which are applicable to a wide variety of language learning tasks, ranging from the more isolated (vocabulary, pronunciation, grammar) to integrative tasks like oral communication and reading comprehension. Studies such as Schmitt and McCarthy [6] have shown that language learning strategies are not inherently 'good', but depend on the context in which they are used, their combination with other strategies, frequency of use, and the learners' proficiency level. One of the first attempts at providing a comprehensive overview of language learning strategies can be found in Oxford .R. [7]

E. Language Immersion

Language immersion is another method for improving in the communication of the second language (also called L2, or the target language). This method uses the target language as a teaching tool by surrounding or "immersing" students in the second language. In-class activities, such as math, social studies, and history, and those outside of the class, such as meals or everyday tasks, are conducted in the target language. Today's immersion programs are based on those founded in the 1960s in Canada when middle-income English-speaking parents convinced educators to establish an experimental French immersion program to enable their children 'to appreciate the traditions and culture of French-speaking Canadians as well as English-speaking Canadians. [8], [9] Based on the French immersion program in Canada Allen, Mary [10] reported that the students in the French immersion programs performed significantly better in reading than other students.

F. Learning disability

Learning disability [11] is a general term that describes specific kinds of learning problems. A learning disability can cause a person to have trouble learning and using certain skills. The skills most often affected are: reading, writing, listening, speaking, reasoning, and doing math. Learning Disabilities can be categorized into Articulation Disorder, Attention Deficit Hyperactivity Disorder, Auditory Processing Disorders, Dyscalculia (Developmental Arithmetic Disorder), Dysgraphia (Developmental Reading Disorder), Dyslexia (Developmental Reading Disorder), Dyspraxia (Motor Planning / Sensory Disorder), Expressive Language Disorder, Receptive Language Disorder, and Visual Processing Disorder.

Auditory processing disorder (APD), also known as central auditory processing disorder (CAPD), is a complex problem affecting about 5% of school-aged children. These children can't process the information they hear in the same way as others because their ears and brain don't fully coordinate. Something adversely affects the way the brain recognizes and interprets sounds, most notably the sounds composing of speech. Kids with APD often do not recognize subtle differences between sounds in words, even when the sounds are loud and clear enough to be heard. These kinds of problems typically occur in background noise, which is a natural listening environment. So kids with APD have the basic difficulty of understanding any speech signal presented under less than optimal conditions. There are many possible causes of APD; they include head trauma, lead poisoning, and chronic ear infections. Sometimes the cause is unknown. Hearing disabilities cause five main problems that affect both home and school activities.

First, Auditory Figure-Ground Problems: This is when the child can't pay attention when there's noise in the background. Noisy, low-structured classrooms could be very frustrating.

Second, Auditory Memory Problems: This is when the child has difficulty remembering information such as directions, listing, or studying materials. Third, Auditory Discrimination Problems: This is when the child has difficulty hearing the difference between sounds or words that are similar. This problem can affect following directions, reading, spelling, and writing skills, among others. Fourth, Auditory Attention Problems: This is when the child can't maintain the focus for listening long enough to complete a task or requirement.

Lastly, Auditory Cohesion Problems: This is when higher-level listening tasks are difficult. Auditory cohesion skills — drawing inferences from conversations, understanding riddles, or comprehending verbal math problems — require heightened auditory processing and language levels. They develop best when all the other skills (levels 1 through 4 above) are intact.

In order to make this research comply with fairness and credibility, an investigation was implemented to see whether these students only had a hearing impairment. A list of all the students' educational profile has been completed.

G. Situated Leaning

Situated learning is a general theory of knowledge acquisition. It has been applied in the context of technology-based learning activities for schools that focus on problem-solving skills.

Lave [12] argues that learning, as it normally occurs, is a function of the activity, context and culture in which it occurs (i.e., it is situated). This

contrasts with most classroom learning activities that involve knowledge which is abstract and out of context. Social interaction is a critical component of situated learning -- learners become involved in a "community of practice" which embodies certain beliefs and behaviors to be acquired. As the beginner or newcomer moves from the periphery of this community to its center, they become more active and engaged within the culture and hence assume the role of experts. Furthermore, situated learning is usually unintentional rather than deliberate.

These ideas are what Lave & Wenger (1991) call the process of "legitimate peripheral participation."This case applied situated learning with information technology; this provides students with a better chance of understanding how each vocabulary are used in a real life natural situation based from that particular location. The vocabularies are classified by town's environment to allow sense-making to occur.

H. Learning System for Students with Hearing Impairment

Sriyaa Niyomthum and Praputsorn Niyomthum [13] categorized learning system for students with hearing impairment as followed: First, oral method is the teaching that emphasizes on oral communication. Oral communication includes speech, lip reading and the use of residual hearing. Second, manual communication involves signs and finger spelling to mediate a message between persons. They are received visually and sometimes tactually upon being expressed manually. Manual communication, when it is a primary form of communication, may be enhanced by body language and facial expressions and other forms of communication. Third, total communication Mayer, P. & Lowenbraun, S. [14] is the use of any means of communication - sign language, voice, finger spelling, lip reading, amplification, writing, gesture, visual imagery (pictures). The sign language used in total communication is more closely related to English. In this research the focus is on total communication because this approach is suitable for bundling with animation technique and related with the teaching method for Thai schools for the deaf.

Mertzani, M. [15], [16] presented Computer-mediated communication (CMC) has high influenced in teaching and learning of Second Language Acquisition. This tool effect positively to traditional class and created new language teaching models focuses on learner centered teaching methodology based on constructivist approach. With the effective outcome, this tool is applied in interactive sign language class and other class for students with a hearing loss.

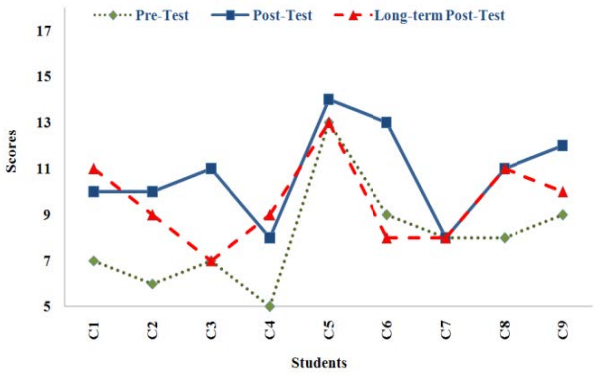
Massaro, D.W., & Light, J [17] and Lecia J. Barker [18] research is based on improving the vocabulary learning with text to speech and speech recognition through computer animated tools called Language Wizard / Player. This tool can encode a speech to word and be written to a speech together to generate visual speech by a speaking avatar, an animated talking face, nicknamed "Baldi". However, the language player doesn't cover the principles of total communication.

The literatures on the theories that were relevant to memory retention had to be reviewed intensively as the author of this work wanted to gain some basic understanding with the topic. The theories dealing with cognitive learning, memory retention, cone of learning, language development, vocabulary language acquisition, language immersion, learning disability, and the learning systems for hearing impairment have all provided a compass for obtaining a better idea and approach to lay out a proper framework.

Before the research team came up with an idea, an assessment was created and implemented to make sure that creating a technological learning tool would be the most effective solution for helping these students with a hearing impairment. The research methodology was as followed.

II. Methodology and Framework

In the first stage, samples of 18 primary students were taken from grade 5 and were categorized as experimental samples. Both groups participated for a short term memory and long term memory test. The first group used total communication with a dictionary composed of animated characters (TCAD method). The second group was learning in the traditional method: classroom instruction, translation of verbal language to sign language, and finger spelling. The results of the short-term memory test in the TCAD class pre-test and post-test and the traditional method demonstrated that both significant students showed an improvement in their English vocabulary. After 9 months, the results of the long-term memory test in the TCAD class post-test demonstrated that students showed an improvement in their English vocabulary score more than the short-term memory post test. The traditional method long-term post-test indicated that students showed an improvement in English vocabulary less than TCAD method.



method

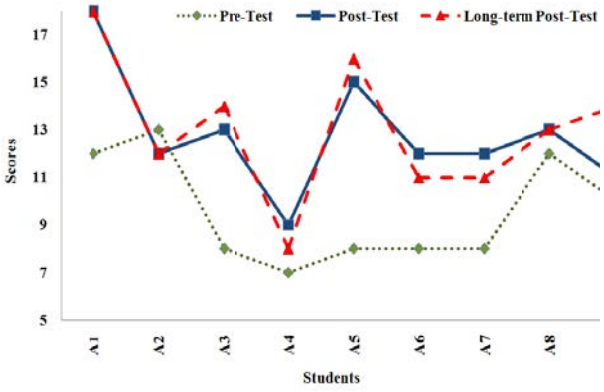


Figure4. Pre-Test and Post-Test results graph of TCAD method

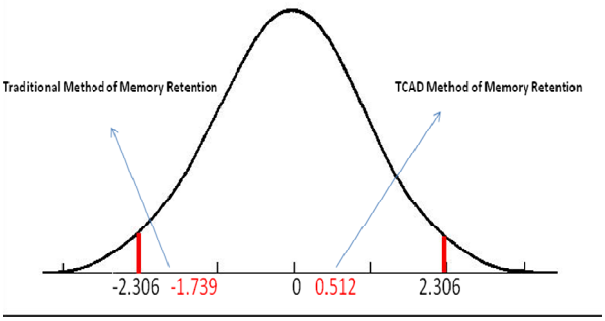


Figure5. Pre-Test and Post-Test paired t-test results of TCAD method

From the results, the graph showed that there was a significant positive English vocabulary memory retention in the TCAD method by using a paired t-test in N<30; DF=8; α = 0.25 (2-tailed); t (from t table) =2.306 and t result between TCAD long-term post test after nine months of teaching. The TCAD short-term posted a test score of 0.512. This proves that the TCAD method supported the English vocabulary memory retention in the long-term memory

Upon proving that the TCAD was valid, the findings in the study were implemented to the entire primary students at the Chiang Mai School for the Deaf. The process revealed that there was an increase in the number of vocabulary among these students. Based on the study, the everyday usage of the TCAD for twenty minutes in the classroom showed a positive result when it is compared between the pretest and posttest. The graphs of figure 6 and figure 7 depict how the students gradually had an improvement of vocabulary.

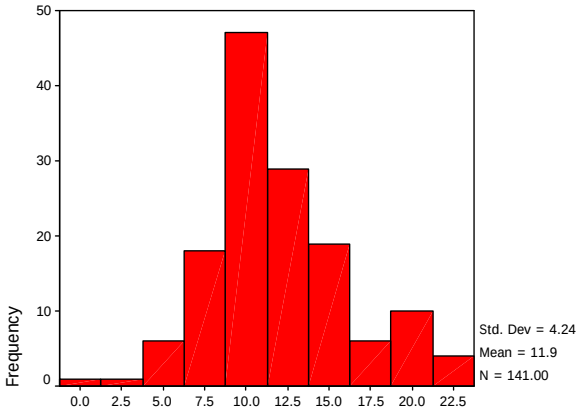


Figure6. Pre-Test histogram results before applying the TCAD method

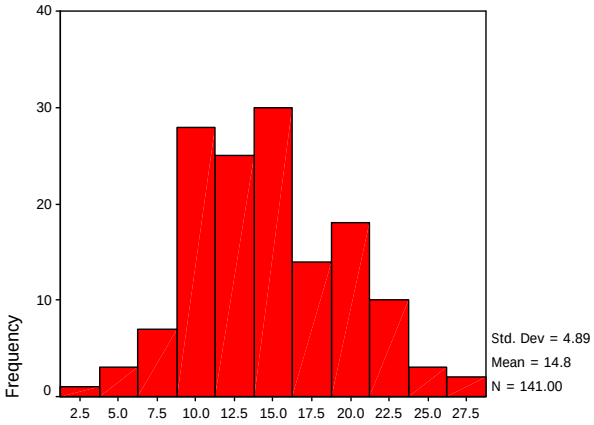


Figure7. Post-Test histogram results of using the TCAD method

Here we see that in figure 6 the pre-test the students would normally obtain an average score of 11.9. These students would take a multiple choice test consisting of English vocabulary words; of which 36 vocabularies were selected from 400. 141 primary students took part in the examination. In comparison to figure 7 these students scored a higher average all together in the post-test session. Thus the TCAD method enabled these primary students to obtain a 8.33% increase overall in the exam in one month.

IV. TOOLS

The TCAD concept was created for helping these students to develop their memory retention. This tool was utilized in the English immersion program. It allowed for activities to be based on encouraging the students to learn a variety of English language vocabularies. This allowed for the learning environment to have an English immersion

experience. The animation characters that created from motion capture technique provided a lively, stimulating atmosphere by applying a multi-approach teaching tactic to give these students an absolute skill development for communication. The teaching methods consisted of sign language, finger spelling, auditory, lip reading, picture captioning, situation learning, reading, and writing. An example is shown below in the figure 8.

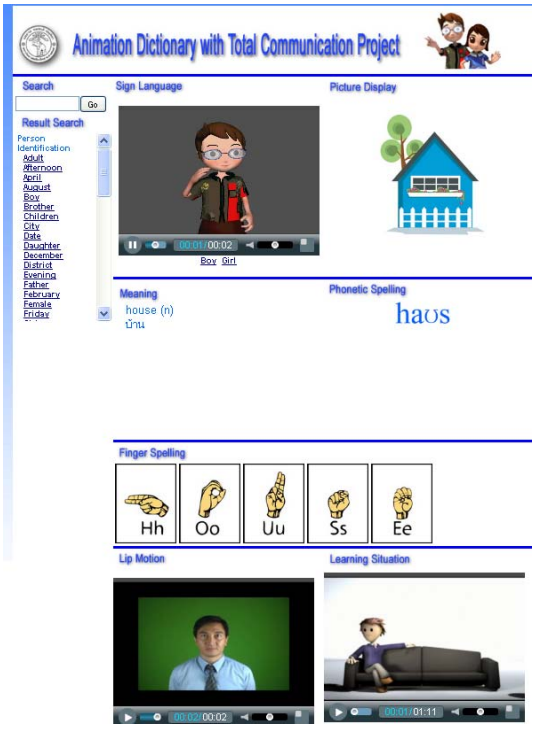


Figure 8. TCAD Screen

Figure 8 demonstrates the total communications with animation dictionary screen composing of eight ways of communicating. The methods are sign language animation, picture caption, meaning, phonetic, finger spelling, lip reading and situated learning. From the user review in the first phase of study the TCAD program extends to bundle the new situated learning part with a situated sign language application for classifying the vocabulary by placing or calling in the word place which is shown below.

Figure 9. Situated Sign Language Main Screen (vocabulary classify by town or location)

Figure 10. Situated Sign Language Screen (Group of vocabulary in "school")

Figure 11. Situated Sign Language Screen (Group of vocabulary in "farm")

Figure 9 - 11 shows the situated sign language screens which is composed of towns or places for classifying the vocabulary group. This screen allows the student to recognize the vocabularies with an easy word group and the application is an interactive tool. The student can click on the picture for leaning the vocabulary that translates into a sign language; both Thai and English words are shown on the top and bottom of the frame.

V. DISCUSSION

The TCAD with a Situated Sign Language can be described as a tool that integrates new material with previously presented information. This tool compares and cross-references new and old ideas of the hearing impaired to produce a convenient way of learning English vocabularies. Before the implementation of the TCAD, we found that the learning tools of the hearing impaired lacked media that would assert itself as a supporting device for English vocabulary learning. This has led to our action in developing the TCAD. However, the TCAD is only composed of eight ways of communicating; they are sign language animation, picture caption, meaning, phonetic, finger spelling, lip reading and situated learning. At times, the students encounter difficulties in comprehending on the lesson being taught. From there on, we continued our work on developing the learning tool that would support students learning by creating a tool for classifying vocabularies that would generate an effortless task in remembering the situated sign language that separates the vocabulary by its related town or location. This tool allowed student to enjoy 20

minutes of learning English vocabulary when they are in the classroom. Test scores also indicate positive results in vocabulary retention.

VI. Conclusion

This research showed the improvement of vocabulary cognition by using TCAD plus situated sign language program. The vocabulary memory of students being taught through this method produced a much higher result. This is due to the ability of the hearing impaired students in perceiving properly through visualization. Therefore the research has been taken as a significant point for designing 3D animation character in vocabularies teaching. The 3D cartoon animation matched students' interest. With the advantages of animation encouraging learners to learn in the active way that is used for creating an intellectual environment, learning through their interest will decrease tension and increase attention. The approach for using animation in improving vocabulary cognition motivates students to learn more than a text based learning style. There are factors such as emotional and intelligence that affect their method of learning. Overall, animation vocabularies immersed in this type of environment will decrease the stressful emotional factors while enabling these students to learn in a more conducive manner.

VII. Acknowledgement

The authors would like to thank Dr. Pradorn Sureephong for fine advisement, Mr. Ravee Phoewhawm and Mr. Pachara Tinamas for technical support, the Graduate School at Chiang Mai University (CMU), Anusarnsunthorm School for Deaf and Mae Fah Luang University. This research was supported by grant from under the program Lecturer Development Scholarship from the Office of the Higher Education Commission, Thailand.

References

[1] Schuman, L. (1996). Perspectives on instruction: Behaviorism, Cognitivism, and Constructivism. Online]. Available:http://edweb.sdsu.edu/courses/edtec540/Perspectives/ Perspectives.html

[2] Sharon J. Derry, "Learning Strategies for Acquiring Useful Knowledge. In Dimensions of thinking and cognitive instruction, ed. Beau Jones & Lorna Idol (Hillsdale, NJ: Erlbaum, 1990), p. 347 – 379

[3] Learning Strategies Development, Queen's University © 2008

[4] Edgar Dale, Audio-Visual Methods in Teaching (3rd Edition). Holt, Rinehart, and Winston (1969).

[5] Edgar Dale (1969), Audio-Visual Methods in Technology

[6] Schmitt, N. and McCarthy, M., editors (1997). Vocabulary: Description, Acquisition and Pedagogy. Cambridge University Press

[7] Oxford, R.(1990). Language learning strategies: What Every Teacher Should Know. Newbery House, Boston.

[8] Julie Sweitzer (2001). Why Immersion? [On-line]. Available:http://www.carla.umn.edu/immersion/acie/vol4/Feb2001_WhyImmers.html

[9] Baker, C. (1993). Foundations of Bilingual Education and Bilingualism. Clevedon: Multilingual

[10] Allen, Mary (2004). "Reading achievement of students in French immersion programs". Educational Quarterly Review, Volume 9, number 4, pages 25-30.Catalogue 81-003-XIE

[11] Learning disability [On-line]. Available: http://www.learningrx.com/learning-disablities.html Accessed May 23, 2008.

[12] Lave, J., & Wenger, E. (1990). Situated Learning: Legitimate Periperal Participation. Cambridge, UK: Cambridge University Press.

[13] Sriyaa Niyomthum and Praputsorn Niyomthum. (1982). Tutoring. Bangkok: Odianstore Publisher.

[14] Mayer, P. & Lowenbraun, S. (1990). Total communication use among elementary teachers of hearing-impaired children. American Annals of the Deaf, 135, 257-263.

[15] Mertzani, M. (2006). Sign Language Learning through Asynchronous Computer Mediated Communication. In C. Vettori (Ed.), Proceedings of the Second Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios, Genova, Italy. May 28 2006. Genova: LREC – European Language Resources Association (ELRA). 64 –69.

[16] Mertzani, M. (2006).Considering Sign Language and Deaf Culture in Computer Mediated Communication (CMC) Environments: Initial Explorations and Concerns. 9th. Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil

[17] Massaro, D.W., & Light, J (2003). Improving the Vocabulary of Children with Hearing Loss Perceptual Science Laboratory, Department of Psychology University of California, Santa Cruz Santa Cruz, CA. 95064 U.S.A.

[18] Lecia J. Barker (2003). Computer-Assisted Vocabulary Acquisition: The CSLU Vocabulary Tutor in Oral-Deaf Education. The Journal of Deaf Studies and Deaf Education p. 187-198 Matters.

[19] Bybee, R.W. & Sund, R.B. (1982). Piaget for Educators (2nd Ed). Columbus, OH: Charles Merrill

A Cooperative Education Learning Process for Improving Students' Competencies in Management and Information Technology

Worawit Janchai[1], Laurent Veillard[2] and Nopasit Chakpitak[1]
[1]College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
[2]Institut Universitaire de Technologie Lumìre, University Lumìre, Lyon2, France

***Abstract*— In this era, information technology (IT) for industrial work has an important role in industry. The people who work in this area should possess both knowledge in IT and that industrial field. The essential thing for developing this cohort is to have the necessary competencies in accordance to the industrial demand; the appropriate learning process and support to obtain real working experience. This paper illustrates the learning process used in the cooperative education (co-op) program in the management and information technology field for improving students' competencies. The co-op program adopted the prototype model from France and Germany under the collaboration project between Thailand and European called "ETHICS-FED". The co-op learning process emphasizes on behavior, work performance and the professional project of the students. The result revealed that the process contributed to the enhancement of students' competencies in behavioral and working performance. In addition, the project became an impressive complement to the education program. However, this research also showed some difficulties from the side of the students, company and university. There are some analysis and recommendation to resolve these difficulties.**

Index Terms*—Cooperative Education; Learning; Competency; Management; Information Technology

I. INTRODUCTION

MMIT or Modern Management and Information Technology is the curriculum at College of Arts, Media and Technology, Chiang Mai University, Thailand. MMIT produces graduate students to be a middle level manager or supervisor in the industrial field. This curriculum teaches students to use information technology knowledge integrated with industrial knowledge for conducting and improving industrial work in many departments such as analyzing the causes of rework, applying enterprise resources planning program in the production department or establishing employee database, and building multimedia program for training in the human resources function.

Nowadays, globalization is making the Thai manufacturing industries facing the challenges of developing itself to be more knowledge base emphasizing on personal development for confronting new technology and innovation [1]. Another reason is that many multinational companies have considered on relocating their factories into new developing countries. These countries have an available supply of cheap labor.

Based on this reasoning, these manufacturing industries have invested in new technology and are applying a strong effort in developing their human resources for leveraging the manufacturing process. New equipments and machines were brought into Thai industry. At the same time, most of the manufacturing companies have been recruiting applicants with an education beyond the tertiary level than in the past [2].

For the middle level manager, the industry needs for them to possess an undergraduate degree from a recognized college or university. This is for the purpose of managing employees who have formal education higher than in the past (raise from primary school to high school diploma or a vocational degree) [2] and coordinating work between the levels of management and operation which requires in-depth skill and knowledge. However, this group particularly lacks working experience and job proficiency.

The problem comes from the educational system in Thailand. The system still emphasizes too much on memorization and content. The education system does not have materials that are relevant to the real world of work as well as meeting labor demand and the country's development policy. The lack of initiative produces students with insufficient skills for creativity, self-learning, problem solving and critical thinking [3].

Moreover, for the employers, the major problem is the issue of recent graduates who are lacking sufficient competencies in the cognitive, behavioral and technical domain to carry out their work [4]. The most serious issue is the behavioral competencies which are considered an essential to function in a manufacturing working environment. The cause is related to the students' inner-self which includes personal mastery (vision) and their mental model [5].

Therefore, the MMIT curriculum has designed the co-op learning process by having students learn

ISBN 978- 974-672-556-9

about theories and applying them into their workplace. The co-op program has been in trend for the last two decades [6]. Its advantage is in contributing to the students' skill and competencies development. The MMIT co-op model was established through the collaboration and technical knowledge transferred from Institut Universitaire de Technology Lumière (IUT Lumière), University Lumière, Lyon 2, France which is the major and Duale Hochschule Baden – Wurttemberg (DH-BW) Vellingen – Schwenningen, Germany as a guideline under the Euro THai Implementation of Cooperative Study - For Economic Development (ETHICS – FED) project. In this model, some activities based on the learning organization theory are added into the learning process for enhancing students' competencies.

II. LITERATURE REVIEW

A. The competencies

McClelland [7] defines competence as "basic personal characteristics that are determining factors for acting successfully in a job or a situation". Boyatzis [8] gave the meaning of competency as "an underlying characteristic of a person which results in effective and/or superior performance in a job". In similarity from Boyatzis, Spencer and Spencer [9] describes competency as "an underlying characteristic of an individual that is causally related to criterion-referenced effective and/or superior performance in a job or situation".

There are many competencies models and each one has different elements. Spencer and Spencer [9] separated competencies into five components; motives, traits, self-concept, knowledge and skill. Motives are "the things a person consistently thinks about or wants that cause action". Traits are the "physical characteristics and consistent responses to situations or information". Self-concept is "a person's attitudes, values, or self-image". Knowledge is "information a person has in specific content areas. Skill is "the ability to perform a certain physical or mental task". Mournier [10] proposed a model that employers use to consider the skills of new employees. He defines skills as "individual and collective abilities to produce" and separated it into cognitive, technical and behavioral skills. The cognitive skill is formed in the educational system. On the other hand, technical and behavioral skills are formed in the workplace. Cognitive skills are "individual abilities to understand, to reason, to give meanings and the capacity to learn more by oneself based on individual knowledge in particular field of scholarship". Technical and behavioural skills are "individual and collective capacities to operate a technology (productive process) and to behave (organisation) in a given work process". Some researcher [11] present competency as individual attribute affected on task performance; it consists of cognitive skill such as technical knowledge, skill and attribute and behavior skill such as principles, attitude, value and motive.

B. Process for learning and developing the competencies

1) The Cooperative education (co-op): Law [12] describes the co-op as "The term cooperative education is generally associated with a school and work program in which student-learners receive supervised payroll experience as part of the school curriculum." The National commission for cooperative education describes it as "a structured educational strategy integrating classroom studies with learning through productive work experiences in a field related to a student's academic or career goals. It provides progressive experiences in integrating theory and practice. Co-op is a partnership among students, educational institutions and employers, with specified responsibilities for each party" [13]. Many researchers agreed that co-op can raise the quality of education through a combination between theory and practice [14]. It allows the student to acquire necessary practical skills through the real working world [15]. For this reason, co-op is an alternative way for many institutions to allow their students to develop their potential and enhance their skill in linking theory and practice. Some educational institutions that use co-op in their curriculum believe that co-op can help students to enhance their attitudes and skills and adapt themselves into the workplace [16]. Institution such as Deakin University, Australia [17] designed a curriculum in the field of information technology by positioning the co-op program as a central element in the education of students.

2) The Learning organization: Main theory for defining activities in the co-op framework is "Learning in Action" [18] because it has many concrete activities contributing to organizational learning. There are four main components;

- Intelligent Learning consists of searching, inquiring and observing.
- Experiential Learning is about reflecting and reviewing the past experience (After Action Review: AAR). This approach discusses the problems, finds the solutions, and recommends topics that are of concerned.
- Experimental learning is about exploring. The outcome is unpredictable. There is a test for the hypothesis which makes the outcome predictable.
- Leading Learning consists of establishing a search team. This is about setting a review board and building a pilot project team.

This research also emphasized on the learning process development by focusing on mastery and mentality of students based on Senge's theory [19], the fifth discipline. There are five disciplines in this concept; Systems Thinking, Personal Mastery,

Mental Model, Shared Vision, and Team Learning. The critical gear is to have students contribute to systemic thinking (understand the cause and effect) and awareness about their future for support in acquiring mastery and a robust mentality.

3) The concept of Personal and Professional Project: In IUT Lumière students will learn academic courses and must attend in a preparation program for cooperative education [20] called **"**Personal and Professional project (PPP)". PPP concept comes from Quebec, Canada. It has four major steps; Exploration: is the discover process, Crystallization: is the consideration or synthesize process, Specification: is the process to make a decision (choose the alternative), and Realization is the implementation or experimental process. The essential aims of PPP are to develop student's sense of responsibilities and intellectual autonomy, to develop their reflexive skills about themselves and their life experiences, to enable students to find information by themselves and to develop their critical views about this information and prepare students to succeed in the school-to-work transition.

III. MMIT LEARNING AND ASSESSMENT PROCESS

MMIT curriculum combines the classroom learning environment with a co-op apprenticeship which transfers knowledge and technology from European country; particularly France and German under the ETHICS-FED project.

MMIT is a four years learning program. The students in the first year learn in the classroom and attend the co-op preparation in the second year. By the third year they start the co-op apprenticeship; this lasts for two semesters of the third year. Within this period, students are still required to attend some classroom lectures. The table below shows the details.

TABLE I. MMIT LEARNING PLAN

Year 1		Year 2		Year 3		Year 4	
Term 1	Term 2	Term 1	Term 2	Term 1	Term 2	Term 1	Term 2
Class room	Class room	Class room	Class room	Class room	Class room	Class room	Class room
-	-	Co-op preparation			Coop I	Coop II	Coop III
-	-	-			Co-op assessment (8 times)		

The MMIT learning plan has two major co-op processes; the co-op learning process and the co-op assessment process.

A. The co-op learning process

This process consists of a co-op preparation period and co-op apprenticeship period (co-op I, II, III)

1) *Co-op preparation period* is the learning process that students go through before entering the co-op apprenticeship. The process was established by adopting the PPP of IUT Lumière combined with many learning activities include observing, inquiring and observation [18]. These activities allow students to participate in university career day to discover their future job. Also, industrial experts offer basic industrial training to acquire quality control knowledge, problem solving by statistical tools, and managing human resources. In addition, company site visits are arranged to observe the real life work. Furthermore, students are matched with companies that they've applied with and take part in the orientation before starting work with that company.

2) *Co-op apprenticeship period (co-op I, II, III)*: students will do a 14 months apprenticeship that is separated into three learning period shown in TABLE II.

TABLE II. CO-OP APPRENTICESHIP LEARNING PERIODAND OBJECTIVE

Learning Period	1st period (month1 - 6)	2nd period (month 7 – 10)	3rd period (month 11 – 14)
Learning objective	Students adapt themselves in work environment, Understand work and responsibilities, Finding the professional project topic	Students have an autonomous in their work, Plan, allocate the resources, and begin the professional project	Students can adapt knowledge in their work, Evaluate, analysis and improve their work, Finish their professional project

The essential part in MMIT co-op learning process is the "*After Action Review (AAR)*" [18]. After students passed their activities or work in company they must "*reflect*" and share their experiences by reviewing the situation on some guideline questions such as

- What did you learn and received from the workplace?
- What problems and difficulties have you encountered?
- What have you already solved and how did you solve it?
- What are the problems which are not resolved and why?
- What will you plan to do next?

MMIT use this activity through the co-op learning period.

The professional project also plays an important part in the learning process. The advantage is scoping the students to have a learning framework in their career path (management and IT in industrial). This project is aimed for improving students' skill and prepares them to be a proficient "modern supervisor" [21] upon graduation. The characteristics of the professional project is demonstrating how IT can be utilized for supporting industrial work such as the use

of an application program for solving rework problems. Other examples of the IT project can be on replacing some HR training programs through multimedia, or producing some databases for support production work.

Figure 1. The students reflected their work/activities.

B. Co-op assessment process

The co-op assessment process is separated into two major assessments; apprenticeship assessment and professional project assessment

1) Apprenticeship assessment adapted from IUT Lumière. The MMIT program is interested in following the behavior and the work performance of the students. There are 7 competencies in each topic shown in TABLE III.

In the apprenticeship assessment meeting, the university advisor is a facilitator that conducts the assessment process on every competency. The student will explain their work and evaluate their own competencies. Then, from the company's mentor point of view, they will inform students to know the level of their behavioral and work performance. To conclude the assessment, three partners will discuss and find the appropriate way to improve the students' competencies and plan for next period.

2) Professional project assessment: the project is assessed by the university advisor and company mentor. There are three periods to assess the project. First, the end of co-op I, considers the feasibility and advantage of the project. Second, the end of co-op II, focuses on the project plan, resources, and project progress. Third, the end of co-op III, results are collected, analyzed and evaluated.

TABLE III. THE COMPETENCIES FOR MMIT STUDENTS APPRENTICESHIP ASSESSMENT

No.	Behavioral perspective	Work performance perspective
1	Enthusiasm	Learning & Understanding
2	Responsibilities	Planning & Managing
3	Adaptability	Knowledge applying
4	Manner	Creativity
5	Communicate	Problem solving
6	Leadership	Follow work instruction
7	Patience	Handle emergent problems

In the apprenticeship assessment meeting, the university advisor is a facilitator that conducts the assessment process on every competency. The student will explain their work and evaluate their own competencies. Then, from the company's mentor point of view, they will inform students to know the level of their behavioral and work performance. To conclude the assessment, three partners will discuss and find the appropriate way to improve the students' competencies and plan for next period.

3) *Professional project assessment*: the project is assessed by the university advisor and company mentor. There are three periods to assess the project. First, the end of co-op I, considers the feasibility and advantage of the project. Second, the end of co-op II focuses on the project plan, resources, and project progress. Third, the end of co-op III, results are collected, analyzed and evaluated.

IV. METHODOLOGY

In this research, the methodology used is the student's individual assessment for testing the teaching model. According to ratings by others method, the assessment was conducted among three parties - the students, university advisor and company mentor. The university advisor assumes the facilitator role when the students are being assessed by the company's mentor. The research collected assessment data from the first co-op students; 12 students from 16 students through 14 months of co-op apprenticeship (November, 2008 – December, 2009). The data was collected in a quantitative and qualitative method by using the student's assessment book as a main tool to follow up on their behavioral and work competencies. For additional data, the researcher interviewed some students and company mentors about the factors impacting the students' work and competencies development.

This research used descriptive statistic to analyze the level of students' competencies which were combined with analyzing the information form interviewed. In addition, the researcher created a "Why-Why analysis diagram" for analyzing the cause and effect of the problem that had occurred in the MMIT co-op learning process. The analysis result will be the support information to improve and develop the co-op learning process in further work.

V. RESULTS AND ANALYSIS

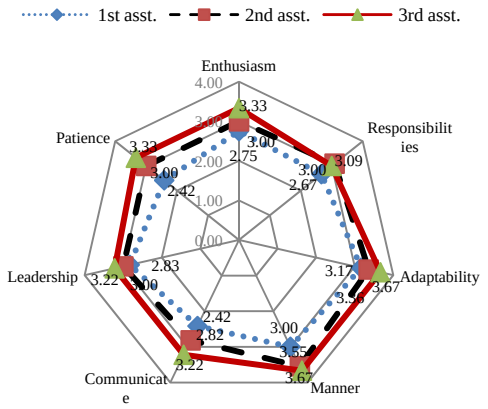
The results showed that the students gradually made an improvement in competencies regarding to their behavioral and work performance. The results illustrate as followed:

A. The evolution of behavior

The standard level for assessing student's competencies are 1 means unsatisfied, 2 means improvement, 3 means good and 4 means excellence.

The main assessor is the company mentor and the university advisor facilitated the assessment.

In figure 2, the average behavioral competencies levels of 12 students tend to have continuous improvement in each assessment. The competencies with a high level since the beginning are adaptability and manner. These students proved that they can adjust themselves to the organizational culture, work as team, and demonstrate respect with other staff.



Behavior perspective				
No.	Competencies	1st asst. n = 12	2nd asst. n = 11	\3rd asst. n = 9
1	Enthusiasm	2.75	3.00	3.33
2	Responsibilities	2.67	3.09	3.00
3	Adaptability	3.17	3.36	3.67
4	Manner	3.00	3.55	3.67
5	Communicate	2.42	2.82	3.22
6	Leadership	2.83	3.00	3.22
7	Patience	2.42	3.00	3.33

Figure 2. The average behavioral competencies level of 12 students

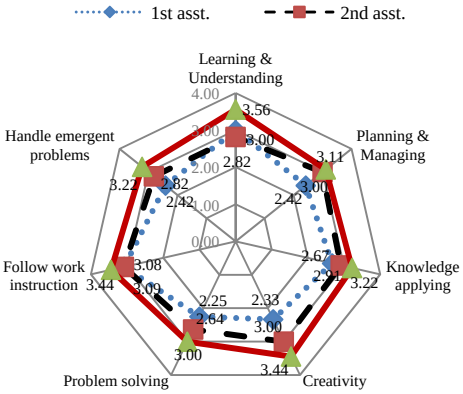
The assessed of leadership, enthusiasm and responsibility competencies improved at first but with a small increase. This could be due to the issue of a lack of leadership. Students feared asking questions or presenting their opinion because they felt unfamiliar with the working environment and with the group they had to work with on the first day. Some students were enthusiastic because of their willingness to do interesting work. Some felt bored with work that didn't provide much of an interest. In the transformation period, where students became employees, some students didn't make an effort to be punctual and handed in an assignment that was considered too late.

At the first time, competencies in the area of communication and patience tend to be the lowest. In terms of communication, the students felt that they were just thrown into the situation of an industrial work and didn't know the proper/technical vocabulary to communicate with their staff. For patience, some students were not mature enough maturity. As a result, they became stressful on the job and felt bored working on a routine job as time goes by.

However, the behavioral competencies improved quite a lot. After the first assessment the students tried to adjust to their behavior according to the company mentor's comment. As a result, some students gradually improved their behavioral competencies. Nevertheless, some of them failed to adapt themselves and we will discuss about this in the analysis part (*D. Analysis the result*) later.

B. The work performance progress

Figure 3 shows the average work performance competencies trend to increase each time that assess 12 students. The topic of following work instruction seems to be the best competency because the students made an attempt and paid attention very much on their assignments. The competencies of learning & understanding received an assess level of 3.00 at the first time but then decreased to an improvement level of 2.82 for the second assessment. However, it rebounded to 3.56 on the 3rd assessment. This can be explained by the students being excited to learn in a new environment when they first arrived but decrease



Work performance perspective				
No	Competencies	1st asst. n = 12	2nd asst. n = 11	3rd asst. n = 9
1	Learning & Understanding	3.00	2.82	3.56
2	Planning & Managing	2.42	3.00	3.11
3	Knowledge applying	2.67	2.91	3.22
4	Creativity	2.33	3.00	3.44
5	Problem solving	2.25	2.64	3.00
6	Follow work instruction	3.08	3.09	3.44
7	Handle emergent problems	2.42	2.82	3.22

Figure 3. The average work performance competencies level of 12 students

when time passes by and also having some knowledge gaps when they work in depth detail. Moreover, they feared asking mentors too much and kept working with a vague understanding of the task at hand. Nevertheless, they improved themselves by making an attempt to find more information to understand the work and asking precise questions to clarify their job.

The knowledge applying, planning & managing, and handling emergent problems competencies are difficult for most of students. They need a lot of skills and work experience to do these kinds of job. On the other hand, some students didn't have a chance to face complex problems and do a creative job. These groups of competencies are difficult to assess and need more time to develop.

C. The Professional project result

The professional project assessment is done by the university advisor and company mentors. All 12 students can complete their project on schedule (6 – 8 months). The professional project can be the learning framework for the students to use IT for improving their job. The students did many learning activities according to "Learning in Action" [18] theory which includes searching, inquiring, observation and experimenting for implementing their project. All of students' professional projects pass the assessment but there are some problems during implementation such as students using an unclear or too complicated methodology for doing the project, as well as writing a few mistake in the analysis and conclusion part. However, the university advisor and company mentors gave students some advice and recommendation to correct it.

D. Analysis the result

From the result of competencies and project assessment, we can see the success of co-op learning process contributed to the students' enhancement of their competencies. Nevertheless, there are some problems that still exist in the co-op learning process. During this research, some information about impact factors on what made it successful and unsuccessful were collected from some students and some company mentors through an interview. The Why-Why analysis diagram was conducted for analyzing the cause of factors that produced both success and failure as illustrated in figure 4 and figure 5.

The "Why-Why" analysis diagram and competencies assessment data provided many interesting information. It can be concluded into many perspectives as follows.

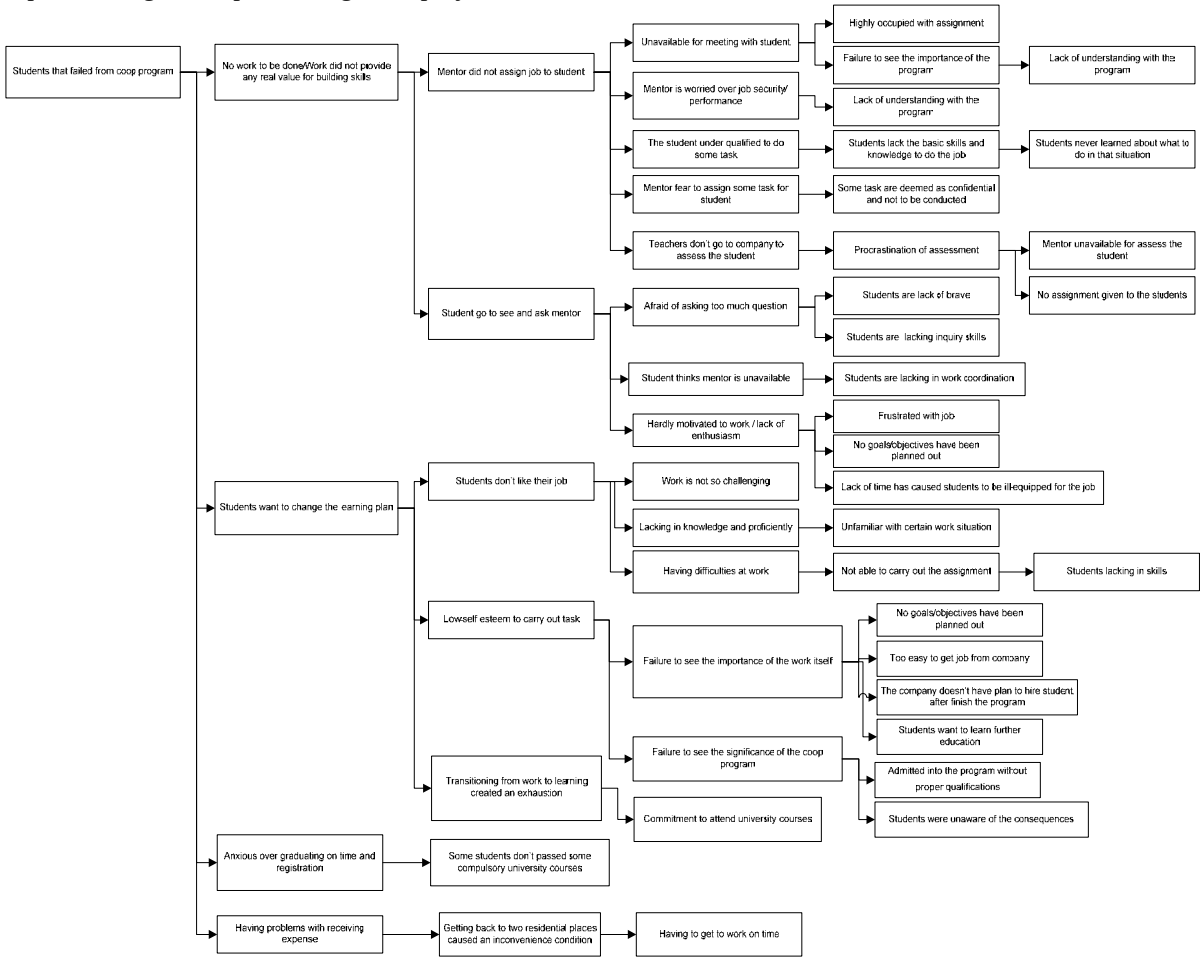


Figure 4. Co-op Why-Why analysis on failure point of view

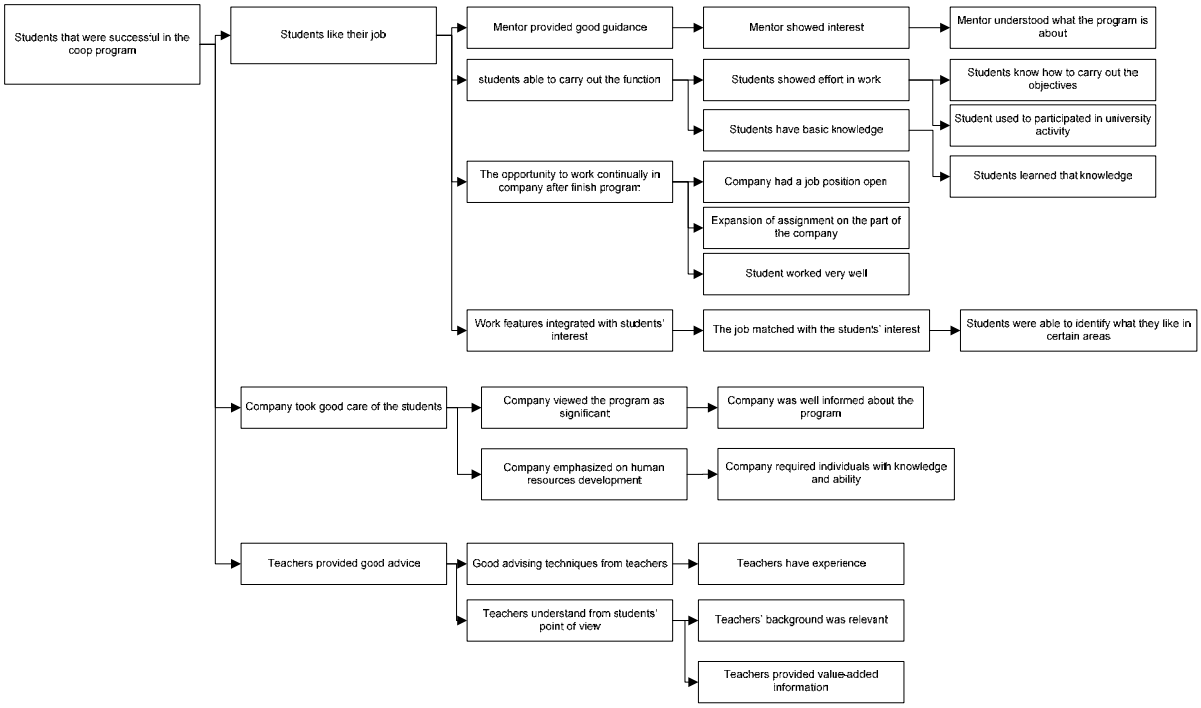


Figure 5. Co-op Why-Why analysis on successful point of view

1) *Student perspective*: The important issue for this perspective is about the students' goal and plan for the future. The students, who can clarify their plan in the future, will know what their skillful job is and how to fit into that position. They can work very well and continue to improve. In contrast, students who don't know exactly what kind of job they prefer, have no plan in the future or think about further education (but haven't made any decision) are mostly confused and aren't interested in landing a good job. Another issue in this perspective is the basic knowledge in the students' field of work. The student can learn and understand fast if they learn about that basic knowledge in the field before. In addition, the students who participated in university activities and also students who do a part time job before the co-op apprenticeship can adapt themselves into workplace quickly.

2) *Company mentor perspective*: the understanding of co-op program of company mentors is the essential issue. In some companies, the mentors think that co-op is the same as normal apprenticeship which have shorter period (3-4 months) and less condense than the co-op program. This provides a negative effect on the students' plan, job details and professional project. The problems occur when the mentors are unclear about the idea of the program.

3) *University advisor perspective*: this part emphasizes on the advisors' experience to advise and follow up students. The students need the academic knowledge and recommendations from their advisor when they face some problems in their work. Hence, advisors who take care of the co-op students should have some related knowledge and experience especially industrial work and IT for assisting students to do their job and complete the professional project.

4) *Company perspective*: the participation and co-op understanding of company's executive level is also very significant for the co-op program. They can allow students to earn a wage and welfare for making students feel as one of the company's employee. Moreover, the co-op program will be complete if the company links the co-op students in their manpower planning. The co-op coordinator in the company is also important in this perspective. This person should make contacts with the university and transfer information into the company. This person should clarify co-op objectives and details as well as having good communication with other related person in their company; for example, mentors who will take care the student.

5) *University perspective*: The most critical point in this perspective is the matching process between students and companies. Each company has a different work culture. The workplace culture can have an effect on the students' enthusiasm, willingness, motivation, and competencies development. The university should be aware of this process by checking carefully on students' attribute and the culture of an organization. The assessment plan is the big problem that should also be revised. The university cannot follow the assessment plan because there is not enough time from the university advisor and company mentor.

6) *Others observance*:

6.1) The students doing the apprenticeship in the same department don't fully pay attention to their job.

6.2) Company mentors who take care of more than one student don't have enough time to advise and make comparison between each student; thus causing an effect on the student's feeling.

6.3) The changing of company mentor or university advisor results on works not being followed up and having to recover lost time with a new advisor or mentor.

VI. CONCLUSION AND FURTHER WORK

From this research, we can conclude that the students develop their competencies both behavioral and work performance through co-op learning process. However, it still has some difficulties that should be improve in further work which includes an enhance related knowledge for students, clarifying co-op concept to relevant personnel in the company; particularly company mentor and also improving university management of the co-op.

As revealed in the analysis part, the students' goal and their future plan are the major factors that have a direct effect on the students' behavioral and work performance. The methodology to develop this part contributes in students having to think systematically and considering things as cause and effect in order to establish them of being aware in their work and their future. In further research, the focus will be on attitude adjustment, systems thinking, and building personal vision that will support them in finding themselves, and planning for their future that will affect their mastery and mentality.

REFERENCES

[1] The office of industrial economic. "The regional industry development plan project phase 2: industrial development plan in northern and middle region area". Ministry of Commerce. 2009. p. 65.

[2] X. Oudin. "Surveys on the Labour Force in Thailand Part I Characteristics of the Labour Force". CELS Working Paper no. 5, 2007.

[3] Waraiporn Sangnapaboworn. "*Higher Education Reform in Thailand: Towards Quality Improvement and University Autonomy*". A Paper Presented at The Shizuoka Forum on "Approaches to Higher Education, Intellectual Creativity, Cultivation of Human Resources seen in Asian Countries". Office of the Education Council. Ministry of Education, Thailand. 12-14 December 2003.

[4] P. Tangchuang, "Thai labour competencies: Lessons learned from Interviewing Employees and Managers", Centre for Education and Labour Studies, CELS Working Papers, No.3, September, 2006, p. 22

[5] W. Janchai, R. Derrouiche, and N. Chakpitak. "A Cooperative Education Learning Framework to Develop Undergraduate Students' Work Competencies". Proceedings of 3rd International Conference on Software, Knowledge, Information Management and Applications. SKIMA 2009. October 21-23 2009 – Fes, Morocco.

[6] A. Gohringer. "University of Cooperative Education – Karlsruhe: The Dual System of Higher Education in Germany". Asia-Pacific Journal of Cooperative Education. 2002. 3(2). 53-58.

[7] G.J. Bergenhenegouwen, H.F.K. ten Horn and E.A.M. Mooijman. "Competence development – a challenge for human resource professionals: core competences of organizations as guidelines for the development of employees". Industrial and Commercial Training. Volume 29. Number 2. 1997. pp. 55–62

[8] D. R. Moore, Mei-I Cheng, and Andrew R.F. Dainty. "Competence, competency and competencies: performance assessment in organizations". Work Study. Volume 51. Number 6. 2002. pp. 314-319.

[9] L. M. Spencer, S. M. Spencer, "Competence at Work: Models for Superior Performance". New York: John Wiley & Sons, Inc. 1993.

[10] A. Mounier, "The Three Logics of Skills in French Literature", NSW Board of Vocational Education and Training, 2001.

[11] R. K. Coll and K. Zegwaard. "Science and Technology Stakeholders' Ranking of Graduate Competencies Part 1: Employer Perspective". Research Report. Asia-Pacific Journal of Cooperative Education, 2002, 3(2), 19-28.

[12] G. F. Law. "COOPERATIVE EDUCATION Handbook for TEACHER-COORDENATORS". 2nd printing. United States of America. 1973.

[13] National commission for Cooperative Education. Available online at http://www.co-op.edu/aboutcoop2.html Last visited March. 1. 2010.

[14] T. Kato "Establishing Cooperative Education as an Integral Part of the Undergraduate Curriculum at Ritsumeikan University" Asia-Pacific Journal of Cooperative Education, 2005, 6(1), 7-12

[15] T. N. Garvan and C. Murphy, "The co-operative education process and organizational socialization: a qualitative study of student perceptions of its effectiveness", Education + Training, Vol.43, No.3, 2001, pp. 281-302.

[16] A. M. Aleisa and M. A. Alabdulahfez " Cooperative Education at the Riyadh College of Technology: Successes and Challenges" Asia-Pacific Journal of Cooperative Education, 2002, 3(2), 1-7

[17] D. Holt "Developing Professional Expertise in the Knowledge Economy: Integrating Industry-Based Learning with the Academic Curriculum in the Field of Information Technology" Asia-Pacific Journal of Cooperative Education, 2004, 5(2), 1-11

[18] D. A. Garvin, "Learning in Action: A Guide to Putting The Learning Organization to Work," Harvard Business School Press, 2000.

[19] P. M. Senge, "The Fifth Discipline: The Art and Practice of the Learning Organization," Random House Business Books, 1990.

[20] Pedagogical Approach presentation by Laurent Veillard, IUT Lumière, University Lumière, Lyon2

[21] N. Chakpitak and O. Tonmukayakul. "Cooperative Education for Industrial Estates in Thailand: A Case study of Lumphun Industrial Estate". Proceedings of 3rd International Conference on Software, Knowledge, Information Management and Applications. SKIMA 2009. October 21-23 2009 – Fes, Morocco.

Continuing Education for New Generation of Higher Military Education

Wg.Cdr. Montatip Vonkdarunee[1], Tirapot Chandarasupsang[2] and Nopasit Chakpitak[2]
[1]Senior Air Officer College, Directorate of Education and Training, RTAF,
Bangkok, 10210, Thailand
[2] College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, 50200, Thailand

***Abstract* - Senior Air Office (SAOC) mission is the training of executive and basic staff function which is essential for the division level. The qualified students must be varied and inter-disciplined in terms of ages, experiences, education level. Due to these differences, individual learning may not be effectively. An evaluation of different components of Kolb's learning model and examination of application of how to improve learning process will be used in order to explain the root cause of problems. After the identification of student learning style, the use of negative reinforcement to assign the role of student to be a leader in group will be used as a solution to the problem. In terms of student outcomes' evaluation, Reflective Learning specializes in Concrete Experience will be applied. From this study, the result shows that individual difference can influence learning outcome.**

***Index Terms- Adult Learning, Individual Difference, Kolb Learning Styles, Negative Reinforcement, Reflective Learning*.**

I. INTRODUCTION

Under the commander in chief's reorganization and specific policy in 2010, Royal Thai Air Force (RTAF) focuses themselves on becoming the professional military. This can be accomplished by developing and strengthening the required competencies of the air personnel in the operation of their daily functions. Furthermore, there is also the need to reduce the working processes while maintaining its efficiency. This is where Knowledge Management (KM) and effective learning process, especially in the new RTAF technology, play an important role as stepping stone for the professional development

The Directorate of Education and Training under RTAF is responsible for the identification of the necessary skills and competencies which are most useful to its personnel as well as the learning processes under the professional development framework. Within the RTAF, there are 4 higher education institutes. These are the Squadron Officer School (SOS), the Senior Air Officer College (SAOC), the Air Command and Staff College (ACSC), and the Air War College (AWC).

The main mission of the SAOC is the training of executives and basic staff functions which are essential for the division level. The qualified students are varied in ages, usually from 38 to54. In addition to that, the basic education can be diversified ranging from the undergraduate to postgraduate level. In other words, within the same SAOC class, students are military personnel and can rank from squadron leader to wing commander, participating in peace support operations (or PS's) in various areas of the RTAF. As a result, these variations imply 2 main conditional problems minimizing the students' achievement.

The first problem is the variety of ages and educational background in the field of adult learning. These can be explained by "Knowles's views of adult learning" as follows:

- Adults are motivated to learn from being in situations in which they see a need to learn. Consequently, adult learning settings should begin with topics that address the adult audience's current learning needs.
- Adults are oriented to the broad range of affairs in life, not to narrow subjects. Thus, adult teaching should be multidisciplinary rather than subject-oriented.
- Adults learn from their experience. Therefore, the most productive adult learning comes from the analysis of adult experience.
- Adults have a deep need to be self-directing. Therefore, teaching adults should involve in setting the agenda for their learning.

Individual differences broaden and harden with age. Therefore, adult teaching should make allowance for differences in style, time, place, pace, focus, and method [1].

Academic development is a different category of discipline where those who have the same reason for existence are understood to change the professional practice of educators. Kolb (1984), supported by Alice Y. Kolb and David A. Kolb (2008) , once said that "learning styles are influenced by personality type, educational specialization, career choice, and current job role and tasks"[2].Also added by Tzu-Chien Liu, Yi-Chun Lin, Kinshuk, Maiga Chang (2008), it is believed that the individual difference learning process may have an influence on the students' learning effects [3].

ISBN 978- 974-672-556-9

Secondly, it is believed that each of these positions has some validity although argued by Marton, Fensham and Chaiklin (1994) that only by a process of integration will be truly able to come to an understanding of the process of learning. Some of the disciplines that need to be considered, in terms of their influence on learning, personality differences, perceptual factors and the nature of memory from psychology, environmental influences from the sociological perspective, with the concerns of the educational practitioner [4].

In summary, due to the individual differences, it is difficult to make a process of learning that gets higher results and better performances. In addition, educators have developed a formal interest in those aspects of the affective domain of learning that relate to character development. It can be referred that the students will be truly creative in their learning only when they really see benefits of the education.

II. LEARNING BEHAVIOR AND THEORY

Adult Learning

In the field of adult learning associated with one of the various age problem of RTAF students, Knowles argues that the andragogical principles are quite different from most of our school systems where the model is that the instructor knows best what is to be taught and learned and where students are expected to learn. Clearly, children in elementary schools don't have the experience to set their own learning agendas. However, somewhere before college graduation, they do develop interests and preferences for an andragogical approach.

Knowles's andragogical message is that effective adult teaching begins with where the students are. Adults will learn faster if what they are studying has an immediate effect on their current situation in life. That is not to say that the instructor cannot alter the students' intellectual whereabouts by adding new information to them, only that the adding will be more effective if it builds on the foundation of interests and understanding already in place [1].

Like, the various age of RTAF students, adult learning can be focused on the application of CoP model to support their achievement both now and in the future.

Individual Differences

It is said that the individual differences moderate the way the individual responds to various situations in different aspects of organizational practices.

Traditionally, what is valued in higher education opinion is the generation or creation of new knowledge. Based on the previous argument, it would then seem that to achieve this aim for all students, the most effective approach would be to adopt a constructivist and therefore inclusive philosophy, then allows to complete freedom in approaching any system in order to facilitate the development of the full range of individual differences. [4].

According to Claudia Voelcker-Rehage's reviewed studies, capabilities remain intact, and older adults are can achieve considerable performance gains. The extent to which the learning capability varies with age, however, has to be considered very carefully [5].

SAOC consists of individual differences that are not only referred to as age difference, but it also includes education level, crop, rank differences. These play an important role influencing learning achievement. Therefore, it may slow down the management of learning process.

Reinforcement Learning

Reinforcement processes were emphasized by Skinner, and were seen as primary in the shaping of behavior. A common misconception is that negative reinforcement is synonymous with punishment. This misconception is rather pervasive, and is commonly found in even scholarly accounts of Skinner and his contributions. To be clear, while positive reinforcement is the strengthening of behavior by the application of some event; negative reinforcement is the strengthening of behavior by the removal or avoidance of some aversive event.

Both types of reinforcement strengthen behavior, or increase the probability of a behavior reoccurring; the difference is in whether the reinforcing event is something applied (positive reinforcement) or something removed or avoided (negative reinforcement). Punishment and extinction have the effect of weakening behavior, or decreasing the future probability of a behavior's occurrence, by the application of an aversive stimulus/event (positive punishment or punishment by contingent stimulation), removal of a desirable stimulus (negative punishment or punishment by contingent withdrawal), or the absence of a rewarding stimulus, which causes the behavior to stop (extinction). [6]

All teachers, even those with seniority, face student discipline challenges every year. Teachers need to understand how they may influence student behavior. Students will choose their own behavior. Teachers may utilize this need by making students feel capable or part of the group. Students who misbehave often times do so for attention, power, or revenge. A student who acts out will generally receive the attention they desire. A student who has the attention of the whole class feels empowered with their ability to derail the lesson. Teachers are not able to directly change a student's behavior but they may

influence the behavior by using positive or negative reinforcement (Skinner, 1968). [7]

Reflective Learning

Reflective learning, initially developed in the field of education and medicine for helping professionals, has been taken up by a number of disciplines in tertiary education because of its relevance for development of new knowledge as opposed to a mere transfer of information from the teacher to the student. Student journals are extensively used in tertiary education "contexts as a means of facilitating reflection, deepening personal understanding and stimulating critical thinking" (Bain, Ballantyne Packer& Mills, 1995 in Bain et al, 1999 p.51).

What is Reflective Learning?

Dewey (1933) defined reflective thought as "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and further conclusions to

Most of this theory relates to perceiving reflection as part of a cycle of learning (Kolb, 1984). Reflection in terms of learning "is a generic term for those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to new understandings and appreciations" (Boud, Keogh & Walker, 1985, p. 7). The concept of reflective practice introduced by Donald Schon (1983, 1987) is the cultivation of the capability to reflect in action whilst doing something, and to reflect on action –after it has been done. This pedagogy creates an alternative to the first learn theory, and then put it into practice that forms the basis of traditional education. It utilizes prior knowledge and focuses on constructing, deconstructing and reconstructing knowledge based on new experiences provided by learning environments which emphasize learner's responsibility for learning, and adult teaching and learning methods.

Journal writing is an intentional reflective design strategy that has been used in traditional (face to face) learning environments to facilitate the integration of new dimensions. "Meaning making, according to constructivists, is the goal of learning processes; it requires articulation and reflection on what we know". (Jonassen, Davidson, Collins, Campbell & Haag, 1995, p.11).

Encouraging Reflective Learning: An Online Challenge

This method of learning was effective and overall student feedback on journal writing was favorable. As with any method some students never developed comfort with the strategy. The exercise certainly allowed the lecturer to establish a closer relationship with the students. Many students who posted their RJ's on a weekly basis said that they enjoyed the immediacy of feedback. This was in contrast to other 12 week courses that required only one or two assessments where the feedback on the first assessment would often come after the submission date for the second assessment and the feedback on the second assessment at, or after the end of the course.

Bain et al.'s (1999) 'Five Point Reflection Scale' (Table IV p 60) is very detailed, and was discussed at length with the students before first RJ submission. Each level represents a degree in complexity from 'reporting' at level one to 'reconstructing' at level five (see Table 1).

TABLE 1.FIVE POINT REFLECTION SCALE

Level 1 - Reporting	The student describes, reports or retells with minimum transformation and no added observations or insights.	3/15
Level 2 -Responding	The student uses source data in some way, but with little transformation or conceptualization.	6/15
Level 3 - Relating	The student identifies aspects of the data which have personal meaning or which connect with their prior or current experience. The student gives superficial explanation of the reason why something has happened or identifies something that they need or plan to do, or change.	9/15
Level 4 - Reasoning	The student integrates the data into an appropriate relationship involving a high level of transformation and conceptualization and seeks deep understanding of why something has happened exploring the relationship of theory and practice in some depth.	12/15
Level5 - Reconstructing	The student displays a high level of abstract thinking to generalize and / or apply learning. The student draws original conclusions from their reflections, generalizes from their experience, extracts general principles, formulates a personal theory, or takes a position on an issue. The student extracts and internalizes the personal significance of their learning and / or plans their own further learning on the basis of their reflections.	15/15

(Adapted from Bain, Ballantyne, Packer & Mills, 1999, p. 60)

The findings suggest that the process of reflection may be actively facilitated through interactive journal

writing. That reflective action extends learning in terms of depth and the personal learning process is stimulated through dialogue with oneself and with the instructor over time.

From this paper the author will use five-point reflection scale to encourage and evaluate the SAOC new group [8].

III. KOLB LEARNING AND ITS APPLICATION

The Kolb learning style model was based on Kolb's experiential learning theory. In this model, Kolb defined learning style on a two-dimensional scale based on how a person perceived and processed information. How a person perceived information was classified as concrete experience or abstract conceptualization, and how a person processed information was classified as active experimentation or reflective observation (Simpson & Du, 2004). Accordingly, Kolb (1985) described the process of experiential learning as a four-stage cycle involving four adaptive learning modes: Concrete Experience (CE), Reflective Observation (RO), Abstract Conceptualization (AC), and Active Experimentation (AE). Kolb (1985) also identified four learning style groups based on the four learning modes: Divergers favored CE and RO, Assimilators favored AC and RO, Convergers favored AC and AE, and Accommodators favored CE and AE.

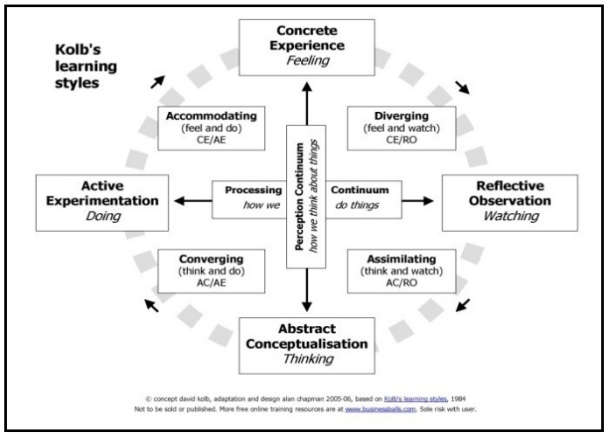


Figure.1. The Kolb Learning style model

"Experiential Learning" (1984): identifies four principal stages. **Convergers** grasp experience through abstract conceptualization and transform their experience through experimentation. The convergers prefer dealing with objects, rather than people, and are often considered unemotional. They are strong in practical application of ideas, can focus on hypo-deductive reasoning on specific problems, and are believed to have narrow interests. **Divergers** grasp experience through concrete modes and transform their experience through reflective observation. They are good at generating ideas, tend to be more "people oriented", and are usually more emotional. They are strong in imaginative ability, good at seeing things from different perspectives, and are believed to have broad cultural interests. **Assimilators** grasp experience through abstract conceptualization and transform their experience through reflective observation. They tend to be less interested in people and more interested in concepts. They are strong at creating theoretical models and excels in inductive reasoning. **Accommodators** grasp experience through concrete experience and transform their experience through active experimentation. They are intuitive and often become impatient when a problem does not conform to their ideas. Their greatest strength is doing things, they are more of a risk taker and they perform well when required to react to immediate circumstances. [9]

SAOC consists of individual differences in which each has their own learning style. If instructors can access each individual learning style and are able to build a learning process to be in accordance with individual learning style, it will improve the learning outcome.

The Relationship of Kolb Learning Styles: Online Learning Behaviors and Learning Outcomes

Hong Lu, Lei Jia, Shu-hong Gong, and Bruce Clark, the researchers study focused on the relationship between Kolb learning styles and the enduring time of online learning behaviors, the relationship between Kolb learning styles and learning outcomes and the relationship between learning outcomes and the enduring time of a variety of different online learning behaviors. The findings of this study were instrumental to instructors and moderators of online courses [10].

First, instructors using online courses should seriously consider the diversity of learning styles when designing and developing online learning modules for different students.

Second, they should provide a large number of electronic documents for students and give enough time to let them absorb knowledge by online reading.

These could be effective methods to improve the quality of online courses.

Kate Henderson, Ksenija Napan, Sylila Monteiro, the researchers' aim of this qualitative study is to examine the application of reflective learning, structured approaches that exist in learning process and may be applied in a program. Furthermore, reflective learning encourages deeper learning and offers a relevant framework for development of professional who will be lifelong learner, committed to improve their practice continually [8].

IV. Case Study

SAOC is used as a case study to confirm that Kolb's learning style and negative reinforcement can be used in learning process for individual differences in order to promote effective learning.

A. General Description of SAOC

One academic year consists of 3 courses, each course contains 96 persons. In terms of student qualification, it varies from 38 to 54 ages, undergraduate to Ph.D., ranking from Squadron leader to wing commander, participating in pace support operations (or PS's) in various areas of the RTAF. Therefore, there is variety in corp.

There are a total of 27 samples where the sample comes from alumni of these courses, 9 persons from each by using learning outcome as criteria. 9 persons are then divided into 3 groups where group A consists of students in the first-three highest score. Group B consists of three medium score and Group C consists of the last-three lowest score.

B. Applying Kolb Learning Style

Kolb learning style assessment was used to evaluate the samples' learning style. The assessment result will be evaluated by Kolb learning style inventory scoring key. A list of database from the learning style inventory scoring key will be plotted in the scoring chart in order to tell each alumni's learning style.

Three pairs of comparison was done as follows: SAOC alumni learning style, old instructional techniques to Kolb's instructional technique.

V. Results and Analysis

The results from alumni learning style were analyzed by Kolb's learning style assessment as shown in fig 1-3.

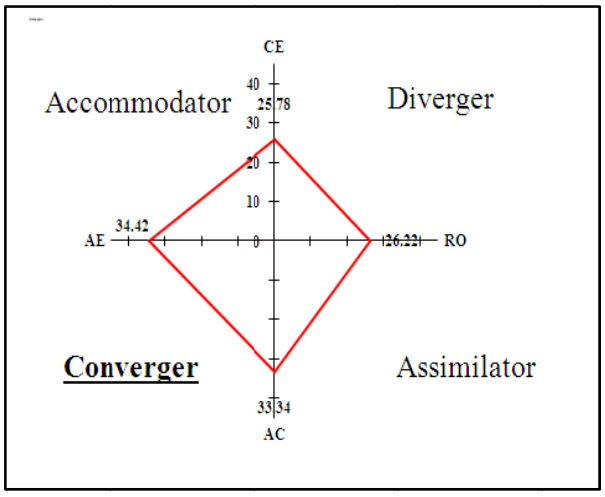


Figure.1. Represented Group A alumni learning style

The data show that Group A alumni scoring chart is a distinct tendency in abstract conceptualization and active experimentation, so they are converger style.

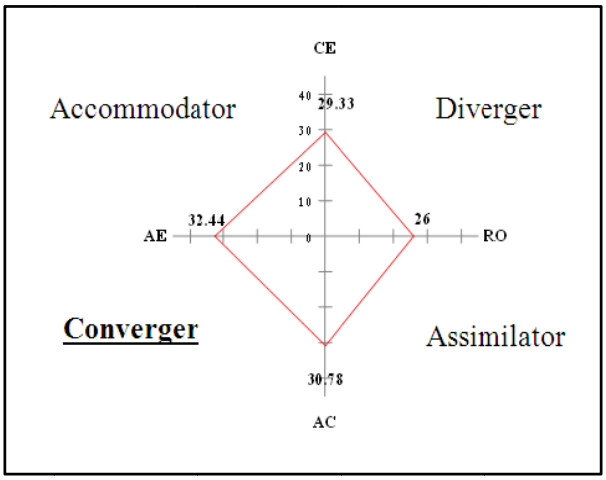


Figure 2. Represented Group B alumni learning style

The data show that although Group B alumni have less higher scores than Group A, the scoring chart is a distinct tendency in abstract conceptualization and active experimentation, so they are converger style too.

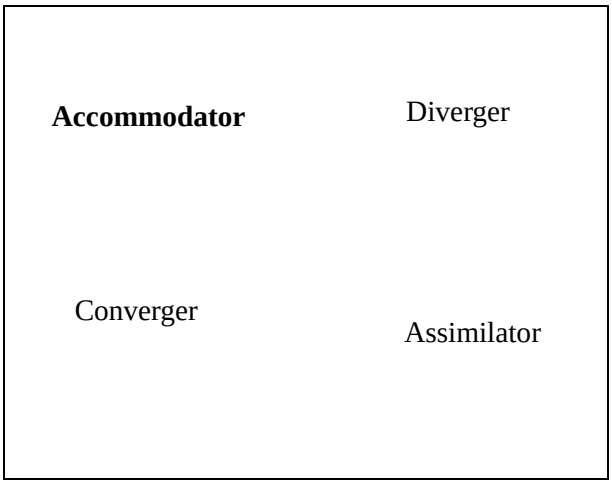


Figure 3. Represented Group C alumni learning style

The data show that Group C alumni scoring chart is a distinct tendency in active experimentation and concrete experience, so they are accommodator style.

When compared the three alumni groups with ages education level and corp. The fact found as in Table 1.

Table 2. Compare the Different Individual of each Group

Gr	CE	RO	AC	AE	Age	Edu	LS	Crop
A	25.7	26.2	33.3	34.8	44.4	M=7, B=2	Conver.	Med.
B	29.3	26.0	30.7	32.4	48.9	M=1, B=3, UG=5	Conver.	Med.& Supp.
C	31.0	26.4	28.7	32.7	51.4	B=5, UG=4	Accom.	Supp.

Table 2 described the process of SAOC alumni learning style as three groups from three courses involving for adaptive learning models; Group A and Group B tended to be Converger while Group C be Accommodator.

Moreover, we found that Group A was the youngest of the three groups, and Group B was younger than Group C

Besides the different of age, when compared to their education we also found that:

Group A : most of them were master degree, Group B : 1 was master degree , 3 are bachelor degree and 3 were under graduation and Group C : 5 were bachelor degree and 4 were undergraduates.

While comparing SAOC alumni learning style, old instructional techniques to Kolb's instructional techniques. The results were in the Table 3.

TABLE 3. COMPARE SAOC ALUMNI LEARNING STYLE, OLD INSTRUCTIONAL TECHNIQUES TO KOLB'S INSTRUCTION TECHNIQUES

Gr.	Learning style	Kolb's Instructional techniques	Old instructional techniques
A B	1. Strong in practical application of ideas 2. Hypo-deductive reasoning on specific problems 3. Unemotional 4. Narrow interests	1. Experiment with new ideas 2. Simulations 3. Laboratory assignments 4. Practical application 5. How question	1. Lecture 2. Gr. Seminar 3. Report 4. Presentation 5. Game simulation
C	1. Greatest strength is doing things 2. More of a risk taker 3. Performs well when required to react to immediate circumstances 4. Solves problems intuitively	1. Work with others to get assignments 2. What would happen if I did this?	1. Lecture 2. Gr. Seminar 3. Report 4. Presentation 5. Game simulation

From the Table 3, it can be explained that group A, B has converger learning style. When compared, Kolb's instructional techniques and SAOC instructional technique, there are some aspects that they are different such as convergers have to be taught by laboratory assignments where old SAOC lacks of.

Learning style of group c is accommodator where this group has greatest strength in doing things, more of a risk taker, performs well when required to react to immediate circumstances, and solves problems intuitively. Students in this learning style should be taught by working with others to get assignments which is totally not related to SAOC instructional technique.

VI. CONCLUSION

This paper applied Kolb's learning style to alumni assessment. It found that alumni who have high learning outcomes tend to be young-age, less experience, high-level education tend to have converger learning style. On the other hand, alumni who have low learning outcomes are likely to be old-age, more experience, lower education level and tend to have accommodator learning style. The result show that SAOC instructional techniques are not related to alumni learning style by Kolb's assessment.

The ongoing working the new instructional techniques taken Kolb's learning styles to represent learning differences, and processed by Negative reinforcement. In the main, Accommodators play a role as a leader of each group. In order to evaluate student outcomes, Reflective Learning specializes in Concrete Experience. It might also be an effective method to improve the quality of SAOC students.

ACKNOWLEDGEMENT

The author wishes to acknowledge the assistance and support of the Senior Air Officer College (SAOC).

REFERENCES

[1] J. G. Clawson, "Adult Learning Theory: It Matters", available athttp://faculty.darden.virginia.edu/clawsonj/pdf/, last accessed Aug 2009.

[2] A.Y. Kolb, D.A. Kolb "Experiential Learning Theory: a Dynamic, Holistic Approach to Management Learning, Education and Development", available at http://www. learningfromexperience.com/images/uploads/ELT-Hbk-MLE&D-LFE-website-2-10-08.pdf,last accessed Apr 2010.

[3] T. Liu, Y. Lin, K., M. Chang, "Individual Differences in Learning with Simulation Tool: A Pilot Study", *International Conference on Advanced Learning Technologies*, 2008

[4] J. F. Glass, "Learning and individual differences: a systems approach", available athttp://www.herdsa.org.au/branches/vic/Cornerstones/pdf/Glass.PDF, last accessed Aug 2009.

[5] C. V. Rehage, "Motor-skill learning in older adults - a review of studies on age-related differences", available at http://www.springerlink.com, last accessed Aug 2009

[6] Wikipedia (2010, Apr 12). B. F. Skinner [Online]. Available: http://en.wikipedia.org/wiki/B._F._Skinner#Theory.

[7] J. Cohen, (2008) "How does Positive and Negative Reinforcement Affect Urban High School Science Students?" Online, available at: http://dspace.nitle.org/bitstream/handle/10090/13976/1344.pdf?sequence=1,last accessed Apr 2010

[8] K. Henderson, K. Napan, S. Monteiro, "Encouraging reflective learning: An online challenge", available athttp://www. ascilite.org.au/conferences/, last accessed Dec 2009.

[9] Trinity College, "Logic, Programming, and Robotics for non-technical students", available athttp://www.scss.tcd.ie/

disciplines/information_systems/crite/crite_web/lpr/teaching/kolb.html, last accessed Sep 2009.

[10] H. Lu, L. Jia, S. Gong, B. Clark, "The Relationship of Kolb Learning Styles, Online Learning Behaviors and Learning Outcomes", available at http://www.ifets.info/journals, last accessed Sep 2009

Development of Knowledge Management for Public Sector Management Quality Award

Paipan Thanalerdsopit
College of Art, Media and Technology
Chiang Mai University
Muang, Chiang Mai, Thailand
jeabpaipan@hotmail.com

Dr.Noppasit Chakpitak
College of Art, Media and Technology
Chiang Mai University
Muang, Chiang Mai, Thailand
nopasit@camt.info

***Abstract*-Thai government has encouraged public sectors to improve their quality management, especially in educational area. Set up in 2005 with the agreement of Thai Ministers, *Public Management Quality Award (PMQA)* consisted of a quality standard of 7 categories is a framework of quality management for Thai public sectors. Also, Thai government would like to let government universities to manage themselves for reducing government expenses and solving the higher education organizations' problem with the bureaucratic system. For these reasons, Thai government universities are encouraged to transform to be affiliated public universities which are managed by universities councils under supervision of Thai Government Act. In addition, affiliated public universities will receive less public funding and must eventually be self financing. It is perceived that affiliated public universities would be run like business corporations, but some problems still exist. So the university councils need to strengthen their management and governance in order to raise the awareness of the importance and necessity of being affiliated public and arouse the motivation and creativity of their people to adapt to the change. How to examine organization by using PMQA is one of the crucial aspects of affiliated public university management. It is proved that management effectively creates a culture of communication and cooperation; positively influences the mentality of knowledge sharing. This research will focus on how to manage PMQA to enhance a better understanding on the policy and decision in an organization and to better involve people to work together to achieve the same goal, so that to improve the efficiency and effectiveness of quality management. College of Art, Media and Technology (CAMT) of Chiang Mai University (CMU) was chosen as a case study and field work of data collection. A conceptual model of *Learning Organization* will be designed to improve PMQA in affiliated public university.**

Index Terms — Public Sector Management Quality Award (PMQA), Knowledge Management (KM), Public University, Affiliated Public University

I. INTRODUCTION

In 1997, economic crisis occurred in Thailand, Thai government has applied for financial assistance from the Asian Development Bank (ADB). One of the suggestions from ADB was to let government universities govern themselves for reducing government expenses and solving the higher education organizations' problem with the bureaucracy system as shown by Nitikraipot [1].

Till 2010 [2], thirteen universities have been encouraged to change into affiliated public universities in Thailand. Among them, ten universities have finished their evolution from public university to affiliated public university, namely Mahachulalongkornrajavidalaya University, Mahamakut Buddist University, Thaksin University, Chulalongkorn University, King Mongkut's Institure of Technology Ladkrabang, King Mongkut's University of Technology Thonburi, King Mangkut's University of Technology North Bangkok, Mahidol University, Burapha University and Chiang Mai University; the other four universities: Mae Fah Luang University, Suranaree University of Technology, Walailak and Payoa University are new universities which were born in a form as affiliated public universities

Dr. Kirtikara [3] from King Mongkut's University of Technology Thonburi Bangkok made a thorough investigation on affiliated public universities in Thailand and studied the whole process of the evolution of being autonomy. He summarized the major features of affiliated public universities as followed:

- More controlled by university council, which is supreme in setting the vision and direction of a university, formulating policy on education and research, overseeing the personnel system, budget and finance. Performance evaluation of faculties, functional units as well as senior administrators is carried out by university council.
- Less budget form the government support, but affiliated public universities are more financially independent and self-supporting. Dr. Kirtikara put forwarded the statistics of

ISBN 978- 974-672-556-9

the national budget report of Thailand, he stated that 20% of the annual funds goes to education, which put Thailand as a country with considerable investment on education, which leads to less competitiveness in the global world.

- More control and management on the working performance of personnel. Public universities personnel are civil servants who are immune to rigorous performance evaluation like employees in the private sectors and enjoy life-long employment until retirement. However, affiliated public universities are regularly evaluated like employees in the private sectors on a contract level. This means, personnel in affiliated public universities are more naturally selected and assessed on the working performance and personal competence by the university council.
- Affiliated public universities have more responsibility and accountability because of a strict control and management of university council.

Compared with the limits of public universities, it is clear to see the advantages of affiliated public universities:

- Increase of income and assets based on technical capability
- More autonomy for universities to earn more income than being given
- New working culture can be established with stakeholders
- More open systems of recruitment are provided to qualified outsiders

Nevertheless, during the process of the transforming to affiliated public university, problems are still found. Two majors problems are: the first one is the ignorance of lectures and staffs in the university who are afraid of losing so-called permanent employment because of the new infrastructures and systems. The second problem is the existence of bureaucratic mentality that blindly obeying the leaders rather than being creative and productive in work.

In 1987, America has established National Quality Awards called *Malcolm Baldrige National Quality Award (MBNQA)* to improve quality management practices. The national governments are playing an active role in promoting and encouraging organizations to embrace quality management practices because quality management is one of the keys to competitiveness [4]. According to the same standard of 7 categories of MBNQA, namely leadership, strategy planning, customer focus, measurement, analysis and knowledge management, workforce focus, process management, results, both private and public sectors in many developed countries in the world have set up their own *National Quality Management Award* shown in Table I.

These awards in different countries borrowed the ideas from MBNQA and revised according to their own national features to implement and have gained success in quality management.

TABLE I. NATIONAL QUALITY AWARD [5]

Countries	National Quality Award	Year of Award
Australia	Australia Business Excellent Award (ABEA)	1988
European Union	European Quality Award (EQA)	1989
Singapore	Singapore Quality Award (SQA)	1994
Japan	Japan Quality Award (JQA)	1995

The Office of the Public Sector Developement Commission Thailand (OPDC) has supported Thai government agencies to apply New Public Management (NPM) by using Public Sector Management Quality Award (PMQA) that was agreed by Thai ministry on June 28, 2005. PMQA is based on a large part of the concepts of Malcolm Baldrige National Quality Award (MBNQA), America and Thailand Quality Award (TQA) which promote understanding of the requirements for Performance Excellence, Competitiveness Improvement, and Sharing of Learning of successful performance strategies [6]. Until 2010, no educational sectors have received this award or met the requirements of 7 categories of PMQA.

Chiang Mai University (CMU) was the first institution of higher education in the north of Thailand. In 2008, CMU changed to be affiliated public university which is governed by the University Council. The vision of CMU is *"to be an Affiliated public, Research-Oriented University, maintaining international standards of Quality Assurance and Academic Excellence". The Strategy Development 2008 -2012* explored that the goal of CMU to be an Excellent University with the aim of strategies based on many strategies. One of the strategies is PMQA which can improve quality management in organization [7]. In 2009, College of Arts, Media and Technology (CAMT), one of 22 faculties in CMU, volunteered to apply PMQA to improve quality management. CAMT has set up a team consisted of the dean, vice deans, assistant deans, head of operational staff and some of CAMT staff who work in PMQA model fields.

The purpose of this study is to design a Model of LO which is one of knowledge management theories for PMQA at CAMT, CMU, hoping to improve the management in CAMT. This model is the beginning of the efforts toward meeting the requirements of PMQA.

ISBN 978- 974-672-556-9

II. LITERATURE REVIEW

A. Quality Management Award in Education Sector

The PMQA Framework depicts the interrelationships among the seven categories of the Criteria and the eighth section of the application (known as the Organizational Profile). The accompanying diagram presents the Leadership triad, which is composed of three interrelated categories (category 1; Leadership, category 2; Strategic Planning, and category 3; Customer and Market Focus), as well as the Results triad, which is also comprised of three interrelated categories (category 5; Workforce Focus, category 6; Process Management, and category 7; Results). These six categories are supported by the category4 of Measurement, Analysis, and Knowledge Management, which supplies the measures and analysis needed to evaluate the triads and their individual categories. The Organizational Profile aids an organization to analyze its environment, business relationships, and challenges. The 7 categories and the relationship among them are shown in the Fig.1

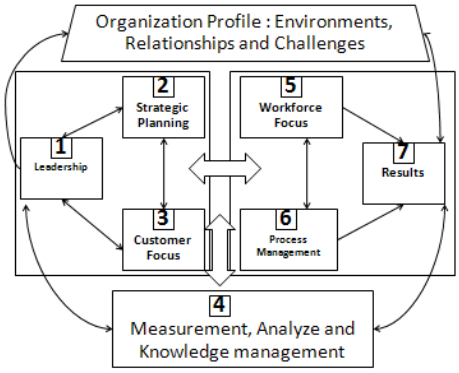


Figure 1. Public Sector Management Quality Award Framework [5]

From 2005 to 2009, Thai public sectors has implementation of PMQA, the number of Thai educational organizations applying for excellence awards increased dramatically on the state and national level, but these institutions have not received the PMQA, but the MBNQA which is the Baldrige Award that PMQA is referred has recipients in education organizations as followed:

TABLE II. MALCOLM BALDRIGE NATIONAL QUALITY AWARD EDUCATIONAL RECEIPTIONIST [8]

Year	*MBNQA Educational Recipients*
2001	The Chugach School District and Pearl River School District and University of Wisconsin-Stout
2003	Community Consolidated School District 15
2004	Kenneth W. Monfort College of Business
2005	Jenks Public Schools and Richland College
2008	Iredell-Statesville School

So that in 2009, the Office of the Higher Education Commission Thailand has set Education Criteria for Performance Excellence which is a version of the PMQA Criteria that focuses on the specific situations of educational institutions because the educational marketplace is growing more crowded every year, many institutions are turning their attention to Performance Excellence as a means of staying competitive, especially for Thai affiliated public universities. Thai Education Sectors need to understand the differences in the criteria that apply to educational groups. Quality management helps organization to understand what to focus on, examine relationships with students and other stakeholders, create student learning outcomes, and evaluate processes to give the utmost value. Sousan and Voss [9] showed that there should have study in qualtiy management in a firm because quality management is very general principle. So this is one challenge for Thai education sectors to start in quality management as the best pratices.

TABLE III: COMPARATIVE ANALYSIS BETWEEN EFQM [10] AND PMQA [11]

Comparative *analysis*	*EFQM*	*PMQA*
History	EFQM was founded in 1988 by leaders from fourteen companies who were convinced that a new membership organization, based in Europe, was necessary to promote higher standards of management through shared knowledge and mutual recognition.	Thai public sectors have implementation of PMQA in 2005 applying for excellence awards to improve quality management practices for advantage of people and nation
Culture	EFQM supports organizations in defining what sustainability means, providing approaches for its implementation and ensuring consistency between apparently conflicting responsibilities toward shareholders, employees and society.	PMQA supports Thai public agencies which promote understanding of the requirements for Performance Excellence, Competitiveness Improvement, and Sharing of Learning of successful performance strategies, also encourages Thai Royal Decree of Good Governance 2003.
Economy	Organizations have higher performance or more advanced in their structure sand processes, by taking part in an exchange of real world practices and experiences.	Thai Public agencies have more high service standard to make people satisfy, Also to exchange knowledge from best practices to other public agencies apply to be successful their organization.
Structures	EFQM Model: the eight fundamental concepts, its widely recognized nine-box structure, and the acclaimed RADAR scoring matrix	The standard of 7 categories, namely leadership, strategy planning, customer focus, measurement, analysis and knowledge management, workforce focus, process management, results.
Awarded Winners	281 organizations Since 1992 - 2009	Non Since 2005 - 2010

ISBN 978- 974-672-556-9

The European Network for Quality Assurance in Higher Education (ENQA) is the umbrella organization for European quality assurance agencies in the field of higher education. ENQA spends a lot of efforts in order for the European Foundation for Quality Management (EFQM). This is a comparative analysis between the Thai and European approaches, with a special focus on the underlying historical, cultural, and economic, structured perspectives and award receptionist are shown in Table III.

From the table III, EFQM was founded by European leader companies, PMQA is referred from MBNQA. The culture of EFQM supports organizations for sustainability; the side of PMQA promotes high performance management in Thai public sectors. Both of them need satisfactions from customers, shareholders, employees and society and use models as a tool for assessment. The economic part, EFQM and PMQA support to be excellent organizations. The EFQM is based on 9 criteria, five of these are 'Enablers' and four are 'Results'. PMQA has seven categories with one organization profile, but the basic their structures are the same with Excellent Key Results: Customer, People and Society Results are achieved through Leadership driving the Strategy that is delivered through People, Partnerships and Resources, and Processes, Products and Services. For 5 years of PMQA, there is no award winner, EFQM has 281 awarded winners in 17 years, comparing 5 years (1988 – 1993) is from started year, EFQM had 8 award winners. How Thai government encourages their public organizations to be the receptionists. This is one of the challenges for Thai public agencies.

B. Learning Organization in PMQA

The Office of Public Sector Development Commission Thailand (OPDC) has promoted knowledge management as creating change management teams through knowledge management within the public sector that change cannot be successful if the persons responsible for driving the change process do not coordinate; consequently, creating change management teams is a vital process to produce resources whereby the change managers can use as a tool to drive the intended change. Additionally, The Royal Decree on Good Governance Article II stated that government agencies are to regularly develop knowledge within their organization in a pattern of the learning organization. Data and information are to be organized for further application to create faster, more accurate and updated work methods. They must also promote and develop capabilities, build visions and change paradigm of the officials within their organization, encouraging efficiency and mutual learning effecting result of performance.

The fundamental objectives of knowledge management and quality management are the same create more organizational knowledge so that improvement can occur. Organizations maintaining a set of quality management practices that support the knowledge creation processes should be more effective at deploying quality management. Quality frameworks like the Baldrige Award increasingly recognize the importance of knowledge as ‘‘Information and Analysis’’ was recently changed to ‘‘Measurement, Analysis, and Knowledge Management’’ but do not incorporated knowledge into the logic of deploying quality management practices. This knowledge perspective presents an opportunity for future refinements of the MBNQA. In addition, deploying quality from a knowledge-based view seems more consistent with the underlying philosophy of quality thought leaders [12].

Garvin [13] and Senge’s theory [14] are used to compare with MBNQA categories and principle Senge’s theory shown table IV as followed:

TABLE IV. LEARNING ORGANIZATION AND MBNQA [15]

MBNQA, Garvin(1991)	Learning Organization of Senge (1992)
Leadership	Building Shared vision
Human resource development	Team Learning Personal mastery
Information and analysis	Working with mental models
Process improvement	
Customer Focus	System thinking
Organization Profiles	
Strategic Planning	

A leader should share planning to employees to know mission, vision and strategy of organization that is “share vision”

Human Resource development should work as a team to learn together; “team learning: and change traditional ways of thinking that matches with “mental models”.

In accordance with using factual information and process management by learning to be open with other people is “mental model”

Customer focus, Organizational performance and strategic planning should observe in wholes picture is “system thinking”

This is a mutual independence between learning organization and framework of MBNQA that Terziovski el al reported on a theoretical framework based on Senge’s principles and the Malcolm Baldrige National Quality Award (MBNQA) Criteria.

III. RESEARCH METHODOLOGY

This research focuses on the Category 4 in PMQA which is Measurement, Analysis and Knowledge Management category. The hypothesis is how to create LO model in PMQA by using CAMT

as case study to support this category because this part likes blood that links all part of PMQA body [11]. The details of research methodologies start from setting up team of CAMT PMQA. The next step is to study the current situation of administration and management by using PMQA framework in CAMT to examine organization. The research tool is in-depth interview with the dean called executive level. Another level is middle level, namely, vice dean and assistant deans. Last level is operational level including heads of administrative, supporting departments and staffs in CAMT. Also, literature reviews in quality management and knowledge management in public fields, especially, high educations are used in this step.

Between September 2009 – January 2010, researchers applied PMQA Model to justify problems of administration and management in College of Art, Media and Technology (CAMT). The CAMT PMQA model was designed the dean of CAMT, Dr.Nopasit Chakpitak as followed:

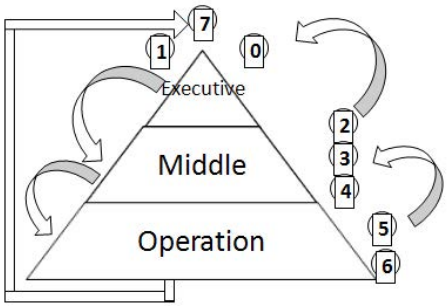


Figure 3. The CAMT PMQA model [16]

This is a model of CAMT PMQA. There are 3 levels namely, executive level, middle level and operational level. In 2009, CAMT applied PMQA standard to examine organization.

From the executive level, the dean answered the PMQA questions of category 7 (Results) because this category shows how organizations design Key Performance Indicators (KPIs) and can help CAMT PMQA committees knew goal of this framework.

After designed KPIs of result, the next step is Organizational Profile which presents environment, relationships, and challenges of organization which asks about all management in organization to help committee to know a whole picture of management in CAMT. Next is category 1 (Leadership).

The middle level answered the question of category 2 (Strategic Planning), category 3 (Customer Focus) and category4 (Measurement, Analysis, and Knowledge Management).

The Last one is the operation level answered the question of category 5 (Workforce Focus) and category 6 (Process Management), but all the process the dean will approve by himself.

Ross R. wrote about designing a Learning Organization (LO) in Senge [17] wrote about designing a Learning Organization: in the first process to design a LO, there are 6 steps. The authors compared these 6 steps with the initial process of LO in CAMT as the table V followed:

TABLE V. COMPARING CAMT PMQA WITH LO STEPS [18]

LO Design Step	CAMT LO
Step1: Establishing the Groups	CAMT sets group of people to apply PMQA in organization call *CAMT PMQA TEAM*
Step2: Divergent thinking	The dean as a facilitator to help people in PMQA team to answer PMQA questions. Center Computer Monitor is a flip chart to show answer. Each staff has their answers.
Step3: Clarity	The PMQA team answers question form PMQA.
Step4: Convergent Thinking	Everyone has opportunity to explain and challenge his or her answers.
Step5: Presentations and Priorities	PMQA team select best answer with all agreement and present these answer in team again.
Step6: Implementation	CAMT can implement their First step of LO in PMQA team

Moreover, the Learning Organization is a concept of Senge[18] which has five components namely, systems thinking, personal mastery, mental models, shared vision, and team learning. Gavin[13] compared Leaning Organization and MBNQA categories. So this concept can adapt to address similar categories of the PMQA as followed:

The dean of CAMT shares his vision, strategies to followers, namely CAMT staffs agree with organizational plan (Share vision). Also CAMT sets team to do PMQA and staffs in organization work together learn to be open with others (Team Learning and Personal Mastery). Furthermore, CAMT uses real information and apply information to process improvement (Metal Model), so mental model needs to balance both of inquiry and advocacy. Moreover, CAMT highlights a circle of influence on Customer focus, Organization Profiles and Strategic Planning (System thinking).

IV. RESULT

"Learning Organization" of Senge is a theory to apply for CAMT PMQA. After finished CAMT PMQA, The dean of CAMT showed challenging how CAMT ensures the quality and availability of needed data, information, software, and hardware for workforce, students and stakeholders, partners and collaborators. The quality of data, information, and knowledge are accuracy, integrity, reliability, timeliness, security and confidentiality by these methods as followed:

- Leader as a designer to cover all of information

- Board of Committee and administrators use information in day-to-day work, for example conference, management, practices and activities in CAMT
- CAMT uses information under CMU regulations.
- Information is audited by original sections, for instance financial information is checked by financial section ,etc
- All staffs in CAMT have personal computer
- There is two quality assurance meetings every year.

Furthermore, there is upload PMQA to disseminate in Knowledge Management System (KMS) in CAMT. So CAMT staff can learn about PMQA. Moreover, some lecturers applied PMQA to teach in Master Degree class.

In a future, CAMT applied results of Category 4 focuses in Risk Management to be the one of planned to use Committee of Sponsoring Organizations of the Treadway Commission-Enterprise Risk Management (COSO- ERM) to reduce risks namely: 1) No or Wrong information Risk 2) Hiring error information Risk 3) Cheating Risk 4) Business Risk 5) Private Risk 6) Disastrously Information risk 7) Error processing Risk 8) Compliance Risk 9) Administrative Risk 10) Lack of Knowledge and Experience risk [16].

This is the first step of setting LO in PMQA of CAMT. People, Process and Technology are used to work together to approve LO in CAMT PMQA.

V. BENEFICIARY

CMU is affiliated public universities on 2008[7]; there is the CAMT as a volunteer to be the example to apply PMQA in organization by using LO theory of Senge[14]. The result of this research showed that *Learning Organization* is to mutual tools with *Public Sector Management Quality Award*. The beneficiary is increase quality management in affiliated public universities to be sustainable organizations and the way to touch PMQA is not difficult.

VI. CONCLUSION

The researcher used concept of "*Learning Organization*" as tools to improve PMQA in affiliated public universities by using CAMT CMU as case study. The result of this research can be a sample for improving in PMQA for education sectors of Thailand. The implementation and evaluation of this model needs any comments with welcomed, Also co-researches are encouraged.

ACKNOWLEDGMENT

Many thanks to my supervisors: Dr.Nopasit Chakpitak, Assistant Prof. Dr.Anchalee Jengjarern, Associated Prof.Dr.Sermkiat Jomjunyong, Dr.Pitipong Yodmongkon with their help and guide for this research between studying Ph.D at Chiang Mai University. Moreover, the Rajamangala University of Technology Lanna and the Graduate School's Chiang Mai University support scholarship for this conference.

REFERENCES

[1] S. Nitikraipot, "Public organization", presented at the meeting of Preparing to Autonomous Public Forum, Lampang, Thailand, 1999.

[2] The Wikimedia Foundation. (2010, Sep. 30). The Wikimedia Foundation. Available: http://th.wikipedia. org/wik

[3] K. Kirtikara, "Thai Public University System in Transition: Some Issues on Management and Financing", presented at the Thai-UK University Presidents Forum, Bangkok, Thailand, Jan. 2002.

[4] P.-M. Lee and H. A. Quazi. (2001). A methodology fordeveloping a self-assessment tool to measure qualityperformance in organizations. J. Quality & ReliabilityManagement, 18 (2), pp.118-141.

[5] The Office for National Education Standards and QualityAssessment(Public Organization). (2009). National EducationAct B.E.2542 (1999) and Amendments (Second NationalEducation Act B.E.2545 (2002). Available: http://www.onesqa.or.th/en/publication/nation_edbook.pdf

[6] The Office of the Public Sector Development Commission. (2008). *The importance and reason of Public Sector Management Quality Award. Manual of Public Sector Management Quality Award 2009 for Department of Thailand*. pp. 3-7.

[7] Chiang Mai University. (2009). Available: University:http://inter.oop.cmu.ac.th/Images/Brochure_CMU/Artwork.pdf

[8] Malcolm Baldrige National Quality Award. (2010). Available:http://http://www.baldrige.nist.gov/Contacts_Profiles.htm

[9] Sousan R. and Voss C.A. (2002). Quality Management re-visited: a reflective review and agenda for future research. J. Operations Management. 20. pp. 91-109.

[10] The European Foundation for Quality Management. (2010). The European Foundation for Quality Management. Available:http://www.efqm.org/en/Home/aboutEFQM/Our models/TheEFQMExcellenceModel/ThereviewoftheEFQM ExcellenceModel/tabid/286/Default.aspx.

[11] The Office of the Public Sector Development Commission.(2010). Munual of Public Management Quality AwardAvailable:http://www.opdc.go.th/pmqa/www/pdf/pmqa-manual-2550.pdf

[12] Linderman, K., Schroeder, R. G., Laheer, S., Liedtke, C., and Choo, A. S. (2004). Integrating quality management practices with knowledge creation process. J. Operations Management . *22*. pp.589-607.

[13] Garvin, D.A. (1991), *How the Baldrige Award really Works. Harvard Business Review*. 69(6). pp.80-93

[14] Senge P.M. (1990). *The Fifth Discipline: Heart and Practice of the Learning Organization*, Doubleday, New York.

[15] Terziovski M. *et al* (2002). Establishing Mutual Dependence between TQM and the Learning Organization: a Multiple Case Study Analysis. J. the Learning Organization. 7(1). pp.23-31.

[16] Chakpitak, N. (2010, January). Category 4Measurement, Analysis and Knowledge Management. (P. Thanalerdsopit, Interviewer)

[17] Senge P.M. (1994). *The Fifth Discipline Fieldbook: Strategies and Tool for Building a Learning Organization*, Doubleday, New York.

[18] Senge P.M. (1992b), *The Fifth Discipline: The Art and Practice of the Learning Organization*, Random House, Australia.

Session 11: Identifications. Modeling and Robotics

ISBN 978- 974-672-556-9

Advancing Towards Rescuer Robot Team

Md. Nazmul Huda, Jason Whittaker, Samuel Oliver Wane, and Hongnian Yu
Faculty of Computing, Engineering and Technology, Staffordshire University, Stafford, ST18 0AD, UK.
m.n.hudah@staffs.ac.uk, s.o.wane@staffs.ac.uk. h.yu@staffs.ac.uk.

***Abstract*—This paper presents architecture for a novel rescuer robot team based on a multihop mobile ad-hoc network (MANET) for disastrous conditions. In hazardous and dangerous circumstances such as earthquakes, cyclones, volcanic eruptions, or other natural or manmade disasters the robot team can perform a rescue mission. In many disasters the total infrastructure is broken: no roads, no mobile networks, no internet and no satellite communication. Then the only access to the disaster area is by helicopters. The proposed rescue team can be dropped into the region and, along with the controller PC will form a multihop MANET and can perform the rescue tasks. Here, a prototype for a single rescue robot is developed which can be controlled wirelessly and also can perform the job of a patroller.**

Index Terms: MANET; rescuer robot; robot team

I. Introduction

Climate change is a major reason for distress in today's world, causing a worldwide increase in disasters. The number of manmade disasters is also increasing along with natural disasters. After any sort of disaster, fast and effective rescue work is essential as after 48 hours the probability of survival is very low. Thus mobile robots are used to decrease the rescue workers' load, to increase the possibility of finding survivors, to minimize the risk to the rescue workers from dangerous environments e.g. poisonous gases, harmful chemicals, and fragile wreckage with the risk of collapsing, and above all to increase the effectiveness of the rescue work. Most recently we have seen the use of robots to repair the oil leakage in the Gulf of Mexico.

In section II a brief review of rescuer robots will be delivered, and in section III an overview of the architecture of the proposed rescuer robot team is presented. Section IV is dedicated to the hardware specifications of a developed single rescuer robot, section V discusses the programming of the robot, and section VI presents the created programs and the purpose of the programs. Section VII discusses the experiment and its application, section VIII describes the modifications that need to be done to the single robot to make it a member of the rescuer robot team and also describes the additional equipment that has to be integrated with the robots to form the tele-operated rescuer robot team. Finally, section VIII is dedicated to the conclusion.

II. Brief Review of Rescue Robots

Many rescue robots have been developed for minimizing the number of casualties from a disaster [1], [2], [3]. Rescuer robots can be classified from different perspectives. From the perspective of mobile mechanisms, the rescuer robots can be divided into the following four classes.

1) Wheeled robot
2) Legged robot
3) Tracked robot
4) Wheeled-legged robot.

TABLE I. Comparison of Mobile Mechanism [4]

Mobile mechanism	*Advantages*	*Disadvantages*
Wheeled Robot	High Mobility / Efficient	Low Obstacle Performance
Legged Robot	High Obstacle Performance	Low Mobility / Low Efficient / Hard to Control
Tracked Robot	High Obstacle Performance/ Large Payload /Cramped Construction	Heavy/ High Energy Consumption
Wheeled-legged	High Obstacle Performance	Hard to Control

Depending on the control of the robot, rescuer robots can be divided into following four classes:

1) Autonomous (Intelligent terrain) robot
2) Semi-autonomous
3) Tele-operated robot
4) Either Autonomous or Tele-operating mode robot

An autonomous robot has to take decisions about its path and thus must have some intelligence. By using intelligence it can direct itself to reach a victim and get information. In [5] a semi-autonomous rough terrain robot is developed which can run autonomously during a temporarily broken connection to the control base. In [6] the development of a semi-autonomous rescue robot is also discussed whereas in [7] a robot has been developed which can perform either in autonomous or tele-operated mode.

Moreover, rescue robots can be further divided into single standalone robots and group rescue robots. The robots we have so far discussed are single rescue robots. [4], [8] and [9] give examples of group rescue robots. From Table 1 we can see that no mobile mechanism has all the necessary qualities. Different mechanisms have their own strong point and constraints. The motivation for using group rescue robots is to use the strengths of different

ISBN 978- 974-672-556-9

mechanisms together. In [4] one parent and two child robots are used to perform a rescue mission. Initially the parent, which has the highest mobility and power efficiency, carries the child robots to the disaster area, releases the child robots, and they perform the rescue task. But one problem with these group rescue robots is that the parent robot cannot cross narrow areas.

III. Overview of The Architecture of The Proposed Rescuer Robot Team

The architecture of the rescuer robot team is shown in Fig. 1. All the intelligent rescuer robots would have a Mini PC (WS635 mini PC) of dimension 14.6 x 25 x 4.2 cm on board. The robots and the controller PC would be able to create a multihop mobile adhoc network (MANET) using the free software olsrd, downloadable from olsr.org. The controller PC would be able to communicate with any robot through multihop MANET. The robots would be both autonomous and teleoperated depending on the demand. Every robot would have a video camera with a microphone which would be capable of moving in all directions. Data collected from all the sensors, encoders and video would be sent to the controller PC. Analyzing the data controller would send the control signal to each of the robots. The robots would be intelligent and would be able to make decisions of their own, and would ask for help if facing a critical problem. The controller PC would integrate the work of all the robots. The robots would also be controlled wirelessly from the controller PC using a graphical user interface or joystick. The camera of one robot may be used to see other robots to make decisions in critical moments.

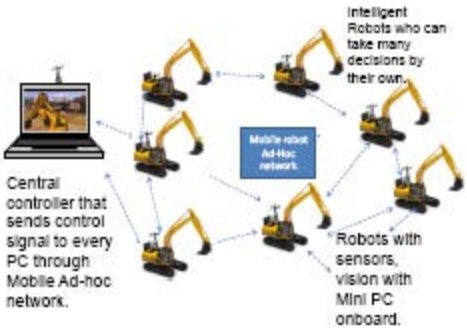


Figure 1. Architecture of the rescuer robot team

A number of Rescuer Robots (5 to 20 as the situation demands) will be dropped into the disaster area by helicopter. The main controller PC will be placed on the helicopter. The robots will be monitored and controlled from the controller.

Figure 2. Rescuer robots performing rescue work

IV. Hardware Specification of a Single Robot

The required components for a single robot are listed below:

1) Traxster II robot chassis kit
2) A serializer microcontroller
3) A Bluetooth communication module
4) A servo motor
5) An ultrasonic ranger
6) A NiCd battery
7) An IP camera
8) A router
9) A voltage converter

A. Traxster II Robot Chassis Kit

This includes an orange chassis with tracks on the either side and two DC gear head motor with encoders. We have chosen a chassis with tracks as it has high obstacle performance and cramped construction which is the requirement of a rescuer robot.

B. Serializer

The serializer is a microcontroller which can be used with the traxster II robot chassis kit and is able to control DC motors, servo motors and is compatible with the .NET framework. The serializer is inserted at the bottom of the chassis.

Figure 3. Underside of traxster with serializer installed

C. Bluetooth Communication Module

A Bluetooth communication module is inserted into the serializer to communicate wirelessly with the controller computer

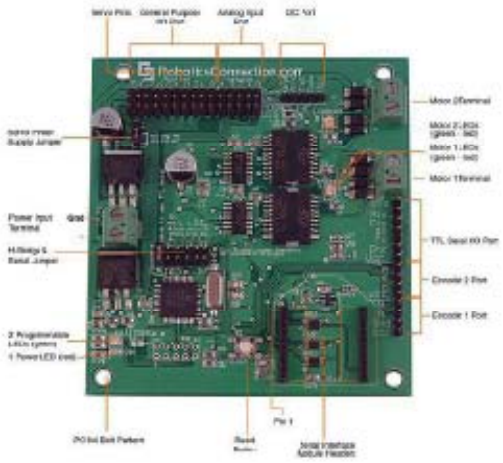


Figure 4: Serializer board layout with labels [10]

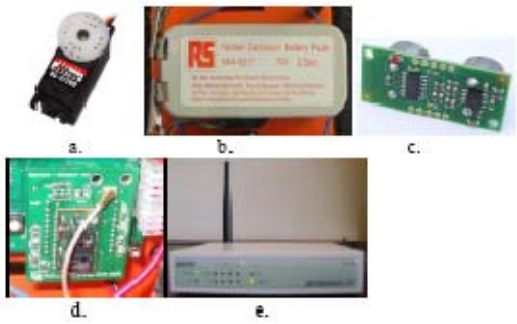


Figure 5. a. servo motor b. NiCd battery c. Ultrasonic ranger d. Bluetooth module e. Router

D. Ultrasonic Ranger/Sensor

An ultrasonic ranger SRF05 is used here to measure distance from an object. Its range is 1-4m and its frequency is 40KHz. It is attached to a right-angled piece of metal which is in turn attached to the servo. The total format looks like a turret.

E. Servo Motor

A servo motor, HS-322HD is used to rotate the ultrasonic ranger in different directions to measure the distance in front, right and left.

F. NiCd Battery

A 12V 2.5Ah nickel cadmium battery pack is used to power the serializer. The servo and DC motors take power from the serializer.

G. IP camera

An IP camera, Panasonic BL-C131 is used to transmit videos with sound to the controller computer. Video with sound is shown by opening a webpage with the IP address of the camera.

H. Router

A Buffalo router, WYR-G54 was used to make the controller PC enable to communicate with the IP camera.

I. A Voltage Converter

A 12V to 9V converter is used to power the IP camera from the 12V battery.

Figure 6. Complete arrangement of a single robot

V. PROGRAMMING OF THE ROBOT

The programs are written and executed on the controller computer and the data and instructions are communicated between the computer and the robot through Bluetooth. Here the program is written in C#. The serializer library contains a lot of commands and some of these commands are used to control the serializer and other components.

At the beginning of each program the serilizer library should be included in the following way which allows the serializer library to be used in the code.

```
using RoboticsConnection.Serializer;
using RoboticsConnection.Serializer.Ids;
using RoboticsConnection.Serializer.Sensors;
using RoboticsConnection.Serializer.Components;
using RoboticsConnection.Serializer.Controllers;
using System;
using System.Threading;
```

To communicate with the serializer the following two commands are used:

```
serializer.PortName = "COM12";
serializer.StartCommunication();
```

Here the number of the COM port at which the Bluetooth is activated is mentioned.

VI. Created Program and Purpose

The robot can be teleoperated or can move autonomously. Both the programs are shown below:

A. Rescuer Robot Program:

When the program is run the graphical user interface shown in Fig. 7 is displayed on the screen of the controller PC.

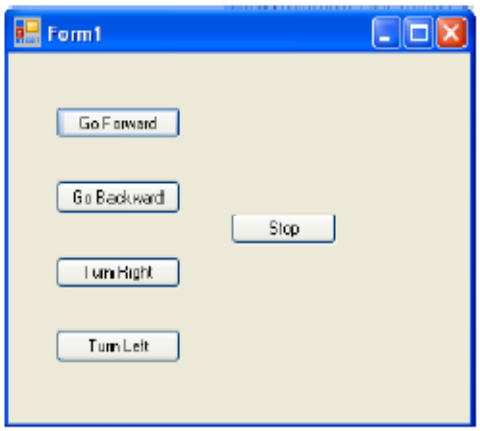


Figure 7. Graphical user interface for rescuer robot program

An operator in front of the controller PC can navigate the robot with the aid of the GUI in Fig. 7 and the received video with sound from the IP camera. The operator can see up/down, right/left of the robot by using the pan and tilt option of the camera and, together with the sound, the operator gain a perfect understanding of the environment of the area and can take the right decisions about which way the robot should be directed.

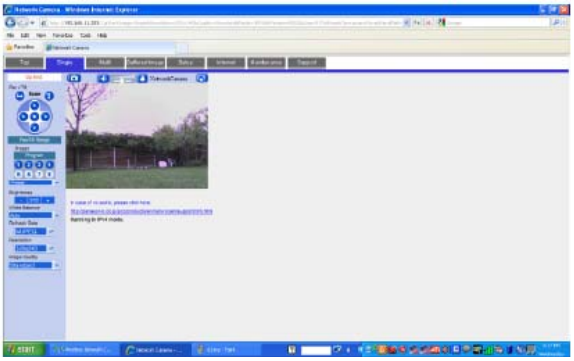


Figure 8. Video received from the IP camera

B. Patroller Robot Program:

When the program is run it opens a CMD window, shown in Fig. 9, and tries to connect to the communication port. Once connected, the turret moves to face the front of the robot (if it does not do so already) and the robot begins to move forward. The sensor takes frequent readings of the distance in front of the robot while it moves forward. The robot detects an obstacle if something is closer than 25.4cm (10 inches). When this happens the CMD window will show the stages that the robot goes through before it begins to move forward again. The order is as follows:

1) Sensor detects obstacle
2) Motors are turned off, stopping the robot
3) The turret turns right and checks the distance
4) The turret turns left and checks the distance
5) The turret faces the front of the robot
6) The robot turns right/ left (Motor turns to the direction with the higher distance)
7) The robot continues to move forward

If the robot finds itself in a situation where all the distances (left, right and in front) are below 25.4cm it declares itself to be trapped. Then the robot goes backward for 2 seconds, turn about half way round and then moves forward. The program is a continuous loop so the robot continues this routine until communication is stopped.

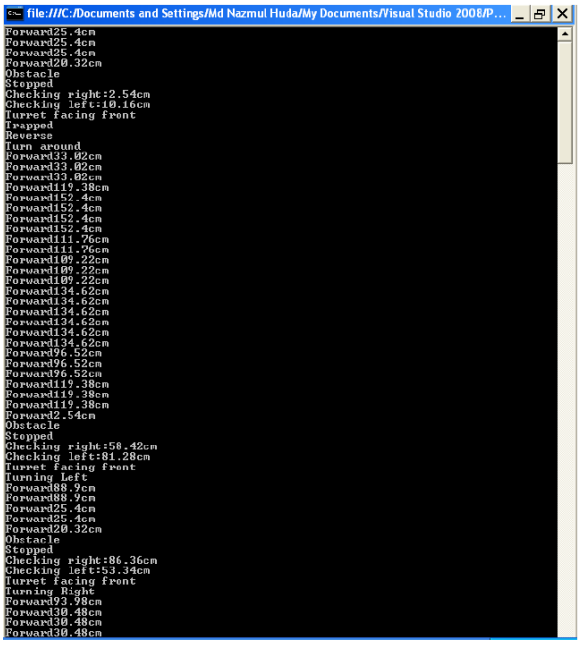


Figure 9. CMD window for Roaming robot program

VII. Experiment and Application

The robot was tested indoors and outdoors. It is capable of moving in roads, in open field with grass, branches and even scree under the grass. It can climb a significant upwards slope and also can go down a significant downwards slope.

Using the GUI of the rescuer robot program and the video with sound the robot can be sent to dangerous places before humans. It can be sent inside mine tunnels after a coal mine disaster to gain perspective of the disaster and the condition of the survivors. The sound received from the camera can help to detect survivors under wreckage. Security officials can use the robot to get information about a criminal hiding in a building without risk to themselves. It can be used to search for poisonous

ISBN 978- 974-672-556-9

gases, chemicals, viruses or radioactivity with the aid of correlative instruments.

Using the patroller robot program the robot can patrol a certain area and an observer can watch from a remote area. If anything suspicious is seen the operator can switch to the rescuer robot program and move toward the suspicious area, and can also follow a suspicious person.

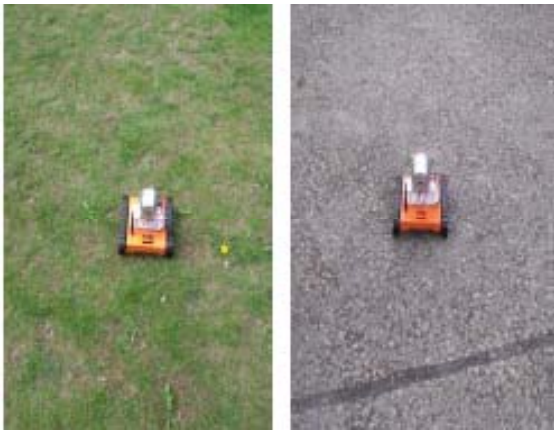


Figure 10. The robot works in field and road

VIII. Required Modification to the Single Robot and Further Equipments

Initially we would develop a prototype of the proposed rescuer robot team with a Mini PC on board each robot and would try to control the robots from the control PC through multihop MANET. We would like to add a lighting system in front of the robot to allow it to see at night and in dark places where the sunlight is blocked. We would replace the battery with a compact Ni MH battery and would put it below the serializer with the help of a belly pad. The Mini PC will be on the top of the chassis. The Bluetooth communication module will be replaced with a usb communication module and will be connected physically to the Mini PC on board by a usb cable. The IP camera will be replaced by a usb webcam with microphone and placed on top of a servo motor so that it can be rotated up to 360 degrees to have a view of all directions. The data from the sensors and the camera will be sent to the corresponding Mini PC, and from the Mini PC, the information will be sent to the controller PC through MANET.

IX. Conclusion

The developed robot worked properly on grass terrains and also on roads. It can be used as video surveillance and also in disaster rescue. This is a good prototype to be used later on. With some modifications this robot will be adopted to develop the rescuer robot team which will have a good impact on the infrastructureless rescue work.

Acknowledgement

Md. Nazmul Huda thanks J.M.M. Kamal of Staffordshire University for sharing his knowledge about ad-hoc network with the author.

This work was supported in part by the EU Erasmus Mundus Project-eLINK under grant EM ECW-ref.149674-EM-1-2008-1-UK-ERAMUNDUS.

References

[1] Masataka Suzuki, Shinya Kitai and Shigeo Hirose, “Basic systematic experiments and new type child unit of anchor climber: swarm type wall climbing robot system”, *Proc. IEEE Int. Conf. on Robotics and Automation*, Pasadena, CA, May 2008, pp. 3034-3039.

[2] Masaki Ohira, Ranajit Chatterjee and Tetsushi Kamegawa,” Development of three-legged modular robots and demonstration of Collaborative task execution”, *IEEE International Conference on Robotics and Automation*, Roma, April 2007, pp. 3895-3900.

[3] Markus Eich, Felix Grimminger and Frank Kirchner, “Proprioceptive control of a hybrid legged-wheeled robot”, *Proc. IEEE Int. Conf. on Robotics and Biomimetics*, Bangkok, February 2009, pp. 774-779.

[4] Yong Yang, Guiyun Xu, Xinyu Wu, Huiwei Feng, and Yangsheng Xu,“Parent-child robot system for rescue missions,” *Proc. IEEE International Conference on Robotics and Biomimetics*, Guilin, China, December 2009, pp. 1427 - 1432.

[5] J. Suthakorn, S.S.H. Shah , S. Jantarajit , W. Onprasert and W. Saensupo,S. Saeung, S. Nakdhamabhorn,V. Sa-Ing, and S. Reaungamornrat, “On the design and development of a rough terrain robot for rescue missions, ” *Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics* Bangkok, Thailand, February 2009, pp. 1830 – 1835.

[6] A. Birk , H. Kenn, “A Rescue Robot Control Architecture ensuring Safe Semi-Autonomous Operation,” {Robocup}-02:Robot Soccer World CupVI, Springer, INIST-CNRS, Cote INIST, February, 2004

[7] J. Pellenz, “RoboCup 2008- RoboCupRescue Team resko@UniKoblenz”, University of Koblenz-Landua, Germany, 2008

[8] Tetsushi Kamegawa, Kenta Saikai and Akio Gofuku, “Development of grouped rescue robot platforms for information collection in damaged buildings”, The University Electro-Communications, August 2008, pp. 1642-1647.

[9] Keiji Nagatani, Kazuya Yoshida and Kiyoshi Kiyokawa, “Development of a networked robotic system for disaster mitigation”, Proc. Field and Service Robotics, August 2008, pp. 453-462.

[10] URL: http://www.activerobots.com [Accessed 14.05.10]

Metamorphosing a Five-link Robotic Manipulator Equipped with Universal Joints into One with Regular Joints Using Direct Collocation

Jyun-jye F. Chen, Dawa, Yeshi, Nidup and Sangay Nidup
Royal University of Bhutan/College of Science and Technology,
Department of Electrical Engineering, Phuentsholing, Bhutan
Email: felipefelipechen@yahoo.com.tw

***Abstract*—This study applies optimal control to differentiate the energy efficiency between two robotic manipulators, one with regular joints and the other with universal joints. The dynamics of the manipulators is described by Lagrange-Euler formulation. Symbolic language programming is used to derive the intricate equations of motion. Finally the method of Direct Collocation and Nonlinear Programming facilitates a robust numerical approach to the solution of the two-point boundary-value problem coming from the Euler-Lagrange equations while solving the optimal control problem. The goals of this work is not just to give a deeper dig on universal joints and regular manipulator joints, but also demonstrates how DCNLP and symbolic language techniques can be used when other problem of interest shares the common ground with the one considered in this paper.**

Index Terms- Robot manipulator; universal joints; optimal control; Lagrange multipliers

I. Introduction

Robotic manipulators have become important tools in many industrial applications. When a manipulator carries out repetitive task, the energy efficiency becomes an issue in its operation. This study focuses on a manipulator with a five-link arm and optimal control techniques are used to explore its least-energy control sequence.

The Denavit-Hartenberg (D-H) convention is adopted in defining the geometry of the arm. The equations of motion(E.O.M.) of the manipulator are described by using the Lagrange-Euler (L-E)formulation [1]. See Section II. Once the E.O.M. are ready, the optimal control techniques pick upright away to search for the solution.

In formulating the necessary conditions for optimality to occur, due to the incomplete boundary conditions given at the initial and final times in the Euler-Lagrange equations [2], it becomes difficult to solve the differential equations (12) and(13), and the optimization problem is trapped by this intricate situation also known as two-point boundary-value problem(TPBVP). Thus the direct collocation and nonlinear programming (DCNLP) method is introduced to take a fling at this hurdle [3]. DCNLP converts the TPBVP into a nonlinear programming problem. See Section III. The tradeoff is that DCNLP requires additional large computer memory capacity and fast computer speed. However these two factors are no longer concerns now-a-day due to the improvement in computer hardware.

A five-link manipulator arm model is built in Section II. It is used as a vehicle in order to verify if the model is correctly and flawlessly built. In Section IV, the same five-link manipulator is gradually metamorphosed into an equivalent manipulator with two universal joints in ten steps. Their cost functions and the times needed to complete the tasks t_f are monitored.

II. Equations of Motion of a Five-link Manipulator and Optimal Control

A. Geometry of the Five-link Manipulator

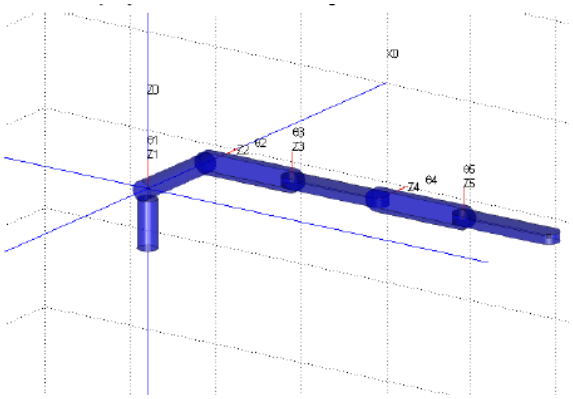


Figure 1. Geometry of the five-link manipulator arm.

The geometry of the five-link manipulator is shown in Figure1. It is laid out in such a way that Links 2 and 4 are capable of moving up and down and Links 3 and 5 are capable of moving left and right. By the Denavit-Hartenberg conventions, the five homogeneous transformation matrices are listed below:

$$ {}^{0}A_1 = \begin{bmatrix} \cos\theta_1(t) & -\sin\theta_1(t) & 0 & 0 \\ \sin\theta_1(t) & \cos\theta_1(t) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1) $$

$$ {}^{1}A_2 = \begin{bmatrix} 0 & 0 & 1 & L_1 \\ -\cos q_2(t) & \sin q_2(t) & 0 & 0 \\ -\sin q_2(t) & -\cos q_2(t) & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2) $$

$$ {}^{2}A_3 = \begin{bmatrix} \cos\theta_3(t) & -\sin\theta_3(t) & 0 & L_2 \\ 0 & 0 & -1 & 0 \\ \sin\theta_3(t) & \cos\theta_3(t) & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3) $$

$$ {}^{3}A_4 = \begin{bmatrix} \cos\theta_4(t) & -\sin\theta_4(t) & 0 & L_3 \\ 0 & 0 & -1 & 0 \\ -\sin\theta_4(t) & -\cos\theta_4(t) & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4) $$

$$ {}^{4}A_5 = \begin{bmatrix} \cos\theta_5(t) & -\sin\theta_5(t) & 0 & L_4 \\ 0 & 0 & -1 & 0 \\ -\sin\theta_5(t) & -\cos\theta_5(t) & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5) $$

where $\theta_1(t)$, $\theta_2(t)$, $\theta_3(t)$, $\theta_4(t)$ and $\theta_5(t)$ are the five joint angles.

B. Equations of Motion

The Lagrange-Euler formulation is introduced in order to describe the relationship between the system dynamics and control inputs.

$$ \vec{\tau} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\vec{\theta}}}\right) - \frac{\partial L}{\partial \vec{\theta}} \quad (6) $$

where $\vec{\tau}$ is the vector of the torques generated by the five joint actuators and L is the lagrangian:

$$ L = \frac{1}{2}\sum_{i=1}^{5}\sum_{j=1}^{i}\sum_{k=1}^{i}\left[Tr\left(U_{ij}J_iU_{ik}^{T}\right)\cdot\dot{\theta}_j\dot{\theta}_k\right] + \sum_{i=1}^{5} m_i\vec{g}\left({}^{0}A_i\cdot{}^{i}\bar{r}_i\right) \quad (7) $$

$\vec{g}$ = [0 0 - 9.81 0] is the gravity acceleration vector and J_i is the inertia matrix as shown in (8).

$$ J_i = \begin{bmatrix} \frac{1}{3}m_iL_i^2 & 0 & 0 & \frac{1}{2}m_iL_i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2}m_iL_i & 0 & 0 & m_i \end{bmatrix} \quad (8) $$

where m_i is the mass and L_i is the length of Link i. The U_{ij} matrices are defined in (9).

$$ U_{ij} = \begin{cases} {}^{0}A_{j-1}Q_j{}^{j-1}A_i & 0 \le j < i \\ 0 & j > i \end{cases}, \quad 0 \le i, j \le 5 \quad (9) $$

Although (6) and (7) may look tidy, it is cautioned that any attempts to derive these two equations with pencil and paper will plague the results with undetectable mistakes when $n > 3$.

According to [5], once $n > 6$, optimal motions modeled without any simplifications are unheard.

This study uses symbolic programming to navigate through this challenging barricade. Equation (6) can be rewritten into such a form:

$$ \vec{\tau}(t) = D(\vec{\theta})\cdot\ddot{\vec{\theta}}(t) + \vec{h}\left(\vec{\theta},\dot{\vec{\theta}}\right) + \vec{c}\left(\vec{\theta}\right), \text{ or} \quad (10) $$

$$ \ddot{\vec{\theta}}(t) = D^{-1}(\vec{\theta})\cdot\left(-\vec{h}\left(\vec{\theta},\dot{\vec{\theta}}\right) - \vec{c}\left(\vec{\theta}\right) + \vec{\tau}(t)\right). \quad (11) $$

III. Optimal Control and Direct Collocation

Once (11) is available, this study applies optimal control to minimize the energy consumption during the operation and adopts direct collocation to solve TPBVP.

A. Optimal Control and the Necessary Conditions

Equation (11) may be rearranged as is shown in (12).

$\vec{x}(t) = \left[\theta_1(t)\ \ \omega_1(t)\ \ \theta_2(t)\ \ \omega_2(t)\ \ \cdots\ \ \theta_5(t)\ \ \omega_5(t)\right]^T$ s the state variable vector. According to [2], the necessary conditions for the optimality to occur are:

$$ \dot{\vec{x}}(t) = \vec{f}_{10\times1}(\vec{x}(t),\vec{\tau}(t),t) \quad (12) $$

$$ \dot{\vec{\lambda}} = -\frac{\partial H}{\partial \vec{x}^*}, \quad (13) $$

$$ 0 = \frac{\partial H}{\partial \vec{\tau}^*}, \quad (14) $$

$$ \vec{\lambda}(t_f) = \left(\frac{\partial \phi}{\partial \vec{x}^*(t_f)}\right), \text{ and} \quad (15) $$

$$ \text{some } x_i(t_0),\ x_j(t_f) \text{ and } \lambda_k(t_f) \text{ are given.} \quad (16) $$

H is the Hamiltonian of the manipulator system $\vec{\tau}^*$ is the optimal control. ϕ is the final state constraint function if any $\vec{\lambda}$ is the Lagrange multiplier vector.

B. Direct Collocation and Nonlinear Programming

The optimal solution is found as long as (12) and (13) can be solved simultaneously. However, due to (15) and (16), the boundary conditions of the differential equations are partially given at the initial and final times and this complicates the procedure in iterating the two differential equation sets. In general, the basis of "indirect method" is to integrate the differential equations, starting at the initial time with available and guessed initial conditions, until the final conditions are satisfied. This method does not guarantee convergence. Instead, this study turns to DCNLP to look for robust approach.

DCNLP discretizes the continuous TPBVP into n segments and hence yields n +1 nodes along its time history as shown in Figure 2. The new variables using DCNLP are denoted as X_i where $\vec{X}_i = \left[\vec{x}(t_i)\ \ \vec{\tau}(t)\right]^T$ Using Hermite interpolation, cubic polynomials are defined for each state on each segment using values of the states at the nodes and the state time derivatives, given by the system differential equations, evaluated at the nodes. DCNLP forces the interpolated derivatives (slopes)

$\vec{x}_c'$ to agree with the differential equations $\vec{f}_c'$ at the center of the segment as shown in (17). See Figure 3.

$$\Delta = f_c - x_c' = f_c + \frac{3}{2\Delta t}(x_i - x_{i+1}) + \frac{1}{4}(f_1 + f_2) \qquad (17)$$

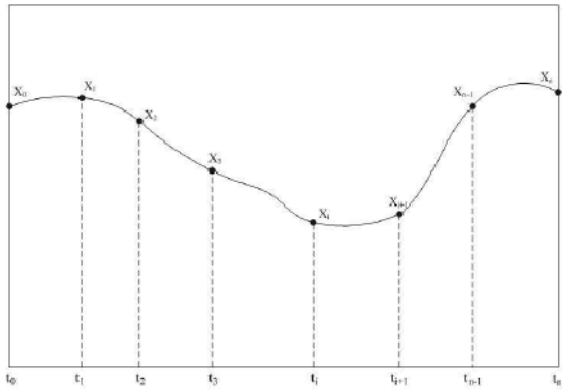


Figure 2. Solution time history discretization.

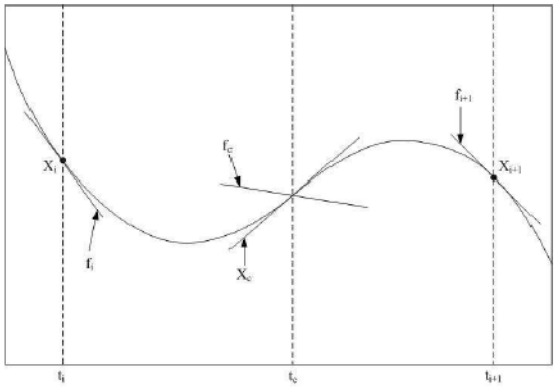


Figure 3. Simpson's system constraint formulation.

As soon as the interpolated slope *xc*′ agrees with the system differential equation at the center of the segment after several iterations, the defect function Δ vanishes. Since there are 10state variables in (11)

and this study segments the trajectory history into 100 pieces, there are 1000 defect functions and they are required to vanish to zero when the iteration converges. The DCNLP method starts the iterations with any arbitrary initial guess of $\vec{x}[k]$s at the $1 + n$ nodes and $\vec{\tau}[k]$s at the centers of each segment. A Jacobian matrix of the defect functions, i.e. $\frac{\partial \vec{\Delta}}{\partial \vec{X}}$ must be supplied before the iteration may proceed. This matrix is often a banded matrix. See Table I.

C. Cost Function

Least-energy control sequence is what this study is after. Therefore the index function is defined as such:

$$J = \frac{1}{2}\sum_{k=1}^{100}\left(\tau_1^2[k] + \tau_2^2[k] + \tau_3^2[k] + \tau_4^2[k] + \tau_5^2[k]\right)\cdot\frac{t_f}{100}. \qquad (18)$$

IV. Simulations and Results

This study applies DCNLP in exploring the behaviors of the five-link robotic arm. In Part A, the model is tested in order to validate if the model is correctly built. In Part B, the same manipulator is compared with a manipulator containing two universal joints. Part C works on Lagrange multipliers.

A. Validating the Model

The configuration of the model is drawn in Fig. 1. Three assumptions are made. There are no frictions in the joints. The links are assumed to be rigid in order to exclude the structural vibrations. The links are slim so that the inertia matrix can be as simple as (8). Since it is an interception, the final velocity of the end-effectors is not specified in Part A. The final time t_f is not specified. The computer has to determine the optimal value of t_f.

In order to intercept the falling target at the final time, the final position vector of the target $\vec{r}_{tgt}(t_f)$ must agree with the final Cartesian coordinate of the end-effectors $\vec{r}_5(t_f)$. The former can be calculated by Newton's law as long as the initial conditions, i.e., $\vec{r}_{tgt}(t_0)$, $\vec{v}_{tgt}(t_0)$ and t_f are known. The information can be found in Table II. $\vec{r}_5(t_f)$ is found by cascading the local coordinate of the end-effectors

TABLE I. Jacobian Matrix of the Constraint and Cost Functions

	$\vec{x}[1]^T_{10\times1}$	$\vec{\tau}[1]^T_{5\times1}$	$\vec{x}[2]^T_{10\times1}$	$\vec{\tau}[2]^T_{5\times1}$	$\vec{x}[3]^T_{10\times1}$	$\vec{\tau}[3]^T_{5\times1}$	…	$\vec{x}[99]^T_{10\times1}$	$\vec{\tau}[99]^T_{5\times1}$	$\vec{x}[100]^T_{10\times1}$	$\vec{\tau}[100]^T_{5\times1}$	$\vec{x}[101]^T_{10\times1}$	t_f
$\vec{f}_{10\times1}[1]$	$\partial\vec{f}[1]/\partial\vec{x}[1]$	$\partial\vec{f}[1]/\partial\vec{\tau}[1]$	$\partial\vec{f}[1]/\partial\vec{x}[2]$	$\partial\vec{f}[1]/\partial\vec{\tau}[2]$									$\partial\vec{f}[1]/\partial t_f$
$\vec{f}_{10\times1}[2]$			$\partial\vec{f}[2]/\partial\vec{x}[2]$	$\partial\vec{f}[2]/\partial\vec{\tau}[2]$	$\partial\vec{f}[2]/\partial\vec{x}[3]$	$\partial\vec{f}[2]/\partial\vec{\tau}[3]$							$\partial\vec{f}[2]/\partial t_f$
$\vec{f}_{10\times1}[3]$					$\partial\vec{f}[3]/\partial\vec{x}[3]$	$\partial\vec{f}[3]/\partial\vec{\tau}[3]$	…						$\partial\vec{f}[3]/\partial t_f$
⋮							⋱						⋮
$\vec{f}_{10\times1}[99]$								$\partial\vec{f}[99]/\partial\vec{x}[99]$	$\partial\vec{f}[99]/\partial\vec{\tau}[99]$	$\partial\vec{f}[99]/\partial\vec{x}[100]$	$\partial\vec{f}[99]/\partial\vec{\tau}[100]$		$\partial\vec{f}[99]/\partial t_f$
$\vec{f}_{10\times1}[100]$										$\partial\vec{f}[100]/\partial\vec{x}[100]$	$\partial\vec{f}[100]/\partial\vec{\tau}[100]$	$\partial\vec{f}[100]/\partial\vec{x}[101]$	$\partial\vec{f}[100]/\partial t_f$
f_{1001}												$\partial f_{1001}/\partial\vec{x}[101]$	$\partial f_{1001}/\partial t_f$
f_{1002}												$\partial f_{1002}/\partial\vec{x}[101]$	$\partial f_{1002}/\partial t_f$
f_{1003}												$\partial f_{1003}/\partial\vec{x}[101]$	$\partial f_{1003}/\partial t_f$
$J = f_{1004}$		$\partial J/\partial\vec{\tau}[1]$		$\partial J/\partial\vec{\tau}[2]$		$\partial J/\partial\vec{\tau}[3]$	…		$\partial J/\partial\vec{\tau}[99]$		$\partial J/\partial\vec{\tau}[100]$		$\partial J/\partial t_f$

${}^5\vec{r}_5 = [1.0 \quad 0.0 \quad 0.0 \quad 1.0]^T$ through the five homogeneous transformation matrices and it yields:

$$\vec{r}_5 = {}^0A_1 \cdot {}^1A_2 \cdot {}^2A_3 \cdot {}^3A_4 \cdot {}^4A_5 \cdot {}^5\vec{r}_5 . \quad (19)$$

Setting $\vec{r}_5(t_f) - \vec{r}_{tgt}(t_f) = \vec{0}$, one creates another three additional constraints, i.e., f_{1001} , f_{1002} , and f_{1003} as listed in Table I. They ensure the end-effectors intercept the target.

TABLE II. BASIC INFORMATION OF THE FIRST TEST CASE.

Link	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
m_i	1.0 kg	1.0 kg	1.0 kg	1.0 kg	1.0 kg
L_i	1.0 m	1.0 m	1.0 m	1.0 m	1.0 m
$\vec{r}_{tgt}(0)$	$[0.0 \quad 0.0 \quad 0.0]^T$ m				
$\vec{v}_{tgt}(0)$	$[1.0 \quad 1.0 \quad 0.0]^T$ m/s				
$\vec{x}[k]$	$[\theta_1 \quad \omega_1 \quad \theta_1 \quad \omega_1 \quad \theta_1 \quad \omega_1 \quad \theta_1 \quad \omega_1 \quad \theta_1 \quad \omega_1]^T$, .				
$\vec{x}[1]$	$[0° \quad 0 \quad 0° \quad 0 \quad 0° \quad 0 \quad 0° \quad 0 \quad 0° \quad 0]^T$				
$\vec{x}[101]$	$[-86.5° \quad 3.7 \quad 105.3° \quad -5.7 \quad -28.8° \quad -3.1 \quad -2.0° \quad 14.2 \quad -23.9° \quad 22.2]^T$				
J	28.633				
t_f	0.81254 s				

θ_i : Joint angle of each link in degrees. $|\theta_i[k]| < \infty$

ω_i : Angular velocity in radians/second. $|\omega_i[k]| < \infty$

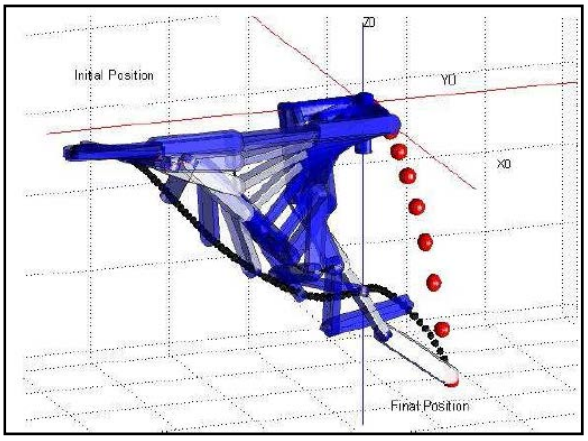


Figure 4. Trajectory of a five-link manipulator intercepting a falling target.

By substituting the final values of the five joint angles from Table II, one obtains $\vec{r}_5 = [x_5(t_f) \quad y_5(t_f) \quad z_5(t_f) \quad 1.0]^T$ from (19). This is the final position of the end-effectors determined by iterations using DCNLP. One finds $z_5(0.8125) = -3.252$m.According to Newton's law, the vertical is placement of a free falling object is -½9.81. t^2_f and it yields –3.238 m. These two numbers are close to each other with an error size of 0.5%. It is caused by the numerical errors accumulated during the iterations. Nonetheless, this model testing ensures that the mathematical model derived in Section II is a correct and reliable one. The trajectory of the interception is shown in Figure 4.

B. Metamorphosing a Universal-joint Manipulator to a Regular Five-link Manipulator

This study takes advantages of the numerical strength of DCNLP to explore the energy efficiency between regular joints and universal joints which are commonly equipped in robotic arms. To produce the universal joints, we may vary the length and mass of each link in such a way:

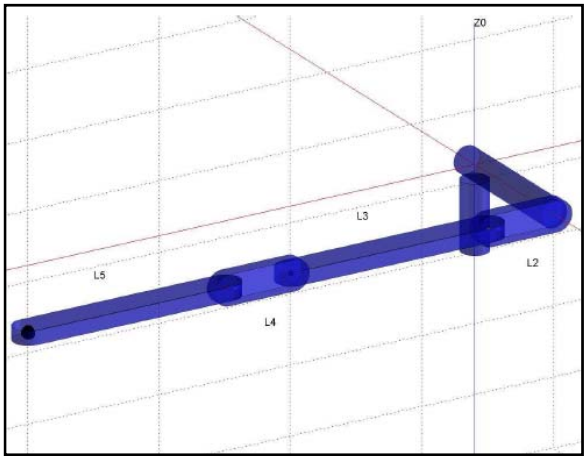


Figure 5. Configuration of the metamorphosing five-link manipulator atα = 0.5

$L_5 = (2 - \alpha)$, $L_4 = \alpha$, $L_3 = (2 - \alpha)$ and $L_2 = \alpha$ where $\alpha \in [0,1]$. Link 5 and Link 4 constitute a universal joint whenα is zero and constitute two regular joints when α is one. See Figure 5. The respective mass of the four links is also distributed in the same way. The end-effectors is also required to come to a full stop at *tf*. This prevents the control algorithm from slamming end-effectors into the target with excessive velocity. Eleven trials are planned by varying the value of α from 0.0 to 1.0, which, in the meantime, metamorphoses a universal joint into two regular joints attached to this simulated five-link robotic arm. Universal joint allows two degree-of-freedom and seemingly may cut down the value of the cost function *J* but the numbers show an ironical story according to the results. In the simulations, the following constraints are applied:

$$|[\theta_i[k]| < 150^o \text{ and} |\tau_i[k] < 40Nt \cdot m . \quad (20)$$

If the state variables and control elements are allowed to vary without proper upper and lower bounds, the motion of the five link manipulator is prone to contain some unanticipated jittering and singularities. The cost and the final time are plotted in Fig.6 and 7. The eleven trajectories are collected together in Fig.8.

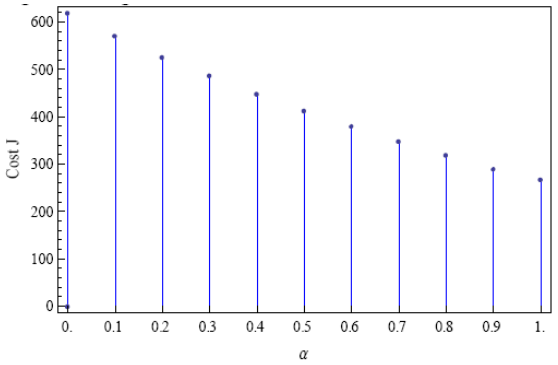


Figure 6. Cost J vs.α

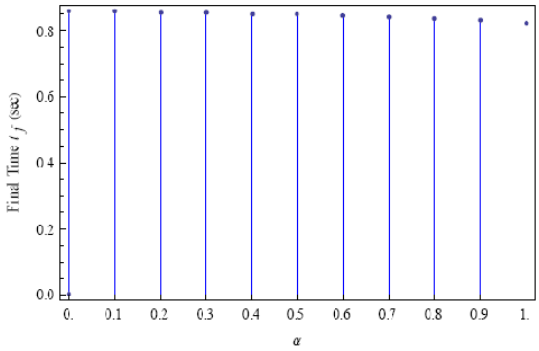


Figure 7. Final time t_f vs.α

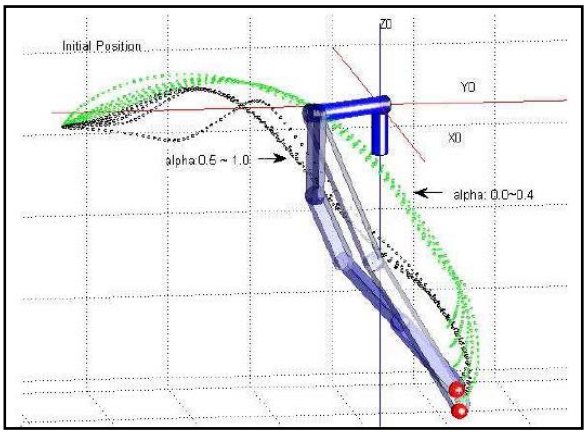


Figure 8. End-effect or trajectories of manipulator metamorphosing from a three-link arm equipped with two universal joints to a regular five-link arm

Figure 8 contains eleven trajectories by tracking the end-effectors of the eleven manipulators. For the sake of clarity, only the two radical manipulator arms are displayed, i.e.,α =1 and $\alpha = 0$. In the figure the eleven trajectories are clearly divided into two families. Due to the over-lengthy links in the family of universal joint, the motions of these manipulators (labeled with alpha 0.0~0.4) do not move as agilely as those labeled with alpha 0.5~1.0.

C. Observing the Lagrange Multipliers

The Lagrange multipliers are plotted against the eleven αcases. Lagrange multiplier represents the sensitivity between constraint and the overall cost function *J*. Large Lagrange multiplier indicates that a tiny violation of the constraint can trigger a large variation and the variation is added onto *J* as an unwanted penalty. A wiggling $\vec{\lambda}_*(k)$ strongly caveats that there are many local minimums populating in the neighborhood of this iteration. Indeed, it takes the universal joint case much longer CPU time to come to convergence. In Figures 10-18, the leftmost curves definitely look more rugged than those rightmost ones.

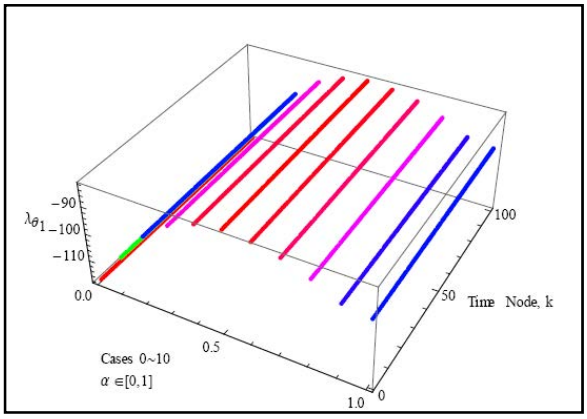


Figure 9. Lagrange multiplier λθ1 vs. α

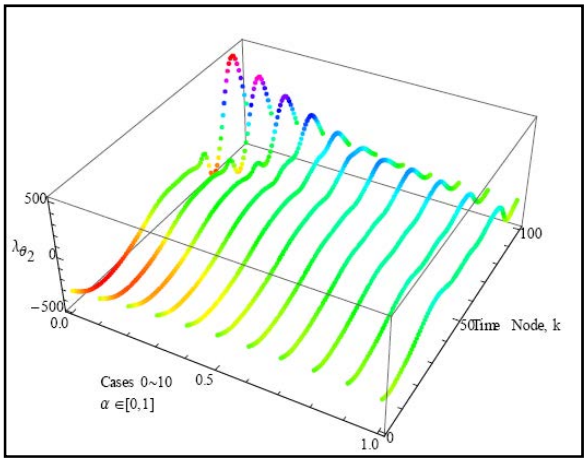


Figure 10. Lagrange multiplier λθ 2 vs. α

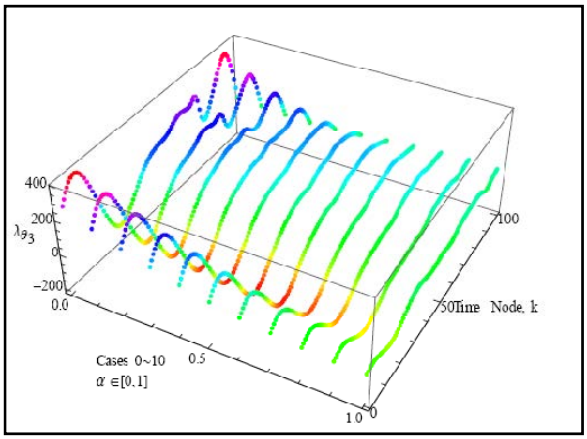


Figure 11. Lagrange multiplier λθ 3 vs. α

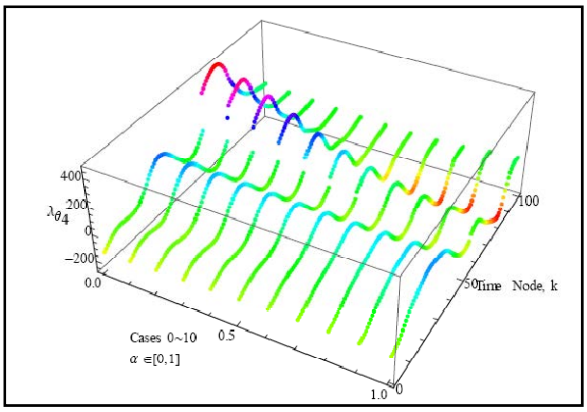


Figure 12. Lagrange multiplier λθ 4 vs. α

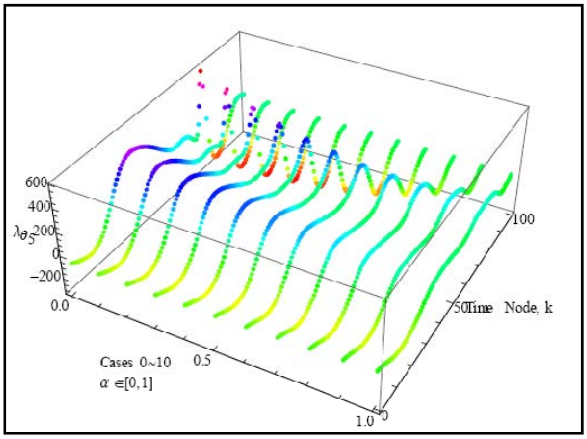


Fig. 13. Lagrange multiplier λθ 5 vs. α

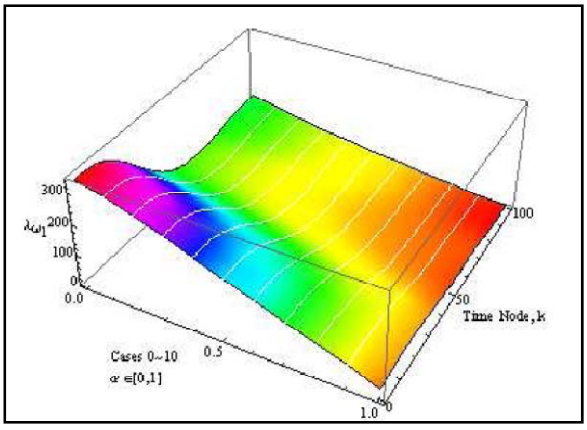


Figure14. Lagrange multiplier λω1 vs. α

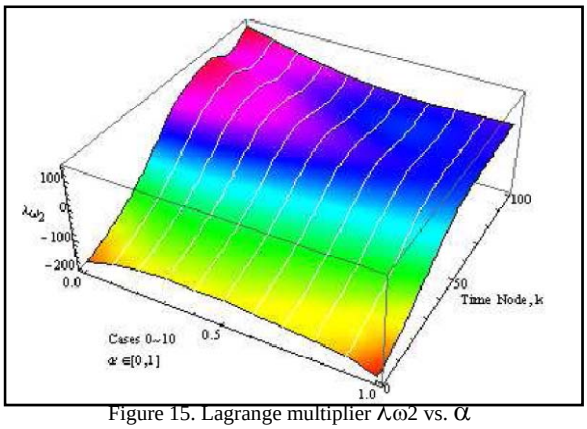


Figure 15. Lagrange multiplier λω2 vs. α

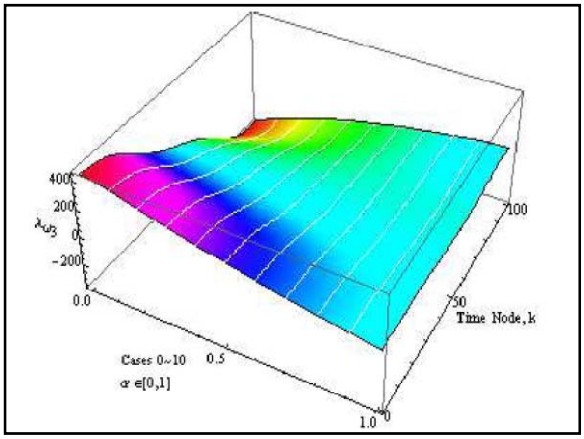


Figure 16. Lagrange multiplier λω3 vs. α

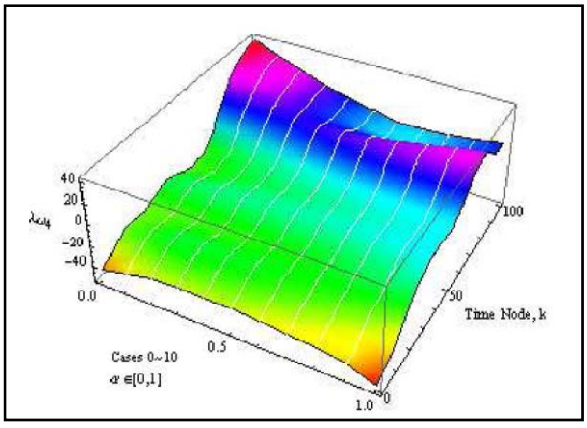


Figure 17. Lagrange multiplier λω4 vs. α

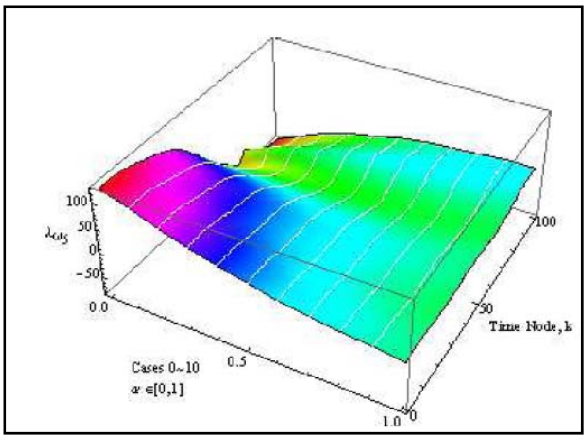


Figure 18. Lagrange multiplier $\lambda\omega 5$ vs. α

V. CONCLUSIONS

In Part A of Section IV, a preliminary validation is carried out before any further explorations are conducted. The difference between the result obtained from DCNLP and the one obtained from Newton's law is only of 0.5%. This justifies an important point that the model is accurately built.

DCNLP is an efficient tool to overrun the two-point boundary-value problems as complicate as this five-link manipulator. See Figure 4. Two-point boundary-value problems used to be roadblocks for many applications and with DCNLP the situation turns out to be breezy and straightforward in solving optimal control problems.

It is believed that a system possesses greater flexibility should yield lower value of cost function. This work metamorphoses a three-link manipulator equipped with two universal joints into a five-link manipulator equipped with five regular joints. In the case of $\alpha = 0$, the advantage of having two universal joints is counteracted by the two over-lengthy links. The value of the cost function of this universal-joint case is $J = 618.4206$. Conversely the regular five-joint manipulator, the case of $\alpha = 1$, is doing a pretty good job in yielding a significantly smaller cost function, $J = 266.8019$. Fig. 6 indicates that universal joint is not necessarily always a good choice in this situation.

The Lagrange multipliers tell stories about local minimum. In Fig. 10~18, the evidences show that a manipulator equipped with universal-joints appears to have a hard time to converge in its iteration since there are many choices of local minimums mined around the solutions found.

ACKNOWLEDGEMENTS

The programming work in this work was done when the first author was at National Penghu University, Penghu, Taiwan in the year of 2008. The purchase of the optimizer software (SNOPT 5.3) was funded by National Science Committee, Taiwan. Project number: NSC—20050524.

REFERENCES

[1] Fu, K. S., Gonzalez, R. C. and Lee, C. S. G., *Robotics: Control, Sensing, Vision, and Intelligence*, McGraw-Hill Book Company, 1987.

[2] Bryson, A. E., Jr. and Ho, Y. *Applied Optimal Control*, Hemisphere Publishing Corp., New York, 1975.

[3] Gill, Philip E., Murray, Walter and Saunders, Michael A., *User's Guide for SNOPT 5.3: a Fortran Package for Large-scale Nonlinear Programming*, Stanford Business Software, Inc., 1998.

[4] Margarita Mediavilla, José Luis González, Juan Carlos Frail and José Ramón Perán, "Reactive approach to on-line path planning for robot manipulators in dynamic environments," in *Robotica*, vol. 20, 2002, pp.375-384.

[5] Park, Jong-keun, "Convergence properties of gradient-based numerical motion-optimizations for manipulator arms amid static or moving obstacles," in *Robotica*, vol. 22, 2004, pp. 649-659.

[6] Sezimaria F. P. Saramago and Marco Ceccarelli, "An optimum robot path planning with payload constraints," in Robotics, vol. 20, 2002, pp.395-404.

[7] Gursel Alici and Bijan Shirinzadeh, "Optimum synthesis of planar parallel manipulators based on kinematic isotropy and force balancing," in *Robotica*, vol. 22, 2004, pp. 97-108.

[8] Dan Zhang, Zhengyi Xu, Chris M. Mechefske and Fengfeng Xi, "Optimum design of parallel kinematic toolheads with genetic algorithms," in *Robotica*, vol. 22, 2004, pp.77-84.

[9] Wei Wang, Yan Zhuang and Wei-min Yun, "Innovative control education using a low cost intelligent robot platform," in *Robotica*, vol.21, 2003, pp. 283-288.

[10] Lyle Noakes and Tomasz Popiel, "Geometry for robot path planning," in *Robotica*, vol. 25, 2007, pp. 691-701.

[11] Ali Meghdari, Davood Naderi and Sohrab Eslami, "Optimal stability of a redundant mobile manipulator via generic algorithm," in *Robotica*, vol.24, 2006, pp. 739-743.

Pectoral Muscle Identification on Mediolateral Oblique View Mammograms

N. Saltanat, M. A. Hossain, and M. S. Alam
School of Computing, Informatics and Media, University of Bradford,Bradford, BD7 1DP, UK.
Saltanat@bradford.ac.uk., M.A.Hossain1@bradford.ac.uk, M.Alam1@bradford.ac.uk.

***Abstract*— In Computer Aided Detection of breast cancer from digital mammogram, elimination of pectoral muscle is a very important and challenging issue. This is because of the fact that pectoral muscle in mediolateral oblique (ML O) mammogram images has common photographic properties with suspicious mass and micro-calcification. Presence of pectoral muscle gives false positive result in automated breast cancer detection. In this paper a new and efficient method using pixel value mapping has been proposed to delineate pectoral muscle region accurately. In this method number of pixel values is reduced by mapping the image first and as a result it becomes easier to identify and segment the pectoral muscle.**

Index Terms- Mammograms; pectoral muscle identification

I. Introduction

Breast cancer is the second largest killer of women all over the world [1]. The reasons behind it are still mysterious. Until the causes of breast cancer are understood, the identification of risk factors for the disease is particularly important. The most effective approach to reduce the mortality from breast cancer is through early detection. Routine screening of women over the age of 40 with X-ray mammography has been demonstrated to contribute to mortality reduction. Studies of the impact of screening have estimated the magnitude of reduction to be between 17-30%[2].

Mammography [2] is a radiographic examination of the breast. The film mammogram can be considered to be a map of optical density that reflects the composition of the breast. From radiographic point of view, the breast consists mainly of two component tissues: fibro glandular tissue and fat. In terms of X-ray attenuation, fat is more radiolucent than fibro glandular tissue; thus, regions of fat appear darker whereas regions of fibro glandular tissue appear relatively brighter on a transilluminated radiograph of the breast. Regions of brightness associated with fibro glandular tissue are referred to as mammographic density and breast cancer risk can be estimated through quantitative analysis of mammographic density [2].

The pectoral muscle in mammograms acts like an additional complexity in automated analysis using CAD systems. The texture of the pectoral muscle may be similar to some abnormalities and thus may cause false positive results in detection of suspicious masses [3]. Elimination of the pectoral muscle can improve automatic breast tissue identification and allow the radiologists to check for the development of cancers in the area overlying the pectoral muscle [4]. Hence, the elimination of pectoral muscle is very important in automated detection of masses in mammographic images.

This paper contains a proposed algorithm to segment the pectoral muscle by mapping the pixel values of the mammogram. This method processes the image first and extracts the region of interest (ROI). Then the pectoral muscle region is identified to segment it from the mammogram.

II. Background Study

There are several existing approaches to detect the mass/density in mammograms. Jahanbin and co-authors [5] detected region of interest (ROI) using speculation filters on mammogram images. The filters were designed based on annotated physical characteristics of spicules. A multiresolution approach was proposed by Dong and co-authors in [6] to classify mammograms using Gabor filters with different frequencies and orientations. Khuzi and co-authors[7] grouped the mammograms of Mammographic Image Analysis Society (MIAS) database into speculated circumscribed or ill-defined. First the image was enhanced followed by segmentation of the ROI using gray level co-occurrence matrices. In [8], the mammogram image was converted into 3D matrix which was then used to convert the image into a binary one. Finally, the glanulometry of tissues was found in an image without segmenting each object. Raman and co-authors [9] chose region grow process for segmentation of a mammographic mass, where a boundary pixel is joined to the current region provided it has the highest gray level among the neighbors of the region. But, whatever may be the technique to detect the mammographic mass; the pectoral muscle always approaches to give false positive results. So, it is very important to eliminate the pectoral muscle before the step to detect the mass

Karssemeijer [10] used Hough transform to detect the pectoral edge where pectoral edge was assumed

ISBN 978- 974-672-556-9

to be oriented straight to a certain direction. Ferrari and co-authors [11] segmented mammograms into skin-air boundary, fibro glandular tissue and a pectoral muscle. The Hough transform linear approximation was refined into a curved pectoral boundary by Yam and co-authors [12]. The pectoral muscle was extracted by region growing in [13]. They all used the Hough transform as their basic technology.

Suckling and co-authors [14] used multiple-linked self organising neural network to segment mammograms into four major components: background, pectoral muscle, fibro glandular region and adipose region. Aylward and coauthors[15] extracted multiple initial points as a first step by using a gradient magnitude ridge traversal algorithm at a small scale, and then resolved the multiple edge definitions via a voting scheme. A semi-automatic method, presented by Saha and co-authors [16], required an input from an operator to locate the pectoral muscle and then delineation was done automatically.

Kwok and co-authors [4] first estimated the pectoral muscle as a straight line and then that estimated line was refined using iterative cliff detection to delineate the pectoral margin more accurately. Finally the pectoral muscle was enclosed in to generate a segmented musk.

In all the reported works one point has not been addressed about the pectoral muscle region; differences in the pectoral region one from another as observed in MIAS database. In some mammograms, the total pectoral muscle region can be segmented into sub-regions. In those cases, the segmentation is done with less accuracy. We propose an algorithm to segment the pectoral muscle by choosing the actual threshold value for masking and for greater accuracy. An iterative method is used to eliminate the extra glandular tissue part if it is added during the masking period. This helps to segment the actual pectoral muscle region.

III. METHODOLOGY

A. Assumptions

The methodology to segment out the pectoral tissue in digital mammogram can be divided into four sections:

1) Mammogram preprocessing.
2) Pixel value mapping and morphological processing.
3) Finding proper threshold value for the preprocessed image and obtaining binary mask.
4) Processing the mask to be fitted for the intended pectoral tissue region and segmenting the pectoral tissue.

This methodology assumes following hypotheses,

1) The pectoral muscle is positioned in the upper left or upper right corner of the mammogram.
2) The grayscale intensity of the pectoral muscle is higher than its surrounding tissues.

The pectoral margin is not a straight line as hypothesized by others [10]-[13] rather its margin that appears on the side of the nipple is mostly concave for upper portion and concave, or convex or a mixture of both for lower portion of the margin. Moreover, the variation of the margin from one mammogram to another becomes so complex that it cannot be generalised with any single geometrical or mathematical model. The other two sides of the pectoral muscle region are almost always straight lines with few exceptions. For the above mentioned reasons, a new approach to segment the pectoral muscle has been proposed here. The method searches the pectoral muscle by region growing, verifies the region if it encapsulates the pectoral muscle exactly, if not then it adjusts the region to encapsulate the desired region.

B. Preprocessing

Before the actual pectoral muscle segmentation process begins it is necessary to prepare the mammograms for that by eliminating all the background objects, i.e., marking stickers that are not part of the breast. To achieve this objective the whole mammogram is converted to binary image from gray image using a very low threshold (=0.1) and the binary image is labeled. As the background of the mammogram is almost black, this binary image contains all the components including the whole breast, marking stickered. Of all the components, the breast has the maximum area and it is easily detectable by computing the component areas. Thus the breast is selected and the image window is cropped to contain only the breast. The orientation of the image is adjusted so that the pectoral muscle region is always positioned in upper left corner.

C. Mapping of image

The smaller image that contains only the breast is mapped to an exponential equation. Let I(x, y) is the pixel value of any coordinate (x, y), then the output pixel value of the mapping would be:

$$I,(x,y) = \frac{e^{I(x,y)}}{40} \text{……..(1)}$$

When the image is converted into binary image using the pixel value of pectoral muscle region as threshold, the output have almost the same shape and position of the pectoral muscle. This is the most important feature of mapping.

D. Morphological operation

To make the regions more distinctive this image is passed through morphological techniques called ‘opening-by-reconstruction’ and ‘closing-by-reconstruction’. Morphological Opening is erosion followed by dilation while ‘opening-by-reconstruction’ is erosion followed by reconstruction.

Again morphological closing is dilation followed by erosion while 'closing-by-reconstruction' is dilation followed by reconstruction. This process sharpens the boundary of the regions and clears the small regions within region that are not useful in the pectoral muscle region segmentation process. The difference of the images before and after morphological operation enables the computer to differentiate the regions more accurately.

IV. Experiments and Results

The mammograms available in mini MIAS database (http://peipa.essex.ac.uk/ipa/pix/mias/) have been used in this work. The mammographic pictures contain four types of components [14]: background, pectoral muscle, fibro glandular region and adipose region. The background is subtracted from others by using a step-by-step process. The image is converted into binary image and the binary image is labeled into major components. And then the largest area of the components is computed. As the whole breast including the pectoral tissue consist the largest area of the labeled components, it is separated from the background easily. A sample of this is shown in Fig.1.

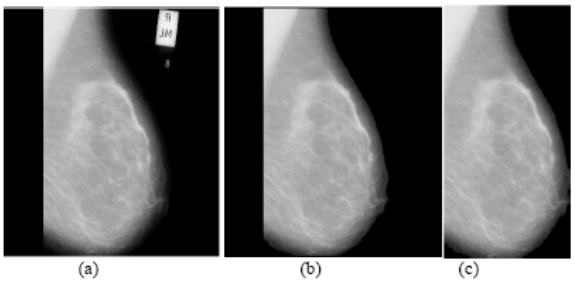


Figure 1. Delineation of the background and medical tags and other unnecessary parts from the mammogram [mdb020.pgm]

Next the image is exponentially mapped according to equation (1). The high magnitude values were converted to the values depicted in Figure 2(a). The output image, Fig.2 (b), has some homogeneous regions where all the pixel values are same. At the same time it divides the image into some regions. A single region encapsulates exactly a major portion of the pectoral muscle and sometime extends to the breast. The regions within pectoral muscle have pixel values such that the region at the upper left corner has the maximum pixel value. The next region has a pixel value lower than that and for the remaining regions the same rule can be applied. When the image is converted into binary image using the pixel value of pectoral muscle region as threshold, the output have almost the same shape and position of the pectoral muscle. This is the most important feature of mapping. After some morphological processing we obtain the picture depicted in Fig.2(c).

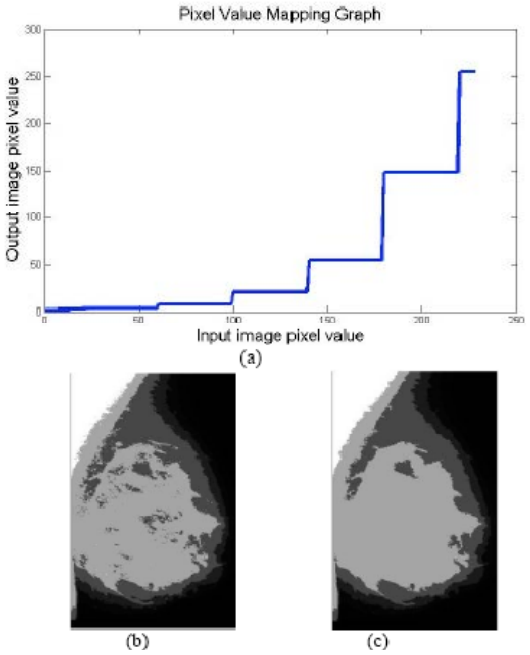


Figure 2. (a). Graph representing the exponential mapping, (b)Mapping the image to get a shape of the pectoral muscle, (c) Output image after the morphological operation[mdb311.pgm]

V. Discussion

From the Figures 2(a)-(c) there are only few pixel values that repeat in the whole image after pixel value mapping and morphological operation, but the boundary of the image is almost intact. At this point the difficulty to get a universal threshold to segment out the pectoral muscle region reduces a great deal. This is because of the existence of only five pixel values as 255, 148, 55, 20 and 7, from which the proper threshold value has to be selected. Regions with other pixel values do not contain much information. The next task is to segment out the pectoral muscle. For this a methodology has been already developed as seen in the flowchart in Fig. 3.

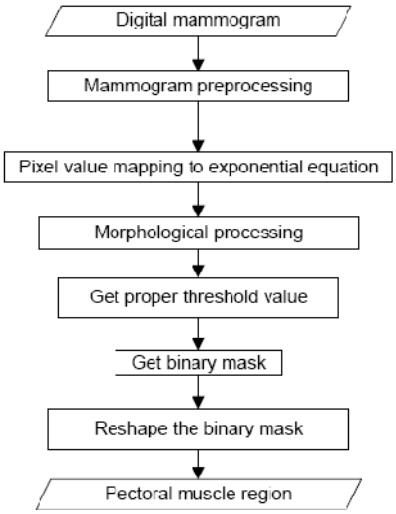


Figure 3. Method to extract the pectoral muscle region from the mammogram

VI. CONCLUSION

In mass detection section, we want to decrease the false positive rate in detecting the mammographic density. In this paper, we have proposed an algorithm to identify and segment the pectoral muscle of mammograms. Here we have processed the image first by eliminating the unwanted objects like marking stickers. And then we have eliminated the background and extracted only the breast portion which is our region of interest. As the first step of our segmentation process, we have mapped the image and reduced number of pixel values. Thus the process of masking the pectoral muscle has been made easier. We have proposed the algorithm to eliminate the pectoral muscle so that it cannot interfere in the detection of exact mammographic mass/density.

The next task is to extract important features from the segmented region of interest of the original mammogram. Here, we want to optimize the process. This optimization can be done by extracting only the most important features with almost 100% accuracy so that they can classify the masses as benign or malignant in the next section.

In classification part, our future plan is to improve not only the accuracy but also required time to evaluate the result. To gain this goal, we want to use Adaptive Neural Fuzzy Inference System (ANFIS) to classify the mammographic density. We think it will optimize the required time with compared to NN or Fuzzy system, because this system is designed with the aid of artificial neural network (ANN) approach combined with the principle of fuzzy logic.

ACKNOWLEDGEMENT

This research has been supported by the EU Erasmus Mundus Project -eLINK (east-west Link for Innovation, Networking and Knowledge exchange) under External Cooperation Window – Asia Regional Call (EM ECW – ref. 149674-EM-1-2008-1-UK-ERAMUNDUS).

REFERENCES

[1] S. H. Landis, T. Murray, S. Bolden and P. A. Wingo, “Cancer Statistics”, CA Cancer Journal 1998, vol.48, pp.6-29.

[2] I. N. Bankman, Quantitative image analysis for estimation of breast cancer risk, “Handbook of Medical Imaging”, San Diego, USA: Academic Press, 2000, pp.323-325.

[3] R. Gupta and P. E. Udrill, “The use of texture analysis to delineate suspicious masses in mammography”, Phys. Med. Biol. , vol.40,pp.835-855, 1995.

[4] S. M. Kwok, R. Chandrasekhar, Y. Attikiouzel, and M. T. Rickard, “Automatic pectoral muscle segmentation on mediolateral oblique view mammograms”, IEEE Trans. On Med. Imag. , vol.23, pp.1129-1140, 2004.

[5] R. Jahanbin, M. P.Sampat, G. S. Muralidhar, G. J. Whitman, A. C.Bovic and M. K. Markey, “Automated Region of interest detection of speculated masses on digital mammograms”, SSIAE 2008, pp.129-132, 2008.

[6] A. Dong and B. Wang, “Feature selection and analysis on mammogram classification”, PacRim 2009, pp.731-735, 2009.

[7] A. M. khuzi, R. Besar, WMD W. Zaki and N. N. Ahmad, “Identification of masses in digital mammogram using gray level cooccurance matrices”, BIIJ 2009, vol.5(3), e17,2009.

[8] R. S. Tomar, T. Singh, S. Wadhwani and S. S. Bhadoria, “Analysis of breast cancer using image processing techniques”, Third UKSimEuropean Symposium on Computer Modeling andSimulation,vol.103.,pp.251-256, 2009.

[9] V. Raman, P. Then and P. Sumari, “Digital mammogram tumor preprocessing segmentation feature extraction and classification” Computer Science and Information Engineering, 2009 WRI World Congress, vol.2, pp.507-511, 2009.

[10] N. Karssemeijer, “Automated classification of parenchymal patterns in mammograms,” Phys. Med. Biol. , vol.43, no.2, pp.365-378, 1998.

[11] R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels and A. F. Frère, “Segmentation of mammograms: Identification of the skin boundary, pectoral muscle, and fibro glandular disc,” in IWDM 2000: Proc. 5thInt. Workshop Digital Mammography, M. J. Yaffe, Ed., Madison, WI,pp.573-579, 2001.

[12] M. Yam, M. Brady, R. Highnam, C. Behrenbruch, R. English, and Y.Kita, “Three-dimensional reconstruction of micro calcification clusters from two mammographic views,” IEEE Trans. Med. Imag., vol.20,pp. 479-489, 2001.

[13] F. Georgsson, “Algorithms and techniques for computer aided mammographic screening,” Ph.D. dissertation, Umeå Univ., Dept.Comput. Sci., Umeå, Sweden, 2001.

[14] J. Suckling, D. R. Dance, E. Moskovic, D. J. Lewis and S. G.Blacker, “Segmentation of mammograms using multiple linked self organizing neural networks,” Med. Phys. , vol.22, no.2, pp.145-152,1995.

[15] S. R. Aylward, B. M. Hemminger and E. D. Pisano, “Mixture modeling for digital mammogram display and analysis,” in Digital Mammography: Nijmegen, 1998, N. Karssemeijer, M. Thijssen, J.Hendriks, and L. van Erning, Eds. Dordrecht, The Netherlands: Kluwer Academic Publishers, Computational Imaging and Vision,vol.13, pp. 305-312, 1998.

[16] P. K. Saha, J. K. Udupa, E. F. Conant, D. P. Chakraborty and D.Sullivan, “Breast tissue density quantification via digitized mammograms,” IEEE Trans. Med. Imag. , vol.20, pp.792-803, 2001.

Definition of a Measurement System Aimed at Murine Model Remote Management

A. Affuso[1], E. Balestrieri[2], M. Ceccarelli[2], P. Daponte[2]
[1]BioGeM – Via Camporeale, 83031 Ariano Irpino (Av), Italy
Ph.: +39 0825881840; Fax: +39 0825881834; E-mail: andrea.affuso@biogem.it
[2] Dept. of Engineering, University of Sannio, Corso Garibaldi 107, 82100, Benevento, Italy
Ph.: +39 0824305817; Fax: +39 0824305840; E-mail: {balestrieri, ceccarelli, daponte}@unisannio.it

***Abstract* - The paper presents the definition of an advanced measurement system in the field of genetic phenotyping enabling (i) to researchers, the remote execution of multiple experiments indifferent laboratories, independently from their location overcoming logistic and resource limits, and (ii) the enhancement of the single Country research capabilities by sharing the experimental data.**

Index terms: measurement system; Murine model; remote management

I. Introduction

For almost a century, mice have been used, more extensively than any other mammalian species, as models of human physiology, disease and toxicity, both in research and testing. The reasons why scientists from a wide range of biomedical fields have gravitated to the murine models are: i) mice are relatively easy to breed and maintain economically in captivity; ii) they have short life-spans, so their long term, post experimental care is normally not an issue; iii) mice are amenable to genetic analysis and manipulation; and iv) they share many physiological features, body systems and developmental and cellular processes with humans [1]. There is, as a result, an enormous amount of information available on the biology, physiology and genetics of the laboratory mouse that have added further impetus to efforts to generate and characterize strains of genetically altered (GA) mice, in order to develop models that improve our understanding of human physiology and disease [1]. However, although the murine models are capable of supplying an irreplaceable and powerful contribution to the biomedical research, they require management systems more and more complex and technologically advanced. Maintenance and breeding technologies of the murine lines aimed at research purposes are continuously evolving also taking more and more into account the well-being of the animals used in the experimental bioscience. Unfortunately, although the facilities are vital to the support of a proper, safe, and humane research effort, they are complex and expensive to build and to operate [2]. For this reason there is a small number of laboratories dealing with the murine model production, especially in those Countries where further logistical problems exist. Aim of the paper is the description of a measurement system aimed at the murine model management, able to integrate laboratory measurement instrument interfaces, automation instruments for the stalling facilities, and systems for biological data processing and analysis in a single environment to make easier industrial research in biotechnology. In this way a stronger linkage between research centers dealing with human disease and those with expertise in animal models can be reached as well as the promotion of the collaborative research. In the following, the murine model development and maintenance both in research and commercial settings are firstly presented highlighting the existing management problems and the current tendency about the research animal facilities requirements. Then, the remote measurement system and telemetry applications in biomedical research are briefly discussed focusing on their advantages, to introduce the innovative measurement system proposed. The main objectives of the measurement system are next presented, along with the chosen web-service oriented architecture. Finally, examples of remote experimental activity are shown, highlighting the interaction among the different subjects involved, (researcher, technician, biologist and software engineers), as well as the remote, automated and the manual operations necessary to carry out the experimental protocol. Some examples of research study cases to be implemented by the proposed system are then proposed, too.

II. Murine Model Development and Production

The development and maintenance of breeding colonies of mice have been an integral part of biomedical research for nearly a century [3].Poorly structured or poorly managed breeding colonies waste both financial resources and animals being counterproductive to research program, too [3].The murine model production can be carried out at noncommercial and commercial level.

Non commercial production of individual stocks, strains, or lines of mice made at research institutions is commonly small in scale. Production may be integrated into the research program in that breeding may be part of an experimental protocol with many

ISBN 978- 974-672-556-9

new lines of mice continually being developed. Only a small number of all animals produced may actually be used in the research program given experimental requirements for a specific genotype, age, sex, or weight of animals. Usually, production is located in one room or area and shares that space with research functions and not uncommonly with differing research programs, many of which do not utilize the same lines [3]. Breeding colony operations are often collectively managed by a team of researchers, with the responsibility for colony maintenance tasks being shared among several individuals [3].

Commercial breeding colonies differ in that there is a large scale production of a limited number of stocks, strains, or lines. In general, a wide range of ages, sexes, or weights are sold because requirements by individual investigators purchasing animals can differ substantially from institution to institution. As with research colonies, multiple strains or lines may be maintained in a single production barrier or area. Depending on the size of the breeding colony, between 2 and10 technicians, with sole responsibility of maintaining the breeding colony are assigned to a single production area or room and have no other responsibilities [3].

Both research and commercial breeding colonies are greatly affected by the costs associated with operating them. Inefficiently managed colonies can overproduce animals that go unused in research program. The cost of operating a breeding colony is driven primarily by labor and the cost of developing and maintaining the production space. Moreover, the complexity of the breeding program will influence costs. The more manipulations, observations, and testing required, the more costly the breeding program and hence the higher the cost per animal utilized will be.

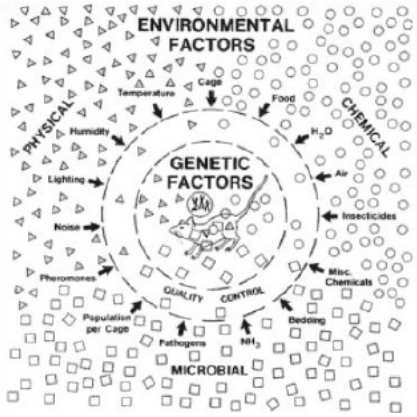


Figure1. Multiple genetic and environmental variables influencing the biological response of the laboratory mouse[4].

Currently, the directions concerning the laboratory animal maintenance provide to carry out facilities able to assure constant environment parameters, rigorously checked and complying with internationally well defined specifications. The biologic response of the laboratory mouse is, in fact, the result of multiple genetic and environmental effects experienced by the animal during the continuum from zygote to death [4] (Fig.1). The environmental factors that affect animal health and well being include cage design and construction materials, bedding material, food and water, available living area, air exchange and air quality, vibration, noise level, electrical and magnetic fields of force, pheromones, microorganisms, parasites and pollutants [5]. The effects of environmental and husbandry factors on animal physiological and psychological function may be subtle, to the extent that effects are not observed, but perturb research results [5]. For these reasons animals used for biomedical research should be kept under conditions that permit as standardized response to experimental parameters as possible. Good control of environmental factors can lead to a significant reduction in the variability of experimental results often seen among different laboratories, or even within the same laboratory [5]. Therefore, today, the equipment used in the animal research facilities, is mainly standardized and consists of shelving purpose made to accommodate cages, in which the air is almost always supplied by circuitries, ventilating singularly each cage, managed by processors allowing the precisely setting of parameters as: temperature, relative humidity, quantity of air put in, quantity of air extracted, number of air changes per hour and partial pressure of the single cage. Moreover, each cage is isolated from the others and everything coming into contact with the animals is decontaminated in different ways, to prevent the infection from spreading into the breeding.

Moreover, currently, the rules concerning animal experimentation recognize and aim at the application of some ethical principles, as those suggested by Russel and Burch, in1959, ratifying principles(*replacement, reduction and refinement*) nowadays universally agreed upon by scientists, academic world and all those who use animal models for biomedical research [6].

Daily procedures to take care of animals are planned, providing for all animal physiological and behavioral needs to assure such well-being levels unimaginable some decades ago. The experimental activities must also be well scheduled, executed by skilled personnel trying to use the minimal needful number of subjects. It is evident that this kind of research requires facilities able to provide a high number of animals genetically modified, as well as their maintenance in controlled conditions, taking care of food and/or medicine administration in agreement with the required protocols. To achieve significant results, is then absolutely necessary to use a large number of resources that in case of genetically modified mouse experiments have to include a system able of allowing the separated colony management.

As quoted above, unfortunately, facilities providing and accommodating tens of thousands mice under microbiological extremely controlled conditions, (Specific Pathogen Free-SPF), are very expensive to build, manage and maintain. For this reason there is a small number of laboratories dealing with the murine model production, especially in countries where further logistical problems exist. The lack of suitable infrastructures, in fact, besides increasing the time and money necessary for the required animal transportation, especially when the research institute asking for the experiment and the laboratory in which the experiment has to be carried out are situated in different countries, also prejudices the research results, since researchers do not have the possibility to know, in every moment, the animal conditions, their history and their characteristics.

III. Remote Measurement Systems in Biomedical Research

At the present time, remote measurement and control systems are implementing in industry, education, R&D, medicine and homeland security technologies [7]. In particular, laboratory activities implemented using the Web as a primary interface medium, have undergone a large development in the last years, mainly due to the rapid evolution of Web technologies. Projects and proposals are now numerous, and involve the activity of many centers and consortia [8].

In the first place, the possibility of remote accessing laboratories is interesting in the field of distance education, where it can be a compensation for the well-recognized difficulties in attaining adequate acquisitions of practical skills and knowledge [8].

In general, a converging interest in the use of laboratories accessible through the Web originates from more situations, when direct involvement in practical activities is difficult or dangerous, owing to potentially harmful or uncomfortable conditions. Moreover, implementing remote measurement laboratories can provide further advantages if practical activities require expensive or rare equipment, or are technically complex and need specialized personnel [8].

A remote laboratory may be shared by a large number of users, located in different places, and at the same time concentrating efforts and resources in a specialized site may boost quality and effectiveness [8].

High-quality distant laboratories may also draw the attention of researchers who need rare instrumentations or carry out particular types of experiments. This might turn are mote laboratory into a centre for specialized services [8] capable of providing remote access to scientific instruments and resources as well as of promoting the collaboration among researchers at remote sites.

In the biomedical research field, new technologies have, in particular, allowed the acquisition of biometric parameters with innovative systems able to transmit data collected at a distance. Today, implantable telemetry has become the gold standard to obtain physiological data from unstressed, free moving animal models. A wide range of vital parameters such as heart rate, blood pressure, EEG, ECG, EMG, temperature, motion activity, and so on, can be monitored in animals [9].

In addition to physiological data, a number of environmental and behavioral data can be simultaneously collected: room conditions such as lighting and temperature, and events such as running wheel, drinking, and feeding activity. External stimuli (light, noise, air jet, etc.) can also be recorded together with the physiological data. And last but not least, valuable video data can be acquired together with the physiological data in order to give a “full picture” of the animals to the investigator [8].

The humane use of laboratory animals requires implementation of the “three Rs” (replacement, reduction, and refinement) of Russell and Burch [6].

The “replacement” alternative is defined by Russell and Burch as any scientific method employing non-sentient material, which may replace methods that use conscious vertebrates. The “reduction” alternative directs the use of experimental designs and statistical methods to support the minimum number of animals required to test the hypotheses at an appropriate power. Finally, the “refinement” alternative focuses upon methods to minimize or eliminate pain and distress and to enhance laboratory animal well-being. Telemetry technology can contribute directly to “reduction” and “refinement” alternatives [10].

IV. Proposed Measurement System

The management of the murine lines, the planning of the experiments needed to characterize them, the acquisition and processing of large amounts of data derived from them, require a structural and organizational support of high quality and technologically advanced. To fulfill these requirements an extremely advanced distributed measurement system aimed at the overcoming of the current problem concerning the lack of a suitable number of animal research facilities as well as the promotion of collaborative research, has been proposed and defined. The system take advantage of telemetry and remote measurement architecture to be efficient and reliable to (i)acquire physiological data from several smart sensors on each animal, (ii) process data acquired, (iii) store and manage the data acquired and the results of calculations with high security,(iv) amend and measure automatically or

semi-automatically environmental conditions of the animals, and (v) carry out some experimental activities within the same laboratory or from abroad thus allowing researchers, working in different locations, institutions or Countries, to accurately monitor some subjects and/or to remotely perform controlled experiments. More in detail, the system has to be capable of obtaining the following objectives:

A. Mouse colony management and execution of basic physiological findings

The management of mouse lines located in a remote facility can be performed by external researchers providing, among others, the possibility to have real-time information about a murine population and to carry out real time measurements on physiological or pathological conditions of subjects under observation.

B. Design and remote control of experimental activities

The procedures and instruments necessary for carrying out experimental protocols with multiple drug treatments and/or monitoring of the treatment groups and/or acquisition of raw data are carried out by the intervention of operators inside the animal house. This has to be made in synergy and under the direct control of external researchers, to prepare the rodents for the experiments and to arrange everything necessary for the acquisition of the experimental data.

C. Acquisition and transmission of biological data using telemetry

The proposed system will be capable of providing several instrumented animal boxes allowing the telemetry of body temperature, heart rate, travel cage, frequency of taking water and/or food, movement and the direct measurement of environmental parameters like air temperature and humidity.

The on-mouse sensors will transmit the measurement raw data to radio-frequency receivers connected to on site data acquisition systems while the temperature and humidity sensors will be connected to the data acquisition systems by means of sensor conditioning circuitries. Each box will be equipped with a network camera transmitting video streams of the animals under examination. According to the specific experimentation the cameras can be sensitive to infrared (IR),or visible radiation and can be chosen to work in very lowlight conditions.

The system will allow the information about present or past murine lines, experimental procedures and the data produced in previous experiments to be shared quickly and easily among the partner laboratories and any external researcher that requires an access. Moreover, the system will allow any entitled researcher from inside or outside the partner laboratories to observe in real time the evolution of each experiment and even the biological and environmental parameters of any animal involved in the experiment as well as to see it by means of the video-camera mounted on its box.

D. Distributive knowledge integrated system for providing data and experimental procedure access

The proposed integrated system can manage the data produced in experiments and provide an easy and immediate access to multiple data sources and related knowledge (semantic annotations), independently from the researchers' physical localization, laboratory, kind of service and data. The large amount of data derived from research and biotechnological development processes, in fact, needs to be:

- Shared among many researchers
- Integrated among different laboratories and platforms
- Analyzed and made explicit to produce new knowledge
- Reused to carry out experiments
- Able to make the users apart from technological details and to mask and/or automate repetitive and boring tasks.

V. Proposed System Architecture

The proposed system is composed by the blocks shown inFig.2. In particular, remote web-services will be provided both for single researcher's activities and for collaborative research programs. The administration services will assure the resource and service access in a simple and integrate way. A wide user community can access to the system clearly and ubiquitously thanks to high standards of security and quality of service, with highly configurable service level agreement, suitable for business transactions.

The system has been designed to maximize the number of automatic procedures dealing with the murine model management during the experimental phase and also before and after this phase, during the maintenance periods preceding and succeeding the experiments. In this way, all the experimental phases can be traceable and the human contribution can be reduced along with the related errors.

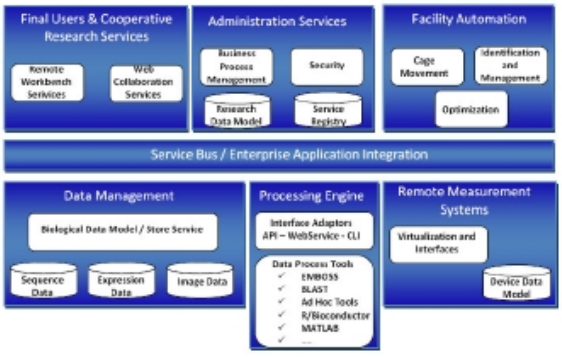


Figure 2. Proposed system block diagram.

Moreover, the human intervention reduction avoid a potential infection source of transmissible diseases also reducing the stress due to sudden and unexpected noises often caused by the operators in the facility areas. It is important to highlight, finally, that systems able to identify, visualize and move mouse cages to the procedural areas, optimize the personnel utilization, limiting at the same time the risk of contracting pathologies. The final users, collaborative research, administration and facility automation services will be integrated to the data management, the processing engine and the remote measurement system. In particular, the proposed system will integrate a collection of data analysis services useful in the experimentation and utilization system phases for the biomedical research processes.

The process engine will allow the design, sharing and execution of scientific workflows, while the remote measurement system will allow the execution of experiments providing a cage-mouse instrument system able to carry out environmental and physiological measurement thanks to wireless or cable and telemetry implanted sensors.

The proposed remote laboratory system will operate as a centre for specialized services by means of a service oriented architecture (SOA). SOA is essentially a collection of services. These services communicate with each other, and the communication can involve either simple data passing or two or more services coordinating some activities. Remote experiments are clear examples of services.

One of the most important aspects is that SOA is not tied to any specific technology; indeed, it can be implemented using a wide range of protocols (COMM, XML-RPC, CORBA, JINI and REST). More concretely, this paradigm is the most suitable for carrying out interoperability between different kinds of services [11].

Nowadays, new tendencies in SOA focus on the use of web service standards with the aim of performing interoperability [10]. One of the major advantages of this approach is the use of XML-based technologies, which provide protocol and language independence. A key component of the SOA infrastructure supporting its service interaction capabilities, and providing the integrated communication, messaging, and event infrastructure to enable them is the Enterprise Service Bus (ESB).

The proposed system will be based on a web service oriented architecture supported by an ESB. In particular, thanks to multichannel interface based adapters the platform can be accessed both by data analysis applications and data acquisition devices as for example imaging and cellular analysis instruments, as well as by generic clients as webpage and mobile devices. Moreover, the system can allow the direct interaction among researchers thanks to typical social networking and collaborative research procedures. The system experimental data management will be integrated with the knowledge included in public databases. The adopted architecture will be capable of promoting the service and process cooperation as well as the collaboration among different research centers, single laboratories and R&D teams. The system will be also capable of designing and providing a set of protocols, processes and data analysis disguised as scientific workflow to researchers and technicians for the distributed system service orchestration. Service semantic annotation methods specifically related to bioinformatics will be used to make easier the phase aimed at the data knowledge representation in explicit form. The portal framework of the system will be structured by means of a distributed architecture, represented by an access portal (fronted layer)and a sequence of distributed remote agents (backend layer)allowing the portal to be interface with heterogeneous distributed situations. The web-focused approach will allow the users to access to whatever service from whatever place, at any moment with every interaction instrument, mobile terminals included. The portal will operate as a real framework putting in communication on the one hand the resources(storage systems, processing, instrumentation) available in the remote laboratory and on the other hand, the user community, not experienced in computer science and needing to analyze data and information in a simple, automatic and intuitive way. Distinctive characteristic of the system is its modularity. The framework will provide a friendly web-user interface integrating basic features as authentication systems (single sign-on), authorization, data management, and advanced features as e-collaboration, 3D remote visualization and accounting.

VI. Remote Experimental Activity

Examples of experimental procedures and data capture that can be performed by the proposed measurement system will include:

- Administration of pharmacological substances or chemicals needed to produce models of disease.
- Any treatment or conditioning of the subjects required to be observed by the researcher staff of the animal house before or during an experiment.
- Possible surgical preparation of the subjects to be submitted to experiment.
- Measurements of physiological parameters (e.g. bodyweight, blood pressure, heart rate, and so on)
- Behavioral analysis.
- Hematological analysis.

A feature set, one or more measurement systems, and real time control for the acquisition of raw or

partially processed data obtained as a result of experimental procedures, performed by the animal facilities skilled personnel, will be required. Moreover, in order to allow optimal management of the experimental procedures from abroad, laboratories will have to be equipped with video monitoring systems accessible, in a safe way, from the remote location of the researcher by means of Internet.

The execution of the remote experimental activity through the proposed measurement system will require the interaction of different subjects, since not all the operation can be conducted in a remote and automatic way (Fig.3). In general, the experimental activity starts by a researcher's request. He/she will have to specify information as the gene modification needed, the number of mice, the requested observation time, the kind of feeding, the pharmacological treatment, and possible intermediate phases. During the experiment the researcher can monitor some physiological parameters as the weight and the movement activity of the mice, and automatically weigh-out the water and food daily quantity, as well as the drug quantity to be inoculated.

The mice preparation and movement operations are manual activities to be performed by a technician. Other manual activities are the surgical operations that must be conducted by a biologist. Finally, a software engineer must supervise the whole structure of the measurement system.

Examples of biomedical research study cases that can be conducted by means of the proposed measurement system are discussed in the following.

A. Evaluation of the postoperative recovery in rats

Postoperative period is a stressful situation during which the animal experiences pain and its ability to perform its normal behavioral repertoire is impaired. Further more recent research has identified several specific pain related behaviors in rats and mice. Under these circumstances both the physical and social environment may affect the way in which the animal copes with this stressful situation which in turn may influence postoperative recovery. Personal observations suggest that socially housed rats need less time to fully recover from telemetry implant surgical procedure. Physical environment like lighting and high temperature may also influence the recovery period as well as the expression of behavior of the animal. In an effort to clarify the animal behavior under different circumstances like social enrichment and different light conditions, radio telemetry transmitters have to be implanted in rats and their vital functions should be monitored such as HR, ECG waveforms, Heart Rate Variability (HRV) and activity of the animal postoperatively and for two consecutive hours. Monitoring of the animals should be performed also in the 3rd and 5th post surgical day.

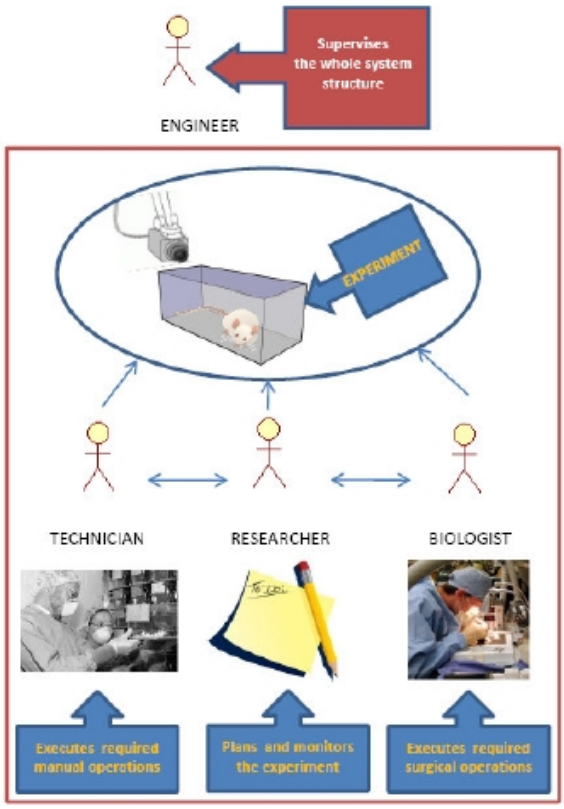


Figure 3. Experimental interaction processes.

B. Remote-screening of phenotype analysis in mice with thyroid related genetic mutations

It has been reported that genetic thyroid disorders in humans is generated from a multiple gene interaction and the same situation seems to exist in mice. In order to analyze the effects of thyroid disorders on clinical conditions objective criteria are needed to identify any phenotypic alterations in behavioral and/or physiological traits.

Behavioral changes that should be analyzed are: activity in cage, frequency of watering and of feeding; sleep/wake cycles. Further more in some cases specific tests in open field, and experiments for memory and learning tests should be carried out.

Physiological parameters to analyze are: body weight and growth curve; body temperature; breathing rate; blood pressure and heart frequency.

C. Analysis of Alzheimer Disease animal model using implantable telemetry

Alzheimer's disease (AD) is the most common neurodegenerative affliction of the elderly, resenting with progressive memory loss and dementia and terminating with death. There have been significant advances in understanding the biology and subsequent diagnosis of AD; however, the furious pace of research has not yet translated into a disease modifying treatment. In the past decade, AD research

has been fundamentally transformed by the development of genetically modified animal models of amyloid-driven neurodegeneration.

These important in vivo models not only replicate some of the hallmark pathology of the disease, such as plaque-like amyloid accumulations and astrocytic inflammation, but also some of the cognitive impairments relevant to AD. In order to carry out a behavioral characterization of the different transgenic models of AD, it is necessary to implant radio telemetry transmitters in mice and to monitor their neural and physical activity. In particular, various physiologic parameters, including temperature, heart rate, ECG, blood pressure, physical activity will be necessary.

VII. CONCLUSIONS

Scientific research is one of the most powerful opportunities to improve and enhance the social and economical development of every Country. The paper has presented the definition of a new high technological service oriented support platform, based on telemetry, capable to enhance the capacity of scientists, involved in preclinical experimental activities, to carry out experiments in facilities located in different Countries and quickly retrieve experimental data, in a cost effective way by using resource sharing and remote access technologies. Moreover, the proposed system architecture can promote the scientific collaboration among researcher coming from different organizations and Countries to obtain better and faster results. This, together with the reduction of the number of animals needed to carry out experimental data, can realize the alignment to the better international ethical standards in theuse of laboratory animals, too.

REFERENCES

[1] N.Bhogal, R.Combes, "The relevance of genetically altered mouse models of human disease", *Alternatives to Laboratory Animals*, ATLA, vol.34, No.4, August 2006, pp. 429-454.

[2] S.Stark, J.Petitto, S.Darr, "Animal research facility", National Institute of Building Sciences, 2009, available athttp://www.wbdg.org/design/animal_research.php

[3] J.G.Fox, S.Barthold, M.Davisson, C.E.Newcomer, F.W.Quimby, A.Smith, "The mouse in biomedical research: normative biology, husbandry, and models", Academic Press, 2006.

[4] M.Suckow, S.Weisbroth, C.Franklin, "The laboratory rat", Academic Press, 2005.

[5] J.Hessler, N.Lehner, "Planning and designing research animal facilities", Academic Press, 2008.

[6] W.M.S.Russell, R.L.Burch, "The Principles of Humane Experimental Technique", London: Methuen & Co., 1959.

[7] V.Hrusha, R.Kochan, Y.Kurylyak, O.Osolinskiy, "Development of measurement system with remote access based on Internet", Proc. Of IEEE Int. Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 6-8 Sept. 2007, Dortmund, Germany, pp.126-128.

[8] L.Fiore, G.Ratti, "Remote laboratory and animal behaviour: An interactive open field system", *Computers & Education*, vol.49, No.4, Dec. 2007, pp.1299-1307.

[9] E.Rieux, "Trends in telemetry monitoring: more data and improved animal welfare", Proc. of Measuring Behavior 2008, Maastricht, The Netherlands, August 26-29, 2008, page 363.

[10] K.Kramer, L.B.Kinter, "Evaluation and applications of radio telemetry in small laboratory animals", *Physiological Genomics*, vol.13, No.3, May2003, pp.197-205.

[11] M.A.Martınez-Carreras, A.F.Gomez-Skarmeta, "Towards interoperability: a wrapper model for integrating remote laboratories in a collaborative discovery learning environment", *Software Practice and Experience*, vol.38, No.15, Dec. 2008, pp.1601-1620.

Study of Diurnal Variation of Global Solar Irradiance Using Pyranometer on Horizontal Surface and Investigation of the Trend of Irradiance at the Solar Noon at Kathmandu, Nepal

K.N.Poudya[1], P.Daponte[2], L.De Vito[2], B.K.Bhattarai[1], B.K.Sapkota[1]
[1] Institute of Engineering, Tribhuvan University, Kathmandu, Nepal
[2] University of Sannio, Benevento, Italy
E-mail: khem100@hotmail.com

***Abstract* - In this paper the results of a measurement investigation on global solar irradiance at Kathmandu (27.7°N, 85.5°E, and 1350.0m) since July 2008 is analyzed. It has been found that the intensity of global solar irradiation is highly significant even during winter season at Kathmandu. The solar irradiance mainly depends upon latitude, zenith angle, aerosols, clouds, altitude, ozone, wind speed, season, local environmental conditions, precipitation, and ambient temperature. It is concluded that there is high level of fluctuation of the measured data of diurnal as well as solar noon value of global solar irradiance depending on local atmospheric conditions. The long term study data will help to develop the modeling of the solar energy estimation for the application of solar energy in Nepal.**

Index terms: diurnal variation; global solar irradiation; Pyranometer; Nepal

I. INTRODUCTION

Nepal is the second richest nations in hydropower after Brazil. However, due to various constraints the rural sector of Nepal is still deprived of a national electrical grid. Renewable energy including the solar energy resources are the most viable and only the option of accessible sources to electrify in the rural sectors. Presently, this source provides electricity to about more than 7% of the rural population [1]. Only the 40% of total population of the country have excess of electricity and rest of them are used only kerosene lamp and wood fire. Thus, the energy demand is fulfilled from the wood fuel, agriculture residue, petroleum product, coal, and only 2% covers the hydro electricity and 1% is taking from the renewable energy resources. In spite of this, there is more than 12 hours load shedding at the dry summer season. As a consequence, it is crucial time to carry out research work in solar energy and other renewable energy resources.

Since 20 years earlier solar photovoltaic systems, solar photovoltaic water pumping systems, solar thermal water heaters are some of the solar technologies that have been used in the country. Solar thermal space heating and illumination in buildings is already in practice. Total capacity of solar photovoltaic (PV) system in Nepal has reached 4.3 MW by the end of December 2007. In addition, there are several solar thermal technologies, such as solar water heating systems; solar dryer, solar cookers are being installed in different parts of Nepal. There is possibility of large-scale development in solar as well as other renewable energy resources technologies in near future in the country. At least government as well as academic institutions should explore the solar energy potential, their data, solar map and solar radiation model. Definitely this approach will give the possibility to plan and implement what type of solar plant development is feasible, such as concentrating solar power, grid connected PV system, water heating PV system, and Remote PV for lighting in particular area [2].

The energy of electromagnetic radiation increases with decrease in wavelength. After the 1980s, academic and public health awareness of global solar radiation and solar ultraviolet radiation both have increased significantly because of the energy crisis as well as ability to harm in human and plants. The study of global solar radiation is dramatically increased after the discovery of declines in global concentration of thin layer of ozone over the Antarctic, climate change, global warming and energy crisis since 1973 in the world. More effort has been made to study the effects of increased or decreased solar radiation including ultraviolet radiation on the surface of the earth [3].

Nepal is located in 27.5^0N latitude which is closer to the solar zone belt. Thus, the average solar radiation is 3.6-6.2 kWh/m^2/day and the sun shines for about 300 days in a year [4].

There was no continuous measurement as well as detail study about the global solar radiation and other types of radiation in this region. Thus, it is very urgent to study the global solar radiation over the long term effects of different factors on solar

irradiance. The data of total radiation and ultraviolet radiation can be made available to the public to increase public awareness so that people can cope the merit and demerits of different type of energetic solar radiations and they will understand the proper applications and as well as precautions.

There are few research works done in the field of solar radiation, and there was not continuous measuring of global solar radiation and ultraviolet radiation in Nepal hence this research will help to investigate the trend of factors effecting global solar radiation as well as ultraviolet radiation in Nepal [5].

II. SOLAR MAPS BY SATELLITE

The difficult topography and scattered settlement solar energy can be one of the best alternative energy solutions for the remote area of Nepal. So it is urgent to develop the solar map on the basis of the validation of satellite data as well as ground measurement data. One of the main objectives of this paper is to find out the ground measurement irradiance data and to find out its deviation from satellite data for the confirmation of accurate solar radiation potential.

The global solar radiation is monitored with different sites of ground and satellite based instruments. In addition, different models can be used to estimate the solar irradiance but the topography of Nepal is unique. Besides this, there is significant climatic variation in every 200 m. altitude due to the variation of temperature, humidity and air pressure. Besides this, the largest inconsistent may occur on clear winter days when it is difficult to distinguish between frost or snow on the ground and low lying fog or clouds by the satellite. Hence it should be prepared the solar map and on the basis of solar map to develop own model of global solar radiation for Nepal.

The spectral irradiance at the surface of the earth depends on different factors such as: radiation from the sun, optical properties of the atmosphere, position of the sun given by the solar zenith angle, mean sun earth distance and the reflecting property of the surface. Among the optical properties of the atmosphere, clouds and aerosols are the complicated factors affecting solar irradiance, which are still not very well understood.

There are many network stations to measure the global solar radiation established in different parts of the world for example Europe, United States of America , Japan, Australia, however, observational data on surface is lacking in the developing countries like Nepal in the ground which is significant to generate the solar energy.

The global solar radiation map is essential on a horizontal surface for use in power systems reliability calculations. The importance of an appropriate model for global solar radiation has increased with the increased use of photovoltaic power generation. The global solar radiation phenomenon is not only regular but also a random behavior. It varies with the state of the atmosphere. The astronomic effects can easily be described mathematical with only some minor simplifications but the atmospheric effects are more complicated to describe [5]. The developing model should be used for any geographical location of Nepal. This is especially useful for developing countries where long-term solar radiation measurement can be hard to obtain. In other hands the satellite data are always available without any expensive cost. For the accurate calculations higher resolution may be needed. The satellite data should be compared with ground-based measurement data and finally deviation is found after the analysis of data the solar map is developed. On the basis of solar map, physical and meteorological parameters the global solar radiation model would be prepared which can be utilized to harvest the solar energy in different geographical locations of Nepal.

III. THEORY

The major depending factors on global solar irradiance are latitude, zenith angle, aerosols, clouds, altitude, ozone layer, wind speed, season, and time of the day, longitude, vertical distribution of atmosphere, inclination, orientation, local environmental conditions, precipitation, and ambient temperature [6].

The intensity of solar radiation falling on a horizontal surface on earth decreases with increases in solar zenith angle. This becomes the incident radiation is proportional to cosine of angle between the beam radiations and normal to the surface leading to cosine effect, since relative path length of the beam increases when sun is lower in the sky [7].

Total solar radiation entering the top of atmosphere is called extraterrestrial solar spectrum which is not the same as observed on the surface of the earth. The different gas molecules which occur at the atmosphere interact with solar radiation and the radiated solar radiation reduced by the absorption, reflection and scattering. Generally, the irradiative energy is inversely proportional to the wavelength. Thus the radiation with lower wavelength is more energetic in terms of solar radiation potential and is mostly responsible for the harmful effects also.

Spectral radiance I is defined as the radiant energy in a given solid angle $d\omega$ per unit time dt. The term spectral radiance is the fundamental radiometric quantity which deals about the frequency, position, distribution, direction and radiation field. It can be expressed as [8]:

$$I_\lambda = \frac{d^4 E}{\cos\theta} \frac{1}{dA} \frac{1}{dt} \frac{1}{d\omega} \frac{1}{d\lambda}$$

Where, I_λ is spectral radiance, $d\omega$ is solid angle, *dA* is area, $d\lambda$ is wavelength of radiated wave, and d^4E is total energy flow within a solid angle.

The projection of surface element *dA* into the normal plane with Ω will be the net energy which passages through the area *dA* in time interval *t+dt* and wavelength interval $\lambda + d\lambda$.

Similarly, the spectral irradiance d^3E is defined as

$dA.n.d\Omega = dA\,Cos\theta$, where, n is unit vector. In terms of mathematical relation is

$$F_\lambda = \frac{d^3E}{(\,dA\,.dt\,.d\lambda\,)}$$, it is measured in watt per meter square manometers ($w/m^2.nm$)

Hence, the global irradiance is related to radiance

$$d^3E = \int d^4E = (I_\lambda)\,Cos\theta\,dA\,dt\,d\omega\,d\lambda$$

$$d^3E = dA\,dt\,d\lambda \int (I_\lambda)\,Cos\theta\,d\omega$$

Then, F_λ may be rewritten as

$$F_\lambda = \int_{2\pi} (I_\lambda)\,Cos\theta\,d\omega$$

On the basis of the above relation the pyranometer measures the global solar radiation directly in watt per meter square so no further mathematical calculation is required.

IV. Instrument and Methods

The portable and calibrated pyranometer model solar 130, Haenni is used to measure the broadband solar irradiance on a planner surface and shadow less open area. There is a sensor that is designed to measure the global solar radiation flux density in watt per meter square from a field of view of 180^0.

This is our preliminary research work. This research takes more than two years. The physical structure of Nepal is unique. There is a drastic climatic variation in every 200m altitude. Due to the zigzag shape of high mountains even the satellite may not cope actual data of global solar radiation thus the ground level data is essential to compare with satellite data for the confirmation of actual solar radiation potential in Nepal.

One of our main objectives of this research work is to develop a model of global solar radiation at near the ground level. For to develop the model we will take ambient temperature, wind speed, wind direction, latitude, longitude, altitude, sunshine hour, precipitation, relative humidity and we should generate the ground measurement data and these data will validate with NASA satellite data and Ev-K2-CNR, (high altitude measurement) data then finally we will come to develop a model of global solar radiation potential. After developing the model we can be utilized to generate the authentic data at the different parts of the country. At the same time, the model can be used for estimation of hourly and daily global solar radiation as well as it can be used to estimate how much amount of solar energy can be obtained at the allocated area. The developing model will be milestone for the solar PV rural electrification where about 93% of people are live without any means of electricity in Nepal.

V. Data and Results

In this research work the calibrated pyranometer is used to measure the global solar radiation at the horizontal surface of open and shadow less area of Department of Physics Institute of Engineering Pulchowk.

Fig.1. reports irradiance versus time trend. This graph indicates that the diurnal variation of irradiance is similar to the trend line as well as the value of solarradiance is sufficient at the winter season to generate the solar energy. This is a good symbol of potential of energy, too. In addition, the value of R^2 is 0.886 which is meaningful with respect to the measured data. Similarly, in Figure2, we found significance result in terms of trend line as well as R^2 thus measured data of solar irradiance at winter season is also perfectly useful to harvest solar energy at Kathmandu. In this graph also there is high level of fluctuation of total solar radiation; it is happened due to the local weather conditions.

In this research, the global solar irradiance at the solar noon time was measured. This data are very much crucial to analyze the sun position and solar zenith angle. In the case of July at Figure 3a, the global solar irradiance gradually increases with increase in date which is clearly shown in trend line. It means that the solar irradiance increases with decrease in solar zenith angle. But in the case of Figure 3b, and Figure 3c, the value of solar irradiance gradually decreases with increase in date. It means that the solar zenith angle increases with decrease in global solar irradiance. In another words the sun is gradually farther distance according to increase in winter date. So the intensity of solar radiation decreases with increase in distance as well as increase in solar zenith angle.

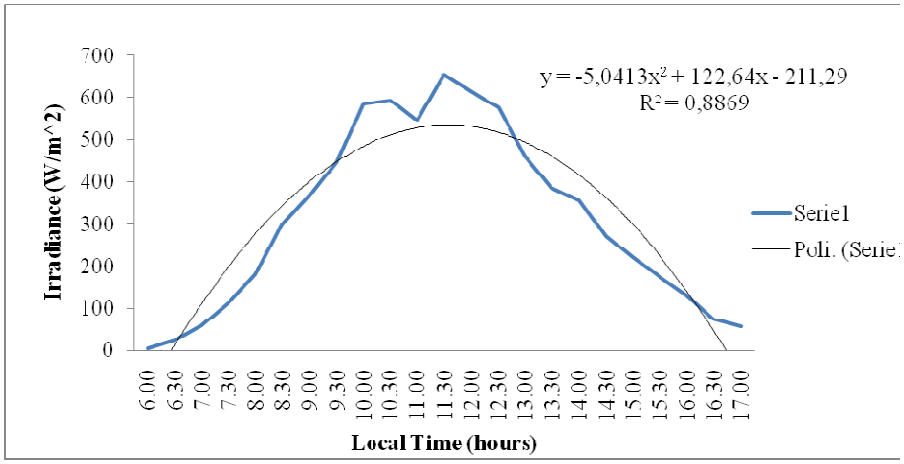


Figure 1. The global solar radiation on September, 2008 at Kathmandu, Nepal.

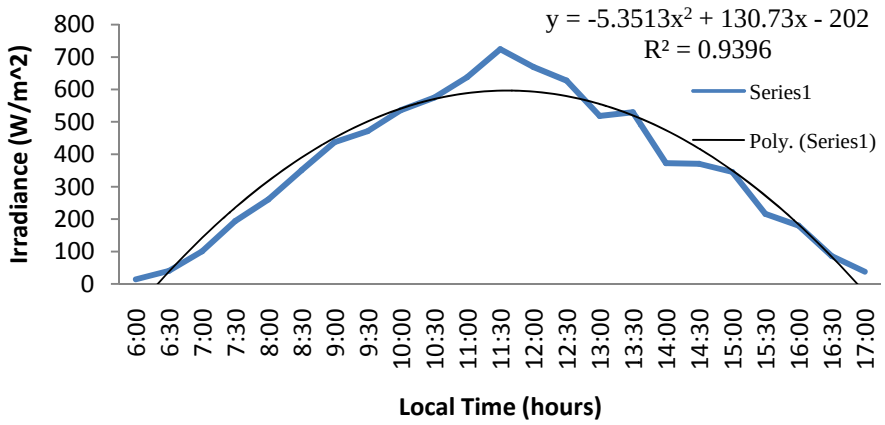


Figure 2. The global solar radiation on October, 2008 at Kathmandu, Nepal.

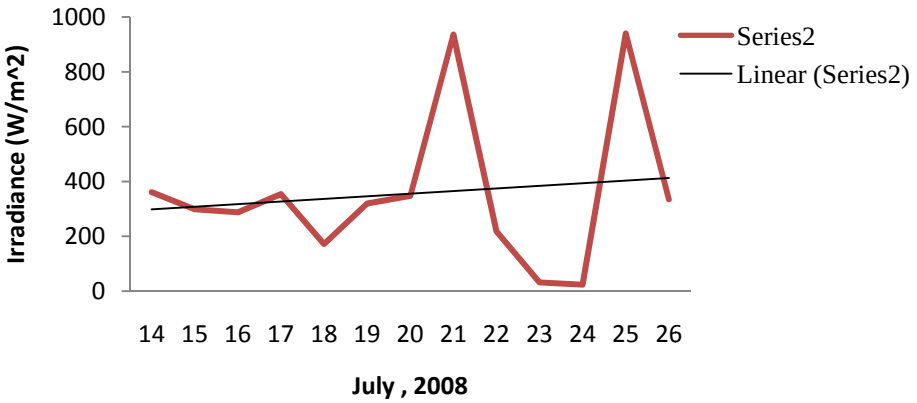


Figure3a. The global solar irradiance at solar noon on July, 2008 at Kathmandu, Nepal

ISBN 978- 974-672-556-9

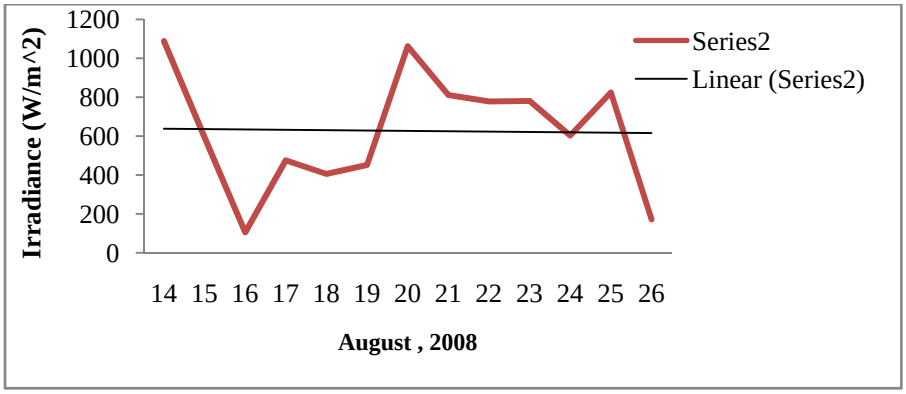


Figure3b. The global solar irradiance at solar noon on August, 2008 at Kathmandu, Nepal.

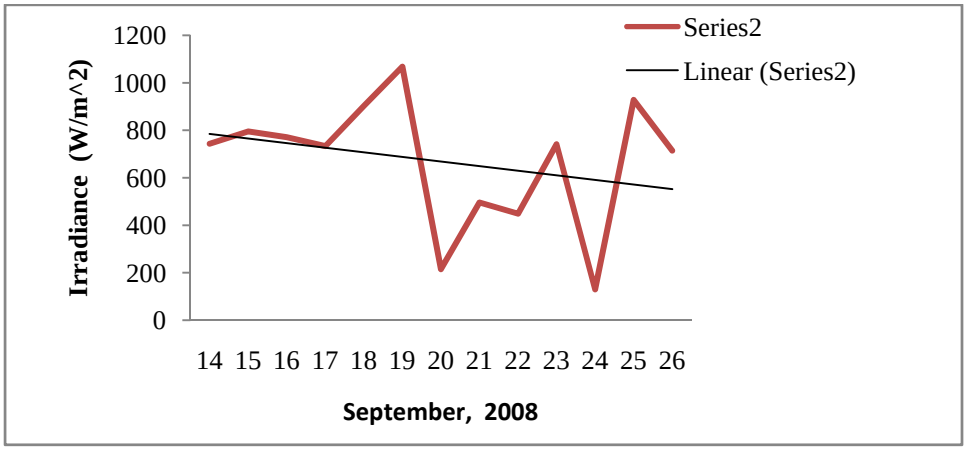


Figure 3c. The global solar irradiance at solar noon on September, 2008 at Kathmandu, Nepal.

VI. Conclusions

One of the main objectives of the research carried out was to investigate the diurnal variation of global solar radiation, in this regards, Figure 1, and Figure 2, show that the global solar radiation is greater than 500W/m^2 even at the winter season.. Similarly, there is high level of fluctuation in irradiance depending on local atmospheric conditions. The experimental results highlighted that the one of the key affecting factor of global solar radiation is cloud. But it is difficult to say what type of cloud and other parameters leads to affects to change the in solar radiation. For the finding of such type of completed condition needs to further research work.

Figure 3a, gives the clear information about the variation of global solar radiation at the solar noon and the irradiance gradually increases with increase in day of the month July but in the case of Figure 3b, and 3c, both show that the trend of solar irradiance gradually decreases with increase in day. This means that the global solar irradiance at solar noon must be decreased with increase in winter season as well as solar zenith angle which are experimentally verified.

It is concluded that the of global solar radiation potential at Kathmandu is sufficient for the purpose of the solar energy technology promotion in coming days which will be milestone to manage the energy crisis at Kathmandu. In addition that, this paper represents a new approach of the problem permitting the evaluation of the hourly average global intensities on a horizontal surface in terms of maximum hourly average, time of day and the day length. Finally, it will help for modeling to project the estimation of solar energy potential especially for rural electrification in Nepal.

Acknowledgements

We are grateful to express our sincere thanks to Solar Radiation and Aerosol in Himalaya Region (SAHR) project of IOE Pulchowk (Nepal) for providing the pyranometer instruments and data, Prof. B. K. Sapkota, Prof. B. K .Bhattarai and Prof. B. Kjeldstad of NTNU, Norway. We are grateful to Ass. Prof.N.B.Amatya, HOD and all the faculty members of Engineering Science and Humanities, IOE Pulchowk, for their support in our research activities.

This paper is a part of PhD research work under eLINK Project for which we would like to express special thanks to Prof. M. Savino, Prof. S. Rapuano of University of Sannio, Italy for their support and encouragement during this research work, Prof. R.Rijal, eLINK local coordinator Kantipur engineering College (KEC), Prof. N. Sharma, Principal of KEC, Nepal and all the Nepal friends who are studying at University of Sannio, Italy.

REFERENCES

[1] M.B.Basnyat, "Rural electrification through renewable energy in Nepal", World Review of Science, Technology and Sustainable Development, vol.1, no.1, pp.74-80, 2004.

[2] "United Nations Environment Program Global Environment Facility Solar and Wind Energy Resource Assessment in Nepal (SWERA)", Final Report (GIS PART) Submitted by: Alternative Energy Promotion Center Government of Nepal Ministry of Environment, Science and Technology Khumaltar, Lalitpur, Nepal, July 2008.

[3] K.N.Poudyal, "UV-A Solar Irradiation on Slant Surfaces at Kathmandu Valley", 1991.

[4] Energy Synopsis Report, Water Energy Commission Secretaries (WECS) Government of Nepal, June 2006.

[5] B.K.Bhattarai, B.Kjeldstad., T.M.Thorseth, and A.Bagheri, "Aerosol climatology in Kathmandu using Sun Photometry", Proceedings of SPIE, Remote Sensing of Clouds and the Atmosphere XI, 2006.

[6] M.Iqbal, "An introduction to Solar Radiation", Academic Press New York 1983.

[7] J.B.Kerr," Understanding the factors that affect surface UV radiation" In: R.S. James, J.R.Herman, W.Gao, (Eds.), Proceedings of SPIE. Ultraviolet Ground and Space Based Measurements, Models, and Effects III, vol. 5156, pp. 1–14, 2003.

[8] B.K.Bhattarai, "Factors Affecting Solar Ultraviolet Radiation, Based on some case studies in Norway and Nepal",March,2007.

ISBN 978-974-672-556-9

Graphic design : Narumon Kiti