ISSN 2225-658X (ONLINE)

International Journal of Digital Information and Wireless Communications

(IJDIWC)

Volume 4, Issue 1
2014

SDIWC
www.sdiwc.net

IJDIWC

ISSN 2225-658X (Online)

Volume 4 Issue 1 2014

TABLE OF CONTENTS

Original Articles

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 1-9
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Risk Management and Information Technology Projects

Amine NEHARI TALET[1], Razali MAT-ZIN[1]and Maaradj HOUARI [2]
[1]KING FAHD UNIVERISTY OF PETROLEUM & MINERALS SAUDI ARABIA
[2]GHARDAIA UNIVERITY ALGERIA
KFUPM P.O. BOX 1431, DHAHRAN 31261, SAUDI ARABIA
nehari@kfupm.edu.sa

ABSTRACT

IT projects management is not free from risks which are created from various sources of the environment. Thus a comprehensive understanding of these possible risks and creating strategic policies to confront them are one of the fundamental requirements for successful implementation of IT projects. The risks faced during the implementation of IT Projects are not just related to financial aspects. IT Project Managers must embrace these fundamental issues with more holistic view, rather than merely focusing on the financial matters. In order to prevent the potential problems from arising or escalating into bigger magnitude, serious attention must be given to it before the implementation of any IT project. The main focus of this paper is to investigate the impacts of Knowledge Management (KM) on Risk Management (RM) in IT project implementation process.

KEYWORDS

Globalization, Information Technology, Project, Uncertainty, Risk Management (RM),

1. INTRODUCTION

Knowledge has been identified as a vital source to innovative companies, a competitive advantage and value creation. It is as a key ingredient for the development of dynamic core competencies and, more generally, as a determinant factor for firms with globalambitions to meet challenges.

Firms are now forced to organize their projects and their structures. Many of them depend on their own resources or external resources to meet their objectives and be better prepare for changes in their surrounding environment.

Furthermore, firms are facing severaltypes of risks every day dueto the changes in the world environment that might introduce new risks. The source of risks differs from the internal or external environment. Uncertainty of the environment and behaviours which exert an influence on the problems in monitoring cooperation results on the level and risk of investment. If this risky environment is not managed appropriately, it might impact negatively on the present and future corporations. For this reason managers have to take into consideration risks that can affect and they have to minimize their impact on the organization.

One of the essential functions of information technology (IT) governance is risk management, which aims at providing a safe environment for E-business because IT projects are characterized by high degrees of risk.

The rapid transformation in information technologies combines changes in business processes to create surprising shifts in cost, the cost benefit relationship, and the feasibility of doing specific things in particular ways. In support of this various IT organizations, concerned with standards have published different risk management methods. These methods have been partially or fully adopted by enterprises using IT, for identifying, analysing, and minimizing risks for their IT activities.

The risks faced by IT projects are not, in essence, financial risks. By understanding these fundamental problems in real terms, rather than through their financial impact, IT project managers can move more quickly to resolve issues before they become major problems that threaten the goals of the project. Most IT

practitioners understand that there are risks other than financial risk in projects.
IT project risk, consists of financial, technology, security, information, people, business process, management, external, and even the risk of success (which occurs when the project is so well done that it draws more transactions than expected and fails to scale to the overload requirements).

2. LITERATURE REVIEW

2.1 Risk and Uncertainty

The well-known economist [**1**]founder of the Chicago School, distinguishes risk from uncertainty by relating risk to a "quantity susceptible of measurement" … "a measurable uncertainty" opposing it to real uncertainty "an immeasurable one" This reference to measurable and quantifiable events is also found within the project management literature. Risk is related to an "identifiable event that will have negative consequences whereas uncertainty relates to the source (of risk)[**2**]. Uncertainty is a situation being the source of risk: "a context for risks as events having a negative impact or opportunities that have beneficial impact[**3**]. Extending the consequence of the identified event to opportunities has the commendable intention of encouraging the management of opportunities, but adds to the confusion when trying to distinguish risk from uncertainty. A risk must contain two elements, namely uncertainty and loss [**4**]. RM refers to strategies, methods and supporting tools to identify, and control risk to an acceptable level [**5**].
Information systems projects always and everywhere around the globe have a reputation for failure, i.e. unused, partially used, cancelled and many other factors. Each project differs from another even if it is for the same system because each project has its own requirements, project management, users, organisation culture, team skills and knowledge, and many other aspects that are linked directly to the organisation and not to the project itself.
The degree of project uncertainty is an important dimension of the context. Many sources of uncertainty can be identified generating different uncertainty levels and needing different management responses. A major source of uncertainty in IT projects is uncertainty regarding the scope or specifications of the project according to[**6**].A proposed project uncertainty profile was anticipated to help determine the degree of uncertainty, from foreseeable uncertainty that can be controlled by traditional risk management techniques to unforeseeable uncertainty and chaos that may be found in some highly innovative projects[**7**].

2.2 Risk and Risk management

Risk and risk management have been studied in a variety of domains, such as Insurance, Economics, Management, Medicine, Operations Research, and Engineering. Each field addresses risk in a fashion relevant to its object of analysis, hence, adopts a particular perspective.
In some situations, risk is equated to a possible negative event as *"… events that, if they occur, represent a material threat to an entity's fortune"*[**8**] Using this definition, risks are the multiple undesirable events that may occur. Applied in a management context, the "entity" would be the organization. Given this perspective, risks can be managed using insurance, therefore compensating the entity if the event occurs; they can also be managed using contingency planning, thus providing a path to follow if an undesirable event occurs.
Some fields, instead of focusing on negative events, are primarily concerned with the probability of occurrence of an event. For example, medicine often focuses solely on the probability of a disease's occurrence (e.g., heart attack), since the negative consequence is death in many cases.
Finance adopts a different perspective of risk, where risk is equated to the variance of the distribution of outcomes. The extent of the variability in results (whether positive or negative) is the measure of risk. Risk is defined here as the volatility of a portfolio's value[**9**].
Other fields, such as casualty insurance, adopt a perspective of risk as expected loss. They define risk as the product of two functions: a loss function and a probability function.
Another important distinction in risk analysis is

the notion of endogenous versus exogenous risk. Exogenous risks are risks over which we have no control and which are not affected by our actions. Earthquakes or hurricanes are good examples of exogenous risks. Although we have some control over the extent of damage by selecting construction standards, we have no control over the occurrence of such natural events. Endogenous risks, on the other hand, are risks that are dependent on our actions.

The present study does not investigate the concept of risk per se; it is focused more on "risk management" as a practice. Project risk management is related to the set of practices and tools generally used to manage project risk. Risks in an organization can extend the range of natural disasters, security breaches, and failings of human resource, third-part vendors, financial chaos, unstable business environments and project failures. Therefore, the different types of risks encountered in common organizations.

Risk is the potential that a chosen action or activity (including the choice of inaction) will lead to a loss (an undesirable outcome). Risk is an uncertain event or condition that, if it occurs, has a positive or a negative effect on at least one project objective, such as time, cost, scope, or quality (where the project time objective is to deliver in accordance with the agreed-upon schedule; where the project cost objective is to deliver within the agreed-upon cost). A risk may have one or more causes and, if it occurs, one or more impacts. Also, risk conditions could include aspects of the project's or organization's environment that may contribute to project risk, such as poor project management practices, lack of integrated management systems, concurrent multiple projects, or dependency on external participants who cannot be controlled.

Furthermore, risk refers to all events, incidences and actions that may prevent you or your organization realizing its ambitions, plans and goals. Risk is surrounding us in our personal and professional lives and it is a potential problem that might happen. However, regardless of the outcome, it's a good idea to identify risk, assess its probability of occurrence and estimate its impact.

RM is generally perceived as a way to reduce uncertainty and its consequences, which in turn will improve the chance of success. RM has the intent to take counter measures that either prevents risks or mitigates the impact of a risk. Several authors, including [**10**] argue that RM should form a primary part of the project management process.

RM is a distinct discipline, which integrates knowledge from a variety of other business fields. It is the discipline in which a variety of methodologies are brought to bear on a specific problem. RM is very important and integral part of any business and well recognized by the project management institutions [**11**]

Most projects or business venture to take place in a changeable environment in which many drawbacks exist that might negatively impact the outcome of project success. A project is considered successful if it meets the requirements determined by the stakeholders, such as security, efficiency, reliability, maintainability, functionality, integration, and other requirements [**10**]. A study illustrated that 35% of deserted projects are not unnecessary until the implementation stage of the project [**12**]. This implies that project managers are doing a poor job of identifying or terminating projects that are likely to fail due to risks encountered during the project life cycle.

The concept of the risk management is applied in all aspects of business, including planning and project risk management, health and safety, and finance. It is also a very common term amongst those concerned with Information Technology. A generic definition of risk management is the assessment and mitigation of potential issues that are a threat to a business, whatever their source or origin. [**13**].

The concept is now fairly universally understood, having been in widespread use for a number of years. It is applied in all aspects of business.

To discuss the definition of the risk management is necessary to explain in advance the meaning of the three main concepts:

Threat is the potential cause of an unwanted impact on a system or organization (**ISO 13335-1)**. Threat can also be defined as an undesired

event (intentional or unintentional) that may cause damage to the goods of the organization.
Vulnerability is a weakness in system procedures, architectural system, its implementation, internal control and other causes that can be exploited to bypass security systems and unauthorized access to information. Vulnerability represents any weakness, administrative process, act or statement that makes information about an asset to be capable of being exploited by a threat.
Risk management is a process consisting of:

- Identifying vulnerabilities and threats to the information resources used by an organization in achieving business objectives;
- Risk assessment by setting the probability and impact of its production, following threats by exploiting vulnerabilities;
- Identify possible countermeasures and deciding which one could be applied, in order to reduce the risk to an acceptable level, based on the value of information resource to the organization.

RM is concerned with identifying risks, understanding risks and drawing up plans to minimize their effect on project. RM can be seen as a series of steps that help a software team to understand and manage uncertainty [**14**]**.**
RM objective is to identify all applicable risks in a project or business or product. It involves ranking the above elements based on their importance, frequency of occurrence, level of impact and then establishes the actions needed to control the identified risks. It is possible for every individual risk aspect to be documented in further details [**15**]**.** Since no one can predict what losses will occur, the objective of RM is to ensure that no risk will occur during the execution of a project in order to minimize losses to an acceptable level. If a loss occurs, then the objective of RM has failed to achieve the objectives intended, which prevent the organization from pursing their goals.
The goal of performing risk management is to enable the organization to maintain at the highest values the activity results. This process should combine as efficient as possible, all factors which can increase the probability of success and decrease the uncertainty of achieving objectives. Risk management should be an evolving process. Particular attention should be given to the implementation of the strategies for eliminating or reduce the risk and their appliance, to the analysis of the past evolution of risks and to the present and future prediction of the events. Management process should be implemented at the highest management level.
The link between uncertainty and failure (or between certainty and success) seems to be well established, but the link between risk management and success is not as clear. Clear indications are presented of the influence of individual project risk management activities on the success of IT projects[**16**].

In IT one of the most important goals of risk management is to accomplish by better securing the informatics systems that store, process, or transmit organizational information; by enabling management to make well-informed risk management decisions to justify the expenditures that are part of an IT budget and by assisting management in authorizing (or accrediting) the IT systems, on the basis of the supporting documentation resulting from the performance of risk management.

2.3 Risk Management Process

Risk management is a permanent cycle process that involves activities for establishing, monitoring and ensuring continual improvement of the organization's activity. This process includes four main activities, which have to be permanent applied and developed:

- *Design the management system* involves identifying business requirements, assessing the likelihood and the impact of the risks, including the implementation of a security policy and selecting the adequate countermeasures for the existing risks*;*
- *Implement the management system* involves applying control measures and work procedures, resource allocation, setting the responsibilities and conduct training and awareness programs;
- *Monitoring, reviewing and reassessing the management system* involve an

evaluation of effectiveness of controls and working procedures, of business changes, of previous incident reports and of existent risks*;*

- *The improvement and update of the management system* involves correcting the identified dysfunctions, or eliminating the unsustainable decisions or applying new control measures.

Risk management encompasses three processes:

1. Risk assessment is the process comprises establishing criteria under which the evaluation takes place (procedure on existing threats and vulnerabilities, and risks associated with, proceedings concerning the impact and likelihood of identified risks, risk assessment procedures, procedures for identifying measures to mitigate or eliminate risks, procedure for selecting the best measures to mitigate or eliminate the risks) and identifying and assessing risks..
2. Risk mitigation refers to determining optimal measures to eliminate or mitigate the risks, to planning, implementing the optimized selected measures, according to the plan, and controlling the rightfulness of the implementation process.
3. Reassessment of the residual risk consist of evaluating the remaining risk after the risk mitigation step and determine whether it is an acceptable level or whether additional measures should be implemented to further reduce or eliminate the residual risk, before the organization can perform work properly.

There are types of RM processes used in organizations today. These are summed up in Table 1. The processes of RM have emerged in a number of existing process models in the area of RM.

Table 1: The processes of Risk Management from previous studies

	Description of Process						
Main Dimension/ RM Process	1	2	3	4	5	6	**References**
Software development process	Risk identification	Risk analysis	Risk control planning	Risk monitoring			[**17**]
Software risk management	Risk Identification	Risk Analyze	Risk Plan	Risk Track	Risk Control		[**18**]
Role in software engineering	Review Define Goals	Risk Identification	Risk Control Planning	Risk Analysis	Risk Control	Risk Monitoring	[**19**]
RM as verification and validation	Risk Identification	Risk Analysis	Risk Planning	Risk Tracking	Risk Control		[**20**]
Software project management	Risk Identification	Risk Analysis	Risk Prioritization				[**21**]
Software development	Risk Identification	Risk analysis	Risk monitoring				[**14**]
Data warehouse system	Goal definition review	Risk Identification	Risk Analysis	Risk Planning	Risk Tracking	Risk Control	[**5**]
RM in insurance field.	Risk Identification	Risk Evaluation	Risk Evaluation	Risk Monitoring			[**22**]
Controlling product development	Identify Risks	Analyze Risks	Prioritize and map risk	Resolve Risks			[**23**]
Project management	RM planning	Qualitative risk analysis	Quantitative risk analysis	Risk response planning	Risk monitoring and control		[**24**]
Software engineering	Risk Identification	Risk Analysis	Risk Planning	Risk Monitoring			[**25**]
Software Risk Management Process	Plan and implement risk management	Manage the project risk profile	Perform risk analysis	Perform risk monitoring	Perform risk treatment	Evaluate RM process	[**26**]
Risk architecture	Risk classification	Risk identification	Initial risk assessment	Risk mitigation	Risk monitoring		[**27**]

RM involves the following stages:

1) Risk Identification is to identify project, product and business risks; 2) Risk Analysis is to assess the likelihood and consequences of these risks; 3) Risk Planning is to draw up plans to avoid or minimize the effects of the risk; and 4) Risk Monitoring is to guarantee the effectiveness of the methods followed and to monitor the risks throughout the project. Figure 1 illustrates the proposed Risk it method [**19**]**.**

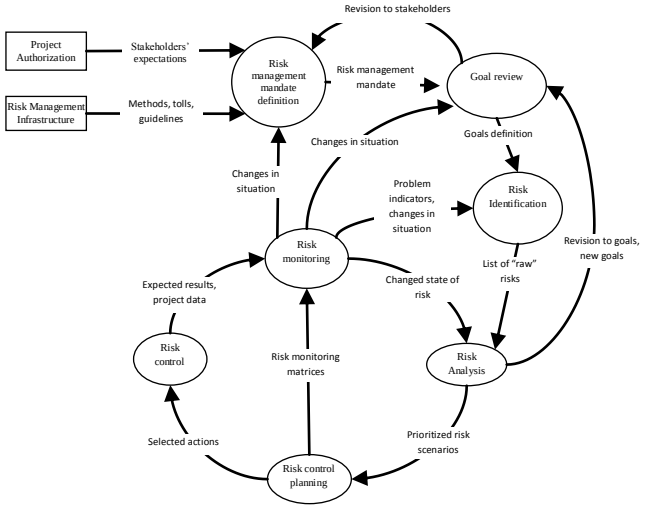


Figure 1: Riskit RM Cycle[**19**]

From figure 1 RM process is divided into Risk Identification, which focuses on a variety of technological content, environmental communication execution and operation approaches, programmatic constraints and the mission duration[**20**]. The second is Risk Analysis of the consequences of the possible risks by scoring their impact on the necessities should they occur.

The third is Risk Control, which is deciding the best suitable and cost-effective measures needed to be executed in order to control the risks such asrisk avoidance, risk reduction, and risk transfer. The fourth is Risk Monitoring, which provides a review of the organization's ability to deal with incidents that might result in business interruption, implementing risk identification, Evaluation and Control measures that minimize the likelihood and severity - both in terms of potential financial and reputation losses of an incident, but can never entirely eliminate the risks. The result is a requirement-driven risk list where failures are listed based on their impact on weighted requirements. Risk planning process has design rules, process controls, testing, modeling, and inheritance. Risk Tracking contains a tool to display the number of report formats to be used by different personnel for different reasons. Risk Control is designed for implementation, which allows the project team to effectively control risk and watch its growth or decline as the design evolves and the results of implementation become available.

RM process is described into six steps:

1) Goal Definition: review the stated goals for the project, refine them and define implicit goals and constraints explicitly.

2) Identification: identify potential threats to the project using multiple approaches.

3) Analysis: classify risks, complete risk scenarios, estimate risk effects and estimate probabilities and utility losses for risk scenarios.

4) Control and Planning: select the most important risks, propose controlling actions and the actions for implementation.

5) Control: implement the risk controlling actions. 6) Monitoring: monitor the risk situation Figure 2 illustrates the RM process [**28**].

.

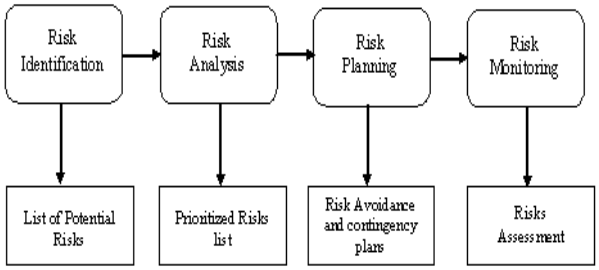


Figure 2: RM Processes [**28**]

RM process contains the following processes: identifying risks using brainstorming techniques to discover any risks that prevent the progress of the project, analyzing risks by the team members to determine if a certain risk is worth migrating or not, prioritizing and mapping risk to establish a seriousness of the risk according to their impact, and resolving risks by implementing plans to prevent risk from occurring [**29**].

2.4 Sources of Risks in IT projects

By examining the literature reviews, IT projects can range from software development, outsourcing, communications and implementing a new security infrastructure. Also, these IT project risks might have a particular management in a different project. Risks are resulting of many factors involved in the projects. Each factor will depend on the type and purpose of the project.

IT projects are characterized by high degrees of risk. The rapid pace of change in technologies combines changes in business processes to create unpredictable shifts in cost,

the cost benefit relationship, and the feasibility of doing specific things in particular ways [**30**].

One classic problem identified in many IT projects occurs when new technologies are developed as the project is underway [**31**].

A survey of more than a thousand Canadian organizations found that the main reason for IT project failure was inadequate risk management and a weak project plan. The risks faced by IT projects are not, in substance, financial risks [**32**]. The financial measures are only indicators of the underlying problems. By understanding these underlying problems in real terms, rather than through their financial impact, IT project managers can move more quickly to resolve issues before they become major problems that threaten the goals of the project. Most IT practitioners understand that there are risks other than financial risk in projects.

IT project risk are broken into nine categories, including financial risk, technology risk, security risk, information risk, people risk, business process risk, management risk, external risk, and even the risk of success (which occurs when the project is so well done that it draws more transactions than expected and fails to scale to the overload requirements[**33**].

Anecdotes, surveys, and field research studies establish that many IT projects fail. Managers abandon some of these failing systems. Other projects that are over budget or behind schedule ultimately result in useful systems [**34**]. However, many more projects continue long after any hope of success has faded[**35**]; [**36**].

IT projects are much more likely to fail than other types of projects, such as building construction projects and states that the main causes for IT project failures are their use of rapidly changing technologies, their generally long development times, and the volatility of user expectations about what the project will yield. Because IT projects generally include all of these characteristics, they are prone to failure, cost overruns, and schedule delays [**37**].

Organizations need to keep IT projects on schedule and costs under control. However, organizations must also encourage managers to respond to changing business needs and exploit technological opportunities before their competitors do so [**38**].

In terms of IT projects, risks can vary, whether it is a software development project, security project, outsourcing project, or specific programming task.

IT projects are known for their high failure rate. Based on in-depth interviews with IT professionals from leading organizations in Western Australia were undertaken to determine how IT risks were managed in their projects. The respondents ranked 27 IT risks in terms of likelihood and consequences to identify the most important risks[**39**]. The top five risks, in order, are: personnel shortfalls; unreasonable project schedule and budget; unrealistic expectations; incomplete requirements; and diminished window of opportunity due to late delivery of software. Furthermore, the respondents overwhelmingly applied the treatment strategy of risk reduction to manage these risks. Additionally, these strategies are primarily project management processes, rather than technical processes. Therefore, this demonstrates that project management is a RM strategy. In particular, managing stakeholders' expectations is a specific risk treatment that helps to manage several key IT risks[**39**].

The risk issues in reference to software development process are not viewed as essential subject of discussion. On the other hand, most organizations hope to implement systems successfully while still assuming their regular business processes. Yet, new systems are not implemented in a space and many authors concur that the first step in developing a business continuation plan is to carry out a risk assessment [**40**].

There are different types of risks and each depends on the kind of IT project involved. Risks can vary whether it is a software development project, or security project, or outsourcing project, or specific programming task. By investigating the effects of risk management on IS/IT project success, it was concluded that project risk management is defined in the literature as being an instrumental action based on rational problem solving.

The purpose of RM is to develop a detail analysis of the project and to establish a comprehensive list of risks.

RM assists project team to make better

decisions, communication and to resolve any risk issues in effective matter.

ACKNOWLEDGEMENTS

The first two authors are grateful to King Fahd university of Petroleum & Minerals for its continuous support **Project: SB111010**

Bibliography

[1] H. Knight. (1921) Risk, Uncertainty and Profit. [Online]. www.econlib.org/library/Knight/knRUP.html

[2] Sicotte, H. & Bourgault, M., "Dimensions of uncertainty and their moderating effect on new product development project performance," *R&D Management*, vol. 38, no. 5, pp. 468-79, 2008.

[3] B. E. Perrott, "“A strategic risk approach to knowledge management”," *Business Horizons*, vol. 50, no. 6, pp. 523-533, 2007.

[4] Parvatiyar, A., & Sheth, N., "Customer Relationship Management: Emerging Practice, Process, and Discipline.," *Journal of Economic and Social Research*, vol. 3, no. 2, pp. 1-34, 2001.

[5] Bruckner, M., List, B., & Schiefer, J., "Risk-Management for Data Warehouse Systems," *Data Warehousing and Knowledge*, pp. 219-229, 2001.

[6] L. Kumar, "Managing risks in IT projects: an options perspective," *Information & Management*, vol. 40, pp. 63-74., 2002.

[7] Meyer, A., De, C., Loch, H. and Rich, M.T., "Managing project uncertainty: from variation to chaos," *MIT Sloan Manage. Review*, vol. 43, no. 2, pp. 60-67, 2002.

[8] M. and Schneider, M. Levin, "Making the Distinction: Risk Management, Risk Exposure," *Risk Management*, vol. 44, no. 8, pp. 36-42, 1997.

[9] E. Levine, "Defining Risks," *CA Magazine*, vol. 133, no. 3, pp. 45-46, April 2000.

[10] Powell, P. L., & Kelin, J. H., "Risk Management for Information Systems Development," *Journal of Information Technology*, pp. 309-319, 1996.

[11] Del Cano, D., & Cruz, P. , "Integrated Methodology for Project Risk Management," *Journal of Construction Engineering and Management*, vol. 128, no. 6, pp. 473-485, 2002.

[12] Ewusi-Mensah, K., & Przasnyski, H. , "On Information Systems Project Abandonment:An Exploratory Study of Organizational Practice.," *MIS Quarterly*, vol. 15, no. 1, pp. 67-88, 1991.

[13] Southern, S. , "Creating risk management strategies for IT security," *Network Security*, pp. 13-14, 2009.

[14] Bandyopadhyay, K., Myktyn, P., & Myktyn, K. , "A framework for integrated risk management in information technology.," *Management Decision*, pp. 437-444, 1999.

[15] Cule, P., Schmidt, R., Lyyttnen, K., & Keil, M. , "Strategies for Leading off is Project Failure," *Information Systems Management*, vol. 14, no. 2, pp. 68-73, 2000.

[16] Bakker, K., Boonstra, A. and Wortmann, H. , "How risk management influences IT project success," in *IRNOP IX Conference 2010*, Berlin, Germany, 2009.

[17] Boehm, B., & Bose, P. , "A collaborative spiral software process model based on theory W," in *3rd International Conference on the Software Process (ICSP94)*, New York, 1994.

[18] Higuera.R & Haimes.Y, "Software Risk Management," Software Engineering Institute Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, Technical Report CMU/SEI-96-TR-012 ESC-TR-96-012, 1996.

[19] J. Kontio, "The Riskit Method for Software Risk Management version 1, 00," University of Maryland. College Park, MD, Computer Science Technical Reports CS-TR-3782 / UMIACSTR- 97-38, 1997.

[20] S., Feather, M., & Hicks, K. Cornford, "DDPTool for life-Cycle Risk Management. Jet Propulsion Laboratory," Californaia Institutte of Technology, IEEE, 2001.

[21] J., Jurison, "Software projectmanagement: The manager’s view. ," *Communications of Association for Information Systems*, vol. 2, no. 17, pp. 2-52, 1999.

[22] Beck, T., Levine, R., Loayza N., "Finance and the Sources of Growth," *Journal of Finance and Economics*, vol. 58, pp. 261-300, 2000.

[23] P. & Merritt, G. Smith, "Proactive Risk Management: Controlling Uncertainty in Product Development," *Productivity Press, New York*, 2002.

[24] Project Management Institute, *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)* , 4th ed., 2008. [Online]. http://www.projectsmart.co.uk/pdf/pmbok.pdf

[25] Sommerville. I, *Software Engineering*, 8th ed. University of St. Andrews, , United Kingdom: Addison-Wesley, 2006.

[26] Software & Systems Engineering Standards Committee of the IEEE Computer Society, "Systems and software engineering — Life cycle processes — Risk management ," International Organization for Standardization/International Electrotechnical Commission, ISO/IEC 16085 IEEE Std 16085-2006, ISO/IEC 16085:2006(E) IEEE Std 16085-2006, 2006.

[27] TOGAF. (2009) The Open Group Architecture Framework (TOGAF). [Online]. http://www.kingdee.com/news/subject/10togaf/pdf/TOGAF_9_ziyuan.pdf

[28] I. Sommerville, *Software Engineering*, 6th ed.: Addison-Wesley, 2001.

[29] Smith, G. & Merritt, M. , *Proactive Risk Management: Controlling Uncertainty in Product Development*.: Productivity Press., 2002.

[30] K. Schwalbe, *Information technology project*

management. Boston: Course Technology, 2007.

[31] H. Havenstein, "IT execs seek new ways to justify Web 2.0," *Computerworld*, vol. 41, no. 33, pp. 14-15, 2007.

[32] B. Whittaker, "What went wrong? Unsuccessful information technology projects.," *Information Management & Computer Security*, vol. 7, no. 1, pp. 23-29, 1999.

[33] McKeen, J. & Smith, H. , *Making IT happen: Critical issues in IT management.* West Sussex, England: Wiley, Ltd, 2003.

[34] G. Anthes, "SEC electronic filing system in high gear," *Computerworld*, vol. 27, pp. 59-67, 1993.

[35] J. Johnson, *Chaos: The dollar drain of IT project failures.*: Application Development Trends, 1995, vol. 2.

[36] M. & Sabherwal, R. Newman, "Determinants of commitment to information systems development: A longitudinal investigation.," *MIS Quarterly*, vol. 20, pp. 23-54, 1996.

[37] G., Lane, S., & Bruton, C Schneider, "Monitoring risk in information.," vol. 13, no. 1, pp. 63-67., 2009.

[38] S. Murthi, "Managing the strategic IT project, Intelligent Enterprise," vol. 5, no. 18, pp. 49-52, 2002.

[39] D., Salm, G., & Love, D. Baccarini, "Management of risks in information technology projects," *Industrial Management & Data Systems*, vol. 104, no. 4, pp. 286-295, 2004.

[40] S. Maguire, "Identifying Risks during Information System Development: Managing the Process," *Information Management and Computer Security*, vol. 10, no. 3, pp. 126-137, 2002.

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 10-23
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Structuring Heterogeneous Big Data for Scalability and Accuracy

Ashraf Gaffar
Dept. of Engineering & Computing Systems
Arizona State University
Mesa, Arizona, USA
agaffar@asu.edu

Eman Monir Darwish
Dept. of Fine Arts
Helwan University
Cairo, Egypt
eman.monir@hu.edu.eg

Abdessamad Tridane
Dept. of Mathematical Sciences
UAE University
Al-Ain, UAE
a-tridane@uaeu.ac.ae

ABSTRACT

Structured data has an inherently great automation value. It renders itself readily for software tools to help store, organize and search effectively. With the growing dependence of data, we face many new problems. While software applications replace each other, and older software has rapidly diminishing value, data has the extreme opposite nature, which we can call "cumulative effect". Unlike software, we see new data as an addition to the old one, so we tend to continuously accumulate date without deleting any thing. Even older data is often archived for it's value, for legal reasons, or just because we can never be sure if we'd need them again. This would not be a problem if we had structured data, as we can automate the storing and retrieval process. However, most of the valuable information lie inside unstructured data, which is extremely difficult to store and retrieve in large scale. Our work shows a new concept of adding structure to highly unstructured, heterogeneous data, which will greatly improve the total process of storing and effectively retrieving them. We use Human-Computer-Interaction (HCI) patterns as a case study to present our proof of concept.

KEYWORDS

HCI patterns, information encapsulation, semi-structured data, information dissemination, information reuse

1 BACKGROUND

The concept of patterns lends itself to human nature. Patterns are an effective approach of disseminating knowledge gained by expert practitioners over the years. Christopher Alexander [1], [2], [3] first introduced them in his work about designing living spaces inside homes and outside home communities. In the pattern community, it is widely accepted that we need to have a look at Alexander's original work to get a real sense of what patterns really are and how they can be used [4]. Several domain experts followed the same pattern approach and started documenting their years of experience in terms of patterns. One major landmark in pattern work is the "Design Pattern" book published by the so-called "Gang-of-Four" [5]. This seminal book made a major breakthrough in disseminating highly sophisticated programming experience from well-known experts to novice programmers by following the pattern approach. This major success motivated several other domain experts to start adopting the pattern approach in wrapping and disseminating their years of experience and delivering it to the novice practitioners in their domain.

The idea of using patterns is fairly simple and straightforward. A pattern provides a solution to a problem in a specific context, allowing pattern users to get directly to a suitable solution to a problem without having to go through extensive theoretical background, analysis, or previous attempts. With their growing success, we ended up with a tremendous amount of patterns, or a plethora. While extremely useful in most cases, as the number of patterns grows exponentially, patterns users started to be overwhelmed as to finding the most suitable patterns and applying them to their work. Unaware of the problem, patterns authors continued to generate patterns as a way of documenting their domain experience, but did not go far enough to consider the environment of how they are being accessed or reused. Very little effort was spend on

understanding how to effectively reuse the plethora of existing patterns [6]. To complicate the problem, patterns –by nature- do not live in isolation. They are typically used in combination, so when they are generated by different authors, the concept of combining them together was a growing problem. It was typically left up to the user to decide on that using their own judgments, and following ad-hoc processes [7]. Redundancies and contradictions were common problems in pattern reuse, resulting in less-than-optimal solution or even erroneous designs.

This particular problem of pattern reuse was identified and discussed in details by different domain experts [8], [9] and [10]. Different tools were build to try and present a possible solution [11], [12], [13]. Unfortunately, most work approached the solution "after the fact". They tried to deal with the heterogeneity of patterns as an inherent nature without trying to generate more structured pattern at a deeper information level.

2 SEARCH BY SYNTAX: A RANKED GUESS WORK

To demonstrate our work, we examined multiple pattern collections, and selected Human-Computer-Interaction (HCI) patterns as case study of both unstructured and large collection of data. HCI patterns are typically visually oriented, so they are presented in a heterogeneous and multi-modal format, including text, code snippets, graphics, flow charts, and often screen shots. While few patterns can be easily understood and applied, previous usability tests on UI designed generated using large number of unrelated HCI patterns have shown several problems in applying HCI patterns in generating new designs [14].

In order to identify the problem, and follow an approach of "separation of Concerns", we can generally split the pattern lifecycle into 2 phases as follows:

1) Pattern Generation Phase: Defines how patterns are generated by the domain experts.
2) Pattern Dissemination Phase: Defines how patterns are communicated from the pattern author to the pattern user.
3) Pattern Reuse Phase: Define how patterns are consumed and reused to generate new designs.

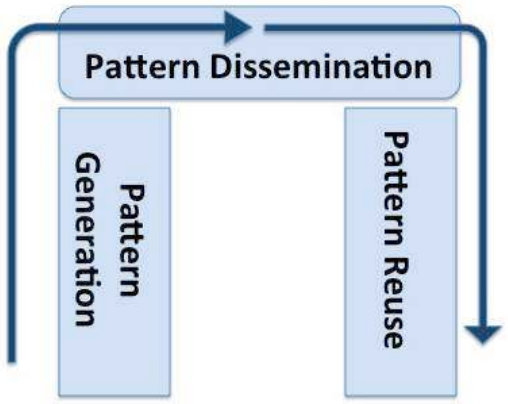


Figure 1: The Pattern Lifecycle

Figure 1 depicts the three phases of the pattern lifecycle.

Defining this lifecycle sheds some light on the main roles associated with patterns. Apparently, the key roles associated with patterns are the "Pattern Author", and the "Pattern User". It is clear that in the overwhelming majority of cases, different people, namely the domain expert, and the novice designer, assume these two roles respectively. Having identified both roles shows the missing link, "who is taking the role of pattern dissemination?" Apparently, the pattern community has relied on distributed responsibility of disseminating patterns, with the majority of the work left to pattern authors. In this regard, pattern authors would present their patterns to the community via conference papers, books, or through their own websites. However, this does not solve the problem. Pattern users still have to "find" the respective patterns, which are often present in multiple disparate sources. This is the neglected task that ends up overwhelming pattern users as they sift through several pattern collections, trying to mix and match them without author's support.

To focus more on the core of the problem, [6] defines ta search behavior, namely the "*Structured Data Continuum*" as depicted in Figure 2.

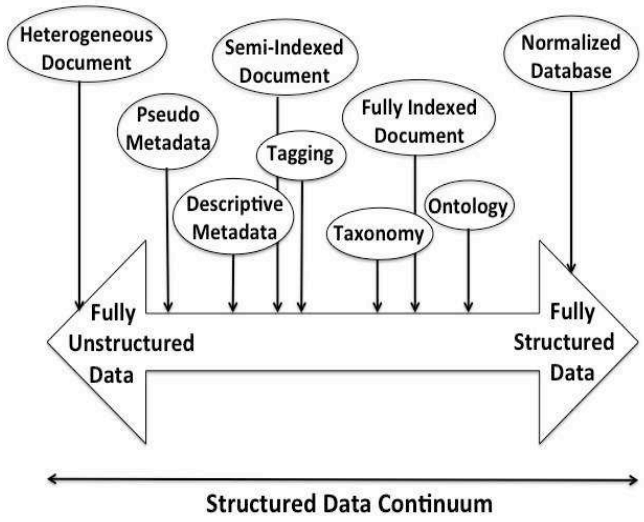


Figure 2: Structured Data

Structured data, presented on the right hand side of the scale can be automatically processed at large scale without any problems. However, structured data are , but they are expensive to create from unstructured data. One common example is database creation. An extensive process of analysis and modeling is required before we are able to generate a well-structured database from daily information.

The other side of the scale the far left hand side) presents the other extreme of data. On this extreme, we typically have rich text documents solely intended for human consumption. They typically have no specific form, and contain multiple modality including text, images, graphs, and even audio and video components. Even when we represent them in an electronic format as in a web page, or an office document (for example MSWord, or MS Power Point), they can be very hard to process at large scale due to the lack of structure. The multi-modal content can be extremely pleasant for a human reader to read and comprehend, regardless of the multiple modality format, or the lack of structure. In fact, we tend to love a heterogeneous document with a coherent context rather than a fully structured excel sheet document for example. One of the main reasons is that a human reader looks for and easily understands the underlying semantics and meaning of a document regardless of how it is presented. Computers, on the other hand, have very hard time extracting the meaning from a text, let alone an image or a video [8].

Software applications use complex algorithms to extract and manipulate the underlying meaning of text sentences. Extensive work is done in the domain of Natural Language Processing (NLP) that is beyond the scope of this paper. To just give a sense of the problem and elaborate more on the semantics challenge, we demonstrate using some examples:

Example 1: If we looked at a common phrase like "*Mr. Brown is wearing a Puma sneakers*". We can easily recognize "*Brown*" as the family name of a person, and we can also recognize "*Puma*" as a common brand of a sports shoe manufacturer. Software applications can have significant challenges trying to understand the actual meaning of those two words, and have to revert to sophisticated algorithms to guess their actual meaning.

Example 2: Searching for a particular picture in a large library of pictures can be a daunting task. If we were matching pixel for pixel of two perfectly identical pictures, it would be easy for a software application to compare them. When it comes to the contents of an image, it is a completely different challenge. Even two identical pictures that have some lighting variations, or are slightly rotated, zoomed in or out, or taken from a slightly different angel can be extremely hard to identify as similar using software. Complex imaging and graphics algorithms, fuzzy logic and pattern matching can help identify the closely similar images [15], [16].

This approach suffers when the images are highly different in pixel contents, but similar in their physical contents. An example would be similar objects laid out differently in two pictures, or two pictures of large-breed dog, and a small-breed dog.

To be able to work our way round such problem, we can identify the "meaning" of the picture contents by adding textual description of the content and linking it to the images. By using standard keyword search of these textual descriptions, we can get sufficiently good search results. The problem with this method is that we need to write a large number of highly descriptive keywords that describe all aspects of the picture,

including objects, layout, context, and even associated emotions and feelings. In this case the meta-data writer has to come up with sufficiently descriptive set of keywords to describe what the picture contents are. Equally important, the user (searching for particular pictures) has to come up with sufficiently descriptive set of search terms of what they're looking for in a picture, and hope for good matches with the existing metadata terms, or at least a common overlap between the two sets. This solution fails when metadata is not descriptive enough or absent altogether resulting in low or no correlation. The failure can also result when the user is unable to find descriptive enough keywords or not sure what they are exactly looking for. We can further argue that this is a way of redirecting the core search activity into a correlated search of auxiliary contents rather than the original one. Users are basically searching textual information (the metadata), and not the original data, and the pictures we get are dependent on the quality and expressiveness of the auxiliary text as well as on the search keywords, and not on the original contents.

The cost of writing additional meta data to images can be prohibitively large, especially if we need to create sufficiently descriptive data to a large number of pictures. A good example is if we need to describe all images available on all websites on the Internet. We can see that the two dimensions "*highly descriptive metadata*" and "*large number of images*" makes it an extremely difficult task. Google recently came with an elegant solution to work around this difficulty. Considering the two dimensions we explained above, Google opted for a great reduction of cost by trading off the highly descriptive metadata generation with the "existing accompanying text". In simple words, Google scanned each image on the Web, and linked it to the text that comes with the page, and then used the accompanying text as a pseudo meta data. The assumption made by Google was that if the webpage contained the word "dog" on it, then it is likely that the images will contain a dog in them. Further ranking criteria like the number of occurrences of the word, the location(s) in the document where the word was used (header, body, caption, etc.), can refine the returned images and rank them for higher fidelity, improving the hit rate.

To check for the validity of our assumptions, we executed a simple search on Google search engine using their Google Image search. We repeated the experiment several times and visually inspected the results. In all cases, we had a high hit rate, where most of the returned images were indeed dog images. We identified four categories of accuracy:

1) Highly-accurate search results: A picture of a dog
2) Moderately-accurate search result: A picture that includes one or more dogs as well as other objects
3) Poorly-accurate search result: A picture that does not include a dog, but is dog-related. Examples are dog-clinics, dog-related charts, or dog-food
4) Wrong search result: A picture that does not contain any dogs or dog-related items. Figure 3 shows one search result of the keyword "dog" with three wrong search results.

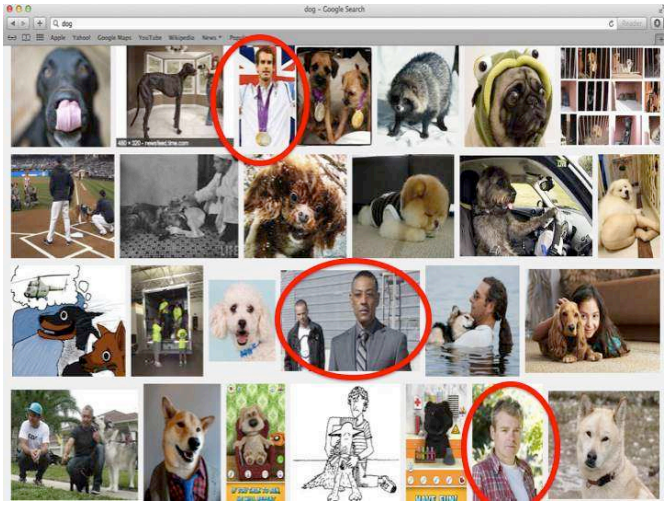


Figure 3: A "dog" keyword search with multiple wrong results

The four categories were observed in our experiment, and we analyzed some of the examples of semantic reasons why the images were identified as dog images.

One of the wrong search results was of a person with the name "Maddog". Nothing else in the

whole accompanying text was dog related. Apparently, the method of pseudo meta-text mistakenly identified the contents as dog related. This was a clear case of category 4, "wrong results".

In another case where a person's face was displayed, the title of the accompanying document was " Florida linebacker Antonio Morrison arrested after barking at police dog". Source: New York Daily News, read more: *http://www.nydailynews.com/sports/college/florida-lb-arrested-barking-police-dog-article-1.1404972#ixzz2eWxm5tyB.* This was a clear case of category 3, "poorly-accurate results". For a person looking for dog-related text, this might be relevant. But for a person looking for a dog image, this is not the best result. However, as it is still dog related, we considered it a category 3.

In another example of a person's face image, the document was reporting a criminal incident of a person charged with the attempt to kill a dog: http://www.koin.com/2013/08/05/deputies-man-attached-explosives-to-dog/. This was also a category 3 search result.

In all cases, the search did not produce a highly accurate dog image as desired. The key issue here is that the metadata (the accompanying text in this case) did include related syntax of a "dog" in all cases, and clearly related semantics of "dog"-related issues in the latter two. However, the search still failed to produce a dog image, as would a human search.

Case Study: Smiling Lady painting

In this experiment, we ran a simple search on the keyword "*Smiling Lady Painting*". This is a particularly interesting problem is searching for artwork. As we got back and analyzed the search results, we identified several semantic differences. There were 5 main types of unrelated search results:

Type 1: The "Smiling Lady" was identified as the well-known enigmatic work of the Mona Lisa by Leonardo Da Vinci, and the "Painting" was identified as a noun, for example at: http://www.destination360.com/europe/france/paris/mona-lisa

Type 2: The 'Painting" was still identified as a noun of fine art artifact, but the "Smiling Lady" was taken into a more generic semantics of any painting that has a smiling lay in it. Two examples are:
http://abstract.desktopnexus.com/get/430280/?t=su1g6iveelahljg2avfa68pg2252b21d338f119
and
http://www.ebsqart.com/Artist/Hiroko-Reaney/5939/Art-Portfolio/Gallery/Other-Art/Smiling-Womans-Face/428398/

Type 3: This search result was still in the realm of fine arts, but the word "Painting" was assumed to be a verb, and the search result was of a smiling lady in front of a painting tripod, actually holding a brush and painting something. The example is shown at
http://www.dreamstime.com/stock-photo-smiling-senior-woman-painting-image16998150

Type 4: A more distant semantic search was that of a lady painting a wall. The word "Painting" was seen as a verb, and the actual painting context was that of a wall, not of an artwork, found at:
http://www.123rf.com/photo_5525670_smiling-beautiful-woman-painting-interior-wall-of-home.html

Type 5: The most distant semantics of the search was that of a lady painting her nails.
http://www.dreamstime.com/stock-photos-smiling-woman-painting-her-nails-image1553133

We can see that the same phrase "Smiling Lady Painting" can result in at least 5 completely different semantics. Figure 4 shows all 5 categories as a search result using Google Image search with the keyword "Smiling Lady Painting".

As we can see, the use of accompanying text to guess on the contents of an image can work most of the time, but it remains a mechanical workaround, and the success rate goes dramatically down with more complex or less descriptive keywords.

Similarly, as discussed earlier, while the comparison of similar but pixel-imperfect pictures can be manipulated using fuzzy logic and non-traditional database approaches with acceptable results, the problem remains as to fully describe the contents of all contents of a complex picture

by a computer in the same intuitive way humans do.

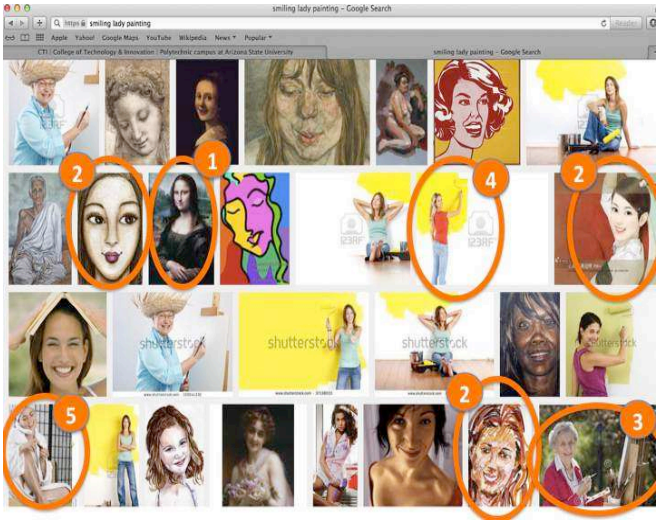


Figure 4: A "Smiling Lady Painting" Search Results

Other approaches to manage unstructured data can be found at Google maps, where they first captured millions of pictures of earth at different heights from ultra high altitudes all the way down to street views, then hired an army of people to manually review, fix, seamlessly connect, and integrate them into one 3-dimensional continuum. In this continuum, you can move in 2-D (north-south and east-west) as well as at the 3rd dimension of zooming in and out. Google further annotated the contents with a gamut of information from street names and directions to related photos and street views. By adding all these integrated structured details, the original pixel-generated images became highly processable by a large number of software application; a perfect example of data reuse.

Amazon.com used a different approach to adding structure to their documents. With the help of millions of users, they annotated the original books contents with additional end-user and system-generated, highly structured contents (e.g. star-ranking, user feedback, related book recommendations, people who bought/looked at this book also bought/looked at those books, etc.).

Further analysis shows an apparent conversion between the two approaches. While Google relied heavily and mainly on hired human force to integrate and annotate data, they later also used the million of users to upload their pictures and used them to further annotate the maps by linking the uploaded pictures to the location where they were take.

Similarly, besides Amazon's heavy reliance on user-created contents on available books, they also included their own compilation of book contents like publisher information, contents scans and chapter examples, and integrated them with the user-created ones.

3 HOW WE SEARCH

We have identified the role of data creator at the "pattern generation" end of Figure 1. Similarly, we can identify the ole of data consumer at the other end, the "pattern reuse" one. We introduce the following data search model to identify three main types of human search behavior.

I) Search to Locate:

This is a common form of search, where the person has already some knowledge about the search object. This pre-existing knowledge can have varying degrees of accuracy related to two main dimensions, the object attributes, and the object location. These two dimensions provide us with four possibilities

- **Clear object and clear location**: The person can have pinpoint accuracy as having seen the object before and has a good idea of where to find it. The search is just trying to locate The object
- **Clear object and vague location**: The person can have accurate information about the object, but has vague idea of where to find it
- **Vague object and clear location**: The person has clear idea about the location of the object. However, she only has abstract idea about the search object itself. The person is vaguely aware of its characteristics or attributes.

The fourth possibility of "vague object and vague location" does not belong to this category as it will have too many unknowns, and the search person's behavior will be greatly different. We believe it

should belong to another category, namely the "search to find", as described below.

The nature of "search to locate" generally has exact keywords, or close enough ones, and will normally end in few iterations. One example is when the user is looking for a particular car model at a particular dealer (clear object and clear location), or when the user is looking for a particular car model at different dealerships (clear object and vague location). The third case of the user looking for all available cars at a particular dealership would resemble the case of vague object and clear location.

II) **Search to Find:**

When we search to find, we do not have a concrete idea of the object we search, or a concrete idea where we might find it (its location). This dual dimensionality of unknowns (unknown object and unknown location) makes the search more complex, and the search domain much larger than the first type. Following the previous example of searching for a new car, this type of search resembles a user looking for a car of any model and make, any year (brand new or used), and at any location (dealership, car resellers, or a private person). We can see the search options in this case are much more than the previous type (search to locate). Moreover, the search behavior of the user can be significantly different. The user will probably look at many different attributes, and will iterate through the search activities several times. The iterations will result in a refinement of the search, and the user will start building an initial idea about both search dimensions: the object, and the location. This type of research will generally take longer than the first type, and may not result in a particular selection; rather it can help build a clear idea of the two search dimensions, or at least one of them, and can eventually evolve into the first type.

III) **Search to Search:**

Unlike the previous two types, in this type of search, the user is not actually searching for any particular object. The user typically has a need, and is not sure how to fulfill it. The user will form an initial idea without having particular object name or attributes. Furthermore, the user will often have no particular search keywords in mind. As the user searches using a wide range of generic keywords, she would get some results that are irrelevant. Those results are typically ignored. Other results would be informative in that it will give the search person an insight into what a search object might be, or even just what some relevant keywords might be. The user behavior in this case would be content-oriented, and not object or location oriented. In content-oriented search behavior, the user will focus on the search result contents and will use them to educate herself about the particular domain, the availability of different objects, the attributes of those objects, the locations, and the like. After several iterations, the user will have identified their needs, and have formed a goal to fulfill those needs. This will result in forming a clear idea about the object to be searched. As we can see, this search will eventually evolve into the previous category of "search to find".

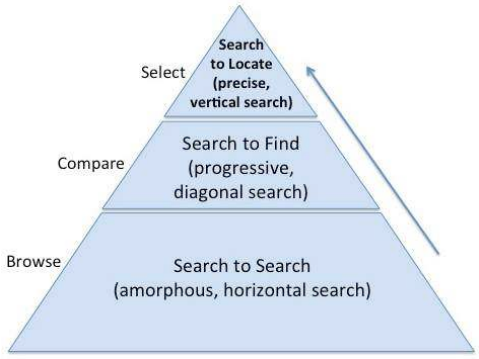


Figure 5: Search Activities Model

One case to demonstrate this type of search is when the user is searching for a Christmas present for a family member, but has no specific idea of what that present might be. As we can see, the object is unknown, the location is unknown, but also the search terms and the knowledge of suitable keywords is absent. If the user used "Christmas gift" as a search object, they would be inundated with infinite number of search results with an infinite range of object types, prices, etc. As the user sifts through some of those results, they would start forming an initial idea of what a Christmas present could be. This type of search

will mainly help the user get an initial idea about an object, leading to the evolution into the second category, search to find. In Academia, we go through this search to search as we –for example- go through literature review. Initially, we sift through several related and unrelated documents, forming an initial idea without any particular eye on any concrete objects. As we sift through and filter out most of the search results, we refocus on selected papers and start building a concrete idea about search objects, relevant keywords, and specific teams' or person's work.

3.1 Search Lifecycle

While the three search types are stated from more specific to more generic, (categories I through III in the previous section), we can more logically connect them in a progressive lifecycle from the most generic search (browse) to the most specific search (select) as per Figure 5.

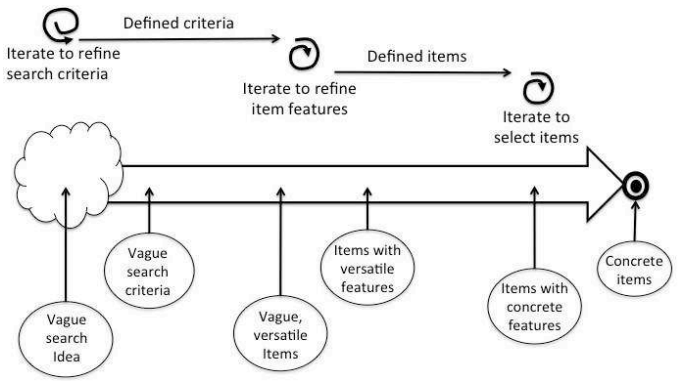


Figure 6: Search Lifecycle

Figure 6 depicts the progression between the three moods of search and some associated artifacts.

4 THE SYSTEM DESIGN

To successfully reuse pattern, we focus on an integrated approach of improving the full lifecycle of data from generation to dissemination to reuse. As we look at the roles associated with the pattern lifecycle (Figure 1), we have identified two main roles: the pattern author, an the pattern user. The fundamental question that arises is: "*Who will be responsible for the dissemination process*".

Pattern Author Role:

From the pattern author perspective, their main role's responsibilities are to formulate their experience and write it own using one of the common patterns format. That will typically include rich, multi-modal document that includes text, images, graphs, screenshots, and possibly a code snippet. This is a significant amount of work, and as we've seen, it is being well-taken care of by pattern authors. However, as we have seen from the literature review above, pattern authors made little effort towards facilitating reuse.

Pattern User Role:

From the user perspective, the biggest challenge for them is to 1) find good patterns, and 2) combine them to generate new design artifacts. Both are challenging tasks for the pattern user. Finding few relevant patterns can be a challenge if we look at the huge number of patterns, mostly represented in reach unstructured, and heterogeneous formats, only suitable for human consumption, but not for machine processing. As discussed earlier, pattern users can only process limited amount of text, so sifting through- and selecting among huge number of patterns become a daunting task. The other challenge is probably equally hard. Combining patterns together in a correct and coherent way would not be easy if it is all presented in isolated patterns or isolated collections of patterns. Research has shown that in several cases, patterns are often applied in a wrong way, or combined in a less-than optimal new design [11] and [16]. Pattern users will have limited capabilities once the patterns are generated in rich document format.

Our approach introduces a new intermediate role of "Pattern Disseminator". This role can be assumed by a human person, or a software agent (Figure 7).

A Human Pattern Disseminator would take the initial pattern artifacts and rewrite them in a structured way that will allow for automatic lookup and integration of patterns. A software agent would interact with the pattern author and –

working at the source- will help pattern authors generate their patterns initially as semi-structured artifacts that render themselves easily to software tools that help the pattern user in finding, filtering, and combining the selected patterns into new design.

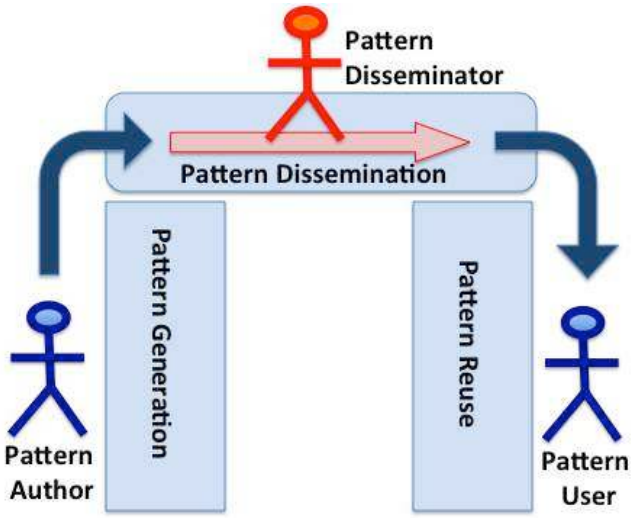


Figure 7: The Role of "Pattern Disseminator"

To validate this approach, we constructed an initial conceptual design with a detailed proof of concept to demonstrate the work. Our system is based on an integrative approach of pattern dissemination that complement the scattered efforts of writing patterns. Following a proof of concept, we built a delivery system in the form of a digital library based on a database of patterns, a transformation logic for any desired algorithms, pattern processing and transformation tools, as well as an interface to offer these functionality to the user. We selected XML as the language to represent patterns for several reasons. XML is becoming the norm for data interchange, and is used extensively in IT systems. Besides the XML language, Several XML-based technologies are emerging to enhance XML-compatible information systems.

The system architecture is shown in Figure 8. The data flow starts at the top right corner in the form of the existing pattern collections. Looking at the data pathways depicted in the figure, we see that the input information is now bypassed from the direct path (from pattern collections directly to user preparations block, marked as A) into the system pre-storage phase, the pattern corpus, and to the rest of the system; marked as B

The User Preparations block, depicted as pathway A, represents the cognitive activity by the user to search for patterns and read through the text, analyze its contents, and figure out how to apply which patterns in the design. This is a fundamental observation in our study. We can evaluate the difference between the two processes, referred to as pathways A and B. If we considered them as representing the dissemination processes between pattern authors and pattern users, we can use the Capability Maturity Model, CMM, of the Software Engineering Institute to briefly evaluate them.

-Process A: This process does not follow a particular approach of dissemination except for relying on users preparations (looking up patterns, understanding them, and applying them in an ad hoc approach). We evaluated this to be a CMM level 1.

-Process B: As suggested by the envisioned system, there is a process in place to help users interact with patterns in a structured way. Moreover, this process relies heavily on feedback, and is constantly changing, as shown in implementing and validating the 7Cs process within the system [17], [18], a CMM level 4

The main modules of the system, as shown in Figure 5 are:

-The Pattern Corpus, a raw collection of patterns before being reformatted and saved in the XMLDB.

-The Data Models, used in the semantics of the XML Rewrites

-The Expert Preparations, used as help in rewriting pattern information to ensure the integrity of the contents and avoid redundancies and inconsistencies.

-The System Process (the 7Cs process), a structured method applied across the system

-The XML Subsystem (XML rewrites, XMLDB and XML semantics/scheme), The core constituent of the system, and the backbone of the three-tiered design.

-The Processing and the Presentation Layers are the middle- and front-tiers respectively.

One of the main aspects of our system is to be able to automatically process the contents of patterns to alleviate the user from this cognitive load. The main purpose is to have a scalable capability to effectively process large number of patterns. Representing patterns in natural language defeats this purpose. XML is much more suitable in this regard due to its machine-readability. It is based on the fundamental concept of automatic processing. Goldfarb [19], the inventor of SGML (the parent, and superset of XML) explains that the vast majority of XML documents will be created by computer programs, and processed by other computer programs, then destroyed. Humans will never see them. The first step is to rewrite patterns in an XML format. A simple XML syntax rewrite (like PLML, Pattern Language Markup Language, [20] can be a small step in this direction, but -by itself- it will not do much good as will be discussed later. To achieve global interoperability, we needed to design the semantics and the behavior modeling behind the XML syntax, according to concrete data models.

Database is a core constituent of the system. An obvious choice for data store is an XMLDB. We implemented part of the system using it. The steps are as follows:

-Phase **1, No-Database System:** By temporarily including the data inside the system (hard-coding the data). We have two major prototypes: In the UPADE project. We developed a systematic approach to glue patterns together, support the integration of patterns at the high design level and automate pattern composition. UPADE [21] generated program elements from an extensible collection of "template" patterns.

In the second prototype, [22], [17], we experimented with prototypes of the interface aspects, and we also hard-coded patterns inside the prototypes. At this stage we were able to test and refine the design and functionality of the system.

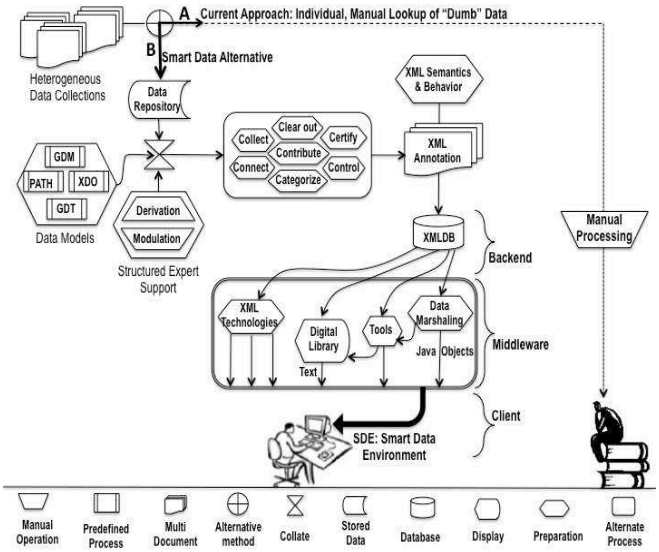


Figure 8: System Conceptual Overview

-Phase 2, Flat-File-Database: To start building an independent database module, external to the system, and to connect it to the rest of the system. We used "XMLSpy" to prototype the XMLDB concept. The XMLSpy allowed us to build an actual, partially functional XMLDB based on flat files as the internal storage medium. The major advantage of this approach is the extreme simplicity of the system, which allowed us to further refine the fundamental concepts of the system without being carried away by the implementation details. The main drawback was its limited scalability and unacceptable performance, as will be explained in step 4.

-Phase 3, Web-based System: To implement the functionality of the system as built around external XMLDB on a Web-based environment. We implemented several aspects of the system using the database we have from phase 2. Using 3- tier system architecture (data- logic-, and presentation tiers), DHTML and Web technologies (CGI), we built a web-based distributed system to test our XMLDB. We implemented the system to interact with pattern users by offering a search interface at the presentation tier.

As the user queries the system for some patterns, the system middle tier (at the web server) processes the request by accessing the data tier at the back end, and returning results by dynamically generating a web page with search results.

The 7Cs Process

One of the main aspects of process-oriented approaches is the dependence on a predetermined process. Gaffar [23] describes the "7Cs Process" as a systematic approach to adding structure to rich but ultra heterogeneous data. An established design process instigates quality design by allowing designers to follow structured methods in their activities. In our approach, we emphasized the need for both dissemination and assimilation processes. In this paper, we briefly summarize the 7Cs dissemination process as completely decoupled from any specific assimilation process [8], [11]. This allows it to offer patterns that can be integrated simultaneously in several assimilation processes. "Free patterns" that do not belong to any process at all are hard to integrate in design. Similarly, "proprietary patterns" that are specifically tailored to manually fit one design process using few specific examples defeat the main purpose of pattern generality and abstraction. We see that a pattern can be integrated in several assimilation processes by properly encapsulating its knowledge and presenting its behavior through a well-defined interface. Any assimilation process can then lookup pattern objects and select the appropriate ones using different search criteria. The selected pattern components can then be integrated in new designs or used to generate code fragments. The 7Cs is "a structured process to replace the huge cognitive load of manipulating HCI patterns with a dissemination system of smart patterns" [23]. The 7Cs process identifies both logical and physical aspects of the system. A logical process focuses on *what* actions and activities need to be done. A physical process complements the logical process by specifying the roles associated with the process, and details *who* is going to do what [24], [25], [26]. As part of the pattern reuse problem is associated with missing roles in the dissemination activities (all left to the user), the 7Cs process addresses both how these activities need to be done, and who should be doing each of them. Briefly, the 7Cs process moves gradually from current unplanned generation and use of patterns into building an automated pattern collection. The process comprises seven steps [23]:

*Step 1) **C**ollect: Place Different Research Work on Patterns in One Central Data Repository*

*Step 2) **C**learout: Change from Different Formats/Presentations into One Uniform Style*

*Step 3) **C**ertify: Define Domain Boundaries and Clear Terminology*

*Step 4) **C**ontribute: Receive Input from Pattern Community*

*Step 5) **C**onnect: Establish Semantic Relationships between Patterns Using a Formal Relationship Model*

*Step 6) **C**ategorize: Define Clear Categories for Patterns that Map them into Assimilation Processes*

*Step 7) **C**ontrol – Build Machine Readable Format for Software Tools*

The ultimate goal of the 7Cs process is to allow designers to use the leverage of tools to process large number of patterns efficiently, and to be able to assimilate them effectively into a new design. Patterns that are both human- and machine-readable are the main outcome of the process of pattern *dissemination* and the first step towards *assimilating* them.

DATA MANIPULATION TECHNIQUES

The XML technology is evolving in a rapid pace. Just few years ago, as we started the project, the XMLDB was not supported by Oracle, an industrial norm in databases; or by Apache, an industrial norm in Web servers. The current and future trends are now determined by the new development, as explained in phase 4 of the system development.

-Phase **4** is to use a more advanced, fully-fledged XMLDB to overcome difficulties we discovered in phase 3. XMLSpy internally implemented the XMLDB as flat files. While invisible to the user, this solution is not scalable. It relies on the file structure of the underlying operating system, and is affected by its performance. Besides, inserting to- or deleting from the DB incurs the penalty of rewriting the whole file. The other major obstacle was that some XML technologies we used are still in their early specification stages, and are immature for full implementation. For example, XQuery is the query language supported by XMLDB. We used it to query our XMLDB. While easy to use, it still does not support insertion or deletion, so we had to implement them manually in phase 3.

In phase 4, we are rebuilding the database to be less dependent on immature XML technologies while still offering all XML functionality upfront. The idea is to use newly emerging commercial- or open source systems that are robust enough to support the full functionality of our XMLDB pattern system, while delegating the internal database implementation to the system, be it native- or non-native XML database. Native XMLDB offers an XML interface to the database, and store the database internally in pure XML format. Non-native XMLDB offers an XML interface to the database, similar to the native one, but transforms it internally to relational database, or to a proprietary database format before storing it physically on the hard disk.

Among the hundreds of commercially available XMLDB today, we compared three common alternatives:

1- ORACLE RDBMS with support to XMLDB. An authoritative commercial application that offers a fully functional XMLDB interface, while translating it internally to an RDBMS (a non-native XMLDB)

2- ZOPE, an open source industrial application that implements XMLDB using internal proprietary RDBMS, but provides fully functional XMLDB interface (also a non-native XMLDB).

3- Apache-Xindice XMLDB, a powerful open source native XMLDB technology, supported by Apache (one of the best known and used web servers).

CONCLUSION

Patterns are useful data encapsulation mechanisms that can help UI designers improve the quality of their work by reusing well-known practices of domain experts. For our purpose of building a proof of concept for programmable data contents, they serve as a good example due to their nature of highly unstructured presentation and highly heterogeneous multi-modal contents. They also rely heavily on reuse –at least in principle- allowing us to focus on the amount of reuse as an indicator of success of dissemination. In the pattern community, knowledge dissemination activities focus mostly on generating patterns, with little or on plan on how to deliver them or how to reuse them in new designs. The proliferation of HCI –Human Computer Interaction- patterns is associated with redundancies and inconsistencies that often confuse and overwhelm the user. More traditional approaches like database domain have provided methods to deal with these problems by allowing data to be organized and reused efficiently using normalized database tables and structured queries. We have proposed and built a proof of concept to complement the process of generating patterns by reducing inconsistencies and conflicts, and by providing patterns in a reusable format. We have designed a process that can help in the assimilation and dissemination of patterns. We have implemented a system that -in applying the process- will help novice users receive patterns in an orderly fashion and be able to effectively use them in their design artifacts. The system relies on an underlying XML database and the process will help feed the patterns into the database and rewrite them in a semantically reusable format. The database and the system allow us to add tools to the interface to automatically process patterns in a user-friendly design environment and to connect them to other XML- or UML-based tools.

7 REFERENCES

1. Erickson, T. Lingua Franca for Design: Sacred Places and Pattern Languages, DIS 2000: Proceedings of Designing Interactive Systems, New York, August 17-19, ACM Press, Brooklyn, NY, USA, pp.357-368 (2000).
2. The HCI Patterns. At http://www.hcipatterns.org/patterns.html, accessed on July 8, (2003).
3. Alexander, C., Ishikawa, S., and Silverstein, M. A. Pattern Language: Towns, Buildings, Constructions. New York: Oxford University Press (1977).
4. Stephanidis, C., ed. Human-computer Interaction: Theory and Practice. Part I. Vol. 1. Routledge, (2003).
5. Gamma, E., Helm, R., Johnson, R., and Vlissides, J. Design Patterns: Elements of Reusable Object-Oriented Software, Addison Wesley, Boston, Massachusetts, USA. (1995).
6. Gaffar, A., Tridane, A., & Darwish, E. M. (2013). The Failure and Success of Unstructured Data Reuse a Pragmatic Analysis. In The International Conference on Digital Information Processing, E-Business and Cloud Computing (DIPECC2013) (pp. 16-25). The Society of Digital Information and Wireless Communication.
7. Gaffar, A., Tridane, A., & Pribadi, K. (2013). Smart Data: A System for Generating Application-Agnostic Interactive Data. In The Second International Conference on Informatics Engineering & Information Science (ICIEIS2013) (pp. 246-254). The Society of Digital Information and Wireless Communication.
8. Gaffar, A., and Moha, N. "Semantics of a pattern system." STEP 2005 : 211 (2005).
9. Moha, N., Qing,, L., Gaffar, A. and Seffah, A. "Enquête sur les pratiques de tests d'utilisabilité." In Proceedings of the 17th international conference on Francophone sur l'Interaction Homme-Machine, pp. 115-122. ACM, (2005).
10. Gaffar, A., Sinnig, D., Seffah, A. and Forbrig, P. "Modeling patterns for task models. In proceedings of TAMODIA." In 3rd International Workshop on Task Models and DIAgrams for user interface design, pp. 99-104. (2004).
11. Gaffar, A., and Seffah, A. "An XML Multi-Tier Pattern Dissemination System." 740-744 (2005).
12. Gaffar, A., Moha, A. and Seffah, A.. "User-Centered Design Practices Management and Communication." In Proceedings of HCII. (2005).
13. Sinnig, D., Gaffar, A., Seffah, A., and Forbrig, P.. "Patterns, Tools and Models for Interaction Design." MBUI 103 (2004).
14. Moha, N., Gaffar, A., and Michel, G. "Remote usability evaluation of web interfaces." Human Computer Interaction Research in Web Design and Evaluation. P. Zaphiris and S. Kurniawan. Hershey, PA, Idea Group Publishing: 273-289 (2007).
15. Gaffar, A. "Design of a Framework for Database Indexes." PhD diss., Master thesis, Department of Computer Science, Concordia University, (2001).
16. Butler, G., Chen, L., Chen, X., Gaffar, A., Li, J., and Xu, L. "The Know-It-All Project: A case study in framework development and evolution." Domain Oriented Systems Development: Perspectives and Practices: 101-117 (2002).
17. Gaffar, A., Sinnig, D., Javahery, H. & Seffah, A., (2003). MOUDIL: A Comprehensive Framework for Disseminating and Sharing HCI Patterns. Position Paper in ACM CHI 2003 Workshop: Perspectives on HCI Patterns: Concepts and Tools, Ft. Lauderdale, Florida, (2003).
18. Metzker, E., Seffah, A. and Gaffar, A., (2003). Towards a Systematic and Empirical Validation of HCI Knowledge Captured as Patterns. Proceedings of HCI International, the 10th International Conference on Human-Computer Interaction, Crete, Greece, vol. 1, Theory and Practice, LEA, (2003).
19. Goldfarb, C. F., and Prescod, P., (2004). The XML Handbook, 5th edition, Prentice Hall PTR, Upper Saddle River, New Jersey, USA, (2004).
20. Fincher, S., & Finlay, J.,. CHI 2003 Report: Perspectives on HCI Patterns: Concepts and Tools; Introducing PLML. Journal of Interfaces, No. 56, September BCS HCI Group publishers, Canterbury, Kent, UK, pp. 26-28. (2003)
21. UPADE,. User Patterns Automated Design and Engineering at http://hci.cs.concordia.ca/www/hcse/projects/, accessed on Mai 12, 2003 (2000).
22. MOUDIL Project, at http://hci.cs.concordia.ca/moudil/, accessed on June 3, (2003).
23. Gaffar, A. The 7Cs: An Iterative Process for Generating Pattern Components, electronic proceedings of HCI International, The 11th International Conference on Human-Computer Interaction, July 22-27, 2005, Las Vegas, Nevada, USA.

24. Hoffer, J. A., George, J. F., Valacich, J. S., Modern system analysis and design. 3rd edition, Prentice Hall (2002).
25. Whitten, J. L., Bentley, L. D., & Dittman, K., C.. System analysis and design methods,5th edition, McGraw Hill Irwin (2001).
26. Ackerman, Mark S., & Thomas W. Malone.. Answer Garden: A tool for growing organizational memory. Proceedings of the ACM Conference on Office Information Systems: (pp.31-39) (1990).

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 24-34
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

A New Approach Towards Integrated Cloud Computing Architecture

Niloofar Khanghahi[1] and Ramin Nasiri[2] and Mahsa Razavi[3]
[1]Department of Computer Engineering, Islamic Azad University, Central Tehran Branch, Tehran, Iran
[2]Department of Computer Engineering, Islamic Azad University, Central Tehran Branch, Tehran, Iran
[3]Department of Computer Engineering, Islamic Azad University, Central Tehran Branch, Tehran, Iran
[1]niloofar.khanghahi.2011@gmail.com
[2]r_nasiri@iauctb.ac.ir
[3]mah.razavi@iauctb.ir

ABSTRACT

Today across various businesses, administrative and senior managers seeking for new technologies and approaches in which they can utilize it, more easy and affordable and thereby rise up their competitive profit and utility. Information Communications and Technology (ICT) is no exception from this principle. Cloud computing concept and technology and its inherent advantages has created a new ecosystem in the world of computing and is driving ICT industry one step forward. This technology can play an important role in an organization's durability and IT strategies. Nowadays, due to progress and global popularity of cloud environments, many organizations moving to cloud and some well-known IT solution providers such as IBM and Oracle have introduced specific architecture to be deployed for cloud environment. On the other hand, using of IT Frameworks can be the best way for integrated business processes and other different processes. The purpose of this paper is to provide a novel architecture for cloud environment, based on recent best practices and frameworks and other cloud reference architecture. Meanwhile, a new service model has been introduced in this proposed architecture. This architecture is finally compared with little other architecture in a form of statistical graphs to show its benefits.

KEYWORDS

Cloud, Cloud Computing, Reference Architecture, ICT Frameworks, Everything as a Service

1 INTRODUCTION

The confluence of technological advances and business development in internet broadband, web services, computer systems and applications has created a complete storm for cloud computing during the past decade. Nowadays, cloud is one of the popular solutions for people who are looking for rapid implementation and/or easy and affordable deployment and development methods [1-7].

Cloud computing is a type of parallel, virtual, distributed, configurable and flexible systems, which refers to provision of applications such as hardware and software in virtual data centers via internet. Cloud computing services are configurable and customers pay fees based on the use of resources and services [1-7].

The most important element of cloud structure is a server which is the brain behind the whole processes in Cloud. Cloud is the major important and nearly new model for access to distributed computing resources [1-7].

Pay per use, scalability, use the Internet technology, self-service based on the demand, high performance, quick to implement, easy to maintain and update are key benefits of cloud computing [1-7].

And the data recovery, lack of control over cloud services, service level agreements, legal problems, different architectures, audit, Reviews and evaluation of the performance cloud computing environment are the major disadvantages of cloud computing [1-7].

Typically there are three models for delivery of cloud services. *Software as a Service (SaaS)*, such that in this model, users use the launched application on cloud infrastructure. *Platform as a Service (PaaS)*, in which users rent platforms or operating systems and they can expand their required programs on it in an on demand policy [1-7]. *Infrastructure as a Service (IaaS)*, which

this model is associated with a virtual engine and users can access to infrastructures with virtual machine [1-7].

Also the decision on implementation of cloud is important. There are four main cloud deployment models. *Public*, the most common model in the cloud deployment model. Large enterprise is owner of a large cloud infrastructure and provides services to users. *Private*, this model simulates a private network. It is suitable for an organization's infrastructure. *Community*, in this model, enterprises which have common policies, goals and concerns share infrastructure of cloud. *Hybrid*, this model is a combination of two or more cloud deployment models. In this model, resource management may be internal or external [1-7].

There are several reference architectures for cloud computing, such as IBM, Oracle, Cisco, HP, NIST, etc. which can be said that, IBM, Oracle and NIST are the most complete and famous. In following section the IBM architecture will be briefly introduced.

Recently, ICT service management Specially ITIL, IBM, which are two powerful frameworks among others and eTOM which is ITU telecommunication framework, etc. has become more structured [8-15]. Therefore using of these frameworks in cloud environments, which is a service oriented environment, can be quite useful.

The rest of this paper is organized as:

Section 2 introduces related works briefly. Section 3 introduces IBM cloud computing architecture as the focal reference model. Later on in section 4 the main purpose of this paper is provided. In this section and Subsections the new architecture of cloud computing based on famous cloud architectures and eTOM, ITIL V3 and IBM reference frameworks are introduced and a comparison between the new architecture and other architectures are presented in a form of statistical graphs. And finally conclusions and future works are listed.

2 RELATED WORKS

Cloud Computing makes the dream of computing real as a tool and in the form of service. This technology has realized service-oriented idea too. Due to needs of organizations to move toward new technologies and reach service orientation and also changes in customer demand, cloud computing has been in the center of attentions and organizations have turned to it. For cloud computing several reference architecture have provided, which few of them are the basis of the proposed architecture in this paper. Some of these reference architectures will be introduced briefly in the following:

- IBM reference architecture, which is the most popular and useful for cloud computing environments [16]. This architecture will be introduced in the next sections.
- HP CloudSystem architecture which is provided by HP Company. This architecture is an integrated system for building and managing services across all cloud environments which combines servers, storage, networking and security together with an integrated approach to automate the application and infrastructure lifecycle. There are three layers of supply, delivery and demand for service delivery therein. The services are provided as three products in those layers, which are called HP CloudSystem Matrix, HP CloudSystem Enterprise and HP CloudSystem Service Provider [17].
- NIST reference architecture for cloud computing is one of the most well-known reference architecture. In the architecture, there are five main roles, which are called cloud consumer, cloud provider, cloud carrier, cloud auditor and cloud broker. In this architecture, there is also an orchestration module which is responsible for the composition of the system components to support their activities and management [18].
- Another important and popular reference architecture is Oracle. In this architecture, cloud management capabilities and policies is developed in five cloud business management, security and policy management, cloud operations, orchestration and design-time categories and also there is cloud computing portfolio to have roadmap for selection is shown in this architecture [19].

- Cisco, Rackspace, Microsoft, etc. are reference architectures for cloud computing environments.

Nowadays, some best practices frameworks have been created in order to standardize and organize services, processes and anything associated with ICT. These frameworks and best practices are too useful and suitable for service oriented and IT-based environment such as cloud computing environments. Some of the most important frameworks are introduced briefly in the following:

- The Information Technology Infrastructure Library (ITIL) is a set of best practices for IT service management (ITSM) that focuses on aligning IT services with the needs of business. ITIL has five core parts; each of these parts covers an ITSM lifecycle stage. This five core parts are Service Strategy, Service Design, Service Transition, Service Operation and Continual Service Improvement [8-13].
- The IBM Process Reference Model for IT (PRM-IT) is a generic representation of the processes involved across the complete IT Management domain. PRM-IT presents a framework that uses eight process categories which are Governance and Management System, Customer Relationships, Direction, Realization, Transition, Operations, Resilience and Administration [14].
- Enhanced Telecom Operations Map (eTOM) is Business Process Framework which defines a model for the telecommunications industry. The model describes business processes required by a service provider and defines key elements and how they interact. This framework fall into three broad sections which are Strategy, Infrastructure & Product, Operations and Enterprise Management [15].
- Microsoft and HP are another reference framework for IT and service oriented environments.

3 IBM CLOUD COMPUTING REFERENCE ARCHITECTURE

The IBM cloud computing reference architecture (CCRA) which is shown in figure 1, introduce the fundamental component of cloud environment.

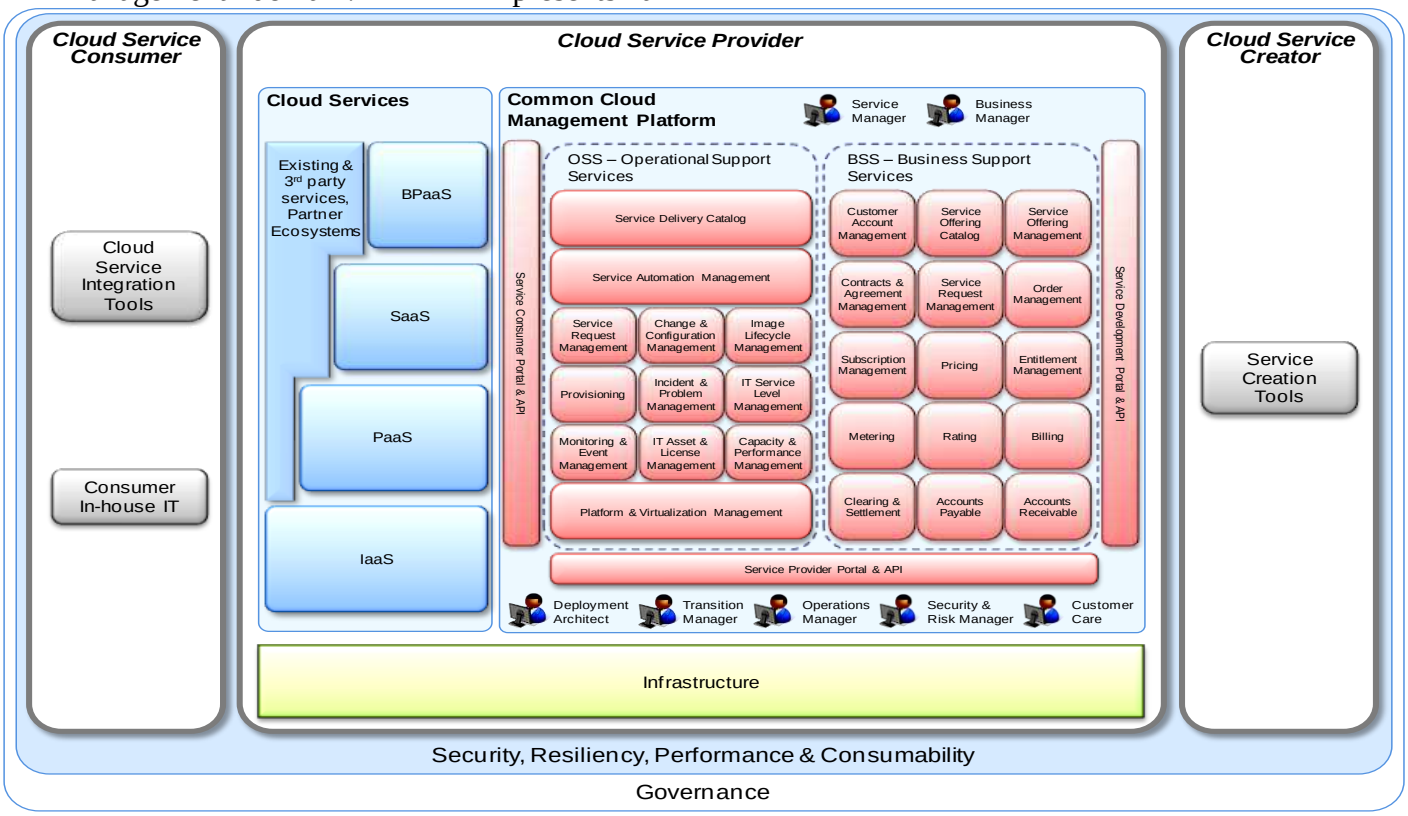


Figure 1. IBM cloud computing reference architecture

It should be mentioned that, the structure of this architecture is modular [16].

In this architecture there are three main roles, which each of these roles is to present a single person or organizations and sub-roles may be defined, based on project scenarios.

A cloud service consumer role consumes cloud service instances. Cloud service provider has responsible to supply and/or provide cloud services. This role and its sub-roles are defined by ownership of a common cloud management platform (CCMP). The third role is cloud service creator which is responsible for creating services in cloud environments [16].

The core components of this architecture are mainly cloud services, common cloud management platform and infrastructure [16].

Cloud services may represent any type of IT capability which is provided for cloud consumers. The management functions for cloud services are defined as part of CCMP. As it can be seen in figure 1, the internal components of CCMP are categorized in operational support services (OSS) and business support services (BSS) and this component include processes and services which are required for cloud management [16].

4 PROPOSED ARCHITECTURE

In this section our proposed cloud computing architecture is presented. It should be mentioned that, the foundation of this architecture is based on the existing architectures and use of best practices of ICT frameworks, such as IBM, ITIL, eTOM, etc.

Figure 2 shows the conceptual and high level model of proposed architecture. In the following sub-sections of this paper, components of this architecture will be introduced briefly.

As can be seen in figure 2, the proposed architecture combines the best component of existing architectures and in addition it has the following advantages:

- A new service layer added to this architecture. This layer offers IT and business management services and processes and other related services and processes to business customers.
- With this new service layer, organizations and business clients are separated from normal client and better services is provided to them.

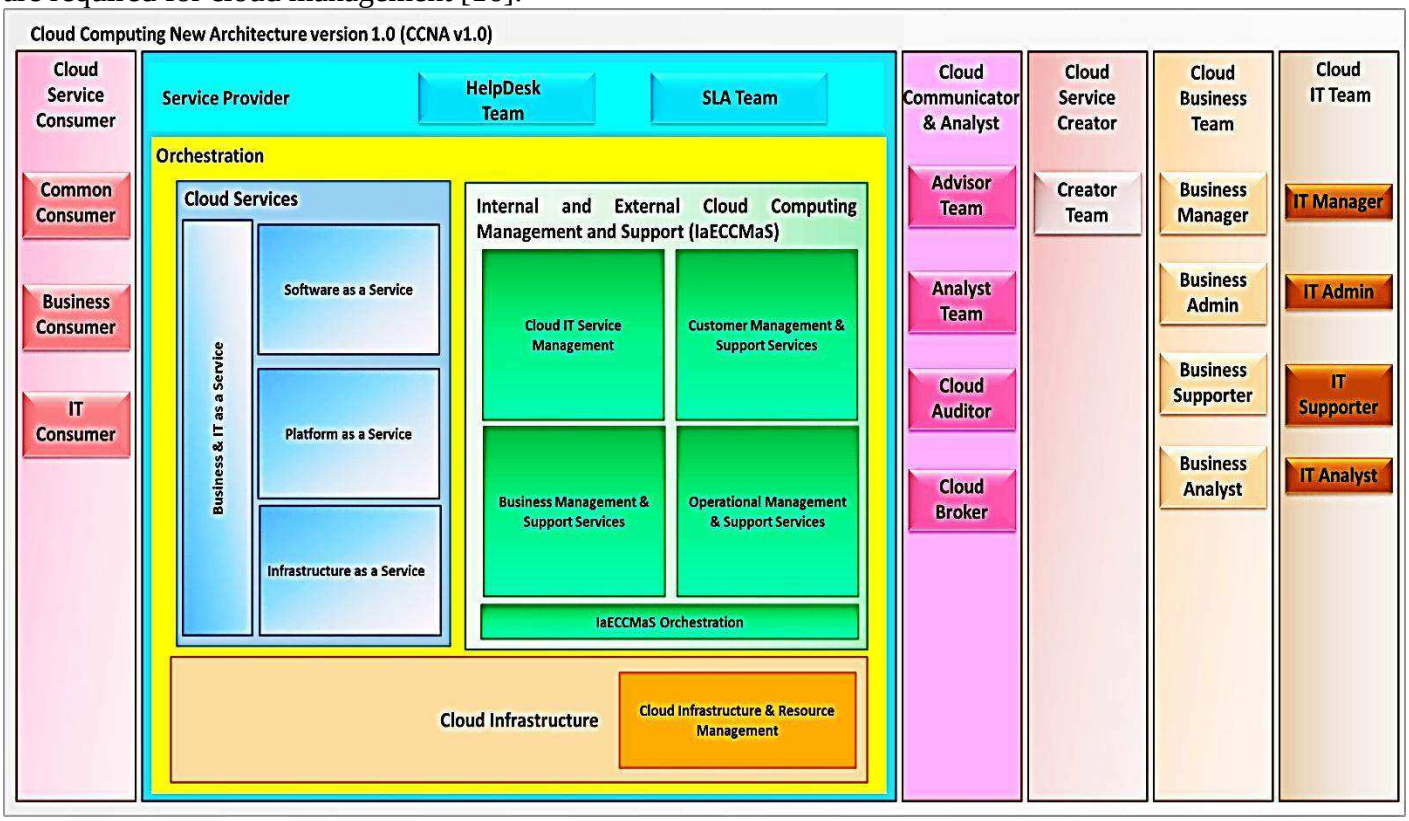


Figure 2. The conceptual and high level model of proposed architecture

- New service layer is enabling more customization and this customization is better for organizations.
- Cloud management layer of this architecture used ICT frameworks and best practices such as eTOM, ITIL, IBM, etc. to manage service layers, infrastructure, customer and everything related to cloud.
- With existing processes and services in I aECCMaS, the management of internal cloud components and services which provided to users will be upgraded and be more structured.

4.1 Components of architecture

As depicted above, the proposed architecture has six major roles and also has four major components. Each major component has sub-components which are services and processes. These components are called as orchestration, cloud services, cloud infrastructure and internal and external cloud computing management and support (IaECCMaS), which would be discussed in each of the following sub-sections.

4.1.1. Cloud Services Component

This component provides services to client in four types. In SaaS, like all common architectures, software are provided for clients. These kinds of services have limited customizable and this feature is not being sufficient for organizations and business customers. Because enterprise customers are more willing to get service packages and do not being forced to prepare hardware, software and other related requirements. In PaaS, a platform is provided to clients and clients deploy their own services or software on top of the platform. In this model, clients will install their favorite software or services and service provider does not involve in it. Service variation in this model is low so this feature is not being favorable for organizations and business customers too. In IaaS, the infrastructure and hardware provided to clients. So due to provide hardware only, it is not being optimal for organizations and business customers.

The new service layer in this architecture is business and IT as a service. In this layer, services, processes, packages and systems that contribute to business and IT management and also other related services provided to clients are shaped in type of packages or services, which could be favorite many organizations, especially IT related organizations.

This layer is a hybrid and this is the advantage of it. This means that cloud services and also organization's specific services can be deployed in this layer. The layer can also design specific services for organizations.

Another benefit of this layer is that organizations can have hybrid package services, which is the combination of Iaas, SaaS and PaaS. B2B services and ERP systems can be considered as examples of services and the package of cloud services, organization specific services and storage media for them may be considered as example of hybrid package in this layer.

It should be mentioned that, clients have only a single interface in order to notify and negotiate with service provider. This interface called service provider help desk. The role of service provider is to support these services and interacts with other roles in order to track and manage anything related to the service and resolve possible problems of customers and service consumers to cloud service consumer directly consume the services of this component. HelpDesk team is unique team that has direct relationship and negotiates with customers and also SLA team is responsible to contracting with customers and does related contracts conditions but not directly. These two teams are part of the Service provider.

4.1.2. Cloud Infrastructure Component

Infrastructure component provide infrastructure, hardware, platform and any related requirements for other components and associated items. It should be mentioned that, infrastructure may be servers, memory, storage, network facilities, etc. Another component in this part is cloud infrastructure and resource management. This component and their processes are responsible to

manage infrastructure and resource and also to orchestrate when providing infrastructure and resource. Figure 3 shows the component, this component involves processes and resources and infrastructures pool. Most of processes in this component are taken from eTOM framework.
These processes are generally in charge of capability definition or requirements to deploy resources or infrastructures, development and retirement of resources or infrastructure, provisioning, allocation, installation, configuration, activation, testing and recovery of specific infrastructures and data centers, etc.
Also performance and trouble tracking and managing of deployed and used or during deployment resources to increase performance and improve resources in order to better serve is another responsibility for this processes.
Resource Orchestration is responsible to undertakes any type of orchestration and any necessary interactions and communications of resources with cloud infrastructure and resource management and other components of architecture in order to resource provisioning and better services.
All roles exception cloud service consumer interact with this component directly or indirectly. As an example, in order to create a service and deploy it in cloud service layers, cloud service creator must inquiries available resources and declares resource requirements with cloud infrastructure component.

4.1.3. Internal and External Cloud Computing Management and Support Components

This component is responsible to support and manage all kind of processes and activities which manage and support cloud service layers and each component related to cloud environments. These processes are classified in Cloud IT Service Management, Customer Management & Support Services, Business Management & Support Services, Operational Management & Support Services and IaECCMaS Orchestration. It should be granted that, OSS and BSS components of IBM architecture are completely supported in these components in addition to more processes. Figure 4 shows this component and its sub-processes and these processes are introduced in sections below. All roles exception cloud service consumer interact with this component directly or indirectly.

4.1.3.1. Cloud IT Service Management Component

All type of processes and services which are related to manage IT services in cloud service layer and internal IT processes in cloud environment should be located in this component. Figure 4 shows processes and services of this component.

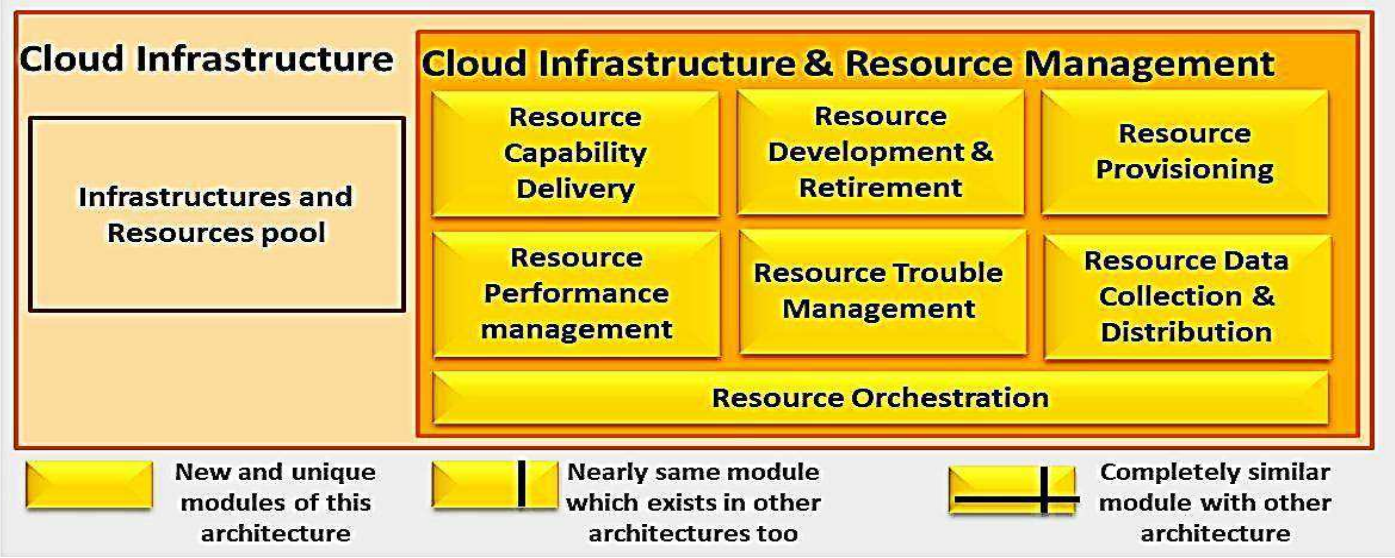


Figure 3: Components of Cloud Infrastructure and Resource Management

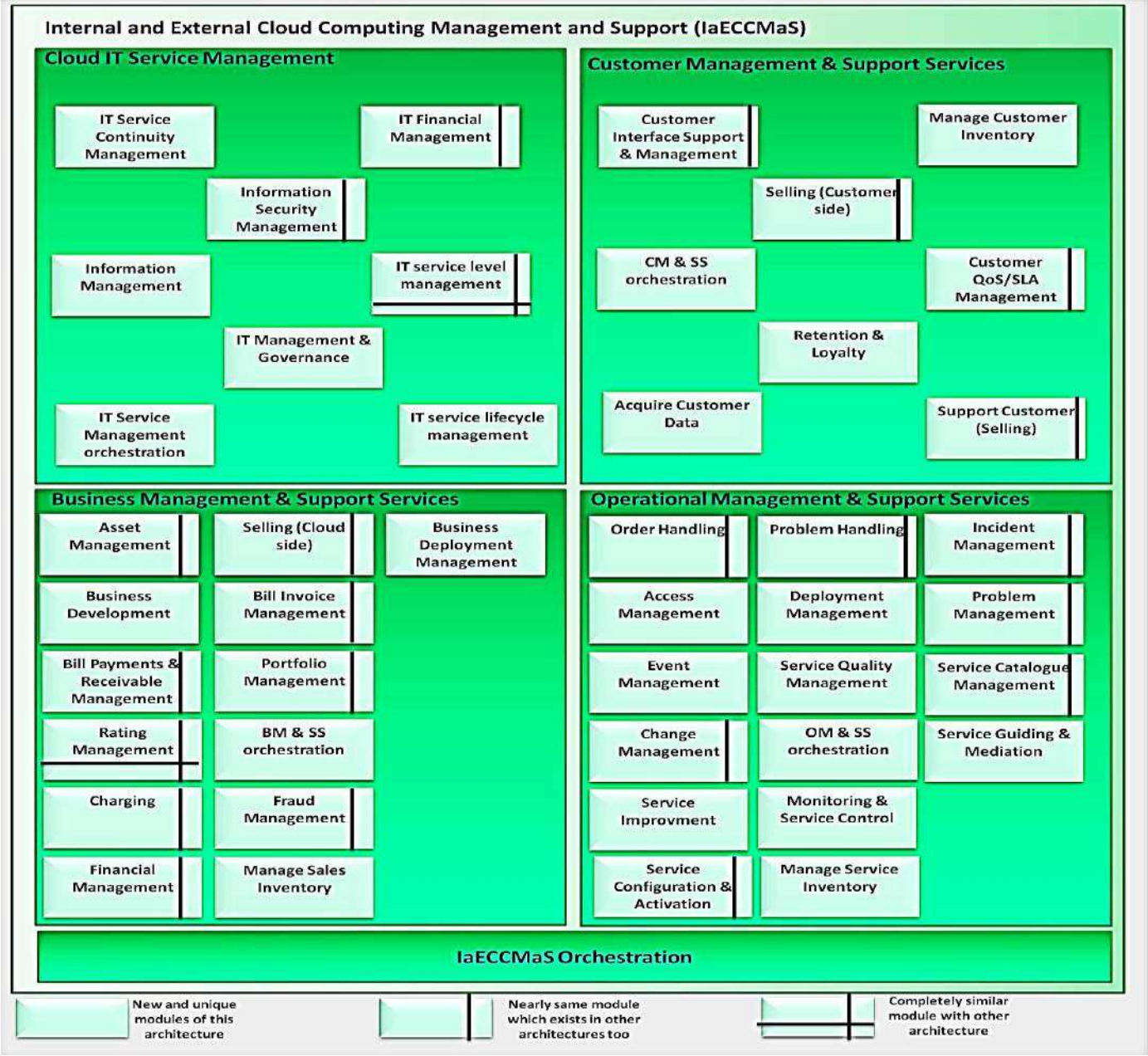


Figure 4: Internal and External Cloud Computing Management and Support components

Most of processes in this component are taken from ITIL and IBM frameworks.

These processes and services are generally responsible to support cloud IT services, establish IT policies, and manage security of IT information in cloud environment, establish affordable financial procedure of cloud IT service, budgeting, manage IT service lifecycle and IT service level, reporting and tracking of IT services and etc. in order to provide better IT services and continual improvement of these kind of services.

It should be mentioned that, information security management process is too important, because customer's information, especially enterprise customer's information should be completely safe. As an example, enterprise which used ERP systems that implemented in cloud Business & IT as a Service, expected that their financial, internal and transactions related information to be completely secure. Cloud service providers are responsible for the security in accordance with the SLA.

Information Management process is specific process of cloud environments which manage information that later will be used in reporting and financial processes.

IT Service Management orchestration processes are responsible to orchestrate all type of information, processes and anything related to IT process with other cloud components. Cloud IT team roles interact with this component directly. As an example, a new IT service which designed by Cloud IT Team and created by Cloud Service Creator in order to consumers or internal cloud use, is managed by IT Service Life Cycle Management processes and these processes also validate, reporting, policymaking and financial review this service too.

4.1.3.2. Customer Management & Support Services Component

All type of processes and services which are related to manage and support customers and there interactions and interfaces in cloud environment will be located in this component. Figure 4 shows the components of these processes.
Most of processes in this component are taken from eTOM framework.
These processes and services are generally responsible to manage customer inventory, monitor and report customer service usage, capture and record all customer data, manage customer interfaces, etc.
Selling (Customer side) processes are responsible of performing all type of sales operations related to client services in order to better quality of service.
Support Customer (Selling) processes are responsible to performing all type of operations related to customer charging, payment and financial functions.
Customer Management & Support Services orchestration processes are responsible to orchestrate all type of information, processes and anything related to customers with other cloud components. All roles interact with this component directly or indirectly. Service consumers are also indirectly interacting with this component by Help Desk.

As an example, when a customer is having problems in using cloud services, they report these problems to service provider HelpDesk by interface. HelpDes team has reviewed problems and resolves it, if it is possible, otherwise problems escalate depending on type or priority and then related team resolve problem by help of processes and services. It should be reminded that, orchestration between processes and services will be done through orchestration components and associated processes.

4.1.3.3. Business Management & Support Services Component

All type of processes and services which are related to manage and support business services in cloud environment will be located in this component. Figure 4 shows the components of these processes.
Most of processes in this component are taken from eTOM and ITIL frameworks.
These processes and services are generally responsible to manage all financial and non-financial assets in cloud environment, account rating, decision about development of services, financial and charging activity related to business in cloud, develop service portfolio, manage bill payment and receivable, prevent fraud, manage sales inventory, etc.
Due to cloud services are provided in internet-based infrastructure and this this environments is too insecure, it is very important to prevent fraud and establish related policies to prevent and manage fraud in cloud environments and services which provided to customers. This is important because, stealing information and disclosed privacy is very important for customers especially enterprise customers. They want full security is satisfied and customers to be concered about this issue when thinking and make decisions about cloud.
Fraud management process in this component responsible to establishing, managing and policymaking activities and capabilities related to manage, detect and prevent fraud in cloud environments.
Business Management and Support Services orchestration processes are responsible to

orchestrate all type of information, processes and anything related to business with other cloud components.
Selling (Cloud side) processes are responsible to performing all type of sales operations related to business and have strong interactions with Selling in customer's component. Cloud Business team roles interact with this component directly.

4.1.3.4. Operational Management & Support Services Component

All type of processes and services which are related to manage and support services operations in cloud environment will be located in this component. Figure 4 shows the components of these processes.

Most of processes in this component are taken from eTOM, ITIL, IBM and Microsoft frameworks.
These processes and services are generally responsible to accepting and issuing customer orders, verifying orders, receiving trouble reports from HelpDesk and resolve it, manage service inventory, administer service improvement, service mediation, service activation, manage provided services and available in near future, supporting services by manage incident, problem, event, change, etc.
Operational Management & Support Services orchestration processes are responsible to orchestrate all type of information, processes and anything related to operational business with other cloud components. All roles specially cloud communicator and analyst interacts with this component directly.
As an example, Cloud Communicator & Analyst, use these processes and services and their operational reports and other related processes of other components in order to service continual improvement and associated analysis for improvements or new services offering.

4.1.3.5. IaECCMaS Orchestration Component

This component is responsible to orchestrate between internal IaECCMaS processes and services, but orchestration between external components and this 5 components in not possible only with this component. This component is only responsible to orchestrate between 4 internal IaECCMaS components.
As an example, this component orchestrate and communicate between selling (customer side) process and selling (cloud side) process through exchange between CM & SS Orchestration process and BM & SS Orchestration process for done cloud services selling steps.

4.1.4. Orchestration Component

Orchestration component is responsible to orchestrate all type of information processes and anything related to orchestration processes which exist in other components, roles, etc.
As should be mentioned that, orchestration include any notifications, change formatting, reporting activities between components through orchestrations processes in other components.
As an example, As an example, this component orchestrates and communicates between Resource Provisioning and Information Management processes through exchange between Resource Orchestration, IT Service Management Orchestration and IaECCMaS Orchestration.

5. Comparison between proposed architecture and other architectures

In this section, the proposed architecture which is presented in section 4 will be compared and evaluated from two following respects:

- Usage of proposed architecture from standard and best practices frameworks
- The difference and similarity of the proposed architecture with other cloud architectures

Cloud environment is service oriented and nowadays large enterprises and many customers have turned to the environment, hence, this environment will be more successful if it is accordance with known and accepted standards and frameworks. Figure 5 show the percent of using standardized processes and services of ICT frameworks in proposed architecture.

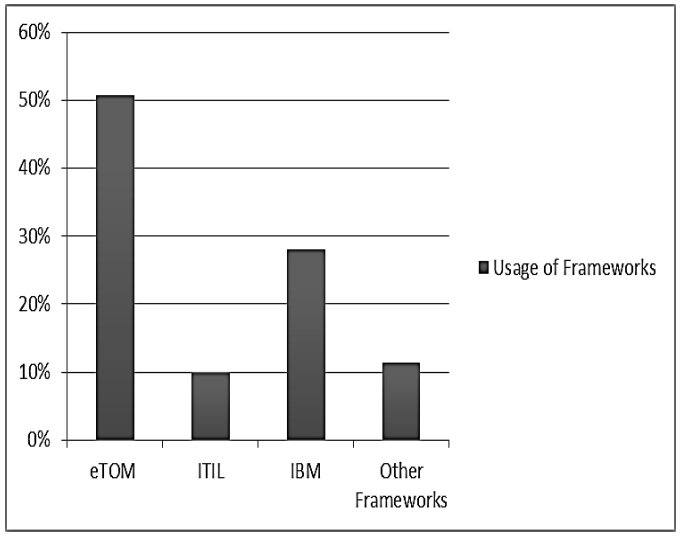


Figure 5: the percent of using ICT frameworks

Also, when the new architecture is proposed, Comparison improvement over the existing architecture Or the extent of its coverage from the existing architecture can be extremely important in acceptance and approval of new architecture. Figure 6 shows, comparison of proposed architecture and known cloud architectures which is NIST, IBM and Oracle.

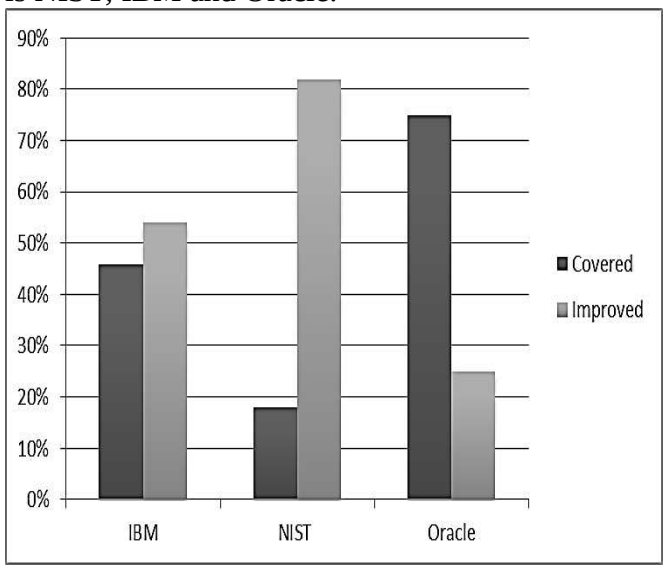


Figure 6: comparison of proposed architecture and other architectures

6. Conclusions and Future Works

The service oriented (SOA) is an absorbing architectural concept which can provide an integrated structure for organizations and moving toward this architecture is one of the main goals of today's organizations. Also nowadays, organizations seeking for standard structures and they use standard frameworks such as eTOM, ITIL, IBM and etc. for the purpose of their internal and external structure synchronous with changes of modern world.

On the other hand, today's cloud computing is the subject of ongoing discussion of societies and organizations. Large organizations moving to cloud environments and sometimes they have been outsourced their costly operations departments such as IT department to cloud, in order to save organizational costs, also provide better services and rise in competition. Therefore cloud architecture would be important and also there are several reference architectures exist for cloud, such as NIST, IBM and etc.

Integrating any new architecture with existing well-known frameworks and best practices might be considered as an affordable effort to transform and help organizations in choosing cloud with confidence.

In this paper, a new cloud computing architecture proposed based on integration of current frameworks and best practices. In this architecture, the first step was to combine the best components of the existing architectures, frameworks and best practices in order to provide a favorite new service layer and bold IT and business in this layer to achieve organizations satisfaction and improve cloud management. Later we augmented few other modules to enrich the architecture to cope with the requirements in real world.

Future research in this area is ongoing by the authors and will be focused on details of architectures components particularly in the management layer of architecture and outcomes will be published soon.

7 REFERENCES

1. Borko Furht, Armando Escalante, "Handbook of Cloud Computing", Springer 2010
2. Abah Joshua, Francisca N. Ogwueleka, "Cloud Computing with Related Enabling Technologies", IJ-CLOSER 2013
3. Michael Hogan, Fang Liu, Annie Sokol, Jin Tong, "NIST Cloud Computing Standards Roadmap", July 2011
4. Peter Mell, Timothy Grance, "The NIST Definition of Cloud Computing", September 2011
5. M.Malathi, "Cloud Computing Concepts", IEEE 2011

6. Niloofar Khanghahi, Touraj Banirostam, "A Novel Intrusion Detection Approach Based on Intelligent Agent in Cloud Computing", ICECE 2013
7. Niloofar Khanghahi, Reza Ravanmehr, "CLOUD COMPUTING PERFORMANCE EVALUATION: ISSUES AND CHALLENGES", IJCCSA, 2013
8. ITIL Advisory Group, "An Introductory Overview of ITIL® V3", The UK Chapter of the itSMF, 2007
9. ITIL Advisory Group, "Service Strategy of ITIL® V3", The UK Chapter of the itSMF, 2007
10. ITIL Advisory Group, "Service Design Overview of ITIL® V3", The UK Chapter of the itSMF, 2007
11. ITIL Advisory Group, "Service Transition of ITIL® V3", The UK Chapter of the itSMF, 2007
12. ITIL Advisory Group, "Service Operation of ITIL® V3", The UK Chapter of the itSMF, 2007
13. ITIL Advisory Group, "Service Improvement of ITIL® V3", The UK Chapter of the itSMF, 2007
14. IBM Advisory Group, "INTRODUCING THE IBM PROCESS REFERENCE MODEL FOR IT V3", IBM Corporation 2008
15. TM Forum Advisory Group, " Business Process Framework eTOM", TM Forum 2012
16. Michael Behrendt, Bernard Glasner, "Introduction and Architecture Overview IBM Cloud Computing Reference Architecture 2.0", IBM 2011
17. Hewlett-Packard Development Company, "Understanding the HP CloudSystem Reference Architecture", HP 2011
18. Fang Liu, Jin Tong, Jian Mao, "NIST Cloud Computing Reference Architecture", September 2011
19. Mark Wilkins, Scott Mattoon, Dave Chappelle, "Oracle® Reference Architecture/Cloud Infrastructure", Oracle, November 2011

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 35-42
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Enriched eTOM Framework in Service Deliver Operation through Alignment with some of COBIT5 Strategic Objectives

[1]Faride Latifi , [2]Ramin Nasiri, [3]Mehran Mohsenzadeh
[1]Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, IRAN
[2]Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, IRAN
[3]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, IRAN

[1]frhlatifi@yahoo.com
[2]Ramin_Nasiri@iauctb.ac.ir
[3]Mohsenzadeh@srbiau.ac.ir

ABSTRACT

eTOM is a standard framework that is defined in telecommunication business processes area. It contains of three phases, 1. Operations 2. Infrastructure, Strategy and Product 3. Enterprise Management. The goal of this paper is to enrich the processes in operations phase included of service assurance and fulfillment and to increase customer retention and loyalty and SLA fulfillment. For this purpose some processes in DSS domain of COBIT5 is introduced and mapped in different area of eTOM operations phase.

KEYWORDS

Telecommunication, Service, COBIT5, eTOM, Framework, SLA.

1 INTRODUCTION

In recent years, types of telecom services are expanding rapidly. On the other hand, present management Frameworks cannot cover and /or support them well. There are the place where few best known standards may be used for IT Service Management to enrich it, such as eTOM and COBIT5 (Control Objectives for Information and Related Technologies).
The Enhanced Telecom Operations Map® (or eTOM for short) is an on-going TM Forum initiative to deliver a business process model or framework for use by service providers and others within the telecommunication industry and consists of three distinct phases: strategic, operational and management [1], [2].
COBIT5 is one of the best if not the only, business framework for the governance and management of enterprise IT. This evolutionary version incorporates the latest thinking in enterprise governance and management techniques, and provides globally accepted principles, practices, analytical tools and models to help increase the trust in, and value from, information systems. COBIT5 builds and expands on COBIT 4.1, the previous version, by integrating other major frameworks, standards and resources, including ISACA's Val IT and Risk IT, Information Technology Infrastructure Library (ITIL®) and related standards from the International Organization for Standardization (ISO) [5].
Each of eTOM and COBIT5 has some benefits and can cover each other to build up a more complete framework, for example COBIT5 can optimize the Quality of IT services and technology that is very useful for telecom service operation domain. In this paper we use this advantage of COBIT5 and thus promote operations phase of eTOM framework [3], [7].

2 COBIT5 FRAMEWORK

COBIT5 is the only worldwide business framework for the governance and management of enterprise IT and provides tools for helping to manage IT and discuss about what must be do not how to do and also is a good governance method to link and imbed in continuity management lifecycle[10],[14]. This evolutionary version incorporates the latest thinking in enterprise governance and management techniques, and provides globally accepted principles, practices,

analytical tools and models to help increase the trust in, and value from, information systems, as well as, COBIT5 provides the next generation of ISACA's guidance on the enterprise governance and management of IT. It builds on more than 15 years of practical usage and application of COBIT by many enterprises and users from business, IT, risk, security and assurance communities. As well as, DSS (Deliver, Service and Supports) domain focused on deliver IT aspects. This domain cover application execution in IT systems and it results [7], [15].

3 eTOM FRAMEWORK

eTOM is a telecommunication business process framework that's service oriented and helps service providers in process management and enhancement [11].

It has three phases that operations phase is the heart of it and contains all customer related operations that categorizes in operations support and readiness, fulfillment, assurance and billing and revenue management vertical process groups. As well as usually named as customer operational processes.

3.1 Customer Relationship Management

It is the first horizontal process group of operation phase in eTOM framework that named as process group level 1 and decomposed to some sub processes:

- CRM SUPPORT & READINESS
- MARKETING FULFILLMENT RESPONSE
- CUSTOMER INTERFACE MANAGEMENT
- SELLING
- ORDER HANDLING
- PROBLEM HANDLING
- CUSTOMER QoS/SLA MANAGEMENT
- RETENTION & LOYALTY
- BILL INVOICE MANAGEMENT
- BILL PAYMENTS & RECEIVABLES MANAGEMENT
- BILL INQUIRY HANDLING
- MANAGE BILLING EVENTS
- CHARGING
- MANAGE BALANCES

It contains of some necessary operations for customer relationship creation and maintenance, Also support customer services to increase loyalty and satisfaction of them, SLA fulfilment and service assurance. On the other hand, DSS02 process group defined in COBIT5 framework is responsible for an effective and timely response to customer requests and resolve all incidents to minimize disruptions through quick resolution of them. Also, based on our analyses DSS02.01, DSS02.02 and DSS02.03 KMPs defined in DSS02 process can be useful in to enrich CRM process group [6], [8]. The results of our first comparision are given in table1.

Table 1 COBIT5 Useful Outputs and Activities for eTOM CRM Domain

Some of KMPs, Inputs/Outputs and Activities of COBIT5 that are useful in eTOM (CUSTOMER RELATIONSHIP MANAGEMENT) process group		
Outputs **Description**	**Activities**	**KMPs**
• Incident and service request classification schemes and models • Rules for incident escalation • Criteria for problem registration	• Define a schemes for incident and service classification and prioritization. • Define rules for incident escalation. • Define criteria for problem registration.	DSS02.01
• Incident and service request log • Classified and prioritised incidents and service requests	• Log all service request and incident. • Prioritize service request and incidents based on SLA service definition.	DSS02.02
• Approved service requests • Fulfilled service requests	• Obtain financial and functional approval if request. • Fulfill service requests by performing some procedures, self help automated menues and predefined requests model.	DSS02.03

Some of advantages of this combination:

- Ensure consistent approaches for handling, informing users about and conducting trend analysis.
- Enable efficient and effective resolution.
- Enable self-help and efficient service for standard requests.
- Handle all service requests and incidents effectively

3.2 Service Management & Operations

Service management & operations as second horizontal process group of eTOM framework that named as process group level 1 and decomposed to some sub processes:

- SM&O SUPPORT & READINESS
- SERVICE CONFIGURATION & ACTIVATION
- SERVICE PROBLEM MANAGEMENT
- SERVICE QUALITY MANAGEMENT
- SERVICE GUIDING & MEDIATION

It is responsible for maintaining knowledge of service such as (access level, type, usage and so on) and all operations to manage service and proposed services. Service planning for short time services and especial customers are some of its duties. On the other h
and DSS05(Manage Security Services) process of DSS domain in COBIT5 is responsible for establish and maintain of information security rules, access license and provide security monitoring. DSS05.01 (Protect against malware) is one of KMP of DSS05 that is responsible for prediction, policy and modify actions in place especially security patch up to date and viruses control) in means of maintaining the IT systems from malwares. Also, DSS04 (Manage Continuity) is important to enable business plan to respond to incidents and disruptions to maintain operations continuity critical business plan and IT service requests in an acceptable level for enterprise and has some KMPs that are very useful in this area such as DSS04.01 (Define the business continuity policy, objectives and scope), DSS04.03 (Develop and implement a business continuity response) and DSS04.08 (Conduct post-resumption review). The results of our second comparison are given in table 2.

Some of advantages of this combination:

- Defining and agreeing on continuity policy and scope
- Continuity planning in the enterprise culture
- Enable continued operation of critical business processes and/or
- Temporary processing arrangements, including links to plans of outsourced service providers.
- Enable continued operation of critical business processes and/or temporary processing arrangements, including links to plans of outsourced service providers.
- Protect against unsolicited information (e.g., spyware, phishing emails).

Table 2 COBIT5 Useful Outputs and Activities for eTOM SM&O Domain

Some of KMPs, Inputs/Outputs and Activities of COBIT5 that are useful in eTOM (SERVICE MANAGEMENT & OPERATIONS) process group		
Outputs **Description**	**Activities**	**KMPs**
• Malicious software prevention policy • Evaluations of potential threats	• Communicate malicious software awareness and enforce prevention procedures and responsibilities. • Regularly review and evaluate information on new potential threats (e.g., reviewing vendors' products and services security advisories).	DSS05.01
• Policy and objectives for business continuity • Disruptive incident scenarios • Assessments of current continuity capabilities and gaps	• Identify critical internal and outsourced business processes and service activities. • Define and document the agreed-on minimum policy objectives and scope.	DSS04.01
• Incident response actions and communications • BCP	• Define the incident response actions and communications. • Develop and maintain operational BCPs containing the procedures.	DSS04.03
• Post-resumption review report • Approved changes to the plans	• Determine the effectiveness of the plan, continuity capabilities, roles and responsibilities, skills and competencies, resilience to the incident, technical infrastructure, and organisational structures and relationships.	DSS04.08

3.3 Resource Management & Operations (Application, Computing and Network)

It is the third horizontal process group in eTOM operations area that named as process group level 1 and decomposed to some sub processes:

- RM&O SUPPORT & READINESS
- WORKFORCE MANAGEMENT
- RESOURCE PROVISIONING
- RESOURCE DATA COLLECTION & DISTRIBUTION
- RESOURCE TROUBLE MANAGEMENT
- RESOURCE PERFORMANCE MANAGEMENT
- RESOURCE MEDIATION & REPORTING

It is responsible for maintaining resources knowledge (networks, IT systems, servers, routers and so on) and manages all used resources to support required or proposed services to customers.
On the other hand in COBIT5 framework, DSS03.05 (Perform proactive problem management) is responsible for optimize resources and follow up its problems. Also DSS04.02 (Maintain a continuity strategy) and DSS04.03 (Develop and implement a business continuity response) can be used to identify resource and it costs and document required resources to support of recovery and continuity procedures according to IT infrastructure and facilities. The results of our third comparison are given in table 3.

Some of advantages of this combination:

- To evaluate the impact over time of a disruption to critical business functions and the effect that a disruption would have on them.
- To identify the possible strategic business and technical options.
- To enable continued operation of critical business
- To raise change requests via the established change management processes.
- To consider recent problems and potential corrective actions.
- To discuss known problems and future planned changes.

Table 3 COBIT5 Useful Outputs and Activities for eTOM RM&O Domain

Some of KMPs, Inputs/Outputs and Activities of COBIT5 that are useful in eTOM (RESOURCE MANAGEMENT & OPERATIONS (Application, Computing and Network)) process group		
Outputs **Description**	**Activities**	**KMPs**
• Problem resolution monitoring reports • Identified sustainable solutions	• Capture change efforts resulting from problem management process activities • Identify and initiate sustainable solutions	DSS03.05
• Business impact analyses • Continuity requirements • Approved strategic options	• Conduct a business impact analysis • Analyse continuity requirements to identify the possible strategic business and technical options	DSS04.02
• Incident response actions and communications • BCP	• Define the incident response actions and communications • Develop and maintain operational BCPs	DSS04.03

3.4 Supplier/Partner Relationship Management

This process group as fourth and last horizontal process group of eTOM operations area that named as process group level1 decomposed to some sub processes:

- S/PRM SUPPORT & READINESS
- S/P REQUISITION MANAGEMENT
- S/P INTERFACE MANAGEMENT
- S/P PROBLEM REPORTING & MANAGEMENT
- S/P PERFORMANCE MANAGEMENT
- S/P SETTLEMENTS & PAYMENTS MANAGEMENT

It contains of primary processes related to purchase in FAB(Fulfillment, Assurance and Billing) that are three vertical process group of eTOM and is responsible for purchase orders and follow up them during the service delivery, problem handling, determine credit bill, verify payment and supplier and partner quality management. On the other hand, use of DSS01.05 (Manage facilities) in COBIT5 framework to

ensure that IT sites and equipment's maintain by authorized people and based on service distances and characteristics proposed from supplier and partner and also use of DSS04.03 (Develop and implement a business continuity response) to ensure that outsourced supplier and partners have efficient and continuity planning in place can be useful in this area of eTOM [4], [6], [9]. The results of our fourth comparison are given in table4.

Table 4 COBIT5 Useful Outputs and Activities for eTOM S/PRM Domain

Some of KMPs, Inputs/Outputs and Activities of COBIT5 that are useful in eTOM (SUPPLIER/PARTNER RELATIONSHIP MANAGEMENT) process group		
Outputs **Description**	**Activities**	**KMPs**
• Facilities assessment Reports • Health and safety awareness	• Make available reports on facilities incidents where disclosure is required in terms of laws and regulations • Ensure that IT sites and facilities are in ongoing compliance with relevant health and safety laws, regulations, guidelines, and vendor specifications.	DSS01.05
• Incident response actions and communications • BCP	• Define the incident response actions and communications • Develop and maintain operational BCPs	DSS04.03

Some of advantages of this combination:

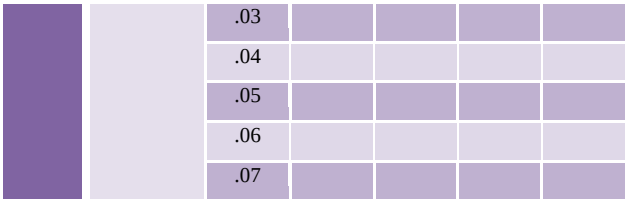
- To enable continued operation of critical business processes and/or temporary processing arrangements, including links to plans of outsourced service providers.
- To ensure that cabling and physical patching (data and phone) are structured and organized. Cabling and conduit structures should be documented (e.g., blueprint building plan and wiring diagrams).

4 An OVERALL VIEW of PROCESS MAPPING ANALYSES

The main purpose of business process mapping is assist to organizations in becoming more efficient, follow up this purpose for telecom companies, in this paper after a deep analysis and comparison between eTOM framework as a telecom business process framework and COBIT5 as a governance framework in part3 and introduce four horizontal process group of eTOM and some useful processes of COBIT5 to enrich it [13]. Some results are obtained and mapped area shown in table5.

Table 5 Combining of COBIT5 DSS Domain with eToM Horizontal Process Groups

COBIT5 DSS Domain			eTOM Horizontal Process Groups			
DSS	P	KMP	CRM	SM&O	RM&O	S/PRM
	DSS01	.01		▲		
		.02		▲		
		.03		▲		
		.04				
		.05		▲		▲
	DSS02	.01	▲	▲		
		.02	▲	▲		
		.03	▲	▲		
		.04				
		.05		▲		
		.06		▲		
		.07		▲		
	DSS03	.01	▲	▲		
		.02				
		.03				
		.04	▲	▲		
		.05			▲	
	DSS04	.01		▲		
		.02			▲	
		.03		▲	▲	▲
		.04				
		.05				
		.06				
		.07				
		.08		▲		
	DSS05	.01		▲		
		.02				



5 ENHANCED eTOM in a VERTICAL POINT of VIEW

As our analysis COBIT5 as an governance framework is useful to demonstrate what activities the enterprise must to do for achieving its goals [13]. As well as it can be combine with eTOM to have efficient IT governance and control framework. On the other hand, combining COBIT5 and eTOM provide a valuable combination for helping telecom operators manage business IT from a business perspective. In figure1 shown a scheme of a vertical mapping between delivery, service and support domain of COBIT5 in right side and eTOM operations phase in left side. On the other hand, eTOM framework contains of four vertical process groups:

5.1 Operations Support & Readiness

This is the first vertical process group in eTOM operation phase that is responsible for support of three other vertical process group(FAB). As our analysis some control objectives of COBIT5 can be useful in this area:

- DSS01.01 **Perform operational procedures**
- DSS01.03 **Monitor IT infrastructure**
- DSS01.05 **Manage facilities**
- DSS03.01 **Identify and classify problems**
- DSS04.01 **Define the business continuity policy, objectives and scope**
- DSS04.02 **Maintain a continuity strategy**
- DSS04.03 **Develop and implement a business continuity response**
- DSS05.02 **Manage network and connectivity security**
- DSS06.03 **Manage roles, responsibilities, access privileges and levels of authority**
- DSS06.05 **Ensure traceability of information events and accountabilities**

5.2 Fulfillment

It is responsible for providing required customer product correctly and timely, offering a suitable resolution for their requirements and also ensure them about resolving their need in that time. As our analysis some control objectives of COBIT5 can be useful in this area:

- DSS02.03 **Verify, approve and fulfil service requests**
- DSS02.06 **Close service requests and incidents**
- DSS02.07 **Track status and produce reports**

5.3 Assurance

It is responsible for ensure customer of continuity service providing and monitor resource status. It manages the SLA and reports service performance to customers as our analysis some control objectives of COBIT5 can be useful in this area:

- DSS01.02 **Manage outsourced IT services**
- DSS06.04 **Manage errors and exceptions**
- DSS06.05 **Ensure traceability of information events and accountabilities**

5.4 Billing & Revenue Management

It is the last vertical eTOM process group and is responsible for bill providing timely and in a correct form for customers, it also resolve customer billing problems and support them [4],[5].

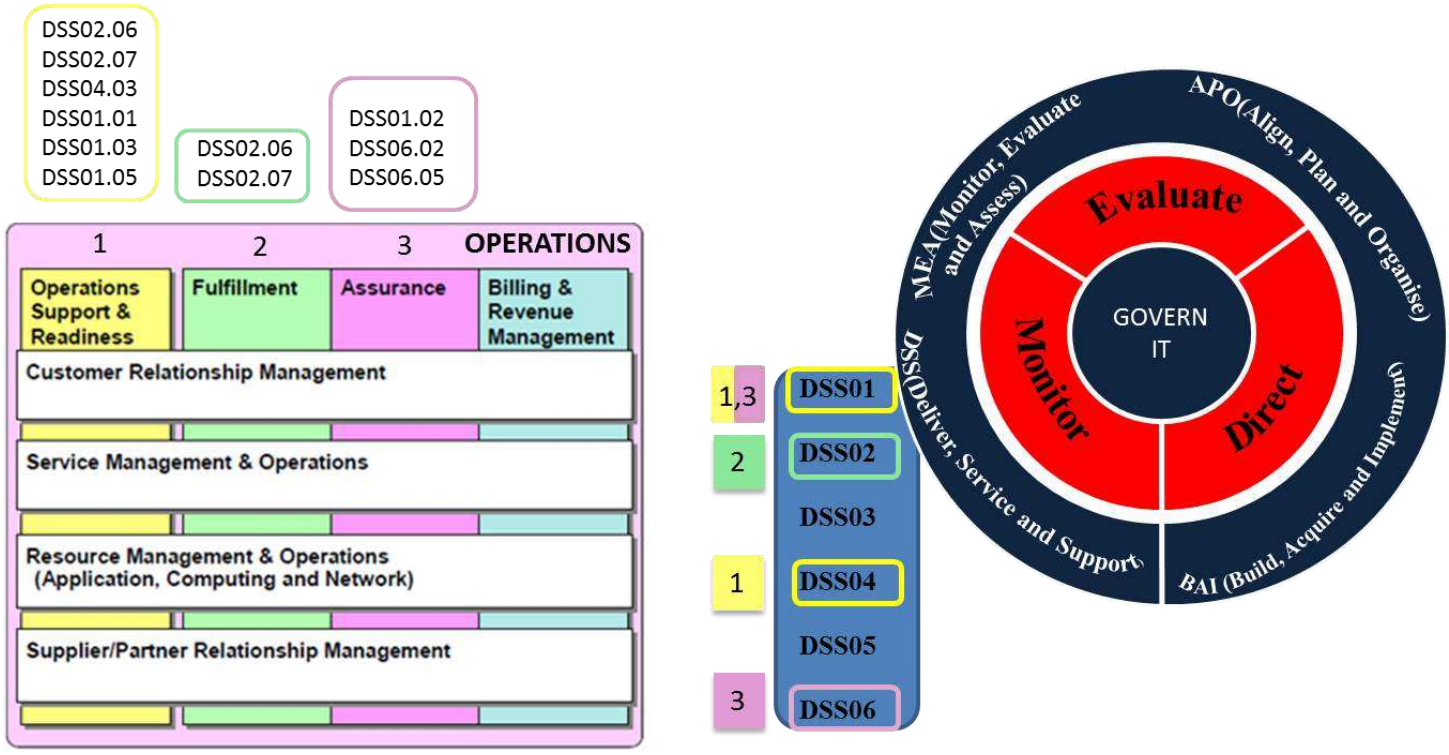


Figure 1 Enhanced eTOM in a Vertical Point of View

6 CONCLUSION and FUTURE WORK

In this paper we propose an enriched eTOM framework that is more powerful than current version. On the other hand, our work is based on a deep study on COBIT5 deliver, service and support domain and combine it to eTOM operations phase, this combine can help to have more powerful eTOM assurance and fulfil process group and increase QoS, SLA and at last customer satisfaction. Future research in this area will be focused on improvement of eTOM in other domain to integrate with suitable processes of COBIT5 framework which would be published later on.

7 REFERENCES

[1] Tele Management Forum, eTOM – The Business Process
Framework, "Use Cases for Commercial Spectrum Management ", 2012.
[2] Tele Management Forum, eTOM – The Business Process Framework, "An eTOM Primer", 2009.
[3] M. Asic, Student, E. Bajrami, Student, A. Tanovic, “Improvement of the eTOM Operations phase through the comparison with ITIL best practices”, 20th Telecommunications forum, pp. 40 – 43, Nov. 2012.
[4] Tele Management Forum, eTOM – The Business Process Framework, "Extended Process Decompositions and Descriptions", 2012.
[5] Tele Management Forum, eTOM – The Business Process Framework, "Concepts and Principles", 2011.

[6] The Book of COBIT5 Enabling Processes available at www.isaca.org.
[7] The Book of COBIT5: A Business Framework for the Governance and Management of Enterprise IT, available at: http: //www.isaca.org.
[8] B. Raouyane, M Bellajkih, M Errais, "IMS Management based eTOM framework for Multimedia service", 14th International Telecommunications Network Strategy and Planning Symposium (NETWORKS), pp.1 – 6, Sept. 2010.
[9] A. Zhuang, X. Qiu, H. Cheng, X. Chen, Z. Gaol," A Management process defining approach for converged services based on eTOM and ITIL", 3rd IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT), pp. 180 – 185, Oct. 2010.
[10] T.Mataracioglu, S.Ozkan," Governing Information Security in Conjuction with COBIT and ISO 27001", International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.4, July. 2011.

[11] M. Valiente, E. Barriocanal, M.Sicilia, "Applying an ontology approach to IT service management for business-IT integration" , pp. 76–87, Vol 28, April. 2012.

[12] J. Lainhart, D. Oliver, "Overview of COBIT 5 Public Exposure Commentary", Vol.1, Jan 2012. Available at www.isaca.org.
[13] J. Mike Jacka, Paulette J. Keller, "Business

Process Mapping: Improving Customer Satisfaction", Second Edition, Jul 2009.
[14] A. Menevse, " Toward Better IT Governance With COBIT 5 ", Vol. 4, Oct 2011. Available at www.isaca.org.
[15] Z. Enslin, "Cloud computing adoption: Control objectives for information and related technology (COBIT) – mapped risks and risk mitigating controls", African Journal of Business Management, Vol.6, pp. 10185-10194, Sept. 2012.

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 43-52
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

A Hierarchical Self-Healing SLA for Cloud Computing

Ahmad Mosallanejad, Rodziah Atan, Masrah Azmi Murad, Rusli Abdullah,
Faculty of Computer Science and Information Technology, UPM, Malaysia
ahmad.upm@gmail.com, {rodziah, masrah, rusli}@upm.edu.my

ABSTRACT

The service level agreement (SLA) is a mutual contract between the service provider and consumer which determines the agreed service level objective (SLO). The common SLA is a plain documental agreement without any relation to other dependent SLAs during the different layers of cloud computing. Hence, the cloud computing environment needs the hierarchical and autonomic SLA. This paper proposes the SH-SLA model to generate a hierarchical self-healing SLA in cloud computing. The self-healing ability contains the SLA monitoring, violation detecting and violation reacting processes. In SH-SLA, the related SLAs communicate with each other hierarchically. The SLA would be able to check its QoS and notify the recent status to dependent SLAs. Furthermore, SH-SLA could prevent or propagate the notified violations by an urgent reaction. Consequently, the service providers have a great chance to prevent the violated SLA before sensing by end users. The SH-SLA model is simulated and the experiment results have presented the violation detection and reaction abilities of the proposed model in cloud computing. Besides, the end users meet the lesser violations in SH-SLA than the common SLA.

KEYWORDS

service level agreement, cloud computing, self-monitoring SLA, hierarchical SLA, self-healing SLA

1. INTRODUCTION

Currently, cloud computing is an excited topic in both business and research area [1, 2]. Many service providers and consumers have dealings among the different layers of cloud computing consist of software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS) [3, 4]. Service level agreement (SLA) is a contract between the service provider and consumer to determine the quality and functionality of agreed services [5]. So, SLA is a fundamental document which covers the service level objective (SLO), their attributes and metrics. This agreement is the basis of relations between service provider and consumer in different layers of cloud computing [6, 7]. Both service provider and consumer need to monitor the agreed services for valdating the SLA [8].

Currently, the most of SLAs are a single XML document which covers functionalities and quality of services (QoS) between specific service provider and consumer. They do not have a connection with other related SLAs while cloud computing is a hierarchical environment. SLAs in SaaS will be failed if its related SLA in PaaS is violated. So, lack of an effective relations between dependent SLAs is a vital challenge which makes the SLA management system inefficient. To have an effective SLA monitoring and a violation reacting system, a hierarchical SLA model is needed based on cloud computing nature [9, 10].

Three types of SLA monitoring systems are available including provider side, consumer side and trusted party side SLA monitoring system [11-13]. Each of these approaches has its own advantages and disadvantages but all of them have a centralized monitoring system. In centralized monitoring system, the specific SLA management center in provider, consumer or trusted party side is in charge for SLA monitoring and violation detecting. The most of SLA monitoring systems are applied in grid computing and service oriented architecture (SOA) environments therefore they are not compatible enough for cloud computing [14, 15]. Furthermore, the most of SLA management and monitoring systems only present the report of SLA violations without any reaction. Likewise, The current structure of SLA and SLA monitoring system is not suitable for cloud computing nature without self-healing SLA feature [16, 17]. In order to have the reliable cloud

computing services, an effective SLA monitoring and the violation reacting model is unavoidable.

In this study, hierarchical self-healing (SH)-SLA model is proposed to enforce the SLA monitoring and violation reacting in cloud computing. In SH-SLA, each SLA has connected to their related SLAs in different layers of cloud computing so each SLA is able to notify its status to other related SLAs. Additionally, each SLA has the ability of monitoring and reacting independently. The SLA can assess the monitoring results itself and notify to the dependent SLAs.

The rest of the paper is organized as follows: The related works are presented in Section 2. Then, the 3rd Section describes the SH-SLA compositions. Section 4 presents the experiment method then the proposed model is evaluated in Section 5. Finally, the conclusion and future work are presented in 6th Section.

2. RELATED WORKS

Many studies have investigated the SLA management systems, but only a few of them covered the SLA enforcement in cloud computing. The most of related works applied the SLA models from other environments such as SOA and grid computing into cloud computing without enough considering to the cloud requiermnets. The main parts of a self-healing system consist of monitoring and reacting procedures.

Some related works such as QoSMONaaS [18] and SLAMonADA [19] focused on SLA monitoring system to detect the SLA violations without considering to the reaction process. These proposed platforms measured the QoS value in running time to release a report about SLA validation. A. Kertesz *et. al*. (2011) and P. Varalakshmi *et. al.* (2011) also proposed the SLA monitoring system to detect the SLA violations [20] [21]. They count the detected violations in order to assess the penalty cost which provider should pay to the service consumer. Some other related works are such as [22], [23] and [24] checked the SLA validation to manage the provider resources effectively. They also did not consider the violation reaction issue in their studies.

There are a few related works to react against SLA violations in cloud computing. FOSII [25], QU4DS [26] and LoM2HiS [27] which are significant in this work. All of these frameworks engaged the monitoring, analysis, planning and execution (MAPE) loop to react against violations [28]. They applied the centralized monitoring and reacting systems in cloud computing, for any ecosystem deploying it. They periodically collected the QoS values from could infrastructure and compare the collected data with SLO. If any violation is detected, the reaction plan is designed based on historical knowledge. Finally the reaction is executed by resource management system. Reaction plan was normally the provider resource reconfiguration or Virtual Machines (VM) replacement. These proposed healing systems are not suitable for agile reactions due to several separated procedures which are to be done sequentially to react against detected violations. Furthermore, a central system is enforced to monitor all SLAs and react against all detected violations. These events are forecasted to be hard for an education ecosystem in terms of maintaining and adjusting to suitable services and platforms.

This literature review has illustrated that the related works and proposed solutions did not suitably cover the cloud computing requirements. The most of related works have transmitted the SLA monitoring framework into cloud computing from SOA and grid computing. Although a few effective SLA monitoring and reacting mechanisms have presented, they have followed the common SLA structure yet.

3. SH-SLA MODEL

Having an effective self-healing SLA in cloud computing, SH-SLA model is proposed including innovative SLA structure, effective SLA monitoring and reacting methods. These abilities need an infrastructure in each service provider to work properly. Although these facilities are completely integrated, they are introduced in next sub-sections separately.

3.1. Self-healing SLA Architecture

To have a hierarchical self-healing SLA, an especial architecture is needed. Each service provider, in different layers of cloud computing, needs to have fundamental components as illustrated in Figure 1. The lower layer (LL) port and upper layer (UL) port are the communication gates to other service providers and service consumers in lower and upper layers of cloud computing. First of all, the negotiation management system specifies the SLA features after the negotiation process. Actually, the SH-SLA is the main output of negotiation management system then SH-SLA should be run. Next, the data collector passes the relevant metrics data from resources and lower layer notifications to the SH-SLA. Finally the running SH-SLA assesses the SLA attributes value and records them to the monitoring warehouse. Moreover, any necessary reaction could be done by the resource manager and notification could be sent to the upper layer consumers.

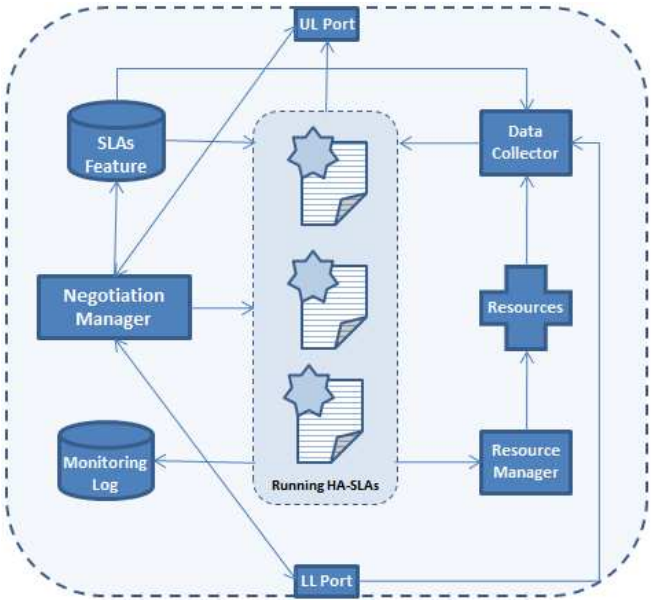


Figure 1. Self-healing SLA Architecture

3.2. Hierarchical SLA in Cloud

This model suggests the hierarchical SLA as a foundation of SLA monitoring and reacting processes. SLAs are the basis of all interactions between service providers and consumers which should be inspected in a monitoring system. Current SLA is a document including the information of service functionalities and QoS. Proposed SH-SLA has two important contributions versus common SLAs: firstly it defines the hierarchical relations secondly this is a self-healing SLA. It is hierarchical because the dependent SLAs are connected to each other. It is self-healing SLA because the both monitoring and reacting functions are allocated inside the SH-SLA. Figure 2 has depicted the common SLA and SH-SLA to present their differences. The current SLA does not have any fields for connecting to the related SLAs while they need to have a hierarchical structure in cloud computing during the different layers. Current SLA, introduced in related works, is an isolated document but SH-SLA is a relational contract.

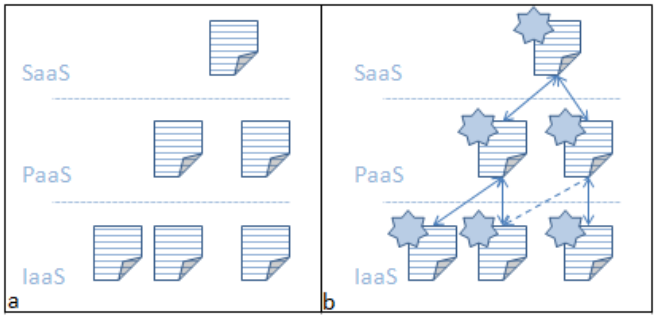


Figure 2. a) Current SLA in cloud computing b) SH-SLA overview

Many SLAs in different layers of cloud computing are contracted. Figure 2 has illustrated some related SLAs during SaaS, PaaS and IaaS which the upper layer SLAs relay to the related lower layer SLAs. If any SLA in IaaS be violated, all dependent SLAs in PaaS and SaaS will be failed. Figure 2a has shown the common SLA in cloud computing which they are not designed to have a hierarchical relation between related SLAs. A few developed frameworks have tried to build these connections by their management system but they are not reasonable when SH-SLA could provide a hierarchical self-healing SLA independently. Figure 2b has displayed the SH-SLA relations which the related upper and lower layer SLAs have been linked in each SLA contents. Moreover some reserved relations are defined for urgent invocation against critical SLA

violation. This reserved relation is illustrated in Figure 1b as a dash line.

3.3. SLA Monitoring Procedures

The SLA monitoring is an important activity for both service provider and service consumer. The service provider utilizes the SLA monitoring systems to manage and economize their resources. On the other hand, service consumer wants to confirm the agreed QoS in SLA. Besides, the SLA monitoring process is the first part of the self-healing SLA to detect the SLA violations. Related monitoring frameworks are discussed in literature review which most of them had a central monitoring approach. In contrast, the proposed SH-SLA has located the monitoring function in each SLA as a part of the distributed monitoring framework. Each monitoring function evaluates the current value of attributes based on their metrics and formula. The monitoring function returns the notification consists of attributes state and their value. They could be recorded in provider or consumer side and also could be used for any relevant reactions.

SH-SLA model changes the passive SLA document to the active SLA identity. Figure 2b insists on self-healing ability of each SLA by a star icon. In each SLA, the star icon indicates to the all operations of SLA such as monitoring and reacting functions. Each SLA could manage, monitor and react by itself. Actually, SLA monitoring and reacting methods are the scope of this research while the SH-SLA structure has the ability of other operations which they will be investigated in future works.

3.4. SLA Reacting Procedures

In proposed method, each service has its own monitoring and reacting instance as shown in Figure 3. SLA contents include attribute, SLO, threshold and related SLAs address. The threshold value is utilized to prevent the SLA violations before happening. The address of related SLAs is used to notify any emergency alert into relevant clients and providers. The attribute listener receives the specific QoS value related to the current SLA from data collector. The QoS value is compared with SLO which is recorded in SLA contents. If the QoS value exceeds the threshold value, the violation reacting procedure will be activated.

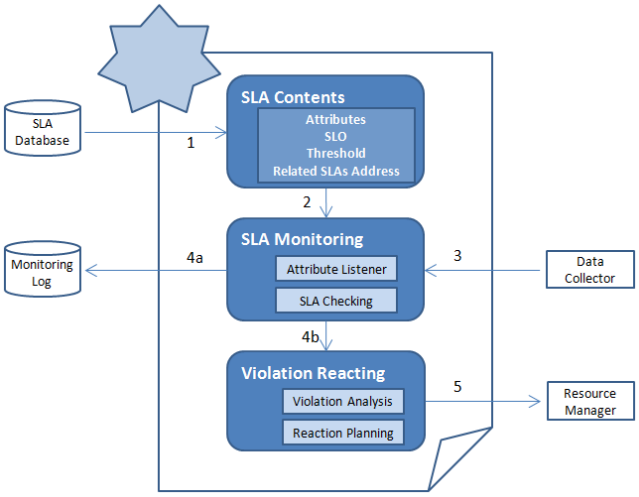


Figure 3. Self-healing components

The violation reaction method is proposed to be as part of HA-SLA. Three Reaction (3R) strategies are defined for violation reaction including: internal resource reconfiguration, external service invocation, and detected violation propagation. If notified QoS from monitoring part exceeds threshold value, it will be propagated into all related consumers and providers as a violated or critical situation. The alerts propagation can help the upper layer providers to react against the notified violation earlier. Sequentially, current provider tries to react against detected violation by internal resource reconfiguration or external service invocation. First, the provider applies the restarting, replacing and reconfiguring methods to the software and hardware resources in order to revive the violated service. If this reaction was unsuccessful, the provider will invoke the external service from another provider which has a same functionality with the violated service. Indeed, the external service is reserved as a spare service to respond the consumers in violation period.

3.5. SH-SLA Developement

The current SLA is a plain agreement document but proposed SH-SLA model is a self-healing SLA including monitoring and reacting

functionalities. The SH-SLA structure code has illustrated in Figure 3 as a first building block of hierarchical self-healing SLA. In this version, some SLA features are avoided because they are not necessary for monitoring and reacting framework and they are out of this paper scope. The SLA attribute structure is defined in lines from 1 to 10 including the definition of metrics, formula, agreed value and threshold value. The threshold value is assigned for agile reaction against any probable violation. Each attribute is able to assess its validity by assessor function as illustrated in Figure 4, line 11. Actually, the array of these attributes is a part of the SH-SLA (Figure4, line 20). Lower layer, upper layer and reserved providers are defined (from 21st to 23rd line) to have an effective hierarchical relation to other SLAs. Figure 4 (from 25th to 33rd line) has shown the functional parts of the SH-SLA consist of monitoring and reacting tasks. Practically, each SH-SLA can monitor itself and react against any violations. This is a significant contribution of the SH-SLA model as a real self-healing SLA.

```
1.  struct Attribute
2.  {
3.     int code;
4.     string name;
5.     Metric[] metrics;
6.     string formula;
7.     float agreedV;
8.     char agreedOperation;
9.     float criticalV;
10.
11.    float assessor()
12.    {
13.      …
14.    }
15. }
16. public struct HA_SLA
17. {
18.    int code;
19.    string name;
20.    Attribute[] attributes;
21.    int[] lowerSLAs;
22.    int[] emergency;
23.    int[] upperSLAs;
24.
25.    Notification[] Monitoring()
26.    {
27.      …
28.    }
29.    Notification[] Reacting()
30.    {
31.      …
32.    }
33. }
```

Figure 4. SH-SLA structure

Normally, the SLA monitoring function checks the value of attributes when the user invokes the service. So the SH-SLA monitoring procedure is running per each service invocation. The monitoring function algorithm is presented in Figure 5.

```
Notification[] Monitoring()
        {
Notification[] notify = new Notification[attributes.Count];
Boolean ReactState =false;

for (int i = 0; i <attributes.Count; i++)
            {
                notify[i] = new Notification();
                notify[i].senderCode= CurrentSLAcode;
                notify[i].AttributeCode = attributes[i].code;
                notify[i].AttributeV = attributes[i].assessor();

if (notify[i].AttributeV  ∉ attributes[i].agreedV)
                {
                    notify[i].state = -1;
                    ReactState= true;
                }
else if ( notify[i].AttributeV  ∈ attributes[i].criticalV)
                {
                    notify[i].state = 0;
                    ReactState= true;
                }
else if (notify[i].AttributeV ∈ attributes[i].agreedV)
                    notify[i].state = 1;
            }

if( ReactState == true )
                    Reacting(notify);

MonitoringLogDatabase.Insert(notify);

return notify;
        }
```

Figure 5. The monitoring function algorithm

The monitoring function takes the metrics value and the SLA status of the lower layer as a parameter from the data collector. After the SLA attributes measuring, the results are notified to the upper layer SLA. Moreover, any detected violation and critical value are passed to the reacting function to prevent or propagate them. The violation prevention method could follow either internal or external reaction strategies. The service provider migration and the resource replacing are the samples of external and internal reactions respectively.

4. METHODOLOGY

The SH-SLA model, described in Section 3, is simulated to evaluate the proposed model. Figure 6 has presented the simulated scenario based on SH-SLA usage in cloud computing. The predefined SLAs from 1 to 4 are the agreements in different layers of cloud computing. Each SLA is a contract about a specificr service between service provider and consumer. SLA1 is a contract

between IaaS provider and PaaS vendor as a client. Service of SLA4 is also located in IaaS but as a reserved service for emergency invocations. SLA2 is a contract between PaaS provider and SaaS vendor as a client. Finally SLA1 is an agreement between SaaS provider and the end user. SLA3 is dependent on SLA2 moreover SLA2 is dependent on SLA1 hierarchically. In this scenario, each SLA includes respond time and throughput attributes. These attributes are selected because they are the popular attributes which many researchers already focused on them such as [29], [30], [31] and [32]. As described, the simulated SLAs are depending on each other hierarchically.

This study tries to evaluate the simulated SH-SLA by within-subjects design of experimental methodology used by Emeakaroha *et al.* [33]. In this method, the effects of both SH-SLA and plain SLA are observed separately on the same data. This experimental design is selected because this research attempts to compare the effects of two different treatments, SH-SLA and common SLA, on the same situation. Finally the number of violated SLA is observed during the invocations.

This study used the throughput and respond time dataset collected from Zheng [34, 35] which its validity is confirmed in his study [17, 35-37]. This dataset includes the throughput and respond time of 5825 services which they are invoked by 339 users. Therefore, 339 throughputs and respond times are collected for each service. On the other hand, the SH-SLA experimental scenario needs the throughput and respond times of only 4 services as illustrated in Figure 6. Finally 339*8 matrix data is captured from original dataset.

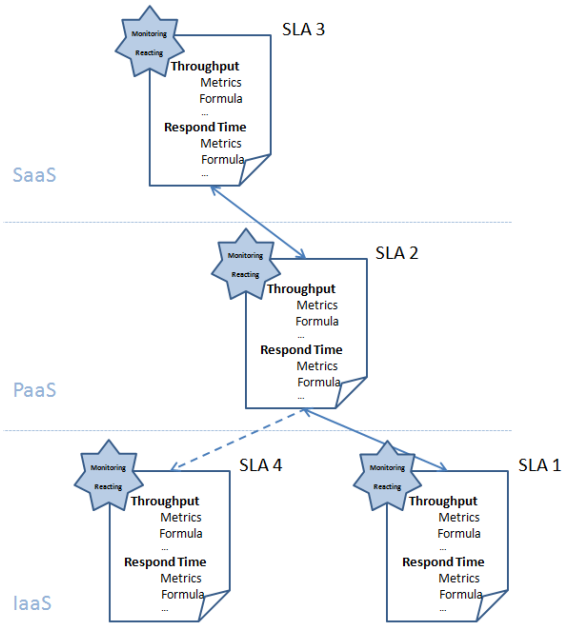


Figure 6. Simulated SH-SLA scenario

The negotiation system is located in service provider as presented in Figure 1 however this process is not a part of this research scope. A predefined agreed value is assumed between service provider and service consumer as a SLO. These agreed values are estimated based on normal attribute value achieved from the dataset. Both SH-SLA and common SLA are configured by the same metrics, formula, SLO and threshold value.

In both SH-SLA and common SLA simulation, firstly the SLA1 should assess the current attribute values as a lowest layer SLA then notify the monitoring results to the upper layer SLA. Afterwards, SLA2 should monitor the current attribute based on SLA1 notification and respond to the SaaS layer SLA. Finally SLA3 assesses the final attribute values based on lower layer attributes value and notify to the end user. This procedure is frequently repeated per each invocation. Moreover, the SH-SLA has the ability of reaction against notified violations while the common SLA only checks the attributes value. In simulated SH-SLA scenario, PaaS user has migrated to the SLA4 when any violations or critical value reported from SLA1.

The SH-SLA and the common SLA violations are observed in this experiment with 339*8 data. Furthermore, this research tried to observe how the increasing of SLA1 violation

affects the end user in both SH-SLA and common SLA. For this purpose, the first experiment is repeated when the value of SLA1 attributes is changing from 100% to -100%. This experiment is started with 100% increased attribute data in SLA1 to observe its effects on SLA3 in both SH-SLA and common SLA. The test is repeated frequently when the SLA1 attributes are decreasing by 10% in each time. The last test is done by 100% deducted SLA1 attributes.

5. EVALUATION AND COMPARISION

5.1. SH-SLA Monitoring Results

After SH-SLA implementation, the experiment has been done and its results confirmed the validity of this model. During the 339 service invocations all attributes of different SLAs are monitored. The Figure 7 has presented the SH-SLA output in the monitoring log database. The output of SH-SLA monitoring for SLA3 is also shown in Figure. SLA3 is an agreement between the end user and SaaS provider. The respond time and throughput are the attributes of SLA3 which they have their specific agreed value. These attributes are assessed based on their dataset value and the notified attributes from the lower layer provider. The respond time of service is presented in Figure 6a for each invocation when the agreed value is less than 1 millisecond and the threshold value is between 0.9 and 1 millisecond. The SH-SLA detected 49 respond delay which they have taken more than 1 millisecond as illustrated in Figure 7a. Moreover, the respond time was in critical range in 18 invocations. On the other hand, Throughput monitoring has depicted in Figure 7b which the throughput should be more than 20 and threshold area is between 25 and 20. During the service invocations, 39 violated throughputs are detected and 26 throughput values are observed in critical area. However, SLA3 is violated if either respond time or throughput is exceeded the agreed value. Totally, the SH-SLA monitoring system detected 83 number of violated SLAs during the 339 invocations which should react against them.

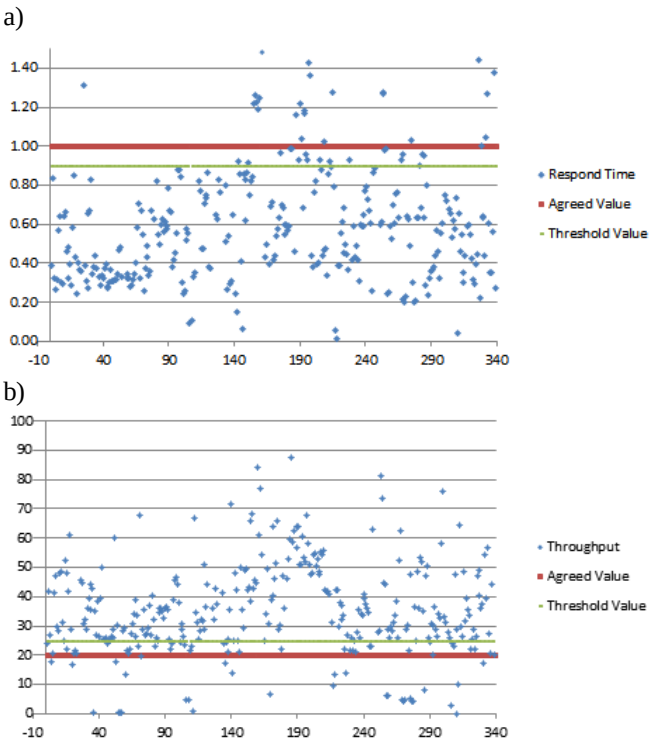


Figure 7. SLA3 monitoring by SH-SLA for a) respond time and b) throughput attributes

5.2. Comparison between SH-SLA and common SLA

The SH-SLA and common SLA are compared regarding two aspects: the number of violated SLAs and the sensation of violation by end user in critical situation.

5.2.1. Violated SLAs

The notified violations in both SH-SLA and common SLA are shown in Figure 8. The number of violations is same in SH-SLA and common SLA for SLA4 and SLA1 because they do not have the ability of reaction and migration to another service provider as shown in Figure 5. On the other hand, the SLA2 has the ability of migration from SLA1 to SLA4 in critical situations therefore SH-SLA can prevent some of the violations before affecting the upper layer. The beneficiations of this reaction are continuing into upper layer and end users as illustrated in SLA3 column. Finally, SH-SLA deducted the violations by 13.68% in SLA2 and the end user sensed 45.75% lesser violations in SLA3 than common

SLA. Therefore the reacting ability of SH-SLA deducted the number of violated SLA2 and SLA3 in comparison with the common SLA.

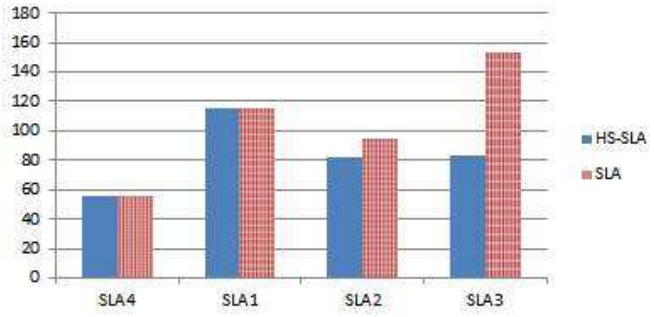


Figure 8. Violated SLAs in SH-SLA and common SLA

5.2.2. End user sensation in tension values

Figure 9 has illustrated the number of violations in both SH-SLA and common SLA at the end user (SLA3) when the lowest layer faults are increased. For this purpose, the tension values are simulated from the real SLA1 attribute value. The experiment is started with 100% increased real SLA1 attributes value. Then, the increased attribute values are deducted step by step by 10% deduction. Finally SLA1 attribute values are arrived to -100% of real value. Figure 8 has illustrated the SH-SLA and common SLA reactions during this changing.

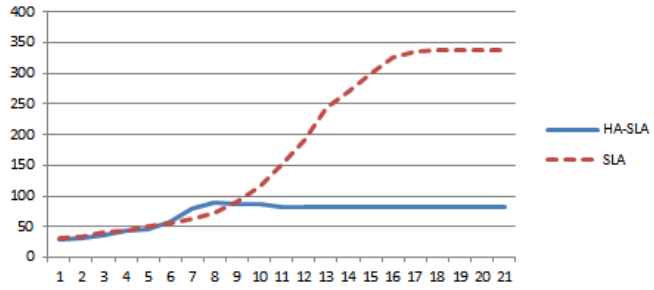


Figure 9. End user sensation in SH-SLA and common SLA when the faults are increasing

The SH-SLA and common SLA have a same result when the SLA1 attribute values are in excellent level (from 1st to 6th). During this period, the SH-SLA does not need any reactions because all SLA1 attribute values are following the agreed value. Surprisingly, the SH-SLA violations are more than common SLA in 7th and 8th periods. In this period, the SH-SLA detected some critical values and migrated to SLA4. In contrast, the critical values of SLA1 were not violated while the SLA4 was in violated attribute. Therefore, the violated SLA in SH-SLA became more than common SLA in this unsuccessful migration period. Finally the violated SLA is dramatically increased in common SLA when the SLA1 attribute values were hardly deducting. During this period, SH-SLA had a long time successful migration to SLA4 so the end users have not sensed the released faults from SLA1. At the final steps, the SH-SLA was stable in 82 violated SLA while the violations of common SLA are increased to 338.

Although in short particular period of time the SH-SLA violations were more than common SLA, this was a rare special situation which could be moderated by effective migration decisions. Moreover, only one migration alternative just for SLA2 is considered in this experiment scenario as an alternative reaction. To have a more reliable cloud services, more reaction strategies can be added to this scenario.

6. CONCLUSION

The most of related works applied the SLA monitoring system from SOA and grid computing to cloud computing while they have the different requirements. This paper proposed SH-SLA model for SLA monitoring based on the hierarchical nature of cloud computing. It also could react against any critical value and SLA violations to prevent them. Proposed model changed the plain SLA to the hierarchical self-healing SLA. Each SLA is able to monitor its QoS and react against violation. So, service providers have a great chance to prevent the SLA violations before sensing by end users. The SH-SLA is simulated to be run in IaaS, PaaS and SaaS layers. The proposed model is validated by within-subject design of experimental methodology. The experiment results indicated that the SH-SLA deducted the violated SLA by 45.75% at the end user in comparison with common SLA. Moreover, SH-SLA was more reliable when the IaaS faults were increasing.

Although the SH-SLA is successfully validated, the reacting decisions should be improved and other preventive strategy should be considered in future work.

AKNOWLEDGEMET

This project is funded by the Exploratory Research Grant Scheme, Ministry of Higher Education of Malaysia. Project no.: ERGS/1/2012/TK06/UPM/02/46.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, "A view of cloud computing," *Communications of the ACM*, vol. 53, pp. 50-58, 2010.

[2] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities," in *High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on*, 2008, pp. 5-13.

[3] T. Dillon, C. Wu, and E. Chang, "Cloud computing: issues and challenges," in *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*, 2010, pp. 27-33.

[4] W.-T. Tsai, X. Sun, and J. Balasooriya, "Service-oriented cloud computing architecture," in *Information Technology: New Generations (ITNG), 2010 Seventh International Conference on*, 2010, pp. 684-689.

[5] S. G. Gómez, J. L. Rueda, and A. E. Chimeno, "Management of the Business SLAs for Services eContracting," in *Service Level Agreements for Cloud Computing*, ed: Springer, 2011, pp. 209-224.

[6] B. P. Rimal, A. Jukan, D. Katsaros, and Y. Goeleven, "Architectural requirements for cloud computing systems: an enterprise cloud approach," *Journal of Grid Computing*, vol. 9, pp. 3-26, 2011.

[7] M. J. Buco, R. N. Chang, L. Z. Luan, C. Ward, J. L. Wolf, and P. S. Yu, "Utility computing SLA management based upon business objectives," *IBM Systems Journal*, vol. 43, pp. 159-178, 2004.

[8] M. Comuzzi, J. Vonk, and P. Grefen, "Measures and mechanisms for process monitoring in evolving business networks," *Data & Knowledge Engineering*, vol. 71, pp. 1-28, 2012.

[9] I. Ul Haq and E. Schikuta, "Aggregation patterns of service level agreements," in *Proceedings of the 8th International Conference on Frontiers of Information Technology*, 2010, p. 40.

[10] O. Mola and M. A. Bauer, "Towards Cloud Management by Autonomic Manager Collaboration," *International Journal of Communications, Network and System Sciences*, vol. 4, pp. 790-802, 2011.

[11] B. S. ARAB and A. A. A. GHANI, "EMPLOYING PERFORMANCE COUNTERS AND SOFTWARE WRAPPER FOR MEASURING QOS ATTRIBUTESOF WEB SERVICES," *Journal of Theoretical and Applied Information Technology*, vol. 34, 2011.

[12] E. Badidi, "A framework for brokered Service Level agreements in SOA environments," in *Next Generation Web Services Practices (NWeSP), 2011 7th International Conference on*, 2011, pp. 37-42.

[13] R. Jurca, B. Faltings, and W. Binder, "Reliable QoS monitoring based on client feedback," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 1003-1012.

[14] I. Ul Haq, I. Brandic, and E. Schikuta, "Sla validation in layered cloud infrastructures," *Economics of Grids, Clouds, Systems, and Services*, pp. 153-164, 2010.

[15] I. Brandic, V. C. Emeakaroha, M. Maurer, S. Dustdar, S. Acs, A. Kertesz, and G. Kecskemeti, "LAYSI: A Layered Approach for SLA-Violation Propagation in Self-manageable Cloud Infrastructures," in *Computer Software and Applications Conference Workshops*, 2010, pp. 365-370.

[16] D. Weyns, B. Schmerl, V. Grassi, S. Malek, R. Mirandola, C. Prehofer, J. Wuttke, J. Andersson, H. Giese, and K. M. Göschka, "On patterns for decentralized control in self-adaptive systems," in *Software Engineering for Self-Adaptive Systems II*, ed: Springer, 2013, pp. 76-107.

[17] Y. Dai, Y. Xiang, and G. Zhang, "Self-healing and Hybrid Diagnosis in Cloud Computing," in *Cloud Computing*, ed: Springer, 2009, pp. 45-56.

[18] G. Cicotti, S. D'Antonio, R. Cristaldi, and A. Sergio, "How to Monitor QoS in Cloud Infrastructures: The QoSMONaaS Approach," in *Intelligent Distributed Computing VI*, ed: Springer, 2013, pp. 253-262.

[19] C. Muller, M. Oriol, M. Rodriguez, X. Franch, J. Marco, M. Resinas, and A. Ruiz-Cortes, "SALMonADA: A platform for monitoring and explaining violations of WS-agreement-compliant documents," in *Principles of Engineering Service Oriented Systems (PESOS), 2012 ICSE Workshop on*, 2012, pp. 43-49.

[20] A. Kertesz, G. Kecskemeti, and I. Brandic, "Autonomic sla-aware service virtualization for distributed systems," in *Parallel, Distributed and Network-Based Processing (PDP), 2011 19th Euromicro International Conference on*, 2011, pp. 503-510.

[21] P. Varalakshmi, K. Priya, J. Pradeepa, and V. Perumal, "SLA with Dual Party Beneficiality in Distributed Cloud," in *Advances in Computing and Communications*, ed: Springer, 2011, pp. 471-479.

[22] H. Liu, D. Xu, and H. K. Miao, "Ant Colony Optimization Based Service Flow Scheduling with Various QoS Requirements in Cloud Computing," in *Software and Network Engineering (SSNE), 2011 First ACIS International Symposium on*, 2011, pp. 53-58.

[23] T. A. L. Genez, L. F. Bittencourt, and E. R. M. Madeira, "Workflow scheduling for SaaS/PaaS cloud providers considering two SLA levels," in *Network Operations and Management Symposium (NOMS), 2012 IEEE*, 2012, pp. 906-912.

[24] X. Wang, Z. Du, and Y. Chen, "An adaptive model-free resource and power management approach for multi-tier cloud environments," *Journal of Systems and Software*, 2012.

[25] C. Muller, M. Oriol, M. Rodríguez, X. Franch, J. Marco, M. Resinas, and A. Ruiz-Cortes, "SALMonADA: A platform for monitoring and explaining violations of WS-agreement-compliant documents," in *Principles of Engineering Service Oriented Systems (PESOS), 2012 ICSE Workshop on*, 2012, pp. 43-49.

[26] A. L. Freitas, N. Parlavantzas, and J. L. Pazat, "A QoS assurance framework for distributed infrastructures," in *Proceedings of the 3rd International Workshop on Monitoring, Adaptation and Beyond*, 2010, pp. 1-8.

[27] V. C. Emeakaroha, I. Brandic, M. Maurer, and S. Dustdar, "Low level metrics to high level SLAs-LoM2HiS framework: Bridging the gap between monitored metrics and SLA parameters in cloud environments," in *High Performance Computing and Simulation (HPCS), 2010 International Conference on*, 2010, pp. 48-54.

[28] M. Maurer, I. Breskovic, V. C. Emeakaroha, and I. Brandic, "Revealing the MAPE loop for the autonomic management of cloud infrastructures," in *Computers and Communications (ISCC), 2011 IEEE Symposium on*, 2011, pp. 147-152.

[29] B. Abrahao, V. Almeida, J. Almeida, A. Zhang, D. Beyer, and F. Safai, "Self-adaptive SLA-driven capacity management for internet services," in *Network Operations and Management Symposium, 2006. NOMS 2006. 10th IEEE/IFIP*, 2006, pp. 557-568.

[30] A. Keller and H. Ludwig, "The WSLA framework: Specifying and monitoring service level agreements for web services," *Journal of Network and Systems Management*, vol. 11, pp. 57-81, 2003.

[31] M. N. Bennani and D. A. Menasce, "Resource allocation for autonomic data centers using analytic performance models," in *Autonomic Computing, 2005. ICAC 2005. Proceedings. Second International Conference on*, 2005, pp. 229-240.

[32] V. Stantchev and C. Schröpfer, "Negotiating and enforcing qos and slas in grid and cloud computing," *Advances in Grid and Pervasive Computing*, pp. 25-35, 2009.

[33] V. C. Emeakaroha, M. A. S. Netto, R. N. Calheiros, I. Brandic, R. Buyya, and C. A. F. De Rose, "Towards autonomic detection of sla violations in cloud infrastructures," *Future Generation Computer Systems*, 2011.

[34] Y. Dai, Y. Xiang, and G. Zhang, "Self-healing and Hybrid Diagnosis in Cloud Computing," *Cloud Computing*, pp. 45-56, 2009.

[35] Z. Zheng. *Distributed Reliability Assessment Mechanism for Web Services*. Available: http://www.wsdream.net/

[36] Y. Zhang, Z. Zheng, and M. R. Lyu, "Exploring latent features for memory-based QoS prediction in cloud computing," in *Reliable Distributed Systems (SRDS), 2011 30th IEEE Symposium on*, 2011, pp. 1-10.

[37] Z. Zheng, Y. Zhang, and M. R. Lyu, "Distributed qos evaluation for real-world web services," in *Web Services (ICWS), 2010 IEEE International Conference on*, 2010, pp. 83-90.

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 53-62
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

A Model for Recalibrating Credibility in Different Contexts and Languages - A Twitter Case Study

Amal Abdullah AlMansour
Department of Computer Science
King's College London
London WC2R 2LS, UK
amal.almansour@kcl.ac.uk

Ljiljana Brankovic
School of Electrical Engineering and Computer Science
University of Newcastle
Callaghan, Australia
ljiljana.brankovic@newcastle.edu.au

Costas S. Iliopoulos
Department of Computer Science
King's College London
London WC2R 2LS, UK
c.iliopoulos@kcl.ac.uk

ABSTRACT

Due to the growing dependence on the WWW User-Generated Content (UGC) as a primary source for information and news, the research on web credibility is becoming more important than ever before. In this paper we review previous efforts to evaluate information credibility, focusing specifically on micro-blogging. In particular, we provide a comparison of different systems for automatic assessment of information credibility based on the used techniques and features, and we classify the Twitter credibility surveys based on the features considered. We then propose a general model to assess information credibility on UGC different platforms, including Twitter, which employs a contextual credibility approach that examines the effect of culture, situation, topic variations, and languages on assessing credibility, using Arabic context as an example. We identify several factors that users may consider in determining credibility, and argue that the importance of each factor may vary with a context. Future work will include both a user study and machine learning techniques to evaluate the effectiveness of various factors for information credibility classification in different contexts.

KEYWORDS

User-Generated Content; Social Networks; Micro-blogs; Credibility; Arabic; Machine Learning.

1 INTRODUCTION

Blogs, wikis, social networking sites, recommendation sites and other types of UGC platforms are becoming a primary source for sharing and spreading news, information, and experiences between people around the world. These services enable its members to easily and freely publish whatever they like. On a downside, this openness and low publishing barriers can lead to the creation of huge amounts of data, containing a substantial quantity of inaccurate content. The presence of misleading, questionable and inaccurate information may have detrimental effects on people's beliefs and decision-making and may create a public disturbance (e.g., Apple shares[1] and Chile earthquake[2]). Consequently, there is significant need to evaluate information coming from UGC platforms to differentiate credible information from misinformation and rumors.

However, assessing the information credibility is a challenging task since "credibility" is a complex concept that is based on at least two key dimensions: trustworthiness and expertise [4, 31]. Both are judged by users consuming information, and therefore, credibility is dependent on users' cognitive states, as users usually invoke cognitive heuristics to assess the credibility [3, 4]. In other words, credibility is determined by the 'subjective judgment and assessment from the users' [2]. Also, credibility is situational and contextual; it varies from one context to another. A person sometimes accepts certain information as credible primarily by relying on the context in which he/she encountered the information [3]. Credibility assessments need to be considered relative to both people credibility judgments and credibility

[1] http://money.cnn.com/2008/10/03/technology/apple/
[2] http://irevolution.wordpress.com/2010/06/30/crowdsourcing-detective/

contexts such as environment, situations, expectations, etc. [3, 4].

Although there is now a variety of literature on credibility evaluation for different UGC domains, most prior credibility models were based on English content, labeled by mainly western users, which limits their applicability. Indeed, credibility perception is subjective: different groups may have different opinions. In this paper we contribute to breaching this gap by developing a framework based on a group personalized credibility measurement for Arab countries audiences.

Among various UGC platforms, we limit our focus to Twitter, which has been considered to be the most pertinent social medium used as news source [5, 6]. First, we report on survey data that examine credibility perceptions among Arab countries audiences. We then go on to identify how users from Arab countries consume micro-blog content, and discuss how to incorporate these findings into classifying Twitter information automatically in different cultural settings. Then, using machine learning methods, mainly supervised learning techniques, we study the effect of the context on the credibility classification results of a given classification technique.

There are different useful features proposed by the previous studies; however, the effect of the message language, topic, and location on these detection features has not been considered. We suggest that assessing information credibility with respect to its context can contribute to understand how and why users carry out with their credibility judgments and argue that the perceived information credibility will change depending on the context.

Our main contribution to the information credibility research is as follows.

(1) We compare different systems for automatic assessment of information credibility based on the used techniques and features.

(2) We classify the Twitter credibility surveys based on the features they considered.

(3) We introduce a new credibility model that deals with different contexts and supports Arab people credibility assessments.

(4) We identify the factors which affect users when evaluating the credibility of information.

(5) By identifying these factors, we enrich existing automatic credibility classification techniques, for example by adjusting features' weights related to specific factors depending on their occurrence and importance to the end users.

It should be noted that in this work we do not introduce a new method for assessing credibility automatically, but rather we propose a general model and empirically study how, and to what extent, the existing automatic credibility measurement can be used to identify credible information in different context.

This paper has two goals: first, it aims to draw a clear view of the techniques and surveys used for micro-blogging credibility; second, it aims to present a model for analyzing credibility in the context of three use cases: using Arabic language, covering different topics, and labelling by Arab viewers.

This paper is organized as follows. Section 2 presents related Twitter information credibility research, along with credibility contexts surveys. Section 3 and 4 present a proposed solution along with future work.

2 RELATED WORK

Most previous research related to this work falls into these broad categories: automatic measurement of Twitter information credibility, Arabic content credibility, and Twitter credibility surveys.

2.1 Credibility definition

We start the literature review by defining the essential and the relevant terms to our research:

- We use Rieh's [1] definition of *credibility* as "people's assessment of whether information is trustworthy based on their own expertise and knowledge". There are numerous other definitions,

including that of Oxford Dictionary (www.oxforddictionaries.com), which defines credibility as the quality of being convincing or believable.

- We define *context* as "the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood" based on Oxford Dictionary (www.oxforddictionaries.com). In this study we count any information related to the circumstances of the assessed message as the context.
- From the two definitions above, we define *contextual credibility* as the user's judgment to believe a message, given all the other information available.

2.2 Automatic measurement of information credibility

Credibility has been studied in UGC domain especially Twitter, the micro-blogging service [21, 23]. Twitter has become a popular source to many people for up-to-date information about what is happening in the world. Its ability to deliver to users "unmonitored" information easily and freely has marked it as a potential place for rumors and spams. Thus, there is a great need for studying its credibility. In assessing information credibility automatically, different approaches rely on different methods and factors to identify information credibility. Some use machine learning, mainly supervised learning methods, and some use graph analysis, while others use only statistical analysis such as feature distribution to predict credibility. Most of existing techniques are categorized in Table 1.

There is a wide range of features proposed in different studies to assess credibility of tweets. Most of these studies rely on Twitter features related to the messages' author and content [7, 8, 9, 10, 11, 12, and 22] for assessing information credibility. Some studies focus on the linguistic features of the content [13, 14, 20, and 21] and others focus on one or two Twitter specific factors and check their validity such as the work by Kang et al. 2013 [12] and O'Donovan et al. 2012 [22], who studied how re-tweet chain length and dyadic exchanges are used as metrics to measure credibility. In reference to tested datasets, some studies use different topics for assessing credibility [7, 8, 9, 13, 14, 20, and 21], while others only rely on a single topic dataset [10, 11, 12, and 29].

Table 1. Assessing Information Credibility Automatically.

Model Used	Used by - Examples
Classifier-based feature: different features	Castillo, Mendoza, & Poblete 2011 [7]; Gupta & Kumaraguru 2012 [8]; Yang et al. 2012 [9]; Xia et al. 2012 [10]; Kang, O'Donovan, & Höllerer 2012 [11]; Kang et al. 2013 [12].
Classifier-based feature: mainly linguistic features	Qazvinian et al. 2011 [13]; Bhattacharya et al. 2012 [14].
Graph-based / Hybrid (Classification with Graph)	Ratkiewicz et al. 2011 [15]; Gupta, Zhao, & Han 2012 [16]; Ravikumar, Balakrishnan, & Kambhampati 2012 [17]; Ulicny & Kokar [18]; McKelvey & Menczer 2013 [19].
Weighting-based feature/ Content similarity with credible source: mainly linguistic features	Assigning scores by certain mathematical functions or algorithms to each feature. Al-Eidan, Al-Khalifa, & Al-Salman 2010 [20]; Al-Khalifa & Al-Eidan 2011 [21].
Statistical analysis	Features distributions by ODonovan et al. 2012 [22].

For the purpose of this study, we will only focus on investigating selective studies with a feature-based approach: classifier-based feature, weighting-based feature and features distributions model. Starting with *classifier-based feature approach*, Castillo, Mendoza, & Poblete 2011 [7] showed that supervised classifiers can be used to automatically judge the credibility of news topics based on different features. The characteristics that were taken into account include: the message itself, the user, all the tweets on the topic and the propagation of re-tweets. The ground truth was subjectively assessed by crowdsourcing website Amazon Mechanical Turk. The results showed that there are measurable differences in the way

messages propagate, which can be used to classify them automatically as credible or not credible. The authors also observed that messages which do not include URLs tend to be related to non-credible news, while messages with negative sentiment terms are related to credible news. The study results were very promising: accuracy equal to 89% for news/chat classification and 86% for credibility levels classification using J48 decision tree classification algorithms. However, it should be noted that their assessment was based on credibility of a trending topic rather than individual tweets.

Gupta & Kumaraguru 2012 [8] covered this point in their work: they used automated ranking techniques to assess credibility of individual tweets. Using supervised machine learning algorithms (SVM Rank) and information retrieval techniques (relevance feedback), they showed that ranking of tweets based on Twitter features can aid in assessing credibility of information in messages posted about an event. Tweets were labeled with the help from human annotators. They applied linear logistic regression analysis to identify the prominent Twitter features: content and user based. For evaluating the performance, they used the standard metric of NDCG (Normalized Discounted Cumulative Gain). Their results showed that only an average of 30% tweets contained information about the event while 14% were spam. Only 17% of the informative tweets were credible.

Another study to detect rumors automatically was carried out by Qazvinian et al. 2011 [13], who classified rumor-related tweets that match the regular expression of the keyword query used to collect tweets on Twitter Monitor. Rumors were already identified by the About.com's Urban Legends reference site. The authors analyzed the users' believing behavior about the rumor-related tweets and identified users that endorsed the rumor versus users who denied or questioned it. They built different Bayesian classifiers on several of mainly linguistic features and then obtained a linear function of these classifiers for retrieval of the two sets. Mean Average Precision was equal to 96% in the rumor retrieval task and 93% accuracy in belief classification task.

Rumor analysis and detection on Sina Weibo - China's leading micro-blogging service has been discussed by Yang et al. 2012 [9]. They collected a set of tweets related to rumor topics published by an official rumor busting account. Two new features were proposed, the client program and the event location in addition to the previously proposed content, user, and propagation features. The authors performed sets of experiments using SVM classifier to study the impact of incorporating these two new features in rumor classification. The classification accuracy before adding the new features were around 72% and with adding new features, classification accuracy was improved to around 77-78%. It was reported that user features were more effective than content to detect rumors.

Kang, O'Donovan, & Höllerer, 2012 [11] studied credibility of topic-specific information on Twitter and presented three computational models for predicting credible topic-specific information on Twitter: social, content-based and hybrid models. They conducted an online study to collect ground truth credibility assessments and to analyze the effects of Twitter context on perceived credibility. This study relies on a single topic "Libya" dataset. Preliminary experiments were performed using Bayesian classifiers (and a range of others) to develop a model based on the features of each prediction strategy. For the full experiment a J48 tree-based learning algorithm was used. These experiments produced similar accuracy results to Castillo, Mendoza, & Poblete 2011 [7] work.

A research that used *statistical analysis approach* to assess credibility has been conducted by ODonovan et al. [17]. The authors studied the distribution of features in four distinct contexts: 8 different topics, credibility levels and two behavioral contexts (re-tweet chains and dyadic interactions). The results of the study in case of topics diverse context showed that feature usage tends to increase in emergency situations. In case of credibility context, the best indicators of credibility included URLs, mentions, re-tweets and tweet length. In case of re-tweet chains context, most notable result was the prominence of the URL feature in the longer chains, occurring in 50% of the long chain context, indicating that

tweets with provenance links to other information do tend to get propagated more often. Similarly, longer tweets in terms of words and characters tend to appear more frequently in longer chains. In case of dyadic context, results showed that dyadic pairs tend to have more words, but shorter words than standard tweets. Also the graph results showed there is a high variance for feature occurrence across different topics, which is an additional motivation to our work.

2.3 Arabic content credibility

A research into Twitter credibility for Arabic content has been done by Al-Eidan, Al-Khalifa, & Al-Salman 2010 [20] and Al-Khalifa & Al-Eidan 2011 [21]. They proposed a system to evaluate the credibility of Twitter Arabic news content automatically using a *Weighting-based feature approach*. The system uses two approaches to assign credibility levels as follows.
(1) An evidence-based method based on the similarity between tweets and verified news sources.
(2) The evidence-based method based on the similarity with verified content, in addition to a set of proposed extra features, where the following formula was used to calculate the credibility score: "Credibility Score = 0.6 Similarity + 0.2 Inappropriate Words + 0.1 Linking to authoritative source + 0.1 Author feature".
Evaluations of the system showed that the first approach is more effective in evaluating the credibility of tweets. However, using this approach, the system rated the tweets to only two credibility levels: Low and High. It should be noted that the above method is only useful for tweets combined with credible external sources and didn't embrace most of prominent features proposed by previous research such as hash-tags, re-tweets, and emoticons. Moreover, there is a need to evaluate the credibility formula and its performance after addition of more features.

2.4 Twitter Credibility Surveys

To identify influential credibility factors, a number surveys has been carried out, most of which are summarized in Table 2.

Table 2. Twitter Credibility Surveys.

Survey Used	Twitter Features Considered
Kang, ODonovan, & Hollerer 2012 [11]	Users' information, followers, and re-tweets.
Canini, Suh, & Pirolli 2011 [23]	Expertise, social status, and word factors.
Pal & Counts 2011 [24, 25]	Author name
Morris et al. 2012 [26]	User name, user image, and message topic.
Yang et al. 2013 [27]	Gender, name style, profile image, location, network overlap, and message topic.
Westerman, Spence, & Van Der Heide 2012 [28]	Number of followers and the ratio between followers and follows.

Kang, ODonovan, & Hollerer 2012 [11] provided a user study to analyze the effect of varying source data: users' information, followers, and re-tweets on perceived credibility rating. Canini, Suh, & Pirolli 2011 [23] measured the extent to which expertise, social status, and visualization such as word factors affect both explicit and implicit judgments of credibility. The results from the user study reported that the domain of expertise factor had a strong influence on credibility judgments, and social status had a smaller influence. Additionally, visualization factor had the smallest influence on both sets of judgments. Neither tweets alone nor word clouds alone provide sufficient information for participants to grant a high credibility rating to a Twitter user, but the combination of presenting specific tweets along with a summary word cloud leads to higher judged credibility.

Pal & Counts 2011 [24, 25] studied how bias due to author's name value impacts the perception of quality of Twitter authors. They found that authors who had more followers received higher "interesting" ratings for their content when their user names were revealed. User names of organizations, rather than individuals, and those which were topically related to the tweet also received higher ratings than those which were not.

In their survey study, Morris et al. 2012 [26] manipulated several features of tweets: message

topic, user name, and user image to assess their impact on credibility ratings. They found that users are poor judges of credibility based on content alone, and are influenced by heuristics (e.g., an item has been re-tweeted) and biased systematically (e.g., topically-related user names seen as more credible). Regarding the topic type, respondents were more concerned about credibility related to news, political, emergency, and consumer oriented tweets. The author of the tweet was generally related to the most important features for the tweet's credibility. These included author influence, topical expertise, and reputation. Content-related features viewed as credibility-enhancing were containing a URL leading to a high-quality site, and the existence of other tweets conveying similar information. Morris et al.'s experiments also indicated that the tweet features might interact with different topical contexts. For example, the difference of default image versus other image types appeared to be the strongest for tweets on the topic of entertainment.

In [27], Yang et al. 2013 ran an online study in which participants from U.S. and China rated the credibility of tweets, with each of the factors of interest manipulated: gender, name style, profile image, location, network overlap and message topic. They found key differences between the two countries. Chinese users show relatively high trust and dependence on microblogs as an information source, greater acceptance of anonymously and pseudonymously authored content, and tend to be more depend on integrate multiple metadata when evaluating microblog credibility.

A study designed to examine the effects of the number of followers and the ratio between followers and followed had on ratings of trustworthiness has been conducted by Westerman, Spence, & Van Der Heide 2012 [28]. They found a curvilinear pattern between the number of followers and the Twitter user's credibility, such that too few or too many followers actually make a Twitter user seem less credible. In addition, the ratio between the number of followers and the number of people one follows has an effect on the degree to which a perceiver judges a target to be competent in a specific subject. That is, if one has many followers, but does not follow many others, that person is regarded as less of an expert.

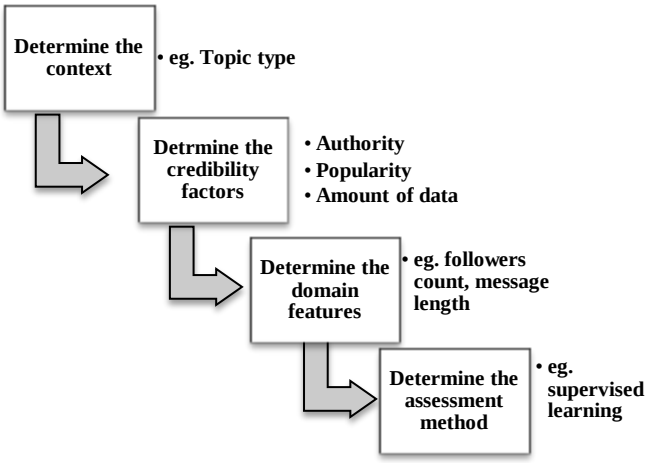
3 PROPOSED SOLUTION AND SYSTEM ARCHTECTURE

In this section we outline our proposed goal from two perspectives. First, we will conduct a user study to explore how people from Arab countries perceive and judge information credibility on Twitter. Second, we will investigate how useful existing methods are for classifying information credibility automatically given different contexts.

3.1 The overall process steps

The process flow for the proposed framework is presented in Figure 1. First, a context has to be determined. In another words, there is a need to specify the topic type and situation criticality. Then, it is crucial to determine what factors have most impact on end users in determining credibility in that specific situation. After that we should determine the available UGC domain features that can be linked to these factors and can be extracted and computed.

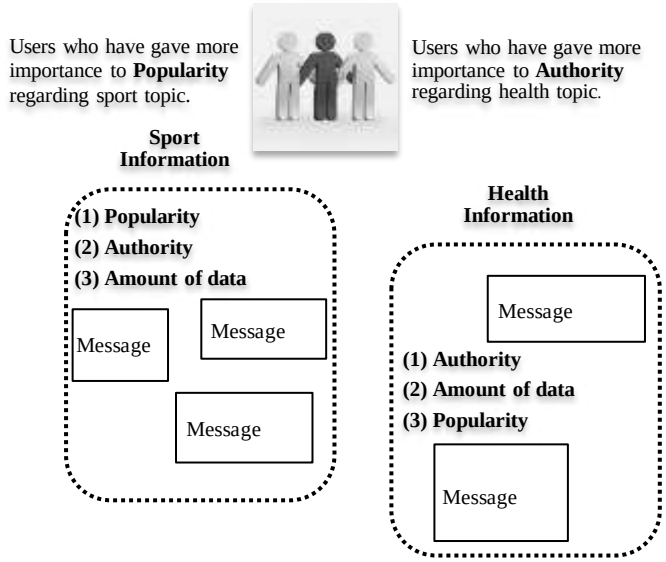
Fig. 1. Process Flow for the Proposed Framework.



3.2 Contexts, factors and features

In the process of modeling information credibility, we will need to identify: the contexts, the factors, and the features. Figure 2 shows how to consider all these steps and incorporated users importance weights.

Fig. 2. Proposed Framework



1. **Contexts:** In this work, we deal with credibility as a context-specific. There are infinite choices for context, so for our analysis we focus on a chosen subset related to the main elements (content, author, and reader):
    a. *Topic type and criticality:* Users may have multiple readings of credibility depending on to the type of information to be evaluated. As an example, consider a situation when a user evaluates health information. In this case, it appears that professional expertise source (Authority) would be more important than any other factor. Additionally, the topic situation should be labeled, for example as critical, normal or insignificant. We argue that in topics related to critical situations such as crises the credibility should be assessed differently. For example, in crises, the most novel information is typically distributed by unknown witnesses who may have weak credibility indicators, e.g., a small number of followers [8].

    H1: Individuals' perceived credibility for message and message's author will vary according to the topic genre.

    b. *Language:* Different languages may have different linguistic features that can affect measuring credibility. A study done by Alarifi & Alsaleh 2012 [30] in detecting Arabic and English spam web pages showed that the distributions of the content features vary according to the underlying language of the examined page. Moreover, content features that are studied in previous work for English tweets, such as the amount of character capitalization does not apply to the Arabic language.

    H2: Messages' features change depending on the language which means different languages will vary in their credibility measurement.

    c. *Location and culture:* We argue that defining the target user group for the credibility assessment is essential. People who are living in conservative locations such as the Arab countries [27] may have different strategies for judging credibility.

    H3: Different communities will have different credibility perceptions.

2. **Factors:** We selected the following credibility factors to use in our model since they have been common factors in several domains and can be adopted for automatic analysis.
    a. *Authority*
    b. *Popularity*
    c. *Amount of data*

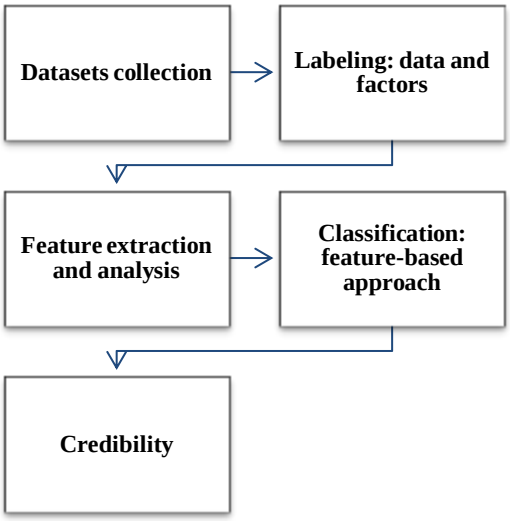
3. **Features:** The most prominent features that indicate the influence of the authors and the quality of the tweets should be identified from user survey. In the user survey, the users will rate the importance of each of the different attributes. The survey also should take in consideration the different contexts and see if users' answers remain the same. Provided below are some of the features that can be extracted and computed:
    a. *Reputation of Authors*: followers count, friends count, list count, statuses count, registration age, verified account, description, follower-friend ratio.

b. Quality of content: message length, unique characters count, words count, swear words count, presence of URL, hash-tags count, re-tweets count, mentions count, presence of sad or happy emoticons, presence of question mark, presence of exclamation mark, presence of pronouns, uppercase characters count. In terms of the aggregated posts on the topic, important predictors include: fraction of tweets having URL, fraction of negative and positive sentiment words, fraction of tweets with an exclamation mark!, fraction of pronounce, users count, hash-tags count.

3.3 Proposed System Architecture

The proposed system architecture consists of four main phases illustrated in Figure 3.

Fig.3. Proposed System Architecture



1. **Datasets collection:** We will collect messages related to different events. We consider Twitter hash-tags as representatives of topics and we will obtain major news topics: hard news topics such as crises (earthquake), politics, health, finance, technology and soft news topics such as art, entertainment, sports, and celebrities.
2. **Labeling:** Using the user study, participants from Arab countries will label the tweets of different topics and situations to measure the credibility degree. Users will not only label the date but also annotate an assessment for the credibility factors.
3. **Feature extraction and analysis:** Messages will be represented by a set of measured features.
4. **Classification:** Using a classification tool, the final datasets will be split into the training and testing data set; the training dataset will be used to develop a classifier, which will then be applied to the testing dataset to evaluate its accuracy in classifying the messages credibility. Two widely used classifiers are decision tree and naïve Bayes. Using weighting-based feature approach [20, 21] as a second technique, each feature is given an experimental weight based on its importance (users survey), and frequency of appearance in a tweet. Then we will use a mathematical formula to calculate the credibility score.

4 CONCLUSION AND FUTURE WORK

In this paper, we reviewed previous research on measuring information credibility related to UGC domain in particular micro-blogging platform. We argue the following:

(1) Even though previous research already proposed a feature-driven approach to assess credibility, it did not investigate the usefulness of these features in informing credibility judgments in different contexts such as different culture, language, topic and situation. Also incorporating user surveys' results in weighting the credibility factors and let them decide their importance then use them to compute credibility score has not been taking lots of attention in previous work.

(2) It is noted that by using different datasets, some of the used features were prominent regardless of the datasets, but many were not. This is a motivation for our work which aims to find the relation between dataset topics and the presence/absence of different features and whether different topics rely on different features. Majority of models in the current literature have used a dataset that include different events, and we aim to investigate whether assessing the information

credibility within a specific topic generates better results than assessing content across different topics.

(3) Unfortunately, not much research has been conducted to assess credibility of Arabic content. The majority of models are only tested on English datasets. Therefore, there is a need to apply previous assessment methods and investigate their usefulness with Arabic content. Also, it was noticeable that linguistic and sentiment features had a great impact on detecting credibility. It is recommended to identify the linguistic features that had been studied before for English content and investigate if some of them should be substituted with other features related to Arabic content (case sensitivity of English words, repeated letters, and word lengthening).

4.1 Recommendation and future work

The next phase in our research project plan is to report the results on (1) user credibility survey, (2) experimental study that examines how the selected features are being distributed within different contextual dimensions. Then, we would like to study (3) the effect of the context on the detection rate of a given classifying technique, and (4) incorporate the user survey results with the experiments results to help us identifying and correlating the credibility factors that influence credibility for each different scenario. Finally, (5) we intend to apply the credibility assessment on other UGC platforms.

5 REFERENCES

1. Rieh, S. Y.: Credibility and cognitive authority of information. In M. Bates & M. N. Maack (Eds.), Encyclopedia of Library and Information Sciences (3rd ed., pp. 1337-1344). New York: Taylor and Francis Group, LLC. (2010).
2. Bian, B.: Research of Factors on Impacting Internet Information Credibility Based on Electronic Commerce Users Demands. In Communication Systems and Network Technologies (CSNT), 2012 International Conference on. IEEE (2012).
3. Hilligoss, B., & Rieh, S. Y.: Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. Information Processing & Management 44.4 (2008).
4. Fogg, B. J.: Prominence-interpretation theory: Explaining how people assess credibility online. CHI'03 extended abstracts on Human factors in computing systems. ACM (2003).
5. Kwak, H., Lee, C., Park, H., & Moon, S.: What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (pp. 591-600). ACM (2010).
6. Cuesta, Á., Barrero, D. F., & R-Moreno, M. D.: A Descriptive Analysis of Twitter Activity in Spanish around Boston Terror Attacks. In Computational Collective Intelligence. Technologies and Applications (pp. 631-640). Springer Berlin Heidelberg. (2013).
7. Castillo, C., Mendoza, M., & Poblete, B.: Information credibility on twitter. Proceedings of the 20th international conference on World wide web. ACM (2011).
8. Gupta, A., & Kumaraguru, P.: Credibility ranking of tweets during high impact events. In Proceedings of the 1st Workshop on Privacy and Security in Online Social Media. ACM (2012).
9. Yang, F., Liu, Y., Yu, X., & Yang, M.: Automatic detection of rumor on Sina Weibo. In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. ACM (2012).
10. Xia, X., Yang, X., Wu, C., Li, S., & Bao, L.: Information credibility on twitter in emergency situation. In Intelligence and Security Informatics: Springer Berlin Heidelberg (2012).
11. Kang, B., O'Donovan, J., & Höllerer, T.: Modeling topic specific credibility on twitter. In Proceedings of the 2012 ACM international conference on Intelligent User Interfaces. ACM (2012).
12. Kang, B., Sikdar, S., Hollerer, T., ODonovan, J. and Adali, S.: Deconstructing Information Credibility on Twitter, 22nd International World Wide Web Conference(WWW), (2013).
13. Qazvinian, V., Rosengren, E., Radev, D. R., & Mei, Q.: Rumor has it: identifying misinformation in microblogs. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, (2011).
14. Bhattacharya, S., Tran, H., Srinivasan, P., & Suls, J.: Belief surveillance with twitter. In Proceedings of the 3rd Annual ACM Web Science Conference. ACM (2012).
15. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F.: Truthy: mapping the spread of astroturf in microblog streams. In Proceedings of the 20th international conference companion on World wide web. ACM (2011).
16. Gupta, M., Zhao, P., & Han, J.: Evaluating Event Credibility on Twitter. SIAM International Conference on Data Mining SDM (2012).
17. Ravikumar, S., Balakrishnan, R., & Kambhampati, S.: Ranking tweets considering trust and relevance. In Proceedings of the Ninth International Workshop on Information Integration on the Web. ACM (2012).
18. Ulicny, B., & Kokar, M. M.: Automating Military Intelligence Confidence Assessments for Twitter Messages.

19. McKelvey, K. R., & Menczer, F.: Truthy: enabling the study of online social networks. In Proceedings of the 2013 conference on Computer supported cooperative work companion. ACM (2013).
20. Al-Eidan, R. M. B., Al-Khalifa, H. S., & Al-Salman, A. S.: Measuring the credibility of Arabic text content in Twitter. In Digital Information Management (ICDIM), 2010 Fifth International Conference on. IEEE (2010).
21. Al-Khalifa, H. S., & Al-Eidan, R. M.: An experimental system for measuring the credibility of news content in Twitter. International Journal of Web Information Systems 7.2 (2011).
22. ODonovan, J., Kang, B., Meyer, G., Hollerer, T., & Adalii, S.: Credibility in context: An analysis of feature distributions in twitter. In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). IEEE (2012).
23. Canini, K. R., Suh, B., & Pirolli, P. L.: Finding credible information sources in social networks based on content and social structure. In Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom). IEEE (2011).
24. Pal, A., & Counts, S.: What's in a@ name? how name value biases judgment of microblog authors. In ICWSM 2011 (2011).
25. Pal, A., & Counts, S.: Identifying topical authorities in microblogs. In Proceedings of the fourth ACM international conference on Web search and data mining. ACM (2011).
26. Morris, M. R., Counts, S., Roseway, A., Hoff, A., & Schwarz, J.: Tweeting is believing?: understanding microblog credibility perceptions. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. ACM (2012).
27. Yang, J., Counts, S., Morris, M. R., & Hoff, A.: Microblog Credibility Perceptions: Comparing the United States and China. In Proceedings of the 2013 conference on Computer supported cooperative work, (2013).
28. Westerman, D., Spence, P. R., & Van Der Heide, B.: A social network as information: The effect of system generated reports of connectedness on credibility on Twitter, Computers in Human Behavior, 28, pp. 199-206, (2012).
29. Mendoza, M., Poblete, B., & Castillo, C.: Twitter Under Crisis: Can we trust what we RT?. In Proceedings of the first workshop on social media analytics. ACM (2010).
30. Alarifi, A., & Alsaleh, M.: Web Spam: A Study of the Page Language Effect on the Spam Detection Features. In Machine Learning and Applications (ICMLA), 2012 11th International Conference on. Vol. 2. IEEE (2012).
31. Wathen, C. N., & Burkell, J.: Believe it or not: Factors influencing credibility on the Web. Journal of the American society for information science and technology, 53(2), 134-144, (2002).

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 63-78
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Factors Generating Risks during Requirement Engineering Process in Global Software Development Environment

Huma Hayat Khan, Mohd. Naz'ri bin Mahrin, Suriayati bt Chuprat
Advanced Informatics School,
Universiti Tecknologi Malaysia,
hkhuma2@live.utm.my, mdnazrim@utm.my, suria@ic.utm.my

ABSTRACT

Challenges of Requirements Engineering become adequate when it is performed in global software development paradigm. There can be many reasons behind this challenging nature. "Risks" can be one of them, as there is more risk exposure in global development paradigm. So it is may be one of the main reasons of making Requirement Engineering more challenging. For this first there is a need to identify the factors which actually generate these risks. This paper therefore not only identifies the factors, but also the risks which these factors may generate. A systematic literature review is done for the identification of these factors and the risks which may occur during requirement engineering process in global software development paradigm. The list leads to progressive enhancement for assisting in requirement engineering activities in global software development paradigm. This work is especially useful for the, less experience people working in global software development.
.

KEYWORDS

Systematic literature review (SLR), Requirement engineering (RE), global software development (GSD), Software, Risks, Factors, distributed software development (DSD).

1 INTRODUCTION

The emergence of globalization concepts has impacted almost every industry, both in positive and negative ways. The word globalization takes into account with multi cultural stakeholders on a single platform. Software industry is also influenced by this globalization by allowing multicultural stakeholders to work together in global platform recognized as global software development environment [1].

Requirement engineering in GSD paradigm is one of the interesting research topics as described by cheng et al. [2]. This research rise for globalization in software industry is due to the number of advantages it has comparing with the traditional software development process. The advantages include; round the clock development, hiring workforce at low cost, maximum chance to the access the highly qualified global pool etc.

The global software development paradigm describes the fact of undergoing changes to many RE activities as the participants are not collocated. The new paradigm of GSD increases the risks of project failure irrespective of its huge number of advantages. To cope up this issue, RE pitfalls due to GSD should be overcome. These pitfalls are mostly due to the differences of culture, languages, knowledge, times zone etc which vary among software development organizations in GSD paradigm. These changing situation factors are the main source of software failure which is specifically influential in RE process as discussed by [3,4].

The goal of this paper is to identify and enlist the factors and the risks generated by these factors during RE process in GSD paradigm. The work compiles the changing situations factors which should be taken care with, to minimize the risks related to various aspects which, may lead to project failure. The information used for this identification is taken from the literature by performing systematic literature review (SLR) [5]. In order to have unique identification of each of the factors generating risks, Grounded theory's [6] constant comparison and memoing steps are adopted.

The rest of the paper is structured as follows. Section II describes the background of the study, section III illustrate the methodology of the study, section IV comprises of results, section V consist of discussion and section VI is the conclusion of the study.

2 BACKGROUND

Wiegers et al. [7] described requirement as a statement which relates to customer needs, objectives, capability or a condition that must be possessed by the product to satisfy and give value to a stakeholder. So we can say that requirement is something which a system must have or satisfy or perform which is being identified by the client side. Now coming towards Requirement Engineering, it is essential to notify that generally, RE is concerned with understanding about what are the things which system must do (the 'what').

A definition by Zave [8] states "Requirements Engineering is the branch of software engineering concerned with the real-world goals for, functions of, and constraints on software systems". Sommerville & Sawyer [9] argue about Requirement Engineering as "the activities that cover discovering, analyzing, documenting and maintaining a set of requirements for a system". By supporting the definitions described above Wiegers [7] recommend that Requirement Engineering cover all the software project lifecycle activities related to the understanding of not only capabilities but also the attributes of a system. Similarly in the same year Deb Jacobs argues about the importance of requirement engineering and says "the cost of incorrect, misunderstood, and not agreed upon requirements affects all of us in terms of time, money, and lost opportunities" [10].

Few years back researchers Fowler [11] argued that "Everything else in software development depends on the requirements. If you cannot get stable requirements, you cannot get a predictable plan". Carmel [12] argues that defining and acquiring the software needs for the new system is challenging and it is one of the crucial phases of software development as discussed by Darke [13]. Davis [4] Anthony [14] explains that it is crucial because it has a direct impact on success and failure of any software. Software requirement specification argued by Greenspan [15] is the outcome document of requirement engineering phase consisting of specified requirements. Continuing to the previous era researchers, there are some more researchers who in 21st century says that this requirement engineering phase is difficult and crucial enough when it is done in co-located environment as described by Damian [16] and, it is further argued by the researchers Damian [16, 17], Espinosa [18], MacGregor [19] that it becomes even more difficult and challenging when different stakeholders, sitting in different geographies having distant cultures, time zones etc. specify requirements.

Requirement engineering process becomes more complicated in globally distributed development paradigm, due to fact of having multiple stakeholders with varying backgrounds, for example. Having requirements common understanding is already a difficult or complex task to takes place within one organization in co located environment, but it becomes even more complex or harder when the stakeholders are having varying tacit knowledge, different time zones as it makes communication much harder. Platform of global software development further complicates requirement engineering due to social and cultural aspects related with not only gathering but also managing requirements [20].

The issue of having more complicated RE process in GSD is due to many reasons. The more risk occurrences in GSD paradigm may be the main contributor in the RE process complexity. Now first of all there is a need to identify the factors which may generate these risks, which ultimately may influence the RE process in GSD. There are numbers of factors which may impact RE process in GSD. These factors may be related to culture, social aspects, technologies etc, as discussed above, which may generate risks for the RE process [21, 22, 23]. These risks may results in project failure and one of the various sources of these risks are the factors such as: technology, culture, human etc leading to changing situation among the software industry working in GSD paradigm. Besides this, authors in their work also identified the sources for requirement engineering risks in global software development environment. They defined them as *communication and*

distance, which occur due to the reason of dispersion of multiple stakeholders across multiple countries and time zones*; knowledge management and awareness*, which occur due to the difficulties of keeping awareness, knowledge cohesion and knowledge coherence when various working groups concurrently access it; *cultural differences*, risks which are derived where people who have different culture interact with each other; *management and project coordination*, these are the risks derived from the organization establishments, definition of roles and coordination of procedures; *tools which support the processes*, risks that are derived due to the lacking of tools that support the requirement engineering process; and *clients*, these are, risks due to the interaction among clients not present in collocated sites [23].

So from the previous literature evidence we come to know that the researchers have not only focused their work to the criticality of RE process in GSD [21, 22, 23], but they have also describe the importance of control of changing situations among software development organizations in order to have successful project. By linking the facts from [3, 21, 22, 23], we come to know the importance of changing situations while performing RE process in GSD. However, to the best of our knowledge, there is a lack of study which describes the list of factors propagating risks during RE process in GSD. Therefore, this study focuses on identification and listing risks with that of factors which may acts as a source of these risks while performing RE process in GSD paradigm.

3 METHODOLOGY:

In our effort to review, we follow the method described in [5]. We have decomposed the research in three parts: *Review planning, Review conduction* and *Results reporting.*

3.1. Review planning:

Review planning deals with the selection of papers for review. Research goals and research questions are identified in this phase. Besides this, keywords with sources, queries, inclusion/exclusion criteria are also identified here.

Research goal and research questions: The goal of this literature review is to identify the factors which may generate risks during requirement engineering process in global software development paradigm.

RQ1: What are the factors which may generate risks during Requirement Engineering process in global software development (GSD) paradigm?

Identifying the keywords: Base upon the research question stated above, we go for keywords which in fact facilitate us in queries construction with that of the selection of relevant papers from the datasets: Software Requirement engineering (RE) risks, distributed requirement engineering risks, software risks factors, distributed requirement engineering risk factors.

Identifying the sources: The databases we considered for the search are: ACM Digital Library, Emerald, IEEE, Springer-Link, Science Direct, Wiley online and JSTOR.

Identifying the queries: The following abstract query for the research questions is used**:** '((software "requirement engineering") AND ("Risks")) AND (“distributed software development” OR “global software development”)'.

Due to the different search interface of each of the search engines, the query is divided into sub queries as some of them does not accept long query.

Identifying the inclusion/exclusion criteria: We have three levels of inclusion and exclusion. First we excluded all those papers which are either table of contents or some information about the full proceedings of conference and workshops etc. The second level is associated with checking papers on basis of keywords. So if the paper does not has any of the keywords (“requirement” OR “requirement engineering”) AND (“risk” OR “risk factors”), then that particular paper is excluded from the dataset. All the papers which must consist of keywords (“requirement” OR “requirement engineering”) AND (“risks” OR “risk factors”) with the other one as mentioned above are included in the dataset. The third level

of exclusion is on basis of repetition; like if a paper is repeated, then it is included only once.
We have divided our study into several steps. *Figure 1* describes the steps taken to identify the factors and the risks generated by those factors.

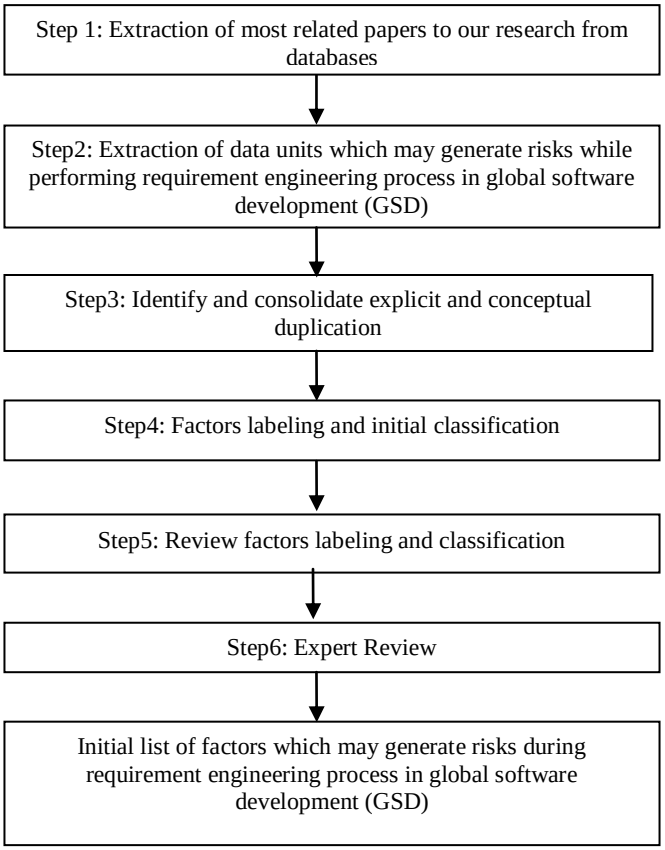


Figure 1. Steps for Generating Initial List of Factors

3.2. Review Conduction:

Step 1 as shown in *figure 1*consists of building the initial data sets and filtration of papers based on inclusion and exclusion criteria. To accomplish this step, we selected seven data sources to get the papers related to our study by using automated query.
Building the initial dataset: The papers are collected on basis of keywords and queries. From ACM we got (113) papers, similarly from Emerald (43) papers, IEEE (67) papers, JSTOR (5) papers, Willey online (67) papers, Science Direct (89) papers and Springer Link (90) papers.
Papers filtration based on the inclusion/ exclusion criteria: We perform filtration on initial data set as described above. These filtrations help us to decide whether particular papers should stay in or excluded from the data set. We filtered papers on basis of their type, keywords and repetition. So on basis of "Type" we excluded all the papers which are either table of contents or any definition document or information about full proceedings of conferences or workshops. After that we excluded papers on basis of keywords. So if the paper does not has any of the keywords ("requirement" OR "requirement engineering") AND ("risk" OR "risk factors"), then that particular paper is excluded from the dataset. All the papers which must consist of keywords ("requirement" OR "requirement engineering") AND ("risks" OR "risk factors") with the other one as mentioned above are included in the dataset. In the last level of exclusion we excluded all the repeated papers and included them only once.

Table 1. Papers Distribution

Data sources	Total	Covered
ACM	47	24
Emerald	4	4
IEEE	26	16
JSTOR	2	1
Willey Online	29	22
Science Direct	28	10
Springer Link	36	28
Total		105

Table1 shows the statistics of papers. Total papers which we find to be included in our research are 172 but after performing filtrations on basis of type, keywords and repletion, the papers covered are 105.

4 RESULTS

4.1. Result Reporting:

This stage of our research comprises of steps from step2 to step6, where we report all the results of our study with that of the comments from the experts related to the factors identified.
Step2 as shown in *figure 1* deal with the extraction of data units from the most related filtered. These data units are actually the identified situational factors, which may generate risks while

performing requirement engineering process in global software development (GSD) paradigm. These identified factors which are large in numbers have to go through a filtration process. This time the filtration is done on basis of conceptual and explicit duplications.

Step3 as shown in *figure 1* consist of steps to remove the duplication (explicit or conceptual). In order to do this, Grounded theory's data coding, constant comparison and memoing techniques are adopted. This theory by Glaser [6] is the methodology for analyzing the data in a systematic. Data coding deals with not only getting raw data but also converting it to a conceptual level as argued by Corbin [24]. In this research, a number of data codes from the multiple data sources are compared via constant comparison for getting the main factors by removing the duplications. Glaser argues about complexity of data coding with constant comparison [6] by saying that ''simultaneously many categories and their properties may be emerging at different levels of conceptualization and different ways of being related by theoretical codes''. So in order to deal with this complexity researchers argue that the memoing process can be a solution [25]. Hence this step consists of scanning the data to find the duplication instances. As there is a direct mapping among the individual data units that are joined or combined in this comparison, so memoing is not required as it shows the thoughts that influenced the joining or combination of data units. Where while combining the data units, careful steps are taken to maintain the source information. So, if two different data units are having same textual depiction or explanation and meaning, they are combined into a single unit – but it is also to be making sure that both sources of data units are apparent and clear in the combined data unit testimony in the main table. Now following the identification and removal of clear duplication among the data units, conceptual duplications in data units are identified. We combine two data units which are not having same textual tags but are having same meaning. In order to do this, memoing is also done which records the thought process behind this. Besides this, as previously done with the removal of clear duplication stage, combined data units sources links are maintained.

Step4 as shown in *figure 1* comprises of factors labeling and initial classification. Here each of the data units (factors) is given a label. These labeling are on basis of factors relatedness to the specific area they are concern with. Memos are used to record the justification for these labeling. Similarly actions are taken to come up with initial classification for data units (factors). This time also, memos draft the motivation or justification that has initiated or created the initial classification. Under this initial classification, factors are present with their appropriate labeling. Similar to the previous steps, memos are used here to record the justification for classification and labeling.

Step5 comprises of constant comparison for evaluating the precision of the factor labels and classifications. The labels are renamed when suitable data units are moved to alternating factors or to completely novel factors – as believed suitable in regard of the rising factors and classification. Similarly some of the classifications are renamed or combined or decomposed as considered appropriate. For this step similar to the previous one, memos are used to draft the thought process with that of the historical trace of the actions. This assists in envisioning the factors and classification sources, and allows an assessment of the impact of each data sources on the main list as it come forward.

Step6 comprises of expert reviews. Once the initial list of situational factors at its final form is created, then it is forwarded to the experts from academia to evaluate the initial list of Situational Factors which may generate risks during RE process in GSD paradigm. The experts are selected from academia on basis of their experience in Requirement Engineering field. More specifically two experts are selected having more than five years of experience in field of requirement engineering. The tasks provided to them is to evaluate the list for its comprehensiveness, as well as to review if the factors are grouped under right classification with that of any recommended modification both at factors and classification level.

The experts recommended some modification in the initial list of situational factors which may generate risks during RE process in GSD. Factors *Interaction tools* and *Interaction medium or technology* are combined and given the name of *"Interaction medium, technology and tools"*, similarly factor *national culture* is combined with *cultural background, social climate* is combined with *social background*, *organization policies and strategies* besides with the *organization person retention strategy* and *organization structure and boundaries* are grouped into single factor named as "*organization structure, policy and strategy"*, *technical expertise* is moved from the classification "*tools, technologies, techniques and standards"* to *"stakeholders"*. *Partner power* and *position in organization* which are previously considered to be same factors are now grouped separately under classification *"stakeholders"*. Under the same *"stakeholders"* classification, *client involvement* is combined with the factor *client commitment*. The recommendation from the experts is included in the list as shown in *table 2*.

Table 2 comprises of four columns: *Classification*, *Situational factors*, *Risks* and *References*. *Classification* and *factors* column contain the grouping of factors under related classification. Similarly the *Risk* column details all the related risks which may be generated by each of the situational factor identified. The last *references* column links the identified list of factors to the literature from where they are taken from.

The initial list of factors which may generate risks as shown above in *table 2* consist of 74 factors grouped under 8 classifications. Each of the classification is named based on the criteria of most frequently used classifications in existing published work. These classifications are further reviewed by the experts for its naming conventions and for the group of factors it contains as described above.

Communication and Distance: comprises of factors; *Interaction skills* (abilities that one can posses for better interaction), *Interaction styles* (formal, informal etc), *Interaction medium* (synchronous or en-synchronous communication mediums), *technology and tools* (web based discussion tools, groupware, search engines etc), *Interaction infrastructure* (copper cable, fiber, or wireless technologies etc) and *Distance* (geographic location).

Cultural, Background, Language, Organizational and Time Differences: comprises of factors; *Language* (language people use for communication), *Cultural background* (one's life experience as shaped by membership in groups based on ethnicity, race, socioeconomic status, gender, exceptionalities, language, religion, sexual orientation, and geographical area), *Work Environment* (the place that one works), *View point* (a technique of composition that provides a vocabulary for thinking about and acting upon movement and gesture), *Time zone* (time differences), *Social background* (how any one is raised), *Organizational culture* (it deals with group behavior, organization values, organization visions, organization norms, organization working language etc), *Political difference* (it deals with the government controls as it is argued that everything is controlled by them. So business cannot be started without visiting them), *Time shifts* (relates to different job timing like evening, morning shift and night shifts), *Inter group culture* (way of dealing, communicating, performing work etc among the team or group), *Organization Structure, policy and strategy* (deals with aspects like task allocation, coordination, task supervision, what is the criteria to choose members, leaders and decisions etc).

Knowledge Management and Awareness: comprises of factors; *Team Awareness* (team members ability to perceive, feel and consciousness etc), *Data repository* (how data is stored and maintained), *Domain knowledge level* (degree of valid knowledge to the specific discipline), *Knowledge management techniques and procedures* (techniques and procedures use to manage knowledge), *Business knowledge* (experience, design and process, files or documents, plans for future activities etc), *Access management* (deals with the access provided for different aspects like data etc), *Tacit knowledge level* (degree of tacit knowledge which is harder to explain but necessary for tasks), *Requirement Engineering practice knowledge* (knowledge related to RE practices), *Configuration management* (task to control and manage changes), *Knowledge management awareness*

(awareness about managing knowledge) and *Requirement management* (deals with ways to manage requirements that is add, delete modify etc)

Management: comprises of factors; *Coordination skills* (deals with ones skill to coordinate with others in the team), *Coordination technique* (techniques which are adopted to have effective coordination like, change of command, effective leadership and supervision etc) *Competence* (deals with ones ability to perform any job or task), *Decisions* (deals with management decisions), *Supervisor sub-ordinate relationship* (it deals with the relationship between the managers or team leads with the one who are working under their supervision e.g. friendly, strict etc), *Management strategy* (strategies related to managing things and work) and *Cooperation approach* (ways to make things balance or in harmony)

Tools, Technologies, Techniques and Standards Selection: comprises of factors*; Technique selection* (it deals with the criteria of technique selection for performing any task), *Standards selection* (it deals with the criteria of standard selection to be considered for task fulfillment), *Tools selection* (it deals with the criteria of tools selection for using in order to complete any task) and *Technology selection* (it deals with the criteria of technology selection for task completion).

Stakeholders: comprises of factors*; Team members competence and experience with in application* (it deals with the abilities and experience of team member related to application they are working on), *Team members motivation level* (how much desirable one to do work), *Team members familiarity with each other* (deals with the concept that whether the team members working together are strangers or they already know each other), *Team members preferences related to project* (it deals with individual attitude towards decision making or evaluation judgment), *Team knowledge exchange ability* (how well the team members involve in the project tasks share or exchange the knowledge among them), *Team members background* (experience, academic qualification, expertise etc), *Personnel trust* (the level of trust among the team members), *Leadership skills* (it deals with the person's ability to lead the people working under him/her), *Personnel/group relationship* (type of relationship among team members like friendly, honesty etc), *Team members decision capability* (ability of team members to take appropriate decisions), *Stakeholders utility values* (it deals with the stakeholder values related to multiple resources and facilities they want from organization they are working with), *Team members level of receiving help with heavy work load* (it deals with the help given to the team members for their heavy work load accomplishment in form of moral support or physical help etc), *Stakeholder priority to situation urgency* (this factor deals with different stakeholders priorities which they give to different situations like it may happen that one conflict or problem is seemed urgent to be resolved but for others it may not be given such priority), *Team members knowledge level* (deals with the knowledge of the team members not specific to the application but related to other aspects also), *Team members international work experience* (whether the team member has some past experience of doing work with people who are from some other countries), *Team member relation to the project* (team members trust and commitment for the project completion), *Partner power* (the authorities, team members from the other organization have and don't have), *Client availability* (deals with the availability of client for discussing and validating things), *Client commitment* (it deals with the client dedication towards project), *Members carriers prospects* (team members job alternatives), *position in organization* (corporate or business title of the team member like manager, executive etc), *Client background* (knowledge, culture, experience, expertise, type (technical or non-technical) etc of the client), *Partner interpretation skills* (deals with abilities to have accurate meaning of things), *Team management capabilities* (deals with the capabilities and abilities related to team management), *Stakeholders common work experience* (have they previously worked together or not) and *Technical expertise* (deals with the stakeholders expertise related to technical aspects.)

Project and Process: comprises of factors*; Project phase distribution* (it deals with the resources in

terms of time, cost, effort etc allocated each of the phase of software project development), *Requirements engineering process* (the set of activities involve for requirement engineering process accomplishment), *Economic process* (deals with the activities undertaken for the production and management of objects wealth), *Process maturity* (deals with the appropriateness and quality of a process use to accomplish task), *Collaboration process* (deals with the activities undertaken in order to realize shared goals) and *Management process* (deals with the planning and controlling the execution of the activities).

Requirements: comprises of factors*; Requirement specification document format* (the standard or format organization follow to generate the documents), *Requirement engineering effort* (amount of exertion expended, work done or attempts), *Requirement representation style* (deals with the styles of requirement representation like visual styles, natural language etc) and *Requirement interpretation* (deals with the understanding of requirements or meaning of requirements).

5 DISCUSSIONS

In this section, we talk about how this work can be helpful for research communities. This work also leads to some possible future work areas. In the end some limitations are also discussed.

5.1. Utilizing Situational Factors:

This work gives the direction towards a comprehensive list of factors. The identified situational factors which may generate risks during RE process in GSD paradigm can be an important reference for researchers, who are working in situational requirement engineering, as it shows the initial list of factors which may generate risks and which directly or indirectly may influence the requirement engineering process in GSD paradigm.

The initial list can also act as a guide for the researchers and practitioners working in situational requirement engineering, to consider and control these factors in order to overcome the risks with changing situations faced by them while performing RE process in GSD paradigm.

Besides with the initial list of factors, this work also contributes in listing the risks against each of the situational factor, hence we also come to have a repository of risks which may be generated by the situational factors during RE process in GSD paradigm.

5.2. FUTURE WORK

The initial list of factors identified is from the literature; hence the list does not contain the data from the industry. So in future a more comprehensive list can be generated covering both the literature and industry.

This work also shows the directions towards the future work of impact analysis of these factors on requirement engineering process in GSD paradigm, hence extending the investigation to a more comprehensive level.

5.3. LIMITATIONS

The scope of the initial list of situational factors is limited to only risks factors and for this only state of the knowledge is concerned and covered. Hence it does not cover the identification of factors from industry. So it is only from state of knowledge not including the state of practice. The steps to overcome this limitation is taken by designing a questionnaire and surveying the situational factors from the industry which will be ultimately included in the final comprehensive list of situational factors (which will be generated after covering and including the factors from various other domains related to RE process) affecting RE process in GSD.

We make every effort to cover all the related papers discuss risks for RE process in GSD, but still it is possible that we may miss any published work. Similarly the paper is forwarded to other researchers in order to deal with biasness about the search protocol used, but still biasness aspect cannot be ignored as well. The threat of misinterpretation can also not be ignored, as it is one of the must factor in every literature review, although we tried our best to overcome these aspect by dealing it carefully. We also cannot ignore the threat related to precision. In our work we tried to have high precision rate but still the maximum precision is not assured.

6 CONCLUSIONS

The goal of this paper is to identify and enlist the factors and the risks generated by them during RE process in GSD paradigm. The work compiles the changing situations factors which should be taken care with, to minimize the risks related to various aspects which, may lead to project failure. The information used for this identification is taken from the literature by performing systematic literature review (SLR) [5]. In order to have unique identification of each of the factors generating risks, Grounded theory's [6] constant comparison and memoing steps are adopted.

In our effort to review we have decomposed the research in three parts: *Review planning, Review conduction* and *Results reporting.* The initial list of situational factors is reported in the paper consisting of 74 factors grouped under 8 classifications. This initial list is an important input to the comprehensive list of situational factors, which is our future work. The initial list can also act as a guide for the researchers and practitioners working in situational requirement engineering, to consider and control these factors in order to overcome the risks with changing situations faced by them while performing RE process in GSD paradigm.

Besides with the initial list of factors, this work also contributes in listing the risks against each of the situational factor, hence we also come to have a repository of risks which may be generated by the situational factors during RE process in GSD paradigm.

7 REFERENCES

[1] Kathryn Fryer, M.G., Global software development and delivery: Trends and challenges. 2008.

[2] Betty H.C. Cheng, J.M.A., Research directions in requirements engineering. In Future of Software Engineering, in FOSE '07. 2007. p. pages 285-303.

[3] Loucopoulos, P. and V. Karakostas, System Requirements Engineering. 1995, New York.

[4] Davis, A.M., Software Requirements: Analysis and Specification. 1989: Prentice Hall. 352 pages.

[5] Kitchenham, B. and S. Charters, Guidelines for performing Systematic Literature Reviews in Software Engineering. 2007.

[6] Glaser, B.G., Doing Grounded Theory: Issues & Discussion 1998: Sociology Pr.

[7] Wiegers, K., Process Impact. 2007. <http://www.processimpact.com>.

[8] Zave, P., Classification of research efforts in requirements engineering. ACM Comput. Surv., 1997. **29**(4): p. 315-321.

[9] Ian Sommerville, P.S., Requirements Engineering: A Good Practice Guide 1ed. 1997: Wiley. 404 pages.

[10] Jacobs, D., Requirements Engineering so Things Don't Get Ugly, in Companion to the proceedings of the 29th International Conference on Software Engineering. 2007, IEEE Computer Society. p. 159-160.

[11] Fowler, M., The New Methodology. 2005. <http://www.martinfowler.com>.

[12] Carmel, E., Global software teams: collaborating across borders and time zones. 1999 NJ, USA Prentice Hall PTR Upper Saddle River.

[13] Peta Darke, G.G.S., User viewpoint modelling: understanding and representing user viewpoints during requirements definition. Information Systems Journal, 1997. **Volume 7**(Issue 3): p. pages 213-219.

[14] Terry Anthony Byrd, K.L.C., Robert W. Zmud, A Synthesis of Research on Requirements Analysis and Knowledge Acquisition Techniques. MIS Quarterly, 1992. **Vol. 16**(No. 1): p. pp. 117-138.

[15] Greenspan, S., J. Mylopoulos, and A. Borgida, On formal requirements modeling languages: RML revisited, in Proceedings of the 16th international conference on Software engineering. 1994, IEEE Computer Society Press: Sorrento, Italy. p. 135-147.

[16] E. Damian, D. and D. Zowghi, RE challenges in multi-site software development organisations. Requirements Engineering, 2003. **8**(3): p. 149-160.

[17] Damian, D.E. and D. Zowghi, The Impact of Stakeholders? Geographical Distribution on Managing Requirements in a Multi-Site Organization, in Proceedings of the 10th Anniversary IEEE Joint International Conference on Requirements Engineering. 2002, IEEE Computer Society. p. 319-330.

[18] J. Alberto Espinosa , E.C., John Wiley, The impact of time separation on coordination in global software teams: a conceptual foundation, in SOFTWARE PROCESS IMPROVEMENT AND PRACTICE. 2005. p. 2004.

[19] MacGregor, E., Y. Hsieh, and P. Kruchten, Cultural patterns in software process mishaps: incidents in global projects. SIGSOFT Softw. Eng. Notes, 2005. **30**(4): p. 1-5.

[20] Hanisch, J. and B. Corbitt. Requirements engineering during global software development: Some impediments to the requirements engineering process – a case study. in Proceedings of the Twelfth European Conference on Information Systems. 2004.

[21] Ågerfalk, P.J. and J. Ralyté, Situational Requirements Engineering Processes: reflecting on method engineering and requirements practice. Software Process: Improvement and Practice,2006. **11**(5): p. 447-450.

[22] Jo Hanisch, T.T., Brian Corbitt, Exploring the cultural and social impacts on the requirements engineering processes - highlighting some problems challenging virtual team relationships with clients. Journal of Systems and InformationTechnology, 2001. **Vol. 5**(Iss: 2): p. pp.1 - 20.

[23] Lopez, A., J. Nicolas, and A. Toval. Risks and Safeguards for the Requirements EngineeringProcess in Global Software Development. in Global Software Engineering, 2009. ICGSE 200 Fourth IEEE International Conference on. 2009.

[24] Corbin, J.M. and A.L. Strauss, Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2008: Sage Publications, Incorporated.

[25] Bryant, A. and K. Charmaz, The Sage Handbook of Grounded Theory. 2010: Sage Publications.

[26] Robinson, W.N., S.D. Pawlowski, and V. Volkov, Requirements interaction management. ACM Comput. Surv., 2003. **35**(2): p. 132-190.
[27] Lopez, A., J. Nicolas, and A. Toval. Risks and Safeguards for the Requirements Engineering Process in Global Software Development. in Global Software Engineering, 2009. ICGSE 2009. Fourth IEEE International Conference on. 2009.
[28] Ramesh, B., L. Cao, and R. Baskerville, Agile requirements engineering practices and challenges: an empirical study. Information Systems Journal, 2010. **20**(5): p. 449-480.
[29] Parviainen, P. and M. Tihinen, Knowledge-related challenges and solutions in GSD. Expert Systems, 2011: p. no-no.
[30] Komi-Sirviö, S. and M. Tihinen, Lessons learned by participants of distributed software development. Knowledge and Process Management, 2005. **12**(2): p. 108-122.
[31] Puhl, S. and R. Fahney. How to assign cost to avoidable requirements creep: A step towards the waterfall's agilization. in Requirements Engineering Conference (RE), 2011 19th IEEE International. 2011.
[32] Omoronyia, I., et al., A review of awareness in distributed collaborative software engineering. Software: Practice and Experience, 2010. **40**(12): p. 1107-1133.
[33] Ramzan, M., et al., Automated Requirements Elicitation for Global Software Development (GSD) EnvironmentSoftware Engineering, Business Continuity, and Education, T.-h. Kim, et al., Editors. 2011, Springer Berlin Heidelberg. p. 180-189.
[34] Perini, A., et al., Evolving Requirements in Socio-Technical Systems: Concepts and Practice Conceptual Modeling – ER 2011, M. Jeusfeld, L. Delcambre, and T.-W. Ling, Editors. 2011, Springer Berlin / Heidelberg. p. 440-447.
[35] Skroch, O., A method to evaluate the suitability of requirements specifications for offshore projects Developing Business Application Systems. 2010, Gabler Verlag. p. 59-78.
[36] Calefato, F., D. Damian, and F. Lanubile, Computer-mediated communication to support distributed requirements elicitations and negotiations tasks. Empirical Software Engineering, 2012. **17**(6): p. 640-674.
[37] Šmite, D. and J. Borzovs, A Framework for Overcoming Supplier Related Threats in Global Projects Software Process Improvement, I. Richardson, P. Runeson, and R. Messnarz, Editors. 2006, Springer Berlin / Heidelberg. p. 50-61.
[38] Lamersdorf, A. and J. Münch, A multi-criteria distribution model for global software development projects. Journal of the Brazilian Computer Society, 2010. **16**(2): p. 97-115.
[39] Boden, A., B. Nett, and V. Wulf, Articulation work in small-scale offshore software development projects, in Proceedings of the 2008 international workshop on Cooperative and human aspects of software engineering. 2008, ACM: Leipzig, Germany. p. 21-24.
[40] Dorairaj, S., J. Noble, and P. Malik, Effective Communication in Distributed Agile Software Development Teams Agile Processes in Software Engineering and Extreme Programming, A. Sillitti, et al., Editors. 2011, Springer Berlin Heidelberg. p. 102-116.
[41] Narayanaswamy, R. and R.M. Henry, Effects of culture on control mechanisms in offshore outsourced IT projects, in Proceedings of the 2005 ACM SIGMIS CPR conference on Computer personnel research. 2005, ACM: Atlanta, Georgia, USA. p. 139-145.
[42] Ning, A., et al., Requirements Engineering Processes Improvement:A Systematic View, in SPW 2005. 2005: Berlin Heidelberg. p. pp. 151 – 163.
[43] Lings, B., et al., Ten Strategies for Successful Distributed Development The Transfer and Diffusion of Information Technology for Organizational Resilience, B. Donnellan, et al., Editors. 2006, Springer Boston. p. 119-137.
[44] Moe, N. and D. Šmite, Understanding Lacking Trust in Global Software Teams: A Multi-case Study Product-Focused Software Process Improvement, J. Münch and P. Abrahamsson, Editors. 2007, Springer Berlin / Heidelberg. p. 20-34.
[45] Šmite, D., A Case Study: Coordination Practices in Global Software DevelopmentProduct Focused Software Process Improvement, F. Bomarius and S. Komi-Sirviö, Editors. 2005, Springer Berlin / Heidelberg. p. 25-46.
[46] Herbsleb, J.D., et al., An empirical study of global software development: distance and speed, in Proceedings of the 23rd International Conference on Software Engineering. 2001, IEEE Computer Society: Toronto, Ontario, Canada. p. 81-90.
[47] Luna-Reyes, L.F., et al., Knowledge sharing and trust in collaborative requirements analysis. System Dynamics Review, 2008. **24**(3): p. 265-297.
[48] Lamersdorf, A. and J. Münch, Model-Based Task Allocation in Distributed Software Development Software Engineering Approaches for Offshore and Outsourced Development, M. Nordio, et al., Editors. 2010, Springer Berlin Heidelberg. p. 37-53.
[49] Jiménez, M. and M. Piattini, Problems and Solutions in Distributed Software Development: A Systematic Review Software Engineering Approaches for Offshore and Outsourced Development, K. Berkling, et al., Editors. 2009, Springer Berlin Heidelberg. p. 107-125.
[50] Ullah Khan, S., M. Niazi, and R. Ahmad, Critical Success Factors for Offshore Software Development Outsourcing Vendors: An Empirical Study Product-Focused Software Process Improvement, M. Ali Babar, M. Vierimaa, and M. Oivo, Editors. 2010, Springer Berlin / Heidelberg. p. 146-160.
[51] Sengupta, B., S. Chandra, and V. Sinha, A research agenda for distributed software development, in Proceedings of the 28th international conference on Software engineering. 2006, ACM: Shanghai, China. p. 731-740.
[52] Mikulovic, V. and M. Heiss, "How do I know what I have to do?": the role of the inquiry culture in requirements communication for distributed software development projects, in Proceedings of the 28th international conference on Software engineering. 2006, ACM: Shanghai, China. p. 921-925.
[53] Valentine, C. and R. Ita. Project Management within Virtual Software Teams. in Global Software Engineering, 2006. ICGSE '06. International Conference on. 2006.
[54] Hossain, E., et al. Risk Identification and Mitigation Processes for Using Scrum in Global Software Development: A Conceptual Framework. in Software Engineering Conference, 2009. APSEC '09. Asia-Pacific. 2009.
[55] Orlikowski, W.J., Knowing in Practice: Enacting a Collective Capability in Distributed Organizing. Organization Science, 2002. **13**(3): p. 249-273.
[56] Calefato, F., F. Lanubile, and C. Ebert, Practice: Collaborative Development Environments, in Global Software and IT. 2011, John Wiley & Sons, Inc. p. 109-123.
[57] Musio, I., IBM Industry Practice: Challenges in Offshore Software Development from a Global Delivery Center Software Engineering Approaches for Offshore and Outsourced Development, O. Gotel, M. Joseph, and B. Meyer, Editors. 2009, Springer Berlin Heidelberg. p. 4-13.
[58] Krishnan, P. and C. Ranganathan, Boundary spanning in offshored ISD projects: a project social capital perspective, in Proceedings of the special interest group on management information system's

47th annual conference on Computer personnel research. 2009, ACM: Limerick, Ireland. p. 225-230.
[59] Lamersdorf, A., et al. A Rule-Based Model for Customized Risk Identification in Distributed Software Development Projects. in Global Software Engineering (ICGSE), 2010 5th IEEE International Conference on. 2010.
[60] Khoshroo, B.M. and H. Rashidi. Distributed development ; A literature analysis on challenges and solutions. in GCC Conference & Exhibition, 2009 5th IEEE. 2009.
[61] Šmite, D., Global software development projects in one of the biggest companies in Latvia: is geographical distribution a problem? Software Process: Improvement and Practice, 2006. **11**(1): p. 61-76.
[62] Casey, V. and I. Richardson, Implementation of Global Software Development: a structured approach. Software Process: Improvement and Practice, 2009. **14**(5): p. 247-262.
[63] Jensen, R.E. and P. Bjørn, Divergence and Convergence in Global Software Development: Cultural Complexities as Social Worlds From Research to Practice in the Design of Cooperative Systems: Results and Open Challenges, J. Dugdale, et al., Editors. 2012, Springer London. p. 123-136.
[64] Berlizev, A. and N. Guelfi, Correct analysis for embedded system modeling: an outcome of east-west scientific cooperation, in Proceedings of the 2008 international workshop on Software Engineering in east and south europe. 2008, ACM: Leipzig, Germany. p. 23-30.
[65] Dai, C.-f. and M.-l. Wang, Distributed requirement elicitation and negotiation based on the hall for workshop of meta-synthetic engineering, in Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human. 2009, ACM: Seoul, Korea. p. 1452-1456.
[66] Al-Ani, B., et al., Trust and surprise in distributed teams: towards an understanding of expectations and adaptations, in Proceedings of the 4th international conference on Intercultural Collaboration. 2012, ACM: Bengaluru, India. p. 97-106.
[67] Dalberg, V., et al., Cross-cultural collaboration in ICT procurement, in Proceedings of the 2006 international workshop on Global software development for the practitioner. 2006, ACM: Shanghai, China. p. 51-57.
[68] Sudhakar, G.P., A. Farooq, and S. Patnaik, Soft factors affecting the performance of software development teams. Team Performance Management, 2011. **Vol. 17**(Iss: 3): p. pp. 187 - 205.
[69] Miller, J.A., R. Ferrari, and N.H. Madhavji, An exploratory study of architectural effects on requirements decisions. Journal of Systems and Software, 2010. **83**(12): p. 2441-2455.
[70] Gorschek, T. and A.M. Davis, Requirements engineering: In search of the dependent variables. Information and Software Technology, 2008. **50**(1–2): p. 67-75.
[71] Šmite, D., N. Moe, and R. Torkar, Pitfalls in Remote Team Coordination: Lessons Learned from a Case Study Product-Focused Software Process Improvement, A. Jedlitschka and O. Salo, Editors. 2008, Springer Berlin / Heidelberg. p. 345-359.
[72] Gutwin, C., R. Penner, and K. Schneider, Group awareness in distributed software development, in Proceedings of the 2004 ACM conference on Computer supported cooperative work. 2004, ACM: Chicago, Illinois, USA. p. 72-81.
[73] Berenbach, B., Impact of organizational structure on distributed requirements engineering processes: lessons learned, in Proceedings of the 2006 international workshop on Global software development for the practitioner. 2006, ACM: Shanghai, China. p. 15-19.
[74] Sakthivel, S., Managing risk in offshore systems development. Commun. ACM, 2007. **50**(4): p. 69-75.
[75] John, M., et al., Human and social factors of software engineering: workshop summary. SIGSOFT Softw. Eng. Notes, 2005. **30**(4): p. 1-6.
[76] Hossain, E., et al. Risk Identification and Mitigation Processes for Using Scrum in Global Software Development: A Conceptual Framework. in Software Engineering Conference, 2009. APSEC '09. Asia-Pacific. 2009.
[77] Šmite, D., et al., Empirical evidence in global software engineering: a systematic review. Empirical Software Engineering, 2010. **15**(1): p. 91-118.
[78] Hossain, E., M.A. Babar, and J. Verner, How Can Agile Practices Minimize Global Software Development Co-ordination Risks? Software Process Improvement, R.V. O'Connor, et al., Editors. 2009, Springer Berlin Heidelberg. p. 81-92.
[79] Siakas, K.V. and B. Balstrup, Software outsourcing quality achieved by global virtual collaboration. Software Process: Improvement and Practice, 2006. **11**(3): p. 319-328.
[80] Richardson, I., et al., A Process Framework for Global Software Engineering Teams. Information and Software Technology, 2012. **54**(11): p. 1175-1191.
[81] Ramachandran, S., S. Dodda, and L. Santapoor, Overcoming Social Issues in Requirements Engineering. CCSIT, 2011. **3**(133): p. pp. 310–324.
[82] Mäkiö, J., S. Betz, and A. Oberweis, OUTSHORE Maturity Model: Assistance for Software Offshore Outsourcing Decisions Collaborative Software Engineering, I. Mistrík, et al., Editors. 2010, Springer Berlin Heidelberg. p. 329-341.
[83] Narayanaswamy, R. and R.M. Henry, Effects of culture on control mechanisms in offshore outsourced IT projects, in Proceedings of the 2005 ACM SIGMIS CPR conference on Computer personnel research. 2005, ACM: Atlanta, Georgia, USA. p. 139-145.
[84] Lin, L., S.J. Prowell, and J.H. Poore, The impact of requirements changes on specifications and state machines. Software: Practice and Experience, 2009. **39**(6): p. 573-610.
[85] Chang, K.T. and K. Ehrlich, Out of sight but not out of mind?: Informal networks, communication and media use in global software teams, in Proceedings of the 2007 conference of the center for advanced studies on Collaborative research. 2007, IBM Corp.: Richmond Hill, Ontario, Canada. p. 86-97.
[86] Kwan, I. and D. Damian, Extending socio-technical congruence with awareness relationships, in Proceedings of the 4th international workshop on Social software engineering. 2011, ACM: Szeged, Hungary. p. 23-30.
[87] Islam, S., et al. Offshore-outsourced software development risk management model. in Computers and Information Technology, 2009. ICCIT '09. 12th International Conference on. 2009.
[88] Betz, S., S. Hickl, and A. Oberweis. Risk Management in Global Software Development Process Planning. in Software Engineering and Advanced Applications (SEAA), 2011 37th EUROMICRO Conference on. 2011.
[89] Anantatmula, V. and M. Thomas, Managing global projects: A structured approach for better performance. Project Management Journal, 2010. **41**(2): p. 60-72.
[90] Macke, D. and T. Galinac, Optimized Software Process for Fault Handling in Global Software Development Making Globally Distributed Software Development a Success Story, Q. Wang, D. Pfahl, and D. Raffo, Editors. 2008, Springer Berlin / Heidelberg. p. 395-406.

[91] Lin, T., Investigating Culture in IT Offshoring: A Literature Review Nordic Contributions in IS Research, H. Salmela and A. Sell, Editors. 2011, Springer Berlin Heidelberg. p. 124-137.
[92] Kauppinen, M., et al., Implementing requirements engineering processes throughout organizations: success factors and challenges. Information and Software Technology, 2004. **46**(14): p. 937-953.
[93] Karlsson, L., et al., Requirements engineering challenges in market-driven software development – An interview study with practitioners. Information and Software Technology, 2007. **49**(6): p. 588-604.
[94] Karandikar, H. and S. Nidamarthi, A model for managing the transition to a global engineering network spanning industrialized and emerging economies. Journal of Manufacturing Technology Management, 2006. **Vol. 17**(Iss:8): p. pp. 1042 - 1057.
[95] Gandhi, S.J., A. Gorod, and B. Sauser, Prioritization of outsourcing risks from a systemic perspective. Strategic Outsourcing: An International Journal, 2012. **Vol. 5**(Iss: 1): p. pp. 39 - 71.
[96] Islam, S. and S.H. Houmb. Integrating risk management activities into requirements engineering. in Research Challenges in Information Science (RCIS), 2010 Fourth International Conference on. 2010.
[97] Mistrík, I., et al., Collaborative Software Engineering: Challenges and Prospects Collaborative Software Engineering, I. Mistrík, et al., Editors. 2010, Springer Berlin Heidelberg. p. 389-403.
[98] Asnar, Y., et al. Computer Aided Threat Identification. in Commerce and Enterprise Computing (CEC), 2011 IEEE 13th Conference on. 2011.
[99] Sadraei, E., et al., A field study of the requirements engineering practice in Australian software industry. Requirements Engineering, 2007. **12**(3): p. 145-162.
[100] Lee, W.-T., et al., Change impact analysis with a goal-driven traceability-
based approach. International Journal of Intelligent Systems, 2010. **25**(8):
p. 878-908.
[101] Crnkovic, I., Component-based software engineering — new challenges in
software development. Software Focus, 2001. **2**(4): p. 127-133.
[102] Lin, L., S.J. Prowell, and J.H. Poore, The impact of requirements changes on specifications and state machines. Software: Practice and Experience, 2009. **39**(6): p. 573-610.
[103] Rodriguez, F., et al., Evaluating Collaboration Platforms forOffshore Software Development Scenarios Software Engineering Approaches for Offshore and Outsourced Development, B. Meyer and M. Joseph, Editors. 2007, Springer Berlin / Heidelberg. p. 96-108.
[104] Lawrence, B., K. Wiegers, and C. Ebert, The top risk of requirements engineering. Software, IEEE, 2001. **18**(6): p. 62-63.
[105] Piri, A., T. Niinimäki, and C. Lassenius, Fear and distrust in global software engineering projects. Journal of Software: Evolution and Process, 2012. **24**(2): p. 185-205.
[106] Krieg, R., Impact of structured product definition on market success. International Journal of Quality & Reliability Management, 2004. **Vol. 21** (Iss: 9): p. pp.991 - 1002.
[107] Pitts, M.G. and G.J. Browne, Improving requirements elicitation: an empirical investigation of procedural prompts. Information Systems Journal, 2007. **17**(1): p. 89-110.
[108] Breaux, T.D., A.I. Antón, and E.H. Spafford, A distributed requirements management framework for legal compliance and accountability. Computers & Security, 2009. **28**(1–2): p. 8-17.
[109] Lin, L., S.J. Prowell, and J.H. Poore, The impact of requirements changes on specifications and state machines. Software: Practice and Experience, 2009. **39**(6): p. 573-610.
[110] Boehm, B. and R. Valerdi, Impact of software resource estimation research on practice: a preliminary report on achievements, synergies,and challenges, in Proceedings of the 33rd International Conference on Software Engineering. 2011, ACM: Waikiki, Honolulu, HI, USA. p.1057-1065.
[111] Ernst, N.A., J. Mylopoulos, and Y. Wang, Requirements Evolution and What (Research) to Do about It Design Requirements Engineering: A Ten-Year Perspective, K. Lyytinen, et al., Editors. 2009, Springer Berlin Heidelberg. p. 186-214.
[112] Kontio, J., et al. Managing commitments and risks: challenges in distributed agile development. in Software Engineering, 2004. ICSE 2004.Proceedings. 26th International Conference on. 2004.
[113] Houston, D.X., G.T. Mackulak, and J.S. Collofello, Stochastic simulation of risk factor potential effects for software development risk management. Journal of Systems and Software, 2001. **59**(3): p. 247-257.
[114] Ocker, R.J. and J. Fjermestad, Communication differences in virtual design teams: findings from a multi-method analysis of high and low performing experimental teams. SIGMIS Database, 2008. **39**(1): p. 51-67.
[115] Hall, R.J. LSS: A Tool for Large Scale Scenarios. in Automated Software Engineering, 2006. ASE '06. 21st IEEE/ACM International Conference on. 2006.
[116] Liu, X., M. Azmoodeh, and N. Georgalas, Specification of Non-functional Requirements for Contract Specification in the NGOSS Framework for Quality Management and Product Evaluation, in Proceedings of the 5th International Workshop on Software Quality. 2007, IEEE Computer Society. p. 7.
[117] Staats, B.R., Unpacking Team Familiarity: The Effects of Geographic Location and Hierarchical Role. Production and Operations Management, 2012. **21**(3): p. 619-635.
[118] Newell, S., G. David, and D. Chand, An analysis of trust among globally distributed work teams in an organizational setting. Knowledge and Process Management, 2007. **14**(3): p. 158-168.
[119] Muethel, M., F. Siebdrat, and M. Hoegl, When do we really need interpersonal trust in globally dispersed new product development teams? R&D Management, 2012. **42**(1).
[120] MacGregor, E., Y. Hsieh, and P. Kruchten, Cultural patterns in software process mishaps: incidents in global projects. SIGSOFT Softw. Eng. Notes, 2005. **30**(4): p. 1-5.
[121] Li, Y., et al., The role of team problem solving competency in information system development projects. International Journal of Project Management, 2011. **29**(7): p. 911-922.
[122] Casey, V., Imparting the importance of culture to global software development. ACM Inroads, 2011. **1**(3): p. 51-57.
[123] Baiqiang, X. and Z. Deming. Engineering safety information in software intensive systems. in Reliability, Maintainability and Safety (ICRMS),2011 9th International Conference on. 2011.
[124] Ebert, C., Requirements Engineering, in Global Software and IT. 2011, John Wiley & Sons, Inc. p. 37-44.
[125] Ferreira, S., et al., Reducing the risk of requirements volatility: findings from an empirical survey. Journal of Software Maintenance and Evolution: Research and Practice, 2011. **23**(5): p. 375-393.

[126] Yang, H., et al., Analysing anaphoric ambiguity in natural language requirements. Requirements Engineering, 2011. **16**(3): p. 163-189.
[127] Ralyte, J., et al. A framework for supporting management in distributed information systems development. in Research Challenges in Information Science, 2008. RCIS 2008. Second International Conference on. 2008.
[128] da Silva, F.Q.B., et al., An evidence-based model of distributed software development project management: results from a systematic mapping study. Journal of Software Maintenance and Evolution: Research and Practice, 2011: p. n/a-n/a.

Classification	Situational Factors	Risks	References
	Interaction skills	Lack of privacy, Lack of efficiency, Lack of knowledge sharing, Ineffective information sharing, Lack of efficient coordination and collaboration, Emergence of group issues, Lack of quality individual decision making, Negative relationship building, Loss of cohesion, Lack of requirement existence and stability, Lack of inadequate user development interaction, Lack of synchronization and act, Inaccurate task allocation, Misunderstanding of remote participants action or spoiling relationships, Chaotic and uneven knowledge transfer	[26] [27] [28] [29] [30]
	Interaction styles	Lack of privacy, Lack of efficiency, Lack of knowledge sharing, Ineffective information sharing, Lack of efficient coordination and collaboration, Emergence of group issues, Lack of quality individual decision making, Inefficient articulation work, Loss of development speed, Requirements misunderstanding, Outcome failure, Negative relationship building, Loss of cohesion, Lack of requirement existence and stability, Lack of inadequate user development interaction, Inaccurate task allocation	[26] [27] [31] [28] [29] [30] [32] [33] [36] [37] [38] [39] [40]
	Interaction medium or technology	Loss of development speed, Lack of trust, Lack of shared team identity, Lack of awareness of members activity, Lack of team members effort coordination, Lack of effective leadership, Lack of effective knowledge sharing, Lack of determination of appropriate task technique,	41] [42] [43] [44] [45] [46] [27] [28] [29] [30] [38] [37]

Classification	Situational Factors	Risks	References
Communication and Distance		Outcome failure, Delay, Negative relationship building, Loss of cohesion, Lack of requirement existence and stability, Lack of inadequate user development interaction, Lack of synchronization and act, Inaccurate task allocation, Chaotic and uneven knowledge transfer, Lack of uniform software development environment, Lack of work awareness	
	Interaction tools	Lack of individual and shared knowledge understanding, Chaotic and uneven knowledge transfer, Lack of uniform software development environment, Lack of work awareness, Lacking proximity	[37] [29] [47] [32]
	Interaction infrastructure	Effort overhead, High defect frequency, Lack of client involvement, Hidden cost, quality work, product, Lacking proximity	[48] [49] [43] [50] [35] [37]
	Distance	Inefficient communication, inadequate communication, Inappropriate knowledge management, Cultural diversity, Time zone difference, Lack of adequate requirement capturing, High rate of integration errors, High rate of organizational differences, Lack of trust, Lack of efficient collaboration process, Lack of effective outcome, Delay, Lack of efficient coordination and communication control, Conflicts over priorities.	[51][52] [53] [54] [55] [29] [30] [56] [36] [57]
Differences with respect to Culture, Background, Language, Organization and Time	Language	Requirement misunderstanding, Lack of quality outcome, Ineffective management practices, Scope creep, Lack of timely project completion, Increase rate of coordination problems, Lack of workflow communication, varying methodologies, Lack of effective communication, Decrease in team productivity, Lack of requirement comprehensiveness, Incorrect reporting from remote team, Verbal contact avoidance, Inaccurate task allocation, Lack of common goals and client involvement	[52] [46] [58] [59] [60] [53] [27] [61] [62] [38] [63] [50] [45]
	Cultural background	Requirement misunderstanding, Lack of quality outcome, Varying meaning for a situation, Lack of effective communication, Decrease in team productivity, Barriers to work ethics, Incorrect reporting from remote team, Lack of quality decisions.	[52] [64] [65] [66] [67]
	Work Environment	Increase in requirement evolution rate, Lack of quality outcome, Lack of work accuracy, Lack of improvisation skills, Lack of information and artifact sharing, Lack of quality requirement document confusion of remote participant actions,	[34] [58] [67] [68] [27] [69] [70] [71]
	View point	Increase in requirement evolution rate	[34]
	Time zone	Lack of coordination, Ineffective management practices, Scope creep, Lack of timely project completion, Lack of workflow communication, varying methodologies, Lack of quality requirement document, Lack of efficient requirement reviews and effective communication, Decrease in team productivity,	[72] [52] [58] [73] [74] [75] [59] [53] [76] [54] [55] [61] [56] [33] [40] [77] [78]
	Social background	Lack of quality outcome, Varying meaning for a situation, Lack of effective communication and social interaction, Project mismanagement	[67] [58][76][54][55][79]
	National culture	Lack of quality outcome, Challenging cooperation, Barriers to work ethics, Misunderstanding of remote participant actions, Lack of social interaction, Incorrect reporting from remote team, Lack of interests, Lack of quality decisions	[58][67] [27] [55] [80] [81] [82]
	Organizational culture	Lack of quality outcome, Challenging cooperation, Ineffective management practices, Scope creep, Lack of timely project completion, Increase rate of coordination problems, Lack of workflow communication, varying methodologies, Lack of efficient requirement reviews, Lack of effective communication, Decrease in team productivity, Inefficient collaboration and communication process, Lack of work accuracy, Lack of improvisation skills, Misunderstanding of remote participant actions, Personal loss, Lack of social interaction, Lack of uniform software development environment, Complex problem escalation	[52] [83] [73] [59] [60] [76] [27] [55] [30] [41] [84] [80] [33] [36] [34] [81]
	Social climate	Decrease in team productivity, Inefficient collaboration and communication process, Lack of social interaction, Project mismanagement, Lack of quality decisions	[85] [76] [79] [82]
	Political difference	Lack of social interaction, Project mismanagement	[79] [34]
	Time shifts	Inaccurate task allocation	[38]
	Inter group culture	Challenging cooperation, Scope creep, Lack of timely project completion, Increase rate of coordination problems, Lack of workflow communication, Lack of effective communication, Barriers to work ethics, Lack of improvisation skills, Lack of social interaction, Lack of information and artifact sharing, Lack of interests, Lack of quality requirement document, Inaccurate task allocation	[67] [83] [27] [34] [33]
	Organization Structure and boundaries	Lack of effective coordination and collaboration, lack of quality outcome, inadequate client involvement, inefficient process for task completion, lack of team motivation, lack of quality decisions, partners week relationships, lack of trust, lack of common goals, weak contractual relations.	[86] [52] [87] [58] [88] [89] [50] [91] [45]
	Organization person retention strategy	Loss of key personnel.	[92]
	Organization policies and strategies	Inadequate client involvement, inefficient requirement engineering process, collecting data without improving requirement engineering process.	[92] [93]
	Team Awareness	Non-effective coordination, lack of efficient information seeking, infrequent communication, lack of exchanging information, lack of maintained awareness, lack of trust, lack of knowledge sharing, delay, un-aware of remote team member skills, unaware of changing requirements, unaware of job responsibilities of remote team members, high defects frequency, lack of control.	[27] [86] [72] [85] [90]
	Data repository	Data loss, lack of requirement specification quality	[94] [95]
	Domain knowledge	Lack of quality outcome, erroneous requirements, budget and schedule overrun, productivity	[73] [87][59]

Classification	Situational Factors	Risks	References
Knowledge management and awareness	level	downfall,	
	Internal knowledge level	Lack of frequent communication, lack of competence	[54]
	Knowledge management techniques and procedures	Lack of frequent communication, low technical efficiency, lack of competence, lack of quality management, lack of project quality, lack of exchanging information, lack of maintained awareness, lack of trust, lack of knowledge sharing, delay, team ineffectiveness, lack of quality documentation, lack of awareness about development project.	[96][60][88][27][33][97]
	Business knowledge	Delay in problem domain clarification, extra cost for rework	[61]
	Access management	Unintended data editing, un/intentional disclosure for personal gain.	[98]
	Tacit knowledge level	Lack of knowledge sharing, lack of synchronization and act	[29]
	RE practice knowledge	Lack of accurate requirements fault modeling.	[99]
	Configuration management	Effort overhead, work unawareness	[32][48]
	Knowledge management awareness	Lack of quality decisions, lack of control, lack of maintained awareness, , unaware of job responsibilities of remote team members, high defects frequency	[27]
	Requirement management	Lack of requirement stability, intrinsic schedule flaws, high frequency of system failure, overlooking crucial requirements, not understanding the needs behind the requirements, overlooking non functional requirements, not inspecting requirements, reducing the solution domain by representing in design form, insufficient change management, lack of careful requirement handling, project completion failure.	[28][100][101][102][103][104]
Management	Coordination skills	Lack of effective collaboration, unfamiliarity with technology, increase cost, loss of development speed, inaccurate task allocation, unawareness about development project, unsuccessful collaboration, lack of trust, lack of personal contact, lack of team involvement.	[52] [27][38] [77] [97] [45][50]
	Coordination technique	Lack of effective collaboration, unfamiliarity with technology, increase cost, loss of development speed, requirements misunderstanding, lack of coordination, misalignment of tools with expectations, unrealistic estimation, lack of effective traceability, delay, lack of work awareness, collaboration, lack of trust, lack of personal contact, lack of team involvement, hidden cost	[27] [30] [52] [45] [50]
	Competence	Lack of skilled analyst role, lack of efficient team performance, lack of organizational performance, lack of contingency, delay, lack of work awareness, lack of understanding the project scope, lack of quality requirement artifacts, scope creep, lack of quality decisions,	[27] [32]
	Decisions	Misalignment of tools with expectations, unrealistic estimation	[27]
	Supervisor sub-ordinate relationship	Lack of proactive transparency	[105]
	Management strategy	Increase in requirement uncertainty, lack of understanding of project scope, loss of key personals	[92]
	Cooperation approach	Lack of understanding activities accelerating the knowledge sharing, lack of trust, troublesome disagreements, less social capital.	[45][50]
Tools, Technologies, Techniques and Standards	Technique selection	Inaccurate estimation, poor quality outcome, lack of quality requirement engineering activities, inadequate customer representation, requirement misunderstanding, , inadequate requirements, lack of propagation of relevant changes to artifacts, scope creep, ineffective communication, lack of shared understanding,	[106] [107] [108] [93] [40]
	Technical expertise	Poor quality outcome, Inaccurate estimation, inadequate customer representation, requirement misunderstanding, , inadequate requirements, lack of propagation of relevant changes to artifacts, scope creep, ineffective communication, lack of shared understanding	[109][110] [108] [48] [31]
	Standards	Low technical efficiency, high frequency of conflicting requirements, lack of standards, lack of shared understanding, lack of effective coordination, lack of group awareness.	[78] [60]
	Tools	Lack f requirement engineering quality outcome, lack of integrated tools, lack of access to requirements history, lack of reporting about fulfillment of preconditions, lack of allowing to requirements documents elaboration, lack of requirements remote negotiation and discussion facilitation, lack of uniform software development environment, lack of trust, lack of requirement management, lack of effective communication, requirement misinterpretations	[78] [27] [30] [93] [80] [111] [40]
	Technology	Lack of team work performance, software project failure, lack of uniform software development environment, lack of early architecture quality, greater frequency of requirements uncertainty, lack of information and artifact sharing, lack of decision making quality.	[112][30][69][84] [113] [33] [91]
Stakeholders	Competence and experience with in application	Requirements conflicts, requirements misunderstanding, lack of quality requirement representation, wrong expectations, undetected activity errors, lack of team performance, lack of efficient collaboration, lack of quality outcome, mistrust, lack of quality management, wrong team setup and adjustment, inefficient requirement engineering process, inaccurate task allocation, effort overhead, lack of quality decisions.	[114] [67] [115] [113]
	Motivation level	Lack of team performance, activity errors, wrong team set up and adjustment.	[114] [82]
	Familiarity with each other	Inadequate communication, increase in staff problems, lack of team effectiveness, mistrust, lack of work team cohesion, lack of effective collaboration.	[116] [85] [117]
	Preferences related to project	Inefficient collaboration, poor quality outcome.	[67] [109]
	Knowledge	Lack of team effectiveness, inefficient collaboration, ineffective communication, lack of	[67] [27]

Classification	Situational Factors	Risks	References
	Exchange ability	requirement rationale understanding, inefficient requirement engineering process.	
	Background	Lack of shared understanding.	[29]
	Personnel trust	Lack of shared understanding, lack of efficient information sharing, lack of efficient collaboration	[67] [41]
	Leadership skills	Inefficient requirement engineering process, lack of efficient information flow tailoring, person becomes bottleneck.	[92]
	Personnel/group relationship	Lack of effective collaboration, lack of quality outcome, on stake security, mistrust, weak team cohesion.	[67] [112] [105] [118][119]
	Decision capability	Lack of accurate estimation, lack of quality outcome.	[92]
	Utility values	Lack of project efficiency.	[67]
	Knowledge level	Lack of team effectiveness, inefficient collaboration, ineffective communication, lack of requirement rationale understanding, inefficient requirement engineering process.	[67] [96] [27]
	International work experience	Lack of efficient collaboration.	[67]
	Relation to the project	Inefficient collaboration, lack of quality outcome, delay.	[67]
	Power/position in organization	Lack of satisfied requirements, inefficient communication, and high rate of requirements conflicts.	[120]
	Client involvement	Lack of client implication, conflicting approaches to requirement engineering process, requirements variability, higher issues with user abilities and concurrences.	[27][76] [121]
	Carriers prospects	Loss of competency.	[67]
	Client commitment	Client misalignment with project goal, requirement variability.	[27]
	Client availability	Lack of participation from clients, higher issues with user abilities.	[27]
	Background	Lack of efficient requirement elicitation and negotiation	[65]
	Partner interpretation	Lack of effect group problem solving ability, measurement scale misconception, inappropriate requirement validation, lack of quality outcome, delay.	[92]
	Team management capabilities	High rate of workforce turnover, lack of remote staff information.	[80]
	Common work experience	Inaccurate task allocation, mistrust, delay, effort overhead, inefficient collaboration and communication.	[67]
Project and Process	Project phase distribution	Problematic overall joint management, problematic responsibility share, extra management needed at each location.	[61]
	Requirements engineering process	Lack of quality outcome, undetected errors, erroneous requirements, budget and schedule overrun, poor communication, not inspecting requirements, attempting to perfect requirements before construction, ignoring non functional requirements	[122] [104] [59] [28] [70]
	Economic process	Unacceptable results.	[123]
	Process maturity	Productivity downfall, inefficient communication, lack of quality requirements documentation, effort overhead,	[42] [82]
	Collaboration process	Lack of communication, collaboration and control.	[52] [90]
	Management process	Lack of quality decision making.	[91]
Requirements	Requirement specification document format	Overlooking crucial requirements, not understanding the needs behind the requirements, ignoring non functional requirements, reducing solution domain by representing it in design form, insufficient change management, lack of synchronization, inconsistent specification, erroneous requirement, flaws in design requirements.	[96][87] [28][124][125][128] [126]
	RE effort	Scope creep, lack of attention to importance concerns, unclear requirements.	[31]
	Requirement representation style	Choosing wrong solution for requirement implementation, reducing the solution domain.	[127][28]
	Requirement interpretation	Effort overhead, inadequate solution, inaccurate requirement capturing and understanding, lack of quality outcome, delay.	[30][126][35] [35]

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 79-94
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Review of Support to Situational Requirement Engineering from Standards and Models

Huma Hayat Khan, Mohd. Naz'ri bin Mahrin, Suriayati bt Chuprat
Advanced Informatics School,
Universiti Tecknologi Malaysia,
hkhuma2@live.utm.my, mdnazrim@utm.my, suria@ic.utm.my

ABSTRACT

Requirement engineering (RE) process clear description can be an important factor for guiding the team members involved in the RE process, which may help organizations not exceeding the estimated schedule and budget for the software project. There can be many reasons of not having efficient RE process such as changing situations among the organizations involved in the RE process. It is certainly one of the various other reasons of having non-efficient RE process. Due to these changing situations the RE process customization accordingly should be performed. As part of our current research project, where we are investigating situational RE in Global Software Development (GSD), first we need to identify the factors which may results in changing situations. This paper we have critically reviewed the domain of "standards and models", and identified the initial list of situational factors from them. These standards and models are related to software engineering, which directly or indirectly discusses RE process. We have adopted the constant comparison and memoing techniques of the Grounded theory for identification of unique initial list of the situational factors. This initial list is a significant consideration for a comprehensive list of situational factors affecting RE process in GSD.

KEYWORDS

Software engineering, Requirement engineering (RE) process, Models, Standards, Global software development (GSD), Situational Factors.

1 INTRODUCTION

Requirements are defined as the description or the explanation about the behavior of the system; it can also be said as the system attribute, or a constraint. We can also describe requirement as what the system should perform. Requirement engineering comprises of activities; requirement discovery, analysis, specification, validation and management. This process focuses on the usage of repeatable techniques in order to come up with complete, consistent and relevant requirements.

We have come to know from the previous literature on and around requirement engineering process, that there are some factors [1] [2], which can affect the software development and more specifically they are factual for requirement engineering. Some of these factors are human related factors, technical factors, management related factors etc. It is also evident from the literature [3] that not only technical issues but besides this social, managerial, economic and organizational issues have an influence on requirement engineering process. These factors may become the source of changing situations, which need to be explored in detail.

It is to notify here that this study is one part of our investigation which deals with the identification of situational factors affecting Requirement Engineering Process in Global Software Development. This domain (standards and models of RE) is one of the ten other domains we have selected for situational factors identification. Because of the huge amount of information from each of the domain they are firstly treated individually in order to analyze them in detail. The related domains used in our research work are taken from work of Paul Clarke et al. [3], who used them for identification of situational factors that can affect software development process. Besides this, defect prediction, task sequencing and metrics are the domains which are taken from their future work. We have used these domains specific for requirement engineering phase in global software development.

In this paper, we have limited our search on standards and models which are on and around requirement engineering process for the identification of the situational factors, which may influence the requirement engineering process.

Table 1. Research domains

Domains	Reasons of selection
Standards and models on and around RE	Although models and standards on and around requirement engineering process do not directly look for identification of the situational factors that may affect the RE process, they occasionally illustrate the situational context in general.
Risk factors for RE process in GSD	This domain is concerned with identification of things results in risk that should be considered for RE process. Everything that is measured as a risk for RE is an alarm that the RE process may require to give consideration in GSD.
RE environmental factors in GSD	Research has identified environment characteristics that explains the RE settings. Such characteristics are associated with situational factors that may affect the RE process in GSD.
RE process tailoring in GSD	This domain is concerned with taking general RE approaches and tailors them to specific context. Therefore, this domain gives some explanation of the significant considerations while tailoring RE process in GSD.
RE process agility in GSD	With the introduction of agile software development, research has been performed to recognize the development process agility with the changing context. RE is the most fundamental step of software development, whose process agility is an important characteristic in GSD where rate of changing situations are at most.
RE body of knowledge	The Requirement Engineering body of knowledge (REBOK) gives a validated repository of knowledge for RE process, which ultimately provides the guidance related to situational context factors that are significant considerations for the RE process.
RE defect prediction in GSD	This domain is concerned with the defect predictions during RE process. Any thing that does not allow accurate defect prediction during RE process is a point of consideration in GSD.
RE task sequencing in GSD	This domain is concerned with aspects of sequencing of tasks during RE process. Any thing that gives rise to varying task sequences and priorities during requirement engineering process is a potential consideration in GSD.
RE metrics in GSD	The domain of RE metrics deals with the quality metrics. Research has shown number of quality metrics; therefore everything that deals with varying quality metrics is important consideration for RE process in GSD.
RE process estimation in GSD	It is concerned with estimating the accurate RE process resources in GSD. Anything that results in not accurate estimation of resources for RE process is potentially a consideration.

Standards and Models of Requirement Engineering Process describe the situational factors in general context. This is one of the reasons that this domain has been included in the list besides with various other domains like; risk factors for requirement engineering, environmental factors for requirement engineering etc as shown in *table 1* which will be discussed in our coming researches independently. These situational factors leads to varying situations among the various organization's requirement engineering team while performing requirement engineering process.

Section II of the paper describes the background of the research undertaken here, while Section III describes the research method which is adopted to identify the situational factors. Section IV describes the results of the research, whereas Section V comprises of discussion part of the research explaining the findings and future work. Section VI consists of conclusion of this research.

2 BACKGROUND

We have focused on the early development phase (Requirement Engineering) of the software life cycle due to the fact of its major effect on software functionality, productivity and development cost. Besides this poorly defines requirement engineering process by organization is also considered as another focusing reason for selection of requirement engineering phase.

We also come across from previous literature about the importance of requirement engineering process effectiveness. Boehm explained the importance of system's requirements by saying that requirement errors are hundred times more expensive to fix comparing with system's implementation errors [4]. Similarly another researcher Lutz [5] comes up with the statement that critical systems 60% errors are requirements errors. Espite by performing a survey on European companies found that 60% people considered requirement engineering problems as very significant [6]. Similarly another researcher Hall [7] identified from case studies of twelve companies that out of 268 problems identified, 128 were related to requirements problem.

So from above it is clearly shown about the importance of Requirement Engineering Process,

as there are number of significant benefits by having improved requirement engineering process and quality requirements. Besides with all this discussion it is also notified from the literature [1,2,3] that there are some factors which act as a barrier for having quality requirement engineering process like; factors related to humans, technical, management, social, organizational and economic etc. these factors vary with the organizational context, which leads to varying situations; ultimately influencing the requirement engineering process.There is a need to explore and identify these situational factors, which can help the practitioners and researchers for having an improved (quality) Requirement Engineering Process. In order to do so, we have selected the domain of standards and models on and around requirement engineering among the various other domains for this part of our research as shown in *table 1*. This domain of standards and models is included in the list with various other domains related to Requirement Engineering process as they can be a good source for situational factors as they describe them in general context. Hence domain of standards and models which are on and around the requirement engineering process are critically reviewed and investigated in this paper for situational factors identification.

3 METHODOLOGY:

For situational factors identification, we have adopted a theory called Grounded Theory.

Grounded Theory: This theory by Glaser [8] is the methodology for analysis linked with the collection of data in a systematic manner which applies certain methods for the theory generation about a substantive area. Grounded theory as argued by Knigge et al. [9] is used to build the theories on the basis of data taken from the social world in such a manner that these theories are grounded in ones daily experiences and actions. These theories also have some philosophical disagreements for its application in practice as discussed by Boychuk Duchscher et al. [10]. However as this research is related with the methods of data analysis used by the theorists therefore this dispute is not relevant concerns. As discussed by Douglas [11] that the theory emphasize on the approach that is systematic for data collection, its handling and its analysis.

This research has also to adopt a likewise systematic approach to handle data and analyze it and therefore uses three main techniques of data management and analysis from Grounded Theory that is coding, constant comparison, and memoing. Coding means to gather or take raw data and raise it to a conceptual level as described by Corbin [12]. They also argue that coding consist of interaction with data by using techniques such as inquiring for the data, doing comparisons among the data, and deriving concepts to position the data, then making those concepts in form of their properties and dimensions. Therefore, it is basically ‘‘the process of defining what the data is about’’ Bryant et al. [13], of ‘‘deriving and developing concepts from data’’ by Corbin [12], where ‘‘codes capture patterns and themes and cluster them under an evocative title’’ by Glaser [14]. For our study, the identifiers and classifications are the data codes that the data sources use to recognize the data and their related concepts. Now the researchers Corbin [12] also argue that the theory relate with ‘‘the analytic process of comparing different pieces of data for similarities and differences’’. This type of analysis as argued by Bryant et al. [13] ‘‘generates successively more abstract concepts and theories through inductive processes of comparing data with data, data with category, category with category, and category with concept’’. In this research, a variety of data codes from the diverse data sources must be compared through constant comparison for their similarities and differences until we get main factors and their categories so coding with such constant comparison becomes complex, as argued by Glaser [14] ‘‘simultaneously many categories and their properties may be emerging at different levels of conceptualization and different ways of being related by theoretical codes’’. Researcher also says that this complexity can be tracked by the memoing process. Researchers Bryant et al. [13] say that theory data coding occurs ‘‘in conjunction with analysis through a process of conceptual memoing, capturing the theorist’s ideation of the emerging theory’’. They also argue that Grounded Theory’s coding and memoing process helps to

lessen uncertainty by stop and capture process for their conceptual ideas about the codes finding.

Our study has been divided into several steps for identification of situational factors which may impact Requirement Engineering process. Figure 1 shown below describes the steps which are undertaken in order to find the situational factors from standards and models. It is a 6 steps process where;

Step1 deals with the extraction of data units. For this we have reviewed the standards and models which are directly or indirectly relating to requirement engineering process and get the data units from there which are considered to be raw data for factors identification.

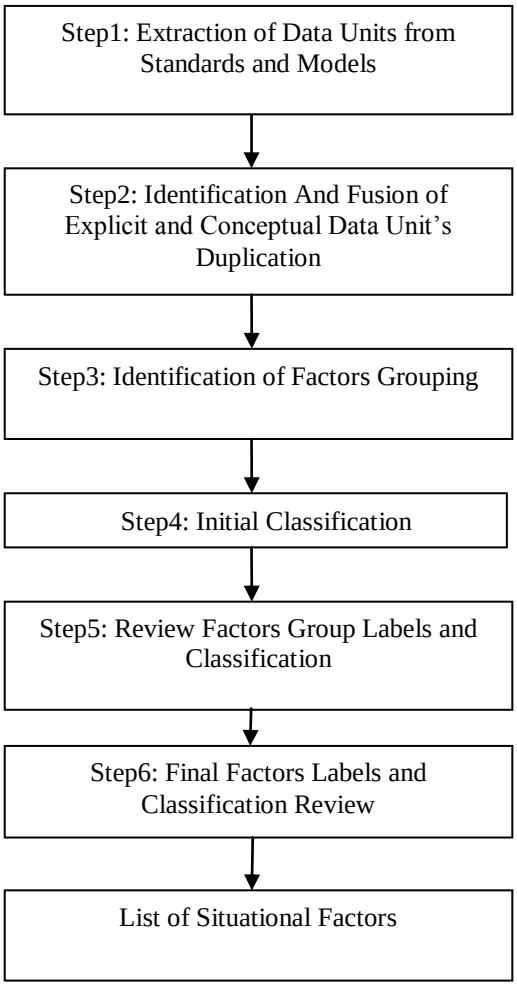


Figure 1. Situational Factors Identification Steps

Step2 search for those data units which are having clear duplications present. This step consists of scanning the data to find the duplication instances. As there is direct mapping among the individual data units that are joined or combined in this comparison, so memoing is not required as it shows the thoughts that influenced the joining or combination of data units. Where while combining the data units, careful steps are taken to maintain the source information. So, if two different data units are having same textual depiction or explanation and meaning, they are combined into a single unit – but it is also to be making sure that both sources of data units are apparent and clear in the combined data unit testimony in the main table. Now following the identification and removal of clear duplication among the data units, conceptual duplications in data units are identified. We combine two data units which are not having same textual tags but are having same meaning. In order to do this, memoing is also done which records the thought process behind this. Besides this, as previously done with the removal of clear duplication stage, combined data units sources links are maintained.

Step3 consists of factors grouping, where every group is given a label. Here each of the data units is combined into factor groups where every group is given a label. These grouping are on basis of factors relatedness to each other and the labels given to them are according to the specific area they are concern with. Memos are used to record the justification for grouping and labeling.

Step4 consists of actions to come up with initial classification for data units. This time also, memos draft the motivation or justification that has initiated or created the initial classification. Under this initial classification, factors groups are present with their appropriate labeling. Similar to the previous steps, memos are used here to record the justification for classification and labeling.

Step5 comprises of constant comparison for evaluating the precision of the group labels and classifications. The group labels are renamed when suitable data units are moved to alternating factors or to completely novel factors – as believed suitable in regard of the rising factors and classification. Similarly some of the classifications are renamed or combined or decomposed as considered appropriate. For this step similar to the previous one, memos are used to draft the thought process with that of the historical trace of the actions. This assists in envisioning the factors and classification sources, and allows an assessment of the impact of each data sources on the main list as it come forward.

Step6 consist of a review for a list of final factors and labels for classification in order to make sure that every individual item is correctly named. Once the initial list of situational factors at its final form is created, then it is being reviewed by experts to have Situational Factors evaluation and modification base upon their comments. The experts are selected from academia on basis of their experience in Requirement Engineering field. More specifically two experts are selected having more than ten years of experience in field of requirement engineering. The tasks provided to them is to evaluate the list for its comprehensiveness, as well as to review if the factors are grouped under right classification with that of any recommended modification both at factors and classification level.

4 RESULTS

This section consists of results explanation against each of the steps mentioned in Section III.

4.1. Extraction of Data Units from Standards and Models

In this step, 189 data units are identified from 16 standards and models which are on and around requirement engineering process and are explained below. Every individual data units correspond to a factor that can affect the requirement engineering process. In addition to this, if we find any classification from the data source (standards and models) then they are also extracted, in order to have raw listing containing not only factors but also the classification from the data source. Nevertheless, this list of raw data is then inspected for data duplication. As a result core concepts are developed with the support of the constant comparison and memoing techniques. Situational factors are identified from the standards and models on and around requirement engineering process, which are explained below.

ISO 9001: This International Standard ISO 9001 [15] is by Technical Committee ISO/TC 176, Subcommittee SC 2 and Quality systems. This mentioned standard is the third edition of ISO 9001 which not only cancels but also replaces the 2nd edition (ISO 9001:1994) with that of ISO 9002:1994 and ISO 9003:1994. ISO 9001 identify requirements for systems of quality management for an organization who a) wants to exhibit its capability to consistently give product that meets up customer requirements b) plans to improve customer satisfaction via valuable system application, together with processes for continuous system improvement and the guarantee of compliance to customer requirements. ISO 9001 requirements are generic and are planned to be appropriate to all organizations, in spite of organization type, organization size and organization product provided for enhancing the customer satisfaction by meeting customer requirements.

Some of the situational factors we have identified from *ISO 9001* are as follows. The standard focuses on some aspects which actually vary among the developing teams like: activity priorities are different in development teams when they work together as they identify the linked activities and manage them in different ways. So in order to have an effective organizational functionality these linked activities should properly be prioritized. Similarly this standard also talks about the nature of the organization and says that it varies as different organization has different nature, (mature and immature) so this varying nature of organization impacts on meeting customer and regulatory requirements. The process to perform any activity also varies with organizational context. Besides this the standard also talks about the varying organizational size, activity type, process complexity, process interaction and personal competence. It says that the documentation is influenced by these factors. We also come across some more factors which influence the system effectiveness like; quality policy, quality objective, availability of resources, management review intervals. Besides these factors, resource management (human resource, infrastructure, work environment, customer related processes) also has an influence on system effectiveness. All these factors vary with among multiple organizations leads to changing situations.

ISO 15504: This is an International standard [16] for assessing the software processes and is developed in corresponding with other standards of software engineering. The purpose of this standard is to have continuous process improvement with that of the capability determination. The scope of the standard is both on comprehensive and modular scale. Processes comprise acquisition, supply, development,

operation, maintenance and support on comprehensive level. On modular level it covers the selection of processes to be assessed and this assess activity is done on capability scale. The standard is the result of the SPiCE-Project (Software Process Improvement and Capability Determination- project). This SPiCE (Software Process Improvement and Capability Determination) is an assessment model which is two dimensional; process dimension and capability dimension. There are various versions of this standard like ISO/IEC 15504-2 situates the least requirements for having assessment that guarantee ratings steadiness and repeatability. Similarly ISO/IEC 15504-3 gives direction for inferring the requirements to perform assessment. ISO/IEC 15504-4 recognizes assessment of process as a doing that can be carried out either as initiative for process improvement or as an approach for capability determination. The reason of process improvement is to constantly progress effectiveness and efficiency of the organizations. The point or reason of process capability identification is to recognize the selected processes strengths and weaknesses. ISO/IEC 15504-5 consists of a Process Assessment Model which is based on the Process Reference Model which is explained in ISO/IEC 12207. The standard focuses on the fact that the organization context in which the process is implemented, competency of the assessor and models for assessment of the process vary. This varying situation influences the process quality.

ISO/IEC 12207:2008: ISO/IEC 12207:2008 standard [17] creates a general framework for software life cycle processes, besides with that of well-defined terminology which can be used by the software industry. It relate to the attainment of systems or software products and their services whether carried out with in or external to an organization. Considering these aspects the system definition is desired to present the software product or services context. It is to notify that the revision join together ISO/IEC 12207:1995 with its two modification and was synchronized with the corresponding revision of system life cycle processes (ISO/IEC 15288:2002) to align equivalent processes of organization and projects. It is also to consider that this standard may be used independently or together with ISO/IEC 15288, and provides a process related reference model that supports capability assessment of process in unity with process assessment from ISO/IEC 15504-2. So we come across this learning that this standard came into being for the following reasons; to include and rationalize revisions, to present a common terminology among the amendment of ISO/IEC 15288 and ISO/IEC 12207, to offer common names and structure of process where applicable between the amendment of the ISO/IEC 15288 and ISO/IEC 12207:2008, to facilitate the users to progress towards fully synchronized standards and to make available a steady standard, while capitalize on backward compatibility.

This standard clearly describes the fact related to project conduction that it is always in context of the organization. Besides this it also focuses on the software project outcomes and says that they are influenced by the business process of the organizations. Similarly the models to develop software are also influenced by the varying projects activity scope, magnitude, complexity, changing needs and opportunities. It is also to be notified that the life cycle stages, policies, process, procedure to support projects needs vary with organizational context. Operational environment and operational conditions also vary with organizational context which make difficult for the team to model the stakeholders requirement as it vary with the organizational operational environment and conditions. Standard describes the fact that the requirement definition process is influenced by human capabilities and skills. Besides this requirement definition process is also influenced by the workplace, environment, facilities and equipment provided. These factors vary with the organizational context and have an influence on requirement definition process.

IEEE Std 830-1998: This standard [18] relates with that of a good software requirements specification (SRS). This suggested practice is intended for software requirements specification to be developed and used to help in the selection of software products for in-house and commercial usage. This practice illustrates approaches for the software requirement specification. Practice is based on a model whose result is an unambiguous and complete software requirement specification document. SRS can be created directly from this practice or indirectly by using a model for a more precise standard. It is also to be notified that this practice does not discover any specific method or tool for preparing a System or Software

Requirement Specification. The factors identified from the standard are related to; requirement specification language, requirement representation tools, process approaches and says that these vary among the organizations which influence the system or software requirement specification.

IEEE Std 1233, 1998: This standard [19] provides guidance for software requirements development, System Requirements Specification (SyRS). It includes requirement identification, requirement organization, requirement presentation, and requirement modification. It is also to be notified that it also addresses the conditions for including operational concepts, constraints related to design, and requirements configuration into the specification. Besides this it also covers the essential distinctiveness and qualities of individual or set of requirements. The factors identified from this standard is on system requirement specification which influence it like; operational and logistical terms, organization political environment, organization market, organization standard, technical policies and organizational culture. Besides this system requirement specification is influenced by customer community and environment. The standard also describes the fact that the techniques and approaches to identify the requirements also varies on basis of organizational context which influence the requirements gathering or definition process.

IEEE Std 1362™-1998: This standard [20] deals with the concept of operations (ConOps) document. It is a user-oriented document that explains proposed system characteristics from the users' perspective. The concept of operation (ConOps) document is used to communicate the system characteristics to the stakeholders and other organizational elements like staffing, training etc. It is also notified that the standard is used to describe not only the user organization(s) but also the user mission(s) besides with that of the objectives of an organization from an integrated systems perspective. It acts as a guide for all software-intensive systems. Irrespective of this; concept in this guide can also be applied for hardware-only systems. The standard is applicable to all software products irrespective of its size, scope, complexity, or criticality. Similar to other standards, we have identified some of the situational factors related to; operational policies and constraints, operational environment (hardware, software, equipment, personnel), organizational structure, responsibilities, skill level, work activities, model of interaction with system and says that these lead to changing situations.

ISO/IEC 15288: It is an International Standard [21] ascertains a process framework for explaining systems life cycle. It explains the processes and related terminology for the life cycle from conception till retirement phase. Besides this; ISO/IEC 15288 also supports the process definition, process control, process assessment, and process improvement. This standard is created for following reasons: to come up with a common terminology between the amendment of the ISO/IEC 15288 and ISO/IEC 12207, to provide common names and structure of process between the amendment of the ISO/IEC 15288 and ISO/IEC 12207, to facilitate the users to progress towards fully synchronized standards and to make available a steady standard, while capitalize on backward compatibility. This standard also highlights the hardware, software elements as the factors vary with organizational context. Besides this the standard also focuses on the human factors, technology factors and processes for the development of product and services as the factors which vary with the organizational context.

ISO/IEC 24748: It is a standard [22] which directs for the application base on ISO/IEC 12207:2008. It deals with the concepts of system, project, life cycle, process, organizations concepts, mainly via reference to ISO/IEC 24748-1 and ISO/IEC 12207:2008. This standard also provides guidance for applying ISO/IEC 12207:2008 in organizations and on projects from the perspective of application, strategy and planning. The standard has three editions and this is the most recent one which is deliberately associated with both ISO/IEC 24748-1 and ISO/IEC 24748-2 (which is a guide to the application of ISO/IEC 15288) in terms of terminology, structure and content. It is also to be notified that ISO/IEC 24748 gives amalgamated and consolidated for system's or software's life cycle management. Its goal is to ensure stability in system or life cycle's concepts, models, stages, processes, application and adaptation.

We have identified some of the situational factors related to; Tools, technologies, process, facilities, equipments and workforce, as standard says that they are required by the development team and organization, which vary with

organizational context. Similarly the project relationship management also varies with that of the organizations like formal or informal for specification as policies and procedures varies.Following are the models on and around requirement engineering process.

BOOTSTRAP: This is a methodology [23] which continues to evolve. At the end of 1993 "The European Software Institute" actually decided to uphold transfer of technology and methodologies related to improvement within Europe, besides this they also decides to utilize Bootstrap as a considered tool to not only assess but also analyze the software processes and to set up improvement related programs. In their work author explains the model elements and tells that how this model can be utilized to assess the need to improve the process besides with its readiness for ISO 9001 certification. We have identified the fact of varying process among organizations from the model as this model highlights the fact of varying process among the organizations by focusing on organizational contexts.

The Personal Software Process (PSP): Watts comes up with "Personal Software Process (PSP)". This model [24] gives engineers with a framework in order to perform software work. The PSP process comprises of methods, forms, and scripts which illustrate software engineers about planning, measuring, and managing their work. The PSP can be used with any design methodology. Besides this PSP can be used with different aspects of software work like; requirements specification, processes definition, and repairing defects. When engineers use the PSP, the objectives of products with zero defect rates, on time or on schedule production and within estimated costs. The factors we have identified from this model are related to; engineer knowledge, process quality evaluation and engineer personal work quality as the varying factors among the organizations and has strong impact and influence on software quality.

Requirement Engineering Maturity Measurement Framework (REMMF): Mahmood Niazi et al. came up with a framework (REMMF) [25] which is based on Sommerville's model [31][32]. The key reason of coming up with REMMF is to effectively compute or evaluate the requirement engineering process maturity which are used by organizations and to support practitioners in evaluating the requirement engineering process maturity. We identified that the framework is based on model focuses the attention towards the requirement engineering problems and says that it varies with the organizational context. It says that communication style, application complexity, lack of training and undefined requirement process is some of the factors which create these problems as they vary with organizational context. Hence also highlights the organizational context aspects.

Software Capability Maturity Model (SW-CMM): The SW-CMM [26] shows sets of suggested practices in form of key process areas (initial, repeatable, defined, managed and optimized) to improve software process capability. Besides this, it not only guides the software organizations to get control of processes use for software development and maintenance but also guides to progress toward software engineering and excellence culture. We identified from the model that it focuses on the varying quality objective factor among organizations for process or product evaluation.

Software Capability Maturity Model Integration (SW-CMMI): The model CMM Integration [27] gives guidance to improve organizations besides with the management of product or service from development till maintenance. CMMI help organizations to appraise the organizational maturity, developing improvement priorities and to implement them via approaches. Initial, Managed, Defined, Quantitatively Managed and Optimizing are the key areas of this model.

We identified the factors related to; quality objectives, models, application domain, organizational structure and size and process vary with organizational context. Besides this we came to know that the model focuses the attention towards the fact that the organization process operates in the business context it means that the process improvement efforts vary with organizational business environment. Similarly the model also describes the fact that the process varies among the organizations because the organizations have different contextual variables like; domain, nature of customers, costs, schedule, quality tradeoffs, technical difficulty of work, experience of the people who are indulged in implementing process. We also identified that the terminologies meaning vary among the development teams, which have influence on the software quality. Similarly it also says that the requirement management is greatly influenced by

some of the factors like; distinct locations of the team members, engineering capabilities, staff and facilities available and staff experience. Model continues to explain some more factors which have influence on requirement process like; techniques due to varying economy and accountability, training techniques or approaches which vary on basis of work environment, organizational context, knowledge, cost and schedule, availability of critical resources and availability of key personals.

Requirements Capability Maturity Model (R-CMM): Sarah Beecham et al. [28] come up with a requirement process improvement model. The model centers the attention on the requirements engineering process as distinct from the Software Engineering Institute's (SEI's) software process improvement framework. This model can be used either together with SW-CMM or separately to assess capability of Requirement Engineering process. The R-CMM comprises an assessment method which helps the user to be aware of their current Requirement Engineering process with perspective of its implementation in opposition to maturity goals. From this model we identified the factors related to requirement engineering and says that it is greatly influenced by the organizational and technical factors like; poor stakeholders communication, staff retention rate, change resistance, resource allocation, skills, application complexity, requirement understanding, requirement traceability and undefined requirement engineering process. Besides this it also focuses attention towards the varying organizational culture and says that it greatly influences the development and support. Model describes the process improvement aspect and says that it is context sensitive like; product nature, its usage, customer perception about product and evolution and organizational structure. These factors vary with context and have an impact on process improvement.

Requirements Engineering Process Maturity Model (Uni—REPM): There are number of models discussed above which particularly focus on requirement engineering process but it is also not to be ignored that all these models are restricted to Bespoke development. This Uni-REPM is a model [29] which via set of activities presents the requirement engineering process maturity. These activities are comprised of seven areas: Organizational support, Requirements Management Process, Elicitation, Requirements Analysis, Release Planning, Documentation & Requirements Specification and Requirements Validation. The model provides support to the organizations to identify their process strengths and weaknesses. Besides this it also guides the organizations to better improve the requirement engineering process.

Similar with other models, we have identified that, this model also describes the processes varying nature with the characteristics of organization or environment. Similarly the requirements methods or techniques vary with usage context like, knowledge type and purpose of requirement. It is also notified that the sources of requirements are influenced by social and political aspects. Besides this the model says that the standard for the requirement documentation vary with organizational context as the organization customs, type of product and development process vary in them.

TRILLIUM: This model is for Telecom Product Development & Support Process Capability. *Trillium* model [30] is on basis of a customer perspective, as apparent in a competitive setting. In this context, capability is to develop a product according to the customer needs, in less cost and with in schedule. This Trillium model is used by the organizations for continuous product development and improvement. This model is based on *Capability Maturity Model* (CMM). We identified the factors from the model related to; organizational culture and says that it influences the development capability, process improvement and selection.

4.2. Identification and Fusion of Explicit and Conceptual Data Unit's Duplication:

189 factors are identified in the identification phase of data unit, which are progressively and systematically refined into 37 factors and 7 related classifications (for this whole cleansing process, wide-ranging notes are maintained in form of memos). This refinement is done by identification and removal of clear duplication and conceptual duplications among the data units. We combine two data units which are not having same textual tags but are having same meaning. These refined factors are shown in *Table 3*.

4.3. Identification of Factors Grouping and Classification:

As shown in *Table 3* the factors are associated with classifications and are mapped to the data sources. The factors naming with that of classification is greatly influenced by the classifications and terminology used in the data sources. Similarly the data duplication is seen at two levels i.e. explicit and conceptual level. So the data units which have duplications at any level is then removed and combined to a single data unit.

For example: organization nature, organization and organization characteristics are combined in a single name that is: 'organization nature' factor. In the similar manner personnel knowledge, education and knowledge are combined into 'personnel knowledge'; similarly changing needs and requirement change are same so they are also combined together in requirement growth etc.

4.4. Reviewing Group Labels and Classification and Generating the Final List

These factors are further decomposed into several sub-factors which are discussed and shown below in *Table 3*.Each of the situational factors identified is organized after the filtration against the explicit and conceptual duplication into respective groups and classification. These factors are further comprised of certain sub-factors which has an impact on them as changes in sub-factors ultimately results in the situational factor which further changes the situation for requirement engineering process.

In this last phase, 37 factors and 7 classifications that were formed via constant comparison and memoing are inspected for potential omissions. The theory includes omissions for entire classification, sub-factors and factors. And after omissions and modifications at classification, factors and sub- factors level 28 factors and 6 classifications are found.

As shown above in *table 3* **"Organization"** has eight factors that is; *Nature:* by nature it means the maturity of the organization i.e. mature and immature, *size:* by size it means the size of an organization i.e. small, medium and large size organizations, *Infrastructure:* by infrastructure it means all the things which are related to organizations setups like; utilities of an organization, working facilities etc., *work environment:* this factor covers all the constraints, laws, policies and conditions related to organizations, *business process:* by business process it means the organization business set of steps and activities, which they use adopt for their business execution, *structure:* this factor relates to the organizational structure like: functional structure, Divisional structure and Matrix structure where each of the structure relates to different working strategies for example in functional structure employees work in departments (engineering, maintenance finance etc) where as divisional is on basis of location, product etc and matrix is combination of both. *Economy:* deals with all the things which actually limit the organization economy like; organization internal resources and external market situation. *Training:* is the factor which covers the aspects related to training which organization's provide to their employee for any specific field or area like; if for example some new tool comes in the market for requirement design then whether the organization provides training provided to its employees for that new tool etc. The sub factors discussed for each of the factor comes under classification "Organization" actually shows that base upon them how these factors vary among organizations. Like for *"nature"* Communication with client, application complexity, and training and un-defined requirement engineering processes are some of the sub factors which actually make any organization nature different from the other one. Similarly *"size"* of the organization varies with the resources the organizations has which makes it small size organization, medium size organization or large organization. Building, workspace, supported services, utilities vary among organizations which make the *"infrastructure"* of the organization different for each of them. Similarly the other sub factors related to "*work environment", "business process", "structure", "economy*" and "*training"* as shown in *table 3* actually varies due to which the specific factor varies among the organizations.

"Software Life Cycle" has five factors which are further decomposed into several sub factors as shown in *table 3. Model:* is the factor which deals with the process models the organization adopt for

completion of any product. This factor is further comprised of sub factors like: scope or magnitude of project, project complexity, changing needs, changing opportunities. These factors in fact vary among organizations which lead to varying organizational process models to be adopted by them in order to complete any product. Similarly S*tages:* is another factor which deals with the organizational software life cycle stages which vary with the organizational context. By organizational context it means; organization structure, skills, capabilities, processes, technologies, culture and level of formality. *Policies, procedures or process* are other factors deals with the organization varying nature on basis of these factors. The sub factors which actually varies and due to them these factors vary among organizations are related to organizational context.

"Resources" is the classification which contains three factors related to the resources of an organization like: Availability, Management and Humans. *Availability:* is the factor which deals with the extent to which the resources are available to an organization. The sub factors which influence the resource availability are organization policy, resourcing plan, economic environment, and schedule. Similarly *Management:* is another factor deals with the resource management and tells the fact that each of the organizations having varying resource management strategies. The sub factors skill, understanding and experience are found to be the one who make the resource management vary among organizations. *Human:* is the factor which is found to be varying among organizations. Here humans are also considered as the organizational resource and the sub factors like: Personal competence, training, skills, experience, responsibility, mode of interaction with system and client are found to be one who make the human resources vary among the organizations.

"Requirements" is the classification which consists of seven factors which vary among organizations related to requirements. *Process:* is the factor which deals with the fact that the requirement engineering process vary among the organizations. The sub factors; quality tradeoff, schedule, application domain, organizational context, human skills, capabilities, workforce, cost, interaction (experience, knowledge, skill), process assessment (tools, competency, models, reference models), activity priority, activity type, process quality (organizational context, engineer quality work), process management (engineer knowledge, engineering capability, staff and facilities available, experience, schedule, plan and effort) are found to be varying among organization which leads to varying process. *Specification:* requirement specification is found to be varying among organizations and the sub factors which results in varying requirement specification are found to be Customer, technical community, environment, communication, policies and procedures. *Utilities:* is found to be a factor which deals with tools, models, standards, approaches and techniques and the sub factors; Behavioral approaches, Style (textual, formal), organizational operational environment, organization condition, usage context, knowledge type, requirement purpose, Organization customs, type of product, development process are found to be varying among organizations leading to varying requirement utilities. *Growth:* it is found that the growth of requirements vary among organizations due to the sub factors of vague requirements or changing requirements. Similarly *conflict:* is another factor which deals with the requirements conflict and it is also found that the conflicting requirements are due to requirements incompleteness, ambiguity, communication mode, knowledge, client type and communication style (formal or informal). *Terminology:* is the factor which deals with the concept of different requirement terminologies meanings among organizations and it is found that the sub factors of culture, language, client type, words and terms vary with organization which lead to varying terminology meaning. Similarly *source:* relates to the requirements sources and deals with the fact that this factor also contributes to the changing situation among organizations. Social aspects, political aspects, relationships, understanding are found to be sub factors which vary among organizations leading to varying requirement sources.

"Management" is the classification which consists of two factors related to organization management people. *Review:* this factor deals with

the management review process for the requirement engineering which is found to be varying among organizations. Communication style, client type, skills are found to be varying sub factors leading to varying management reviews. *Commitment:* deals with the fact that the management commitment varies with organizations and it has impact on the requirement engineering process. The sub factors; values, skills and leadership styles are found to be varying among organizations which lead to varying commitment level.

"Risks" is the last identified classification which consists of four risks related factors which can impact the requirement engineering process. There are four type of risks which are identified like; *requirement:* this risk factor is found to be a source of changing situations among organizations. The sub factors which actually lead to this varying situation are uncertain requirements, vague requirements and infeasible requirements. Similarly *technology:* deals with changing situations due to changing technology. Now the sub factors which results in such changing situations is unavailable technology. *Estimation:* is identified as another factor which can generate risk and due to which the situation can be changed among organizations. The sub factor resulting in this changing situation is unrealistic estimation of cost, schedule and effort. Skills and understanding are also found to be sub factors which can lead to estimation risks. *Staff:* is the risk related to organization workforce. This factor is found to result in varying situations among organizations and the sub factor identified for this varying situation is inadequate staff.

In order to further check the accurateness of the list of situational factors, sub factors and classification, it is forwarded to two experts from the academia. Both of them are having above ten years of experience in requirement engineering field. They recommended some very important changes which are then incorporated in our study by performing modifications like: they said that sub factor *"organizational context"* should further elaborated, so it is then modified into *organization structure, skills, capabilities, processes, technologies, culture and level of formality*.

Similarly there were some sub factors which comes under same heading but are separately written like *organization technology, tools, environment, workplace facilities* all come under organizational context hence they are then combined together and modified in *organizational context* word. Similarly the recommendation from the experts to combine the factors *Tools, Models, Techniques and Approaches, Traceability and Standards* into a single factor is done by combining them together and giving the name of *utilities.*

The *client* classification is removed from the list and combined with the factor *"source"* which comes under classification "*Requirements"*. Besides this its factors *relationship and understanding* are included in the list of sub factors of "*source"* (one of the factors comes under "Requirement" classification). Similarly *values* and *leadership* styles are the two sub factors included with the sub factors of "*commitment"* (a factor coming under "*Management*" classification). After the final review phase, we come up with situational factors consisting of 6 classifications and 28 factors that may influence the requirement engineering process as shown in *Table 3*. Table also consists of listing of 144 sub-factors – which are factors components, which are carefully consolidated from the various data sources.

5 DISCUSSIONS

In this section, we talk about how this work can be helpful for research communities. This work also leads to some possible future work areas. In the end some limitations are also discussed.

5.1. Utilizing Situational Factors:

The situational factors which are identified from the standards and models on and around the area of requirement engineering process are the important reference for researchers, who are working in situational requirement engineering standards and models as it shows the list of factors which are identified from the existing standards and models deals directly or indirectly to requirement engineering. Specifically it is useful for the researchers who are working in requirement engineering process improvement

with perspective of changing situations. The initial list can also guide the researchers and practitioners working in situational requirement engineering, to consider the standards and models as a significant consideration for dealing with changing situations faced by them while performing RE process. Besides with the initial list identification, this work also contributes in describing all the related standards and models on and around RE process in detail by explaining its purpose and the domain in which these standards and models are used. But the main contribution which is the initial list of factors identified, represents the technical (tools, techniques etc) and non-technical (motivation, values etc) aspects hence fully covering the requirement engineering process from the standards and models data source.

5.2. FUTURE WORK

The list of factors identified is from one of the domain from the areas which can affect requirement engineering process by changing the situations for the process. The domain we have searched for this study is standards and models which are on and around requirement engineering process. The other domains can be risks factors affecting requirement engineering, environmental factors affecting requirement engineering, body of knowledge etc as shown in *Table 1,* which are also to be investigated for the identification of the situational factors.

Our upcoming research is focusing to identify the situational factors from other domains; risks related to requirement engineering, environmental factors, requirement engineering process tailoring etc. hence this study will act as a basis for the upcoming research and will also become one of the various other important inputs for the comprehensive list of situational factors which affect requirement engineering process in global software development (GSD).

5.3. LIMITATIONS

We have given care in selection of the data source but it can become its limitation as the scope of the situational factors list is limited to the standards and models related to requirement engineering directly or in-directly (like we have also included the standards and models which deals with software development besides with the one which are directly dealing with requirements engineering). This limitation will be overcome by completing our future work (covering all the related domains as describe in *table 1*).

Another limitation of the study is that it does not cover the identification of factors from industry as the study comprises of the situational factors which are identified from the state-of-the knowledge. The steps to overcome this limitation is taken by designing a questionnaire and surveying the situational factors from the industry which will be ultimately included in the final comprehensive list of situational factors affecting RE process in GSD.We make every effort to cover all the related standards and models on and around RE process but still possible, that we may miss any related standard and model. Similarly the paper is forwarded to other researchers in order to deal with biasness about the search protocol used, but still biasness aspect can not be ignored as well. The threat of misinterpretation can also not be ignored, as it is one of the must factor in every literature review, although we tried our best to overcome these aspect by dealing it carefully. We also cannot ignore the threat related to precision. In our work we tried to have high precision rate but still the maximum precision is not assured.

6 CONCLUSIONS

This paper shows the evidence that the situational context is involved in finding the most suitable requirement process definition for any product. So we do not come across any study which identifies the situational factors which affect requirement engineering process. Its absence motivates us to identify the situational factors from various domains related to requirement engineering which can lead to changing situations during requirement engineering process. For that we have selected one of the various domains that are standards and modes directly or indirectly related to requirement engineering process for this study.

In order to come up with the unique situational factors we used the constant comparison and memoing phases of Grounded theory [8]. So by following this theory we have organized our work in six steps. Each of the step lead to some contribution to our study. In the end we come up with the list of unique situational factors which are

filtered on basis of explicit and conceptual duplications. After which they are grouped with various labels and then classified accordingly.

We think that the list of situational factors consisting of key situational elements related to requirement engineering process represents a strong initial input from the domain of standards and models of RE process towards a comprehensive list of situational factors, which we will have after having factors from all the related domains of requirement engineering process.

7 REFERENCES

[1] Locke, E.A. and G.P. Latham, *A theory of goal setting & task performance*. 1990: Prentice Hall.

[2] Chatzoglou, P.D., Factors affecting completion of the requirements capture stage of projects with different characteristics. Information and Software Technology, 1997. **39**(9): p. 627-640.

[3] Paul Clarke, Rory V. O'Connor, The situational factors that affect the software development process: Towards a comprehensive reference framework, Information and Software Technology 54 (2012) 433–447

[4] Boehm, B. W, The economics of software maintenance. In Proceedings of Software Maintenance Workshop (Washington, D.C.), 1983, 9-37.

[5] Jantunen, S., "*Exploring software engineering practices in small and medium-sized organizations*", Proceedings of the 2010 ICSE Workshop on Cooperative and Human aspects of Software engineering (CHASE'10), pp.99-101, 2010.

[6] Espiti. 1996. Software process improvement on the right road with ESPITI—The ESPITI European Survey Results. ESPITI Newsletter Issue 2. Available at: http://www.cse.dcu.ie/cse/international/trispin/News2.html#espiti.

[7] Hall, T., Beecham, S. & Rainer, A. *Requirements problems in twelve software companies: An empirical analysis.* IEE Proceedings—Software, (2002) 149(5), 153–160.

[8] Glaser, B.G., Basics of Grounded Theory Analysis: Emergence Vs. Forcing. 1992: Sociology Press.

[9] Knigge, L. and M. Cope, Grounded visualization: integrating the analysis of qualitative and quantitative data through grounded theory and visualization. Environment and Planning A, 2006. **38**(11): p. 2021-2037.

[10] Boychuk Duchscher, J.E. and D. Morgan, *Grounded theory: reflections on the emergence vs. forcing debate.* J Adv Nurs, 2004. **48**(6): p. 605-12.

[11] Douglas, D., *Grounded theories of management: a methodological review.* Management Research News, 2003. **26** (5): p. 44 - 52.

[12] Corbin, J.M. and A.L. Strauss, Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2008: Sage Publications, Incorporated.

[13] Bryant, A. and K. Charmaz, *The Sage Handbook of Grounded Theory*. 2010: Sage Publications.

[14] Glaser, B.G., Doing Grounded Theory: Issues and Discussions. 1998: Sociology Press.

[15] Weissfelner, S. (1999). ISO 9001 for Software Organizations. In K. Emam & N.

[16] ISO/IEC 15504, Joint Technical Committee ISO/IEC JTC, *Software and system engineering*, 2004.

[17] ISO/IEC 12207:2008, *Systems and software engineering — Software life cycle processes*, Software & Systems Engineering Standards Committee of the IEEE Computer Society, 2008.

[18] IEEE Std 830-1998, *IEEE Recommended Practice for Software Requirements Specifications*, Software Engineering Standards Committee, 1998.

[19] IEEE Std 1233, 1998 Edition, *IEEE Guide for Developing System Requirements Specifications*, Software Engineering Standards Committee of the IEEE Computer Society, 1998.

[20] IEEE Std 1362™-1998 (R2007), *IEEE Guide for Information Technology—System Definition— Concept of Operations (ConOps) Document*, Software Engineering Standards Committee of the IEEE Computer Society, 2007.

[21] IEEE Std 15288-2008, *Systems and software engineering — System life cycle processes*, Software & Systems Engineering Standards Committee of the IEEE Computer Society, 2008.

[22] IEEE Guide— Adoption of ISO/IEC TR 24748-3:2011, Systems and Software Engineering—Life Cycle Management, Software & Systems Engineering Standards Committee, 2012.

[23] Hasses, V., Messnarz, R., Koch, G., Kugler, H. J., and Decrinis, P. 1994. Bootstrap: Fine tuning process assessment. IEEE Software 11, 4, 25–35.

[24] Humphrey, W. 1989. Managing the Software Process. Addison-Wesley, Reading, Mass.

[25] Niazi, M., Cox, K., & Verner, J. (2005a). An empirical study identifying high perceived value requirements engineering practices. In Fourteenth International Conference on Information Systems Development (ISD'2005). Karlstad University, Sweden August 15–17.

[26] Paulk M.C., and Chrissis M.B., The 2001 High Maturity Workshop (CMU/SEI-2001-SR-014), Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, January 2002.

[27] Ahern, D. M., Clouse, A., and Turner, R. 2001. CMMI Distilled. Addison-Wesley, Reading, Mass.

[28] Beecham S., Hall T., and Rainer A., "Building a Requirements Process Improvement Model," Technical Report No. 378. University of Hertfordshire: Hatfield, 2003.

[29] Gorscheck, T., Svahnberg, M., & Kaarina, T. (2003). Introduction and application of a lightweight requirements engineering process evaluation method. In Proceedings of the Requirements Engineering Foundations for Software Quality (REFSQ'03) (pp. 83–92). Klagenfurt/Velden, Austria.

[30] Coallier, F. 1999. Trillium: A model for the assessment of telecom product development and support capability. In Software Process Improvement, R. B. Hunter and R. H. Thayer, Eds. IEEE Computer Society Press, Los Alamitos, Calif.

[31] Sommerville, I. and Sawyer, P. 1997. Requirements Engineering: A Good Practice Guide. Wiley, Chichester.

[32] Sommerville, I., & Ransom, J. (2005). An empirical study of industrial requirements engineering process assessment and improvement. ACM Transactions on Software Engineering and Methodology, 14(1), 85–117.

Table 2. Situational Factors from Standards and Models

Classification	Std & Models/ SF	ISO 9001-2000	ISO / IEC 15504	ISO / IEC 12207	IEEE 830-1998	IEEE 1233-1988	IEEE 1362-1998	ISO/ IEC 15288-2008	ISO / IEC 24748	BOOTSTR	PSP	REMMF	SW-CMM	SW-CMMI	RW-CMM	REPM	TRILLIUM
Organization	Nature	x							x			x				x	
	Size	x												x			
	Infrastructure	x		x		x	x	x	x					x			
	Work environment	x		x		x	x		x		x			x		x	
	Business process			x		x								x			
	Structure						x							x			x
	Economy							x						x			
	Training												x	x			
SW Life – Cycle	Model		x						x					x			
	Stages			x													
	Policies			x													
	Process			x						x							
	Procedures			x													
Resources	Availability	x												x			
	Management	x											x				
	Human resource	x					x	x						x			
Requirement	Process	x		x	x			x	x	x				x		x	x
	Project Complexity	x		x								x	x	x			
	Specification				x				x				x				
	Tools				x				x								
	Models			x										x			
	Techniques & approaches					x			x							x	
	Growth			x									x				
	Traceability												x				
	Conflicts	x															
	Terminology													x			
	Source					x								x			
	Standard					x										x	
Client	Related process	x															x
	Relationship								x							x	
	Understanding	x											x				
Management	Reviews	x															
	Commitment	x						x									
Risks	Requirement												x	x			
	Technology								x					x			
	Estimation										x			x			
	Staff													x			

Table 3: Classifications of Factors and Sub-factors

Classification	Situational Factors	Sub-factors
ORGANIZATION	Nature	Communication, application complexity, training, un-defined requirement engineering process
	Size	Resources (Small size organization, medium size organization, large organization)
	Infrastructure	Building, workspace, supported services, utilities
	Work environment	Conditions, politics, market, standard, technical policy, cultural policy, physical environment, operational policy, quality policy, quality objective
	Business process	Poor stakeholder communication, staff retention, change resistance, lack of training, lack of skills, poor resource allocation
	Structure	Functional, divisional and matrix
	Economy	People, material, infrastructure, capital, customer behavior, competitive position.
	Training	Knowledge, cost, schedule, work environment
SOFTWARE LIFE CYCLE	Model	Project scope or magnitude, project complexity, changing needs, changing opportunities.
	Stages	organization structure, skills, capabilities, processes, technologies, culture and level of formality
	Polices	Organization structure, culture.
	Procedures	Organization structure, skills, capabilities, culture and level of formality
RESOURCES	Availability	Organization policy, resourcing plan, economic environment, and schedule
	Management	Skill, understanding, experience
	Human	Personal competence, training, skills, experience, responsibility, mode of interaction with system and client
REQUIREMENT	Process	Technological differences, quality tradeoff, schedule, application domain, organizational context, human skills, capabilities, workplace, environment, application domain, facilities, equipment (HW, SW), facilities, tools, workforce, cost, interaction (experience, knowledge, skill), process assessment (tools, competency, models, reference models), activity priority, activity type, process quality (organizational context, engineer quality work), process management
	Specification	Customer, technical community, environment, communication, policies, procedures
	Utilities (tools, models, standards, approaches, techniques)	Behavioral approaches, Style (textual, formal), organizational operational environment, organization condition, usage context, knowledge type, requirement purpose, Organization customs, type of product, development process
	Growth	Requirements change, vague requirements.
	Conflict	Requirements incompleteness, incorrect requirement, ambiguous requirement, communication mode, communication style, understanding, knowledge, client type
	Terminology	Culture, language, client type, words and terms.
	Source	Social aspects, political aspects, Relationship with client (formal or informal), policies, procedures, communication style, client type, feedback.
MANAGEMENT	Reviews	Communication style, client type, skills
	Commitment	Values, skills, leadership style
RISKS	Requirement	Uncertain requirements, vague requirements, infeasible design
	Technology	Unavailable technology
	Estimation	Unrealistic estimates (cost, schedule, effort), skills, understanding
	Staff	Inadequate staff

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 95-102
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Developing Usable Software Product Using Usability Risk Assessment Model

Jayaletchumi T. Sambantha Moorthy, Suhaimi bin Ibrahim, Mohd Naz'ri Mahrin
Universiti Teknologi of Malaysia (UTM)
Kuala Lumpur, Malaysia.
tjaya771129@yahoo.com, suhaimiutmkl@gmail.com, mnazrim@ic.utm.my

ABSTRACT

Usability is an important factor in ensuring development of quality and usable software product. Ignorance and unawareness about the concept of usability and failure to address usability during software development process has led to usability problems in software product. Many efforts has been suggested in literature to overcome usability problem in software products but current practices faces challenges in reducing these usability problems. Alternatively, the concept of risk management can be used to control usability problems even though these problems cannot be eliminated totally. The concept of risk management is important to deal with usability problem before it occurs. Unfortunately, there is still lack of proper definition of usability risk and a proper model to identify, analyze and prioritize potential usability risk during Software Development Lifecycle (SDLC). This paper presents comprehensive study on the need for Usability Risk Assessment Model to reduce usability problems in software products.

KEYWORDS

Usability; Usability Problem; Usability Risk; Risk Management; Software Risk Assessment.

1 MOTIVATION

Software quality has emerged as important aspects of software development process, as it might result in serious consequences such as financial loss and reputation loss [1]. Quality improvement after development of software is not recommended because it only increases the cost and is almost remaking the product [2]. Poor software quality has been identified as the key factor for success or failure of a software product [3][4][5].

Usability has been recognized as an important quality factor in various quality model [6][7][8] and has always been present even in the very first model of software quality known as *McCall Factors Criteria Metrics (FCM)* proposed by McCall in [9].

Study has shown that ignorance and unawareness on most referenced usability standard, ISO 9241-11 among industries [10] and governmental organizations [11] in Malaysia, has created various usability problems in software products [12]. Existence of usability problems are also reflected when usage level of software products is low [13]. A survey by Malaysian Administrative Modernization and Management Planning Unit (MAMPU) in 2011 had reported that the usage of online services in government is only 40% of the overall online services provided [14] and this result shows that users are exposed to higher usability problem in using online services[15] [16] [17]. Example of reported usability problems are such as high number of broken links and slower accessibility speed [49]; less usability activities in product designs; limited skills and knowledge on usability among the designers and management; unawareness on various activities of usability engineering life-cycle and inappropriately used usability methods [16]. Generally, existence of usability problem could lead to failure of a system.

To increase the rate of successful software project, it is crucial to reduce usability problems. Usability professionals have expressed that by integrating usability closely with software development process [18], usability problems can be reduced. There are many efforts to integrate a formal usability process, standard, techniques and practices into software development process to

improve the interaction and quality of the systems [19][20]. However, software developers face difficulties when new usability practices are introduced [21]. Some usability approaches are only integrated in requirement and design phase [22]. In fact, their practical implementation is largely missing. Usability practices as well are not part of requirement engineering [19], so developers are often given an incomplete, confusing, and sometimes contradictory requirement. As a result, many development teams are facing difficulties in avoiding and minimizing usability problems.

Besides this, various usability evaluation activities such as inspection, empirical testing, and metrics for usability standards in computing has been integrated into software development process to measure and improve usability of software [23]. However, it only evaluates a completed system and does not intervene at earlier stages of development process [24]. An International standard, ISO 13407 [25] also had proposed a framework for integration of usability in all phases of software development process. Even so, current usability engineering practices had failed to reduce usability problems in software products.

Alternatively, studies have shown that problems in software products can also be controlled using Software Risk Management methods, even though these problems cannot be eliminated totally [26][27]. Using Software Risk Management, problems in software products are dealt before it occurs, so that risk control activities can be planned and implemented as needed.

Tim Altom [28] had suggested that usability should be portrayed as risk management and not as an abstraction, to avoid later arguments, encourages discussion and allows management to see the benefit of usability in software products. Study has also shown that usability problems can be considered as a significant usability risk factor [13]. However, there are great ignorance on managing usability risk compared to managing other risks such as technology risk, market risk and money risk [32]. Furthermore, there is little effort in identifying, analyzing and prioritizing potential usability risks at earlier phases of development process. If development team continues to develop software products without identifying, analyzing and prioritizing usability risks, the chances of producing less usable software products are higher.

In relation to this, development of Usability Risk Assessment Model could guide development team on identifying, analysing and prioritizing potential usability risks that could arise during SDLC in order to produce more usable software products with less usability problems.

This paper is structured as follows. Section 2 reviews existing studies on usability, usability problem, usability risk, software risk assessment processes and software risk assessment models. A conceptual view of proposed Usability Risk Assessment Model is illustrated in Section 3. Section 4 explains on contribution and Section 5 includes conclusion and future work.

2 RELATED WORKS

Existence of usability problems creates many quality problems in software and contributes to its failure. Risk management approach can be used to overcome usability problem in software by identifying and analysing usability risk earlier in the SDLC.

2.1 Usability

Usability is considered as one of the significant factor of software product quality. Even, five most mentioned quality models, McCall [9], Boehm [6], FURPS [30], Dromey [8] and ISO 1926/2001 [7] had defined usability as significant factor in improving product acceptability and reliability, increasing user satisfaction, and it is also financially beneficial to companies [31]. Usability is the best factor that balances between technical and human aspects of a software product which is important in defining quality [32]. Hence, it's important to ensure usability characteristics are integrated well in software products to ensure its success.

2.2 Usability Problem

Many studies had revealed that usability problems causes quality problems in software product [33][34]. Potential problems lying in interface design, operating process or product structure also contributes to usability problems which resulted in lower effectiveness, efficiency and difficulty of use for end users [35]. Websites as well faces huge usability problems which has been identified as factors for lower usage level [15][36][37][38].
Discussion above shows that it is extremely important to reduce frequency of usability problem to improve user-experience and quality of software product [39].

2.3 Usability Risk

The term usability risk was first introduced in the context of e-commerce and World Wide Web services [29]. Generally, usability risk is a chosen action or activity that leads to a loss or an adverse outcome which could impact the usability of a software product. It is related to user acceptance and meeting user's requirement. Usability risk contributes to negative user experiences which lead to software product failure [40]. However, search for the term 'usability risk' in the literature has shown that this term is not widely used and some studies related to mobile application has mentioned this term in their studies [40][41][42]. Most of the current researches are associated with usability problem and not usability risk.

Usability problem can also be perceived as a risk factor in producing usable software product since it is a threat to an optimal user experience [13]. User experience, in general includes many aspect such as human factors, design, ergonomics, Human Computer Interaction (HCI), accessibility, marketing as well as usability. Hence, usability problem is recognized as a risk factor for producing usable software product.

Since studies on usability risk is still lacking, we believe that it is important to define usability risk as a single entity, either in form of checklist, model or in other appropriate form, as a guidance for development team in developing a usable software product.

2.4 Software Risk Assessment Processes

For several decades, risk management has been a popular approach in non-software domain and it has been adopted in software domain since the last few years [43].

According to Boehm [44], two main processes in software risk management are software risk assessment and software risk control. The first process in software risk management is software risk assessment with activities of risk identification, risk analysis and risk prioritization. The second process in software risk assessment is risk control and it involves risk management planning, risk monitoring and risk resolution.

Usability Risk Assessment Model focuses only on the software risk assessment activities namely, Identification. Analyze and Prioritization. Furthermore, since usability risk has not been identified as an entity previously, it is more appropriate to conduct the process of risk assessment before risk control can be performed.

2.5 Software Risk Assessment Models

Some popular published software risk assessment models are Software Risk Assessment Model (SRAM) [45], Software Risk Assessment and Estimation Model (SRAEM) [46], Risk Identification, Mitigation and Avoidance Model for Handling Software Risk (RIMAM) [47], and Software Risk Assessment and Evaluation Process using Model Based Approach (SRAEP) [48].

All four assessment model explains the perspective of software development and not software product itself. These models are lacks in practical guidelines. There is also a lack in standard framework or model for assessing and managing software product risk based on attributes of quality, particularly usability. This creates extensive need for Usability Risk Assessment Model.

3 CONCEPTUAL MODEL

As the needs for a Usability Risk Assessment Model is clearly described in the previous section, a proposed model of usability risk assessment is suggested in this section. Figure 1 illustrates the proposed model of usability risk assessment.

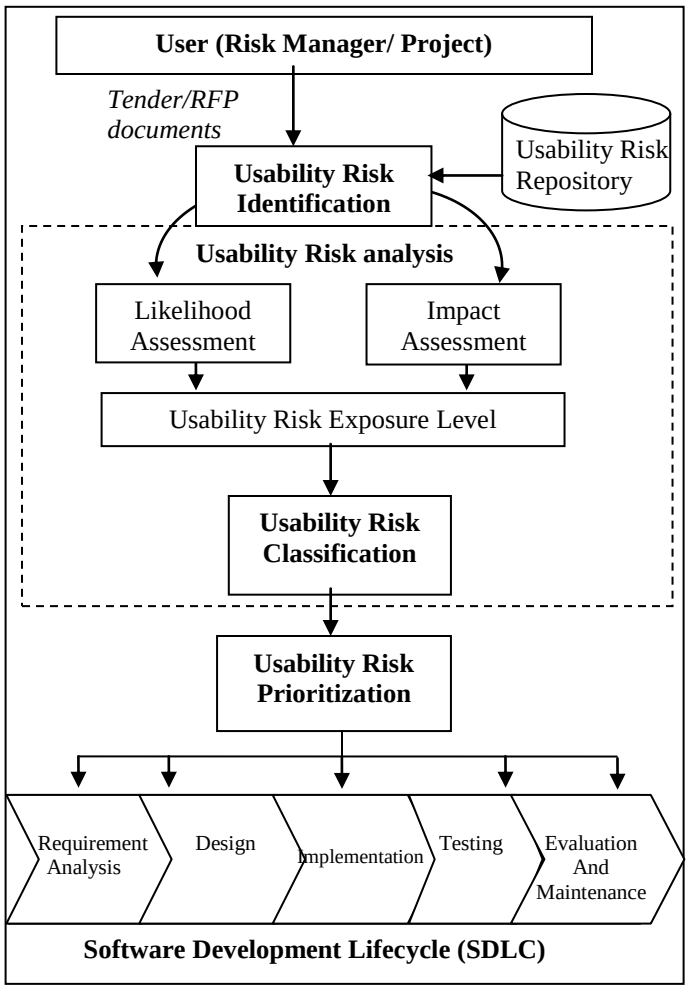


Figure 1: Conceptual view of Usability Risk Assessment Model

In Figure 1, important elements and activities in the proposed Usability Risk Assessment Model are shown and its description is as given below:

(i) Usability Risk Identification

This model facilitates user in identifying potential usability risk in a software development projects by analysing initial tender, Request for Proposal (RFP) document or any other documents that clearly stated the initial intention of the software development projects. Potential usability risks are selected from the list of usability risks stored in *Usability Risk Repository*. Identification of potential usability risks is done with two steps:

a) Identification of usability risks mentioned in the literature. Search in several digital databases such as *Institute of Electrical and Electronics Engineers (IEEE), Association for Computing Machinery* (ACM), Springer, Science Direct, Web of Science and Google Scholar will be conducted to identify existing usability risks in government applications and systems.

b) Identification of usability risk factors from usability attributes defined in a quality model which will be used to construct questionnaire to identify usability risks from the industry.

In 2012, Aman Kumar, Arvind and Hardeep [2] suggested that factors affecting the quality of software can be identified from attributes defined in software quality models. This study considers attributes of usability as factors in producing usable software products. These usability attributes are subjected to risks that a software product might have troubles with. Scenario which affects the ability to achieve these attributes is considered as potential usability risk during software development process. Based on usability attributes defined in quality model, namely *Efficiency, Satisfaction, Comprehensibility and Safety*, a list of usability risk factors will be derived and used as constructs to design and develop a questionnaire which will be distributed to the industry to collect information on usability risk, from the perspective of industry.

Potential usability risks identified from a) and b) will be reconciled so that risks with the same meaning are considered only once. The output will be a list of potential usability risk that could occur during SDLC.

(ii) Usability Risk Analysis

Once potential usability risks have been identified, these risks will be analyze using risk analysis technique to determine the likelihood and impact of each usability risks towards phases in SDLC.

Likelihood and impact of each usability risks are identified using the Delphi method. Classification of usability risk based on SDLC phases is also determined from this method.

Delphi is a method to iteratively gather experiences, knowledge, and opinions of the experts on an issue or development process under study by using normally by interview or survey [50]. It uses a number of questionnaire rounds, feedback from responses of experts and gives opportunity to the experts to modify their responses and anonymity of responses are assured,

In this study, using Delphi method, a three-iteration questionnaire survey will be conducted with experts. The first round gives opportunity for the experts to suggest new potential usability risks and determine the likelihood of each usability risks. The second round determines the impact of each usability risks to SDLC phases**.** The third round involves the activities of experts to classify and suggest mitigation plan for each usability risks.

Experts are selected based on their experiences in software development and/or risk management. Experts with more than 10 years of experience in dealing with software development projects or who are seek advice from when dealing with software development projects, are chosen. There is no rules on the number of experts that is required in a Delphi survey since the decision on the number of experts needed is taken after consideration of factors such as time and expenses [51]. The representation of experts is assessed with the qualities and experiences acquired than its numbers [53]. Study has suggest that a suitable expert size from 4 to 3000 [50][51][52] but for the context of this research, five experts will be chosen due to time factor.

Initially, each expert will be given a brief explanation on the goal of the session and how the outcome of the session contributes to the development Usability Risk Assessment Model. Then, the first round questionnaire will be distributed to the experts to collect information on likelihood of each usability risk for analysis. Since Delphi method is done in an iterative manner, the information from first round of survey will be fed back to the experts for comments and as a basis for the second round which is to determine the impact of usability risk to SDLC phases. This process is repeated in third round until common understandings between experts are achieved. All questionnaires in this survey is developed using five point Likert scale.

The advantage of this method is that all knowledge from experts is evaluated and commented by other experts, producing collectively agreed information on likelihood, impact, and classification and mitigation plan of usability risks on SDLC phases.

Then, the exposure level for each potential usability risk is calculated based on likelihood and impact of each usability risks. Each usability risk will be classified according to phases in SDLC and correlation analysis will be done to determine relationship between potential usability risk and phases in SDLC.

(iii) Usability Risk Prioritization

Prioritization of each potential usability risk is done by sorting respective risk exposure level in descending order to know the impact of usability risk in each phase in SDLC. Higher the risk exposure, more priority and attention should be given to that particular usability risk because it has more impact in creating less usable software product.

Prioritization is important because it gives insight of critical and noncritical usability risk based on SDLC phases, among usability risks and in overall software product development.

5 CONTRIBUTIONS

The proposed model the proposed model could help project managers, quality managers, risk management team and software development team in identifying and describing potential usability risk that impacts and effects SDLC phases and delivery of software development objectives. By eliminating these risks, customer's expectations on quality can be met while saving the costs and time.

Theoretically, this model contributes to the area of knowledge in Software Engineering Body of Knowledge (SWEBOK) and Project Management Body of Knowledge (PMBOK) as the concept of usability risk assessment has not been explored before and would contribute to development of more usable software products.

A list of important elements and activities in software risk assessment models is identified and used as a basis for development of Usability Risk Assessment Model. Gaps and lacking in current software risk assessment models are identified too and will be overcome in the proposed model.

Usability Risk Assessment Model also consists of two other sub contributions:

- List of potential usability risk which influences SDLC to increase understanding and awareness on existence of potential usability risks among development team.

- Exposure level of each usability risk which describes the importance to overcome usability risks.

Usability risk is also incorporated to SDLC phases to show the impact of usability risk during software development and this incorporation serves as guidance for development team to develop more usable software product.

6 CONCLUSION AND FUTURE RESEARCH

This paper proposes a concept in ensuring high usability in software products by using the concept of risk management. The idea of this concept is that if usability risks can be identified and managed well, the overall chances of reducing risk of failure and producing usable software product could be increased. Since studies on usability risk is still lacking, this will be a good approach for development team to be responsive towards potential usability risk that need be managed during software development process.

In future, suitable methods to implement each activities suggested in the proposed Usability Risk Assessment Model will be determine in detail. Current practices in developing a risk assessment model will also be investigated for this purpose.

7 REFERENCES

1. O. K. Hussain, T. S. Dillon, F. K. Hussain, and E. J. Chang, "Understanding Risk and Its Importance in Business Activities," in *Risk Assessment and Management in the Networked Economy*, Springer Berlin Heidelberg, 1-29, 2013.
2. A. K. Sharma, A. Kalia, and H. Singh, "An Analysis of Optimum Software Quality Factors," *IOSR Journal of Engineering*, vol *2*(4), 663-669, 2012.
3. J. Verner, J. Sampson, and N. Cerpa, "What factors lead to software project failure?,"in IEEE *Second International Conference on Research Challenges in Information Science, 2008, RCIS 2008,* 71-80, 2008.
4. R. Kaur and J. Sengupta, "Software Process Models and Analysis on Failure of Software Development Projects," *arXiv preprint arXiv:1306.1068,* 2013.
5. K. Schwaber, and J. Sutherland, *Software in 30 Days: How Agile Managers Beat the Odds, Delight Their Customers, and Leave Competitors in the Dust.* Wiley, 2012.
6. B. W. Boehm, J. R. Brown, H. Kaspar, M. Lipow, G. J. MacLeod, and M. J. Merrit, "*Characteristics of software quality,"* ,*1*, North-Holland Publishing Company, 1978.
7. *Software Engineering-Product Quality-Part 1: Quality Model*, ISO/IEC 9126 -1, 2001.
8. R. G. Dromey, "A model for software product quality," *IEEE Transactions on Software Engineering," 21*(2), 146-162, 1995.
9. J. A. McCall, P. K. Richards, and G. F. Walters, "*Factors in software quality,"* General Electric, National Technical Information Service, 1977.
10. I. Hussein, M.Mahmud, and A. W. Yeo, "HCI practices in Malaysia: A reflection of ICT professionals' perspective," *2010* IEEE *International Symposium in Information Technology (ITSim), 3*, 1549-1554, 2010.
11. A. W. Yeo, P. C. Chiu, T. Y. Lim, P. P. Tan, T. Lim, and I. Hussein, "Usability in Malaysia," In *Global Usability*, Springer London, 211-222, 2011.

12. S. K. Dubey, and A. Rana, "Analytical Roadmap to Usability Definitions and Decompositions," in *International Journal of Engineering Science and Technology*, *2(9)*, 4723-4729, 2010.
13. A. Inversini, L. Cantoni, and D. Bolchini, "Connecting Usages with Usability Analysis through the User Experience Risk Assessment Model: A Case Study in the Tourism Domain," In *Design, User Experience, and Usability. Theory, Methods, Tools and Practice,"* 283-293, Springer Berlin Heidelberg, 2011.
14. Modernization and Management Planning Unit (MAMPU), "Kajian Baseline Perkhidmatan E-Kaunter & Paperless Government," 2011.
15. W. A. M. Isa, M. R. Suhami, N. I. Safie, and S. S. Semsudin, "Assessing the usability and accessibility of Malaysia e-government website," *American Journal of Economics and Business Administration*,*3*(1),40-46, 2011.
16. T. Jokela, "Performance rather than capability problems. insights from assessments of usability engineering processes," in *Product Focused Software Process Improvement*,115-127, Springer Berlin Heidelberg, 2005.
17. A. Seffah, M. Donyaee, R. B. Kline, and H. K. Padda, "Usability measurement and metrics: A consolidated model," *Software Quality Journal*,*14*(2), 159-178, 2006.
18. G. Lindgaard, "Usability testing and system evaluation: A guide for designing useful computer systems," London: Chapman & Hall,. 221-246, 1994.
19. J. Heiskari, M. Kauppinen, M. Runonen, and T. Mannisto, "Bridging the gap between usability and requirements engineering," in *17th IEEE International Requirements Engineering Conference*, 303-308, 2009.
20. H. Fischer, "Integrating usability engineering in the software development lifecycle based on international standards," in *Proceedings of the 4th ACM SIGCHI symposium on Engineering Interactive Computing Systems*, ACM, 321-324, 2012.
21. P. Carlshamre, and M. Rantzer, "Business: Dissemination of Usability: Failure of a Success Story," *Interactions*, vol 8(1), 31-41, 2001.
22. A. Seffah, R. Djouab, and H. Antunes, "Comparing and reconciling usability-centered and use case-driven requirements engineering processes," in *Australian Computer Science Communications*, IEEE Computer Society, 23, 5, 132-139, 2001.
23. S. R. Humayoun, "Incorporating usability evaluation in software development environments, " *KI-Künstliche Intelligenz*, vol *26*(2), 197-200, 2012.
24. C. Lallemand, "Toward a closer integration of usability in software development: a study of usability inputs in a model-driven engineering process," in *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems*, ACM, 299-302, 2011.
25. Human-centred design processes for interactive systems, ISO 13407, 1999.
26. R. S. MBA and J. C. Arias, "Review of Risk Management Methods," *Volume 4-Number 1-January 2011-Semiannual Publication*, vol 4(1), 59, 2011.
27. Z. Jun-guang, and X. Zhen-chao, "Method study of software project risk management," in *International Conference on Computer Application and System Modeling (ICCASM)*, IEEE, 8, 8-9, October 2010.
28. T. Altom, "*Usability as risk management,"* Interactions, *14(2)*, 16-17, 2007.
29. A. B. Platt, "The usability risk," in *Proceedings of the 18th IEEE Symposium on Reliable Distributed Systems*, IEEE, 396-400, *Proceedings of the First Asia-Pacific Conference on Quality Software, IEEE*, 297-305,1999.
30. R. B. Grady, and D. L. Caswell, "Software metrics: establishing a company-wide program," 1987.
31. M. Simes-Marques and I. Nunes, "Usability of interfaces," 155-170, 2012.
32. M. Sangeetha, C. Arumugam, P.G. Sapna, and S. K. A KM, "Cooperative approach to ensure software product quality," *International Journal of Computer Science*, *8*. vol. 8 (4), 598, 2011.
33. S. U. Farooq and S. M. K. Quadri, "Quality Practices in Open Source Software Development Affecting Quality Dimensions," *Trends in Information Management*, *7*(2), 2012.
34. L. Hua, and Y. Gong, "Usability Evaluation of a Voluntary Patient Safety Reporting System: Understanding the Difference between Predicted and Observed Time Values by Retrospective Think-Aloud Protocols," in *Human-Computer Interaction. Applications and Services*, Springer Berlin Heidelberg, 94-100, 2013.
35. L. Liang, X. Deng, and Y. Wang, "Usability Measurement Using a Fuzzy Simulation Approach," in *International Conference on Computer Modeling and Simulation, 2009. ICCMS'09*, IEEE, 88-92, 2009.
36. S. Lee and J. E. Cho, "Usability Evaluation of Korean e-Government Portal," in *Universal Access in Human-Computer Interaction, Applications and Services*, Springer Berlin Heidelberg, 64-72, 2007.
37. M. Abdul Aziz, W. A. R. Wan Mohd Isa, and N. Nordin, "Assessing the accessibility and usability of Malaysia Higher Education Website," in *2010 International Conference on User Science and Engineering(i-USEr)*, IEEE, 203-208, 2010.
38. E. N. Asiimwe and N. Lim, "Usability of government websites in Uganda," *Electronic Journal of e-Government*, vol *8*(1), 1-12, 2010.
39. J. Sauro. (2010, September). *How common are usability problems* [Online]. Available: http://www.measuringusability.com/problem-frequency.php
40. P. Ketola, "Integrating usability with concurrent engineering in mobile phone development," Academic Dissertation, Department of Computer and Information Sciences, University of Tampere, Finland 2002.
41. C. Jones, "*Assessment and Control of Software Risk,"* Yourdon Press619, 1994.
42. B. Jin, S. Ko, J. Mun, and Y. G. Ji, "A study for usability risk level in physical user interface of mobile phone," in

Usability and Internationalization. HCI and Culture, Springer Berlin Heidelberg, 327-335, 2007.

43. C. M. Subhas, K. Vinod, and K. Uma, "Different techniques for risk management in software engineering: a review," ASAC 2006 Eric Sprott School of Business Carleton University Banff, Alberta, 2006.
44. B. W. Boehm, "Software risk management: principles and practices," *Software, IEEE*, *8*(1), 32-41, 1991.
45. S. W. Foo and A.Muruganantham, "Software risk assessment model," in *Proceedings of the 2000 IEEE International Conference on Management of Innovation and Technology, 2000. ICMIT*, IEEE,*.2*, 536-544, 2000.
46. D. Gupta, and M. Sadiq, "Software risk assessment and estimation model," in *International Conference on Computer Science and Information Technology, 2008. ICCSIT'08*, IEEE, 963-967, 2008.
47. B. Shahzad and A. S. Al-Mudimigh, "Risk Identification, Mitigation and Avoidance Model for Handling Software Risk," in *2010 Second International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN)*, IEEE, 191-196, 2010.
48. M. Sadiq, M. K. I. Rahmani, M. W. Ahmad and S. Jung, "Software risk assessment and evaluation process (SRAEP) using model based approach," in *2010 International Conference on Networking and Information Technology (ICNIT)*, IEEE, 171-177, 2010.
49. Isa, W.A.R.W.M., Suhami, M.R., Safie, N.I., & Semsudin, S.S. Assessing the Usability and Accessibility of Malaysia E-Government Website. *American Journal of Economics and Business Administration*, *3*(1), 40-46, 2011.
50. Lilja, Kari K, Laakso, Kimmo, & Palomki, J. *Using the Delphi method.* Paper presented at the 2011 Proceedings of PICMET'11: Technology Management in the Energy Smart World (PICMET), 2011.
51. Thangaratinam, Velayutham, & Selvambigai, Shakila.. *Health technology assessment in maternal and perinatal medicine: delphi survey of practice, systematic reviews of evidence and meta analyses.* University of Birmingham, 2011.
52. Schmidt, Roy, Lyytinen, Kalle, Keil, Mark, & Cule, Paul. Identifying software project risks: an international Delphi study. *Journal of management information systems*, *17*(4), 5-36, 2001.
53. Powell, Catherine. The Delphi technique: myths and realities. *Journal of advanced nursing*, *41*(4), 376-382, 2003.

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 103-113
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Critical Factors and Comparative Analysis that Influencing the Registration of Domain Name

Norul Bahriyah Shaari, Assoc. Prof. Dr. Harihodin Selamat
Advanced Informatics School
University of Technology Malaysia Kuala Lumpur,
Jalan Semarak, 54100 Kuala Lumpur, Malaysia
cutecahaya@gmail.com, harihodin@ic.utm.my

ABSTRACT

Growth and increasing use worldwide domain is inconsistent with growth of world population. Although in terms of domain usage statistics across the world increased, but the distribution of the domain name registration unevenly between countries around the world. There are a number of users of the gap between the domain names countries around the world. The technologies domain name system (DNS), marketing strategic, domain names services and other factors such as economy in the country, culture, environment, educational, geographical location, political, internet activity trends and so on, are also influenced the use and registration of the domain name. Malaysia is also one of the countries categorized as a developing country that is still minimal use of domain names compared to the rate of population, consumer internet and social media. This paper also presents the comparative analysis is the item-by-item comparison critical factors that influence the registration of domain name. This paper shows how perspectives can be used for comparative analysis, summarizes a soundness proof for critical factors that influence the registration of domain name..

KEYWORDS

Domain Names System (DNS), marketing strategic, critical factors, domain usage, statistics, comparative analysis

1 INTRODUCTION

The Domain Name System (DNS) is a critical component of the global Internet infrastructure. Throughout its history, its design and administration has experienced significant dynamic changes as the Internet itself has evolved. Future directions for DNS use and abuse are explored, along with challenges in its future governance. DNS evolution and its influences from political, legal, psychological, sociological, and technological perspectives. [1].

One of the keys to the tremendous success of the Internet is the Domain Name System (DNS). The DNS enables users to identify interesting websites in the vast realm of cyberspace by connecting domain names to specific Internet Protocol (IP) addresses. The DNS makes domain names distinguishable. Since its invention, domain names are no longer necessarily close substitutes. The ability of an individual domain name to distinguish the connected website from the millions of other available websites is the source of its potentially high economic value. [2]

During the month of April 2006, a little more than 35 million domain names were registered. Of these names, 32.7 million were used - most again and again - but never permanently registered. This paper discussed about the expectation factors will influence the growth of domain names. Compilation of information based on latest information from research paper, statistic, report, journal, conference paper, online news, computer magazine, website and blog. Research done by identifying factors with four main factors which is DNS technologies, domain names marketing strategies, domain names services and other factors.

2 LITERATURE REVIEW

DNS has always been a critical part of the Internet’s modern infrastructure which is one of

factors influencing the registration of domain name around the world. While it replaced the simpler system of HOSTS.TXT and may one day be replaced itself, there will always be a need for a directory service to translate complicated network identification into a readily human-readable format. As such, it is important that the development of DNS be monitored and studied. DNS is unique in the demands that are placed upon it whereas many other Internet protocols and systems are far more independent of government and cultural influences, DNS must represent these. With the increase in the proliferation of the Internet, the larger network population and stronger commercial influence also place demands on DNS. The influences that may affect DNS research and development are also in flux—while that of the international community is on the rise, the exclusive power of the United States government over its regulation and administration is waning, and technological limitations that would have at one time proven insurmountable will increasingly become tractable problems. [1]

Another critical factor influencing the registration of domain name around the world is internet governance. Currently, an important aspect of the Internet is governed by a private sector, international organization called the Internet Corporation for Assigned Names and Numbers (ICANN), which manages and oversees some of the critical technical underpinnings of the Internet such as the domain name system and Internet Protocol (IP) addressing. ICANN makes its policy decisions using a multi stakeholder model of governance, whereby a "bottom-up" collaborative process is open to all constituencies of Internet stakeholders. The debate over how the Internet's domain name system is governed may have a significant impact on future debates on how other Internet policy areas are governed on a worldwide basis. The ultimate success or failure of ICANN, and the multi stakeholder model of Internet governance it represents, could help determine how other Internet policy issues—such a Cyber security and privacy—are addressed.

Based on paper "This Factbook, published by CIRA provides an overview of the domain name industry. Researcher: published by Canadian Internet Registration Authority (CIRA) Year: 2012" reported the latest .ca domain name growth, trend, market and internet economy. Also give the result survey about .ca domain name which is the result showed the more than half of survey respondents say they pay attention to domain extensions when registering a name. Internet users by 2015, the majority of them will access the Internet via a mobile phone. In many countries, fixed line broadband is often either prohibitively expensive and/or unavailable; whereas mobile access is increasingly available and relatively inexpensive. For Canadian businesses to take advantage of the business opportunities offered by new online and mobile markets, they will have to shift the way they communicate with their customers by adopting mobile-ready web sites and applications. That the proven and based on the report the latest technologies is one of factor influencing the registration of domain name around the world.

DNS technologies the most critical factor to influencing the registration of domain name around the world. The high costs and inflexibilities associated with traditional, hardware-based DNS and IP addresses computing appliances, most organizations have implemented existing by utilizing unsupported open source software or generic Microsoft server production industry standard servers. Virtualization is an excellent way to streamline these inefficiencies. [3].

A DNS hosting service is a service that runs Domain Name System servers. DNS hosting service is optimal when the provider has multiple servers in various geographic locations that provide resilience and minimize latency for clients around the world. There are a number of utilities for administering, monitoring, and troubleshooting both DNS servers and clients. These utilities include: The DNS consoles; Command-line utilities which can be used to troubleshoot DNS problems; Logging features Performance monitoring utilities; Windows Management Instrumentation (WMI) and Platform Software Development Kit (SDK).

WHOIS (pronounced as the phrase who is) is a query and response protocol that is widely used for querying databases that store the registered users or assignees of an Internet resource, such as a domain name, an IP address block, or an autonomous system, but is also used for a wider range of other information. The protocol stores and delivers database content in a human-readable format. [4]

At least one major attempt will be made to create an alternative DNS, backed by a government or state. Additionally, at least one large-scale commercial venture will do the same. The government entity may succeed, but the commercial venture will fail unless it is also backed by a major government, if for nothing more than sheer lack of profit, unless it fulfils a specific niche market, such as some network built on the Internet for a special purpose such as high security. Alternatively, attempts by governments to control DNS and the Internet, such as copyright-related domain name seizures executed by the United States in 2011 [5] or attempts to seal off parts of the Internet in politically volatile regions [6], may lead to the adoption of a peer-to-peer based DNS.

Uneventful IPv6 DNS conversion will be relatively painless, as many, if not most, major DNS software packages support the IPv6 version alongside IPv4. As such, any well-maintained site with updated software may very well need only some slight reconfiguration to provide full IPv6 functionality. A far more pressing concern is in the actual deployment of IPv6; in short, DNS is the least of the problems that IPv6 adoption should be concerned with.

The DNS architecture will remain a standard if IPv4 to IPv6 conversion is difficult, converting from DNS to a completely new system will probably not be much better, and have far fewer short-term benefits that are visible to the end user. Any changes will have to be client transparent, as there is far too much software written with DNS in mind to make a switch over feasible except in the most extreme circumstances. IPv6 DNS is designed to address many of these issues, so if and when the conversion of the main Internet to IPv6 is activated, many flaws should, with luck, become irrelevant [7].

DNS will never be perfect. Almost all systems have flaws. Even if the oft-cited IPv6 version of DNS corrects all the major structural flaws in the IPv4-based DNS, it remains under the radar and relatively new, whereas IPv4 DNS is ubiquitous and has been around for over two decades for analysis and dissection by would-be attackers. Furthermore, national governments will always squabble, as will agencies that govern systems like the DNS, whether they are government backed, corporate backed, independent, or otherwise. Ultimately, there will always be problems with the DNS, even if we fix all of those that are in existence; it is, in the end, a never-ending cycle, which, with luck, will continue to induce a net strengthening of the system as a whole.

Another critical factor to influencing the registration of domain name around the world is domain names marketing strategies*:* The strategies to promote and increase the domain name registration. There is no scientific method to determine a precise value for any domain name, there are some considerations that go into determining a reasonable ballpark value for that domain name you want. Normally, the domain name price based on new registration domain name and normally it's about one year period.

The bundled discount (such as buy-one-get-one-free), volume discount, contract year discount, special discount to registrar with a special status, fixed price discount, domain category discount, seasonal discount and etc. Domain auction lists often include names that have been previously unavailable at a wide range of price points. Good domain names are becoming harder to find. With over 30 million com, net and org domains registered, most of the best domain names have been taken. The expired domain names are names that were previously registered, but where the registration has not been renewed, or where the registrant has defaulted on payment. Also the alternative domains, strictly from a user

standpoint, improperly thinking out how each product should be categorized can cause many products from being found by the shoppers. They have to consider carefully any variations that someone else might use, including: Alternate spellings; Misspellings; Abbreviated /Long-form versions; phonetically similar versions; Plural / singular versions. Premium Resale Domains are priced higher than unregistered domain names based on a variety of criteria including the number of characters in the domain, the number of years the domain has been registered, relevancy and popularity of the keyword, and the traffic it generates.

Domain names services will be given the good impact the domain name registration. The Mobile Web refers access to the World Wide Web, i.e. The use of browser-based Internet services from a handheld mobile device such as a Smartphone, a feature phone or a tablet computer, connected to a mobile network or other wireless network. E-commerce (e-Commerce Website & Merchant Account), the Internet has created a new economic ecosystem, the e-commerce marketplace, and it has become the virtual main street of the world. Providing a quick and convenient way of exchanging goods and services both regionally and globally, e-commerce has boomed. Also the email hosting services usually offer premium email at a cost as opposed to an advertisement-supported free email or free webmail. All above factors will give the positive impact to registration domain name around the world.

Affiliate marketing is a type of performance-based marketing in which a business rewards one or more affiliates for each visitor or customer brought about by the affiliate's own marketing efforts. Domain Names Search Engine: 90% of local shoppers search online first for nearby businesses. Ongoing submissions to top search engines including Google, Yahoo Bing and etc. A social networking service is a platform to build social networks or social relations among people who, for example, share interests, activities, backgrounds, or real-life connections which are given positive impact of domain name registration. Same with the .XXX Domain Names: This sTLD (sponsored top-level domain) is designed specifically for the global adult entertainment industry as a trusted brand, globally recognized and extolling responsible and safe behaviour.

Others factors to influencing the registration of domain name around the world is growth of population**.** According to OECD/World Bank population statistics, the world population grew by 27%, or 1.423 billion people, between 1990 and 2008. Many of the world's countries, including many in Sub-Saharan Africa, the Middle East, South Asia and South East Asia, have seen a sharp rise in population since the end of the Cold War. Based on report cities in UK domain names and how the population make increase the total domain name registration. The report informs "In 2001 the 50 cities of England were used in over 22,000 domain names. This averaged 441 aces per city.

The story of the commercial Internet in Malaysia began in 1990, when the Malaysian Institute of Microelectronic Systems (MIMOS, now MIMOS Berhad) launched JARING (Joint Advanced Integrated Networking), the first Malaysian ISP. The Malaysian government has been an enthusiastic supporter of Internet technology since the early nineties.

To achieve an informed and computer-literate society, IT and Internet access has two widespread. The online population's example in Malaysia is "infants" and it must be admitted that the level of Internet knowledge is better compared with other developed countries. The e-commerce and e-business scene in Malaysia and around the Asian region is beginning to blossom. E-commerce technology coming to the market and the growing number of internet users buying through the net stimulate the opportunity to expand the marketplace by deploying a cost effective and efficient solution. Malaysia enterprises have to re-act fast to the rapidly changing environment. [8]. The business environment is culturally different in the country. The convenience of time and spatial are not reasons for Malaysian consumers to buy from

retail shops instead of shopping online. A growing number of shopping complexes are opening in Malaysia. The largest mall in Malaysia - Mid Valley Megamall has more than 430 shops in five and a half floors. This might be the major competitor to the Internet portal sites. [4].

The main obstacle that prevents Internet users from transacting over the Internet is security. This issue is not only the major concern among Malaysian consumers but of users worldwide. The Internet is still a patch work of private security solutions for parties who already know each other. [8]. The potential of international domains is also enormous. The benefits of international domains are larger audience, demand localization, brand protection and availability.

The economic contribution of .my is the level of total value added to Malaysia's gross domestic product in the industry. This provides a measure of the industry's relative importance to the overall economy. One key reason for domain name registrations is the establishment of a new business, which is less likely to occur during times of economic uncertainty. Based on the Economic and statistical analysis of the .au domain range by .au Domain Administration Ltd and AusRegistry Pty Ltd, December 2011, The economics contribution of .au domain names which is the level of total value added to Australia's gross domestic product by the industry. This provides a measure of the industry's relative important to the overall economy. In this case, the economic contribution estimates measures the activity generated through assisting Australian businesses and individual in getting online, rather than the economic activity generated through their online presence. [9].

Future trends in domain registration also give the impact of registration domain name. In 2005, ICANN's Generic Names Supporting Organization (GNSO) began a policy development process to consider the introduction of new generic Top-Level Domains (gTLDs), based on the results of trial rounds conducted in 2000 and 2003. In 2008, the ICANN Board adopted a raft of specific GNSO policy recommendations for implementing new gTLDs. Amongst the key considerations for the introduction of new top-level domains identified were: Expanding the domain name space to accommodate the introduction of both new ASCII and internationalized domain name (IDN) top-level domains will give end users more choice about the nature of their presence on the Internet. In addition, users will be able to use domain names in their language of choice. Demand for additional top-level domains as a business opportunity. The GNSO Committee expected that this business opportunity will stimulate competition at the registry service level. [10]. Same with future trends in internet activity: E-commerce refers to consumers purchasing goods and services from online stores rather than in a traditional physical store; The increased use of m-commerce (purchasing using mobile phones); The role of search engines - Search engines are increasingly used by consumers as the starting point for research in a purchasing decision [10]; Social networking - Social media like Facebook, Twitter and etc. are becoming an intrinsic part of modern social life. [10]

The geographical distribution of registrants closely resembles the distribution of the Malaysian population as a whole, with large of activity focused around Kuala Lumpur, Johor, Perak, Melaka, Kuching and etc. Registered .my domain names, by registrant remoteness will categorize: Major Cities of Malaysia; Inner Regional Malaysia; Outer Regional Malaysia; Remote Malaysia; Very Remote Malaysia.

The business environment is culturally different in country. In the United States, stores are not generally a block away from houses as they are in Malaysia. Not only are retail shops nearby, but also shops' and restaurants' business hours extended through a distinctively longer period compared to those found in western countries. As such, the convenience of time and spatial are not reasons for Malaysian consumers to buy from retail shops instead of shopping online. Buying a burger from a shop located two doors away is easier than ordering through the Internet, so why should consumers choose to buy online?

Consumer attitudes and behaviour have a big influence on making the decision to buy from the Internet. In Malaysia, the majority of buyers like to do brick-and mortar shopping. One of the shopping malls located in Kuala Lumpur, Suria Mall has 450,000 visitors per week. To visit a shopping mall has become a kind of Malaysian "weekend activity"; it is a method for some Malaysians to release pressure or spend time with family and friends. They may not necessarily go there to buy anything; it could be for various services that are available in shopping complexes like salons, movies, and bowling alleys. A growing number of shopping complexes are opening in Malaysia. The largest mall in Malaysia - Mid Valley Megamall has more than 430 shops on five and half floors. This might be the major competitor to the Internet portal sites. [4]

The possibility exists that other countries or organizations may start their own domain name registries for their own purposes. This is not the limit for potential political interference in domain name registration. Moves to censor the Internet in western countries such as Italy [11], Australia [12], and the United States [13] join other well-known censorship initiatives in other countries such as China [14]. That the governments of these influential countries seem to be pushing for such movements in their own sphere of influence makes it quite possible that they may move their interests abroad and attempt to exert pressure on ICANN to modify their policy to better fit their demands. ICANN is under the employ of the U.S. Department of Commerce as a contractor and is a private organization with non-profit status dedicated to maintaining the coordination of aspects of the Internet such as the DNS [15]. These services are vital, but ultimately ICANN's authority is derived by the mutual consent of the Internet community. That authority theoretically could be revoked at any time, and in many cases it would take only a relatively limited amount of legislation to entirely deprive ICANN of power in a country, and possibly many countries. In order to prevent the emergence of alternative domain registrars backed by the resources of a large country, ICANN may need to at least partially acquiesce to such interests.

Third party registrar corruption will reach critical levels: Third party registrars do not seem to have the same spirit of community that helped to build Internet from scratch. Although it is arguable that organizations like ICANN are no longer in possession of this quality, it is more likely that they at least retain some of the cultural mindset within the organization, not to mention some of the veterans; as such, many of the lower-level, third party registrars will continue to attempt to extract as much profit from their position as possible, even at the possible long-term detriment of the Internet at large. Eventually there will likely be some critical turning point that leads to heavy re-evaluation of the entire system.

Increasing governmental influence, which is as of 2011, legislative action in several countries has indicated that DNS may encounter influence by governments as a method of filtering out undesirable Internet sites, as a result of pressure from both political and corporate forces. This may create considerable problems for its continued acceptance as a standard, as it is likely that the marketplace will gain support for a replacement resistant to external changes, regardless of its legality [16].

Such competition may place considerable strain on the primary implementation of DNS to remain relevant volume 30 Article 21 339 and address the needs of many on the Internet, though it may be questionable how much support such a shift in naming technology could actually gather if it should retain a reputation as contraband or be challenging for a user to install and utilize. It may also endure the abuses DNS already struggles with, as well as additional, unanticipated abuses that may accompany any new technology used in potential replacements. Although highly unlikely, a worst case scenario may result in a period of considerable ambiguity if no single DNS implementation maintains universal global acceptance.

The Uniform Domain Name Dispute Resolution (UDRP) will change significantly and has worked so far, but not without problems.

Some shortcomings of the procedure include requisite "bad faith" is ill-defined; complaining parties (often trademark holders) seem to have bias in their favour; the UDRP is not legal arbitration nor binding, allowing litigious intervention; parties such as large, corporate interests can more easily afford associated costs; and English dominates the process [17]. The UDRP has existed for over a decade [18], providing sufficient experience to learn where it needs improvement [17]. Given the increase in corporate influence on the Internet, as well as public awareness, it is likely that there may be a struggle, with corporate interests gaining the upper hand and possible changes due to backlash; however, given the legal position of the UDRP, it may ultimately end up a supplement to the court system as opposed to an attempt at manifesting a final authority as originally intended.

Summarize from the explanation above, refer at Table 1: List of the critical factors that influencing the registration of domain name..

TABLE 1: LIST OF THE CRITICAL FACTORS THAT INFLUENCING THE REGISTRATION OF DOMAIN NAME

Factor	Details Factor
DNS Technologies	• Domain Name Processes • Domain Name Life Cycle • Hardware & Network Environment • Registry & Registrar System • DNS Hosting • Manage & Monitoring DNS Entry • Database • Domain Names System Security (DNSSEC) • Internet Protocol Version 4 (IPv4) • Internet Protocol Version 6 (IPv6) • WHOIS • Internationalized Domain Name (IDN) • Internationalized Domain Names in Applications (IDNA) • ANYCAST • Domain Availability Checker • Electronic Numbering (ENUM) • The Extensible Provisioning Protocol (EPP) • Next-generation network (NGN) • Domain Name Dispute Resolution Policy (DRP) • Secure Domain Names • DNS Fragmentation
	• Uneventful IPv6 DNS Conversion • The DNS Architecture Will Remain a Standard • DNS Will Never Be Perfect
Domain Names Marketing Strategies	• Price • Promotion • Bid Domain Name • Partner with Social Network • Expired Domain Name • Alternative Domains • Premium Resale Domain Names • Public Registration • Private Registration • Global Registration and Global Partner
Domain Names Services	• Mobile Web site • E-commerce • Email Hosting • Affiliate Program • Design Services • Domain Names Search Engine • .XXX Domain Names • Social Network Services • Training Services • Branding Domain Names Promotion • Business Partnership of Domain Name Business
Other Factors	• Growth of Population • DNS Education • Education • Internet Governance • Government Support • IT Knowledge • Attitude Problems • Internet Security Issues • The Shopping Culture • International Domain • Native Language • Economic Contribution • Global Economy • Future Trends in Domain Registration • Future trends in internet activity • Geographic Location • Internet Connection, Usage & Value • Active Domain Names at Entry DNS • Green Technology • The Culture • Popularities Registry & Registrar • Mobile Technologies • Regulator Rules & Procedure • Management of domain names • Political • Third Party Registrar Corruption Will Reach Critical Levels • Increasing Governmental Influence • UDRP Will Change Significantly

3 COMPARATIVE ANALYSIS

Basic statistical methods is the method used in the process. organize, analyse and draw conclusions about the factors that affect the increase in domain name registration as well as the Malaysia and global. These is the quantitative because the data collected is not properly structured. Data collection will be factors in a random order and it is raw data. Thus the foundation will organize statistical methods in the order form and make it more meaningful calculations to interpret the data accurately. Measure of central tendency measurement focus with "Mean" which showed that the average increase in the factors that influence the registration of domain name. This method is a most convenient way to describe the overall performance for the purpose of inter-country performance. This method is also useful for calculating the deviation of the effect of the factors that influence registration of domain name.

A. Quantitative Data Collection

Data collection of critical factors that Influencing the registration of domain name is randomly generated. A comparison is made between several countries from selected ccTLD list. Compilation of information based on latest information from the research paper, statistical reports, computer reports, journals, conference paper, online news, computer magazine, website, blog and etc.

B. Mean, Median, Mode (Dimensions Central Tendency)

This refers to an average value of a set of scores of critical factors that influencing registration of domain name. There are three types of measures of central tendency of critical factors that influencing registration of domain name, the mean, median and mode.

- Mean shows the average score values of critical factors. It also reflects the overall performance for the purpose of performance comparisons between the critical factors or group critical factors that influencing registration of domain name. Min is also useful when calculating the standard deviation and the deviation score of critical factors;

- Mode is the score of critical factors with the highest frequency in a distribution. Symbol statistics Mod yield M0;

- The median is the middle score of critical factors in a distribution of the ascending or descending order. The median can be calculated by dividing a set of scores of critical factors organized into two equal parts.

The calculation of the mean is important because it facilitates our statistical formula to calculate the standard deviation, Z scores and scores of T critical factors. For non-accumulating data, the mean is calculated by summing all the scores of critical factors in the set, then divide this amount by the number of scores of critical factors. For example, the mean scores of 70, 85 and 100 in the test are:

$$(70 +85 +100)/3 = 85$$

Mean data calculated using the following formula:

$$Mean, \ddot{X} = \Sigma X/N$$

Note: $\ddot{X}$=Mean, Σ=Total, X=Score and N = Total Score

C. Comparative Analysis the Registration of Domain Name

Mean widely used because it takes into account all scores in the distribution and it is very accurate. The basic comparative analysis for critical factors that influence registration of domain name use based on mean calculation and the analysis ranking of registration domain name around the world. The comparative analysis based on listed below:

- Number of Global Domain Names: By the midpoint of 2011, there were slightly more than 214 million domain names registered globally.

- Domain Names Growth Rates: Generic TLD's, which had enjoyed growth rates well above 20 percent annually before the economic downturn hit, experienced a continuing slowdown in growth throughout the downturn, bottoming out at about eight per cent through 2009 before recovering slightly to 10 per cent in 2010, a pace that was maintained through the first half of 2011.

- ccTLD Top Ranking: The world's most popular ccTLDs have enjoyed robust growth rates over the past five years although, in almost every instance, those growth rates have tapered off over the past year or so.

- Domain Names Trend: One of the keys to buying valuable domain names is to acquire names that are meaningful and would be desired by others. Owning a name that is either the generic category killer for a popular trend or a name that represents part of that category is usually a good buy.

- Standard Factors Domain Name: The standard factors are number of domain name, gain, lost, net and rank of domain name around the world.

- Critical Factors that Influencing the Registration of Domain Name: The analysis based on mean calculation of critical factors that influencing the registration of domain name.

- Life Cycle Domain Name: The comparative analysss based on mean calculation of life cycle domain name which is the life cycle domain name include 7 phases as listed below:

    - Available: A domain names are available to the public;

    - Active: Once the domain name is registered it comes in active state and can be renewed for a period not to exceed a maximum of years;

    - Expired: If the domain is not renewed on or before its expiry date, the domain name is expired;

    - Redemption Period: Redemption period generally lasts for 30 days during which the domain can still be renewed by the original owner normal registration fees;

    - Pending Delete: If the domain is still not renewed by the owner during the redemption period then domain name goes into pending delete;

    - Deleted and Active: The domain name is deleted and available to anyone for registration.

4 METHODOLOGY

A field research will use mixed method research strategies. The strategy begins with a quantitative method and follows it up with qualitative methods. The first research strategies will use as below steps which is focus on literature review approach as steps below:

- Phase 1: Searching information of the critical influencing the registration of global domain name

- Phase 2: Gathering, reading and analyzing the literature from Phase 1 with collection of information and capture the critical factors of the growth global domain name

- Phase 3: Get results from Phase 2 can be categorized into two, namely:

a) Comparative Analysis;

b) The details explanation of standard deviation critical factors;

- Phase 4: Collection data based on interviews using a structured questionnaire and structured questionnaires in which data is collected via a postal survey. Pilot study need to start with get the data sampling, cluster sampling. data reduction, content validity index (CVI), instrument development process and etc.

- Phase 5: Get the suitable algorithm from Calculation of statistics standard deviation critical factors and get the result the most critical factors that influence the registration of domain name

4 CONCLUSIONS

The result after Phase 1 & Phase 2, a research model is proposed to guide future study of critical factors to influencing of domain name registration. The model illustrated in Figure 1, includes the critical factors will impact of influences the domain name registration in DNS technologies such as hardware and network environment, secure domain name, DNS fragmentation, DNS architecture and etc. This paper not discussed all existing DNS technologies such as domain name process, life cycle, registry and registrar system, EPP, DNSSEC, ENUM, IDN, ANYCAST, NGN, DRPIPv4, IPv6 and etc. because all the technologies are direct impact to influencing registration of domain name. Figure 1 shown the conceptual model for encapsulates the impact of domain names marketing strategies, domain names services and other factors that influence the registration of domain name.

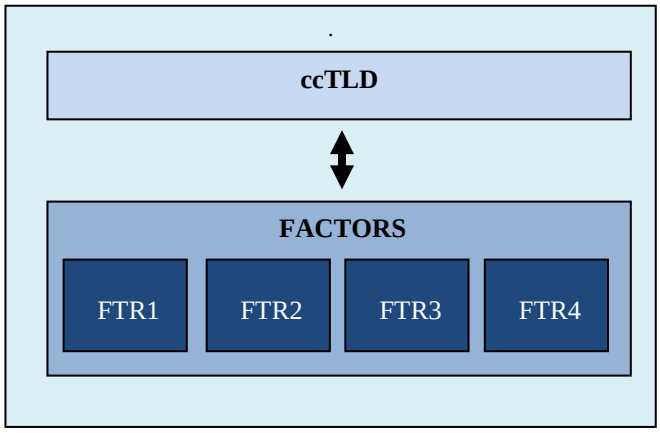


Figure 1. The conceptual model for encapsulates the factors that influencing the registration of domain name.

Note:

ccTLD: Country Code Top Level Domain

FACTORS: Factors to influencing the registration of domain name around the world.

FTR1: Factors of DNS Technologies
FTR2: Factors of Domain Name Marketing Strategies
FTR3: Factors of Domain Names Services
FTR4: Other Factors

The comparative analysis activities very important for get the initial result for this research. The comparative analysis use simple mean calculation with several types as below:

- Number of Global Domain Names
- Domain Names Growth Rates
- ccTLD Top Ranking
- Domain Names Trend
- Standard Factors Domain Name Critical Factors that Influencing the Registration of Domain
- Life Cycle Domain Name

5 REFERENCES

[1] The Domain Name System – Past, Present and Future – Michael Brian Pope, Merrill Warkentin, Leigh A.. Muthhler, Xin(Robert) Luo – Volume 30, Article 21, pp 329-346, May

[2] Sebastian Baum, 2004, Domain Name Conflicts in Germany – An Economic Analysis of the Federal High Daigle, 2004, WHOIS Protocol Specification RFC 3912.

[3] Nixu Software, Virtualized Domain Name System and IP Addressing Environments White Paper, September 2010

[4] RFC 3912, WHOIS Protocol Specification, L. Daigle (September 2004).

[5] Kravets, D. (2011) "U.S. Faces Legal Challenge to Internet-Domain Seizures," Wired, June 13, http://www.wired .com/threatlevel/2011/06/domain-seizure-challenge/ (current June 23, 2011).

[6] Greenemeier, L. (2011) "How Was Egypt's Internet Access Shut Off?" Scientific American, Jan. 28, http://www .scientificamerican.com/article.cfm?id=egypt-internet-mubarak (current June 20, 2011).

[7] Carli, F. (2003) "Security Issues with DNS," SANS Institute InfoSec Reading Room, http://www.sans.org/reading_ room/whitepapers/dns/security_issues_with_dns_1069?show=1069.php&cat=dns (current Jan. 8, 2009).

[8] Drivers and Impediments to E-Commerce in Malaysia, John Paynter and Jackie Lim, Department of Management Systems and Information Systems, University of Auckland Private Bag 92019 Auckland, New Zealand

[9] Economic and statistical analysis of the .au domain range by .au Domain Administration Ltd and tAusRegistry Pty Ltd, december 2011.

[10] Economic and statistical analysis of the .au domain range .au Domain Administration Ltd and AusRegistry Pty Ltd December 2011.

[11] Warner, B. (2007) "A Geriatric Assault on Italy's Bloggers," Times Online ,http://technology.timesonline.co.uk/tol/ newstech_and_web/the_web/article2732802.ece (current Jan. 8, 2009).

[12] Bryant, N. (2008) "Australia Trials National Net Filters," BBC News, http://news.bbc.co.uk/2/hi/technology/7689964.stm (current Mar. 30, 2009).

[13] Bambauer, D. (2011) "De-lousing E-PARASITE," Info/Law, http://blogs.law.harvard.edu/infolaw/2011/11/05/delousing-e-parasite/ (current Nov. 6, 2011).

[14] Zittrain, J., and B. Edelman (2003) "Internet Filtering in China," IEEE Internet Computing (7)2, pp. 70–77

[15] Fuller, K.E. (2001) "ICANN: The Debate over Governing the Internet," Duke Law & Tech Review, http://www.law.duke.edu/journals/dltr/articles/2001dltr0002.html (current Mar. 30, 2009).

[16] Bambauer, D.E. (2011) "Orwell's Armchair," University of Chicago Law Review, forthcoming, http://ssrn.com/ abstract=1926415 (current Nov. 6, 2011).

[17] Diéguez, J.P.C. (2008) "An Analysis of the UDRP Experience: Is it Time for Reform?" Computer Law and Security Review (24)10, pp. 349–359.

[18] InterNIC (2002b) "InterNIC FAQs on the Uniform Domain Name Dispute Resolution Policy (UDRP), InterNIC, http://www.internic.net/faqs/udrp.html (current Sept. 1, 2009).

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 114-123
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Lightweight Priority Scheduling Scheme for Smart Home and Ambient Assisted Living System

Hamdan Sayuti, Rozeha A.Rashid*, Mu'azzah A. Latiff, A. H. F. A. Hamid, N. Fisal, M. A. Sarijari, Alias Mohd, Kamaludin M. Yusof, Rozaini Abd Rahim
Universiti Teknologi Malaysia
81310 Johor Baharu, Johor, MALAYSIA
hamdan@fkegraduate.utm.my, {rozeha, muazzah, sheila, adib_sairi, alias, kamalmy}@fke.utm.my, {hadifikri, kapten06}@gmail.com
*corresponding author

ABSTRACT

Smart Home and Ambient Assisted Living (SHAAL) systems utilize advanced and ubiquitous technologies including sensors and other devices that are integrated in the residential infrastructure or wearable, to capture data describing activities of daily living and health related events. However, with the introduction of these technology-orientated services come a number of challenges, which to date are still largely unsolved. The management and processing of the large quantities of data generated from multiple sensors is recognized as one of the most significant challenges. Therefore, a simple yet efficient data scheduling scheme is proposed in this paper to manage incoming data packet from the system based on their application types and priorities. The performances of this lightweight context-aware scheme are investigated in a real SHAAL setting under two scenarios; centralized and distributed set-ups. The experimental results show the proposed scheme offers a promising solution for guaranteeing higher throughput to the high priority data while giving sufficient access to low priority data without introducing much delay impact.

KEYWORDS

Smart Home; Ambient Assisted Living; Wireless Sensor Network; Priority Scheduling

1 INTRODUCTION

The emerging demographic change towards an ageing population introduces dramatic need to improve the health and quality of life of the elderly and provide assisted living, aided by technology while maintaining a high degree of autonomy and dignity. Pilot studies have shown that facilitating elderly people with their essential demands for independent life and sovereignty, added together with self-care and self-management technologies can enhance their health outcomes. Information and Communication Technologies (ICTs) can play a major role in order to help achieve the above goals.

A smart home is not only a modern home of the future, but rather an integration of various devices at home to a network that controls all of the functions at once. The Malaysian agency known as IDC Energy Insights and Real Estate and Housing Developers Association Malaysia (REHDA) has reported that by 2015, the worldwide market share for smart home would reach a staggering 10.2 billion USD while in Malaysia, a steady 2.22% growth rate in real estate sector is a promising figure for SHAAL to take off [13]. From their survey, in average, over 40% candidates have a positive acceptance to the Smart Home idea.

A Smart Home and Ambient Assisted Living (SHAAL) system is a residential setting equipped with a set of advanced electronics, sensors and automated devices specifically designed for care delivery, remote monitoring, early detection of problems or emergency cases and promotion of residential safety and automated living [1], [2], [3], [4], [5], [6]. It typically makes use of Wireless Sensor Network (WSN) technology in order to monitor physical or environmental conditions [2]. The parameters include temperature, humidity, sound, vibration, pressure, motion, pollutants, images and even biometrics parameters such as ECG, blood pressure, SpO2 and etc. These sensed data would be locally processed before being transmitted to a remote location for further analysis and interpreted before any decision making takes place. In general, SHAAL system can provide

care for out-patient, elderly, disabled people or individuals needing assistance with activities of daily living and wishing for independent living as well as secure, comfortable and convenient living environment [3], [4].

Surveys on home automated health-care system projects and frameworks [7], [8], [9] illustrate that the sensor commutations for identifying the environmental data has attained a considerable maturity, but still there exists a gap related to data resources management. Moreover, the priority for each data differs, i.e., the health and security data are considered more significant than others. Thus, the data management is crucial when different packets of data reach the base node at the same time [10]. A robust and reliable framework is needed for effective data scheduling based on the set priority of different applications. Authors in [12] have demonstrated the scheduling of sensor data to provide an insight into the use of sensor based system and improving the level of automated support within a system context. However, the management of data is based on module synchronization. In this paper, we propose a lightweight priority scheduling scheme to manage the sensor data based on application types and their respective priorities. The performances of this proposed context-aware scheme are analyzed in a real SHAAL setting. The findings show the delay-throughput tradeoff using this scheme is within acceptable limit in addition to its feasibility for practical deployment.

The rest of the paper is organized as follows. Section 2 describes the general system model for SHAAL. The concept of priority scheduling is introduced in Section 3 while Section 4 presents the performances of the proposed lightweight data scheduling scheme in SHAAL system in terms of throughput, delay and system architecture. The conclusion of this paper is outlined in Section 5.

2 SMART HOME AND AMBIENT ASSISTED LIVING (SHAAL) SYSTEM

The proposed SHAAL system is based on Machine-to-Machine (M2M) communication and also cloud computing in a cloud network. M2M communication is proposed to enable electronic devices (smart sensors) to communicate and share the data via wireless networks. Within the deployed WSN as shown in Figure 1 and home area network (HAN), M2M communication for heterogeneous devices normally adopts IEEE 802.15.4 and IEEE 802.11 radio technologies while Wide Area Network (WAN) implements cellular based (GSM, 3G) or internet based cloud communication technology [5]. Base controller/station serves as data center where the sensed or collected data can be visualized and analyzed. The base station also acts as a gateway to allow remote access to the system. This would allow web-enabled devices to access the system through Internet while GSM network allows access using cellphone. The information can be relayed to home owners and authorized personnel through current gadgets such smart phone, laptop, tablet and others.

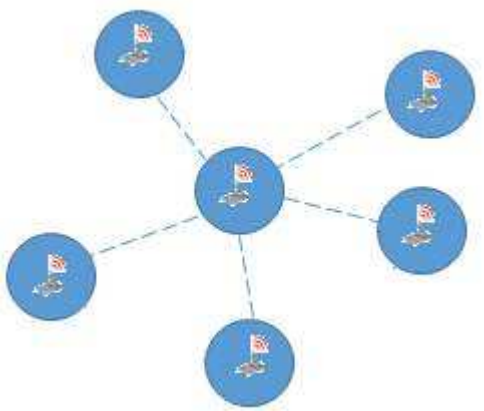


Figure 1 WSN deployment using star topology

While the Smart Home system described in [6] deals with energy management system, the SHAAL platform at the UTM-MIMOS Center of Excellence in Telecommunication Technology, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, concentrates on smart home applications such as automated control of electronic appliances and security solution which include door access system and web-based log-in record. Figure 2 shows a few of the devices used in our smart home module.

Component	Specification
Light	Bulk Lamp
Appliances	Socket 3 Pin
RFID Door	RFID reader, magnetic lock
Alarm	Buzzer
Camera	C328R CMOS Camera
WSN Mote	TelG(Atmega644PV, WiseOS)

Figure 2 Smart Home devices and respective specifications

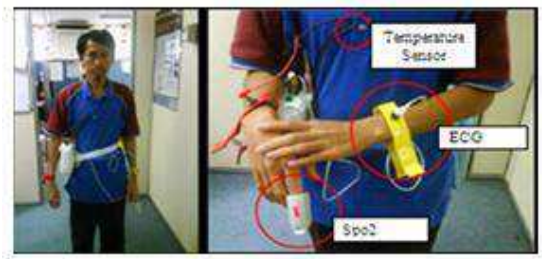


Figure 3 Ambient Assisted Living contraptions

In addition, the AAL framework is integrated into the existing platform by utilizing wearable body sensors as illustrated in Figure 3. The communication within this wireless sensor based SHAAL system is facilitated by self-designed WSN motes called TelG motes; each embedded with self-developed WiseOS as the operating system. The 5.5 cm x 3.5 cm mote, as presented in Figure 4, uses IEEE802.15.4 technology and operates in the ISM band of 2.4 GHz. It is composed of low power components in term of processing, communication and sensing. It is also designed to have small footprint for easy deployment, cheap maintenance and ubiquitous functionality.

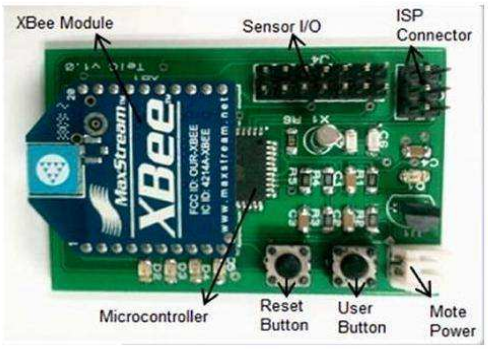
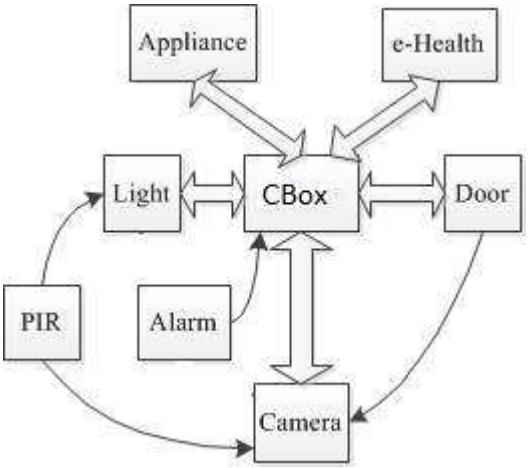
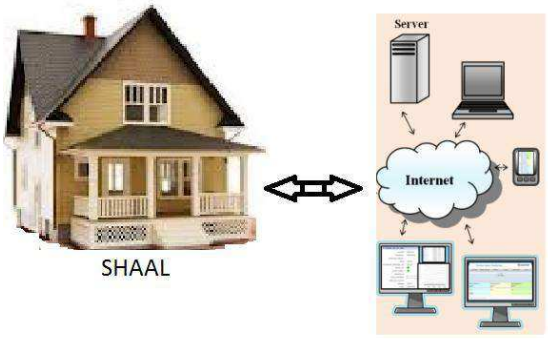


Figure 4 WSN based TelG mote

The sensor motes automate the process of collecting data which will be transmitted to the home main controller (i.e., the CBox) and consequently processed and stored in a server. Vital physiological data such as blood pressure, electrocardiogram (ECG) and temperature, collected from respective sensors worn by the person with related health problems, will be delivered to the remote authorized medical personnel or center via internet gateway for monitoring of health status and storage, respectively. TelG mote and WiseOS operating system is designed to be "ready out-of-box" for rapid application development. This MCU + RFIC design enables wide arrays of sensors to be fitted and a whole range of applications can be deployed. Although this generic design of packet oriented communication between radio and processor has several shortcomings such as limited low-level access of the radio parameters as pointed in [17], for SHAAL system, this generic mote provides adequate functionalities to support the application requirements. Figures 5(a) and 5(b) summarize the overall system functionalities.



(a)



(b)

Figure 5 (a) SHAAL functional state diagram (b) SHAAL System

This paper focuses on the complexities of managing different sources of sensor-based information with the overall goal of improving the levels of automated support. The scheduling

for resource allocation will be based on the required Quality of Service (QoS) for each application. In this work, high priority will be assigned to AAL application, followed by Smart Home services such as security, control and monitoring of electrical/electronic devices.

3 PRIORITY SCHEDULING SCHEME

Priority scheduling is an abstract data type, normally in the form of queue or stack data structure, in which each element has an additional "priority" associated with it. Priority scheduling is used to determine which processes need to be implemented first. Processing is conducted in accordance with the priority given; high priority is served before lower priority element. If two elements have the same priority, they are served according to their order in the queue. It is widely used in CPU processor and the running of devices in our SHAAL system employs similar mechanism, where priority scheduling is required to optimize the performances. Data processing nodes that act as sinks can implement scheduling algorithm to guarantee applications' QoS requirements. For example, the scheduling algorithm can be implemented in middle tier network in K-HAS as proposed in [15]. The implementation of the scheduling algorithm can also be extended into routing protocol mechanism. In the case of flat network architecture, an efficient scheduling algorithm in routing protocol is required when the data collected has different priority. In the literature, most of the routing protocol only concern with energy efficiency and scalability such as chain forming technique using Ant Colony Optimization (ACO) in [16] for data gathering. With recent developments in Cognitive Wireless Sensor Network (CWSN), context aware routing protocol has become a standard requirement and priority scheduling can be used to achieve this objective.

There are two types of priority packet scheduling, non-preemptive and preemptive [11], [12].

A non-preemptive packet scheduling scheme is based on the deadline of arrival of data packets to the base station (BS). The packet is classified as First Come First Serve (FCFS), of which the scheduler processes data in the order of their arrival times at the queue, i.e. a round-robin. In contrast, preemptive scheduling may interrupt a running task for some time and the task will be resumed once the priority task completes its execution. The preemptive may apply FCFS if two or more packets arrive with same priority. It means that in priority based scheduling, if there are two packets with the same priority in the queue, the first packet in queue will be processed followed by the next packet in queue.

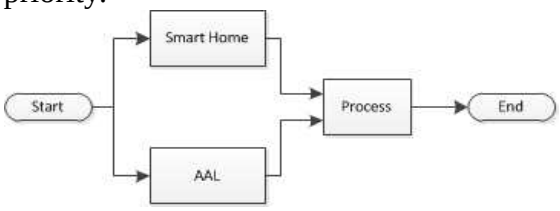
4 PROPOSED LIGHTWEIGHT SCHEDULING SCHEME FOR SHAAL

In this section, the proposed priority packet scheduling scheme to manage data packets in our SHAAL system will be further explained. A case study is presented for the proof of concept of the proposed scheme.

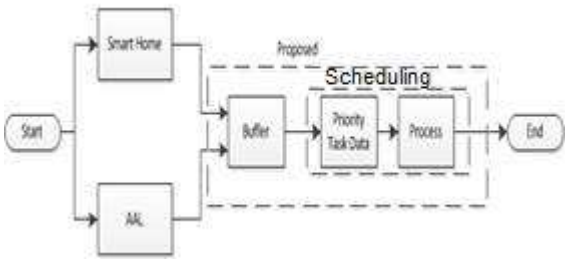
4.1 General Working Principle of SHAAL

By default, the data packet will be processed by first-come first-serve (FCFS) in the CBox using the conventional method as in Figure 6(a). This will affect the delay of high priority packet if the total amount of data packets in queue is too many. Within a SHAAL system, the data packet of AAL is given high priority, because it contains the health information. Hence, it is important to ensure its processing takes place with minimum delay. If the system grows too large where more new functionalities are introduced, its performance might be overwhelmed with the enormous amount of data to be processed.

The model of the proposed solution is shown in Figure 6(b). In the design, a buffer handles the incoming packet and outgoing packet before the packet is forwarded. The next task is more important as it will schedule the data packet by its priority.



(a)



(b)

Figure 6 Data Scheduler Model (a) conventional model and (b) proposed model

Figure 7 illustrates the proposed priority based scheduling mechanism. In this scheduling scheme, specific applications such as e-health and security are given higher priority than other applications such as lighting and temperature controls. The former is denoted in term of 'E' and the latter by 'T'. The execution for high priority applications is enabled by hardware or software interrupt. For instance, low priority application/task T1 in Figure 5 is being executed by the scheduler but preempted at t1 by high priority application E1. E1 would run till completion before resuming task T1 at t2. As illustrated, T1, T2 and T3 are being executed in an order by sequence and can only be preempted by an event of high priority application.

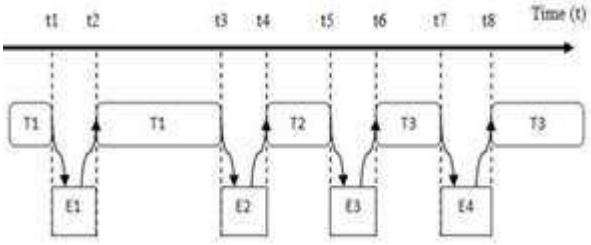


Figure 7 Priority Application/Task Data Flow

For example, an alert or alarm data which is considered high in priority and therefore will be processed immediately and subsequently relayed to the authorized personnel. This guarantees the reliability of high priority data.

From the proposed algorithm, it is clear that the scheduling algorithm can also be used in other applications with data that have different priorities. For instance, WSN has two well-known protocols that are being used in sports application which are ANT/ANT+ protocol and Bluetooth Low Energy (BLE) protocol. In [18], the authors concluded that ANT/ANT+ can only be used for application with low traffic characteristics with no QoS enforcement despite having no interference from the neighboring nodes. This is due to the nature of ANT protocol Time Division Multiple Access (TDMA) scheme which allows only one node to transmit at one time. By implementing priority scheduling to the system, it is possible to alleviate this problem where nodes with low priority have less timeslots allocated and nodes with higher priority have more timeslots allocated for them.

4.2 Architecture Model

SHAAL system implements WSN for data transmission. Data control in WSN depends on two types of architecture; centralized and distributed. The centralized architecture only allows communication between sensor modules via the base station (BS) as illustrated in Figure 1. This type of architecture is common in wireless networks such as WiFi and ad-hoc network. The advantages of this architecture include easy maintenance and debugging as well as monitoring since the entire event inside the system is logged to the main controller.

In the distributed architecture, the communications between nodes are not restricted to be carried out through the BS. It is possible for each node to communicate directly and independently to another node in its cluster as shown in Figure 8. Among its benefits are scalability and ease of expansion. The expanded system can be realized not just by adding more sensor modules, but also by introducing new BS node in a new network which communicates with another BS node. This concept describes hierarchical network architecture [14] and is presented in Figure 9.

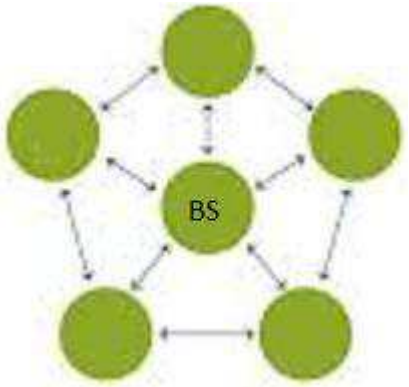


Figure 8 Distributed WSN Architecture

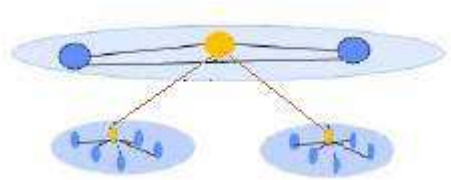


Figure 9 Hierarchical Network Architecture

Both architectures are utilized in our SHAAL system for performance analysis.

4.3 Case Study

The proposed data scheduling scheme is applied in a case study with a set-up as shown in Figure 10. Sensor modules like camera, light, alarm and AAL-related sensors such as temperature, ECG and SPO2 sensors, are placed in specific location and the performances of data transmission are studied in two scenarios; centralized and distributed experimental set-ups.

In both scenarios, CBox is responsible as an events logger as well as a gateway to the outside world i.e. remote terminals. These functionalities provides the mechanism required for SHAAL system to facilitates notification of any event occurred to authorized personnel via remote terminal as well as receiving commands to be executed by SHAAL system.

a) Distributed Approach

In this particular set-up, each sensor node exposes their services directly to each other without the CBox coordination. For smart home (SH) or home automation application, sensor nodes in turn negotiate with each other to form an application that has been programmed and only communicate with CBox for event notifications or receiving a command in a multi-hop manner if required. For AAL on the other hand, the CBox would act as a data processing node where the vital signals would be processed before being transmitted to the remote authorized medical personnel or center via internet gateway. In this scheme, SH nodes also act as a relay node for data transmission i.e., Light status and captured ECG data from a wearable device on a moving person.

b) Centralized Approach

CBox in the centralized scheme would act as data processing node for both SH and AAL. In this scheme, all nodes are coordinated by CBox. In home automation application, each node would register their services to the CBox and request services provided by other nodes through CBox to perform the programmed tasks. CBox in other words are responsible for resource allocation in the SH system. AAL related sensor nodes in this scenario are required to transmit its data directly to the CBox for processing. Note that in centralized scheme, multi-hop communication is not allowed.

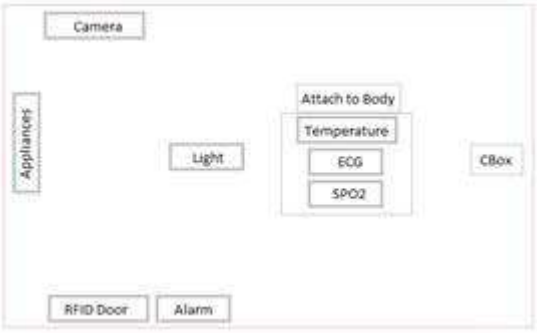


Figure 10 Topology of scenario for SHAAL system

The assigned priority of applications within our SHAAL system for both set-ups is shown in Table 1.

Table 1. SHAAL Priority

Application	Module	Priority
Smart Home	Light	2
	Appliances	2
	Door	2
	Alarm	1
	Camera	2
AAL	ECG	1
	Pulse	1
	SPO2	1
	Temperature	1

Based on the algorithm presented in Figure 11, the task's priority will be given by reading data packet priority and the task will be classified by *p1* or *p2*. This packet will queue in buffer *buff* and the base station/main controller module will read all tasks in the buffer. The task with the priority *p1* will be processed first and follows by the task with priority *p2* if the preemptive scheduler is enabled.

Algorithm: **Priority data scheduling scheme**

```
while task_k received by module_i do
 if task_k priority = p1 then
   process task_k
 else
   if hastask = true then
     queue until task_k finish
   else
     process task_k
   end if
 end if
end if
```

Figure 11 Priority Data Scheduling Scheme

The packet throughput (P_t) measures the capability of system to carry the data and defined as the following:

P_t = Total Packet Receive per second (P_r) / Total Packet Send per second (P_s) (1)

The higher the packet throughput, the better the system is. Our SHAAL system consists of two different types of data, of which are generated from Smart Home and AAL systems, respectively. The AAL data is given higher priority than smart home data. In other words, the proposed scheduling aims to produce a higher packet throughput for AAL system.

Figure 12 shows the average packet throughput in the system with no priority scheduling scheme. The measurement is based on the conventional model of Figure 6(a) and taken over one minute time duration. When the *data*1 packet starts execution, the processing task carries on even if a higher priority *data*2 packet than the currently running packet arrives at the ready queue. Thus *data*2 has to wait in the ready queue until the execution of *data*1 is complete after 10 seconds. The processing of *data*2 will start right after.

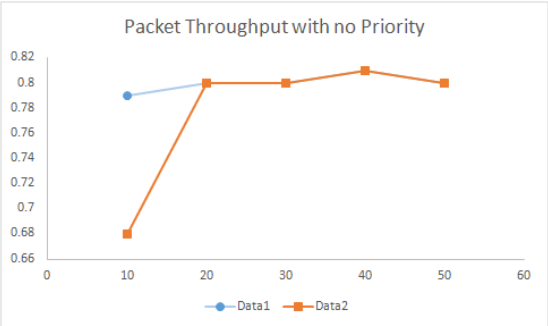


Figure 12 Average Packet Throughputs with no Priority

For SHAAL system with priority scheduling scheme, the ECG data has the chance to be processed first and can pre-empt the light data by saving the context of the light data if they are already running. Figure 13 shows the recorded average packet throughput for the system with priority scheduling under centralized architecture scenario. It can be observed that high average throughput is achieved for the ECG data from the recorded measurement over one minute period. Although the recorded throughput for light data is low, it should be reminded that the measurement taken is the average throughput over one minute time duration of when it is pre-empted by ECG data and only resumes its transmission after ECG data completes its execution. Moreover, the light data usually corresponds to ON/OFF switches.

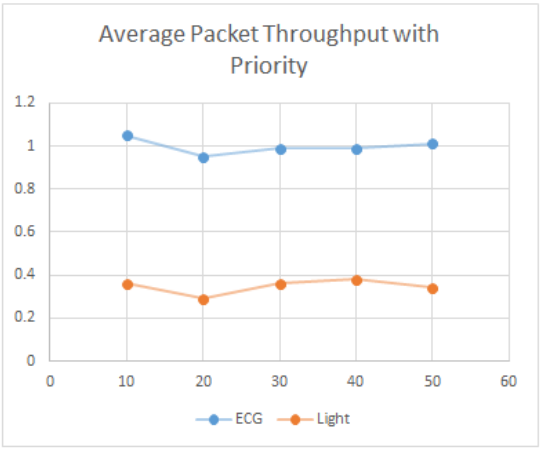


Figure 13 Packet Throughputs with Priority

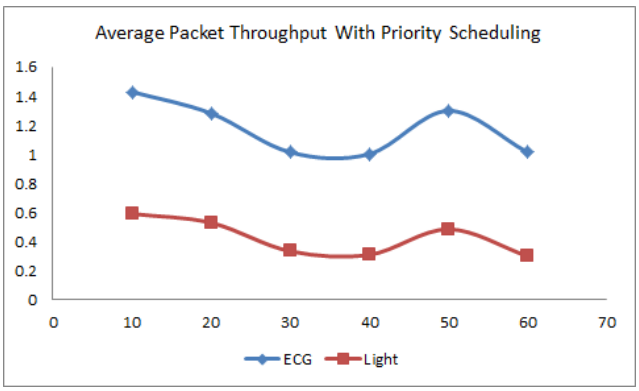


Figure 14 Average Packet Throughputs with Priority in Distributed Mode

Figure 14 shows the average packet throughput using priority based on distributed architecture approach. It can be observed that the achieved throughput is higher compared to the one achieved under centralized approach. The reason is due to a higher amount of packet in the

network as the sensor module is set with the capability to operate with 2-hop communication. As a result, the packet with higher priority gets more chances to be processed in the system.

The delay is the waiting time for a task to be executed. The delay will affect the performance and reliability of a system. In a real system, delay is inevitable, but the length of the delay must be in an acceptable range for system application. In our SHAAL system, delay is measured in term of packet delay (P_d). The calculation of the packet delay is shown in (2).

P_d = Packet Time Schedule (P_{ts}) – Packet Time Receive (P_{tr}) (2)

The lower the packet delay is, the better the system performance. The experiments from this point onwards are based on centralized approach.

Figure 15 shows the packet delay performance for SHAAL system with priority scheduling. The ECG data has lower delay than Light data. This is because the ECG data is processed immediately as it has higher priority, where the Light data is being queued until the processor finishes his current task. In our experimental set-up, the recorded maximum delay for ECG and Light data is 4 ms and 6.5 ms, respectively.

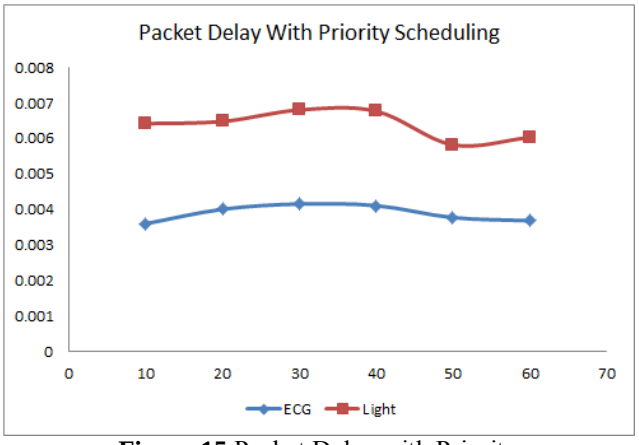


Figure 15 Packet Delay with Priority

Figure 16 shows the packet delay for system with no priority scheduling over a period of 60 seconds. The delays are quite similar for both ECG and Light data. This is due to the equal probability for both ECG and Light data packets to be processed which depends on FCFS method. The trend line for the packet throughput is the same as in Figure 12. The average P_d for both ECG and Light data is 1.7 ms in the experimental set-up.

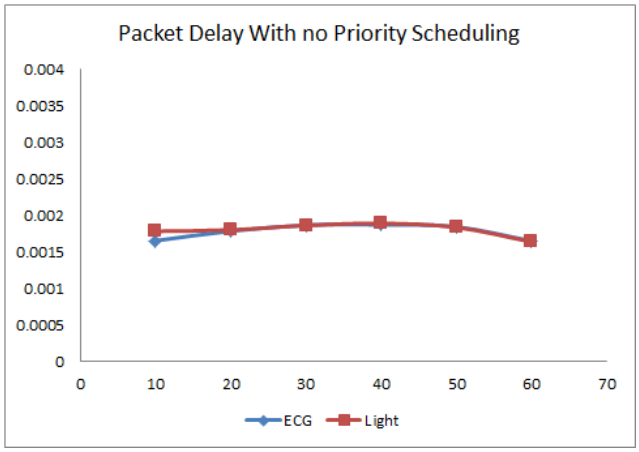


Figure 16 Packet Delay with no Priority

A comparison of P_d performance for high priority ECG data in systems with priority and without priority scheduling is illustrated in Figure 17. Although P_d is 2.2 times higher for system with priority scheduling as compared to the one with FCFS, we consider this value to be practically acceptable for our SHAAL system. With priority scheduling, the reason for the increase in delay time is because the system needs to check and confirm the priority level of the incoming packet. Hence, the time consumed to process the packet is affected. However, this is compensated by an increase in throughput for high priority data. In contrast, under FCFS method, data only needs to be processed without undergoing any priority check as well as no throughput gain.

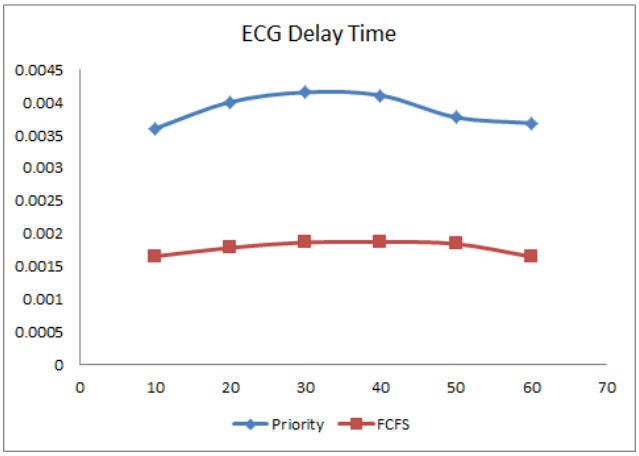


Figure 17 ECG delay with Priority and without Priority

From the performance analysis, we discover that the impact of delay on the throughput achieved for system with priority scheduling

needs to be considered. It can be seen that the proposed priority scheduling scheme is feasible for practical implementation as its lightweight nature delivers a higher throughput for high priority data at the expense of incurring a slight increase in system delay. Although the throughput gain is higher in a distributed architecture, we expect the system delay also to be further compromised.

5 CONCLUSION

The creation of a smart environment within the home not only offers an opportunity to embed sensors in order to acquire information about a person within their own home, it also allows us to process the information collected and subsequently interact with the person while aiming to enhance their underlying quality of life. To date, research conducted within the AAL domain has provided a deeper insight into factors associated with the design and development of technology to support independent living. Nevertheless, a number of challenges still remain in all of the three main components within the smart home: sensors, data processing and environmental control. In this paper, the tradeoff between delay and throughput is addressed in a case study utilizing the lightweight priority scheduling scheme. The findings show that the proposed scheme demonstrates promising solution to support decision making to facilitate SHAAL system in a real setting deployment.

6 ACKNOWLEDGEMENT

The authors would like to thank all those who contributed toward making this research successful. The authors wish to express their gratitude to Ministry of Higher Education (MOHE), Malaysia Research Management Centre (RMC) for the sponsorship,and Telematic Research Group (TRG), Universiti Teknologi Malaysia for the financial support and advice of this project.(Vot number Q.J130000.2623.08J96)

7 REFERENCES

1. J. Cheng and T. Kunz, ”A Survey on Smart Home Networking,” Carleton University, Systems and Computer Engineering, Technical Report SCE-09-10, September 2009.
2. M. Friedewald, O.D. Costa, Y. Punie, P. Alahuhta, S. Heinonen, “Perspectives of ambient intelligence in the home environment,” Elsevier.B. Notari. 1993. *Catalysis Today*. 18: 163.
3. H. Martin, A. Bernardos, L. Bergesio, and P. Tarrio, “Analysis of key aspects to manage wireless sensor networks in ambient assisted living environments,” Applied Sciences in Biomedical and Communication Technologies, 2009, ISABEL 2009. 2nd International Symposium on, pp. 1 –8.V. A. Zeitler, C. A. Brown. 1957. *J. Phys. Chem*. 61: 1174.
4. R. A. Rashid, S. H. S. Arifin, M. R. A. Rahim, M. A. Sarijari, N. H. Mahalin, “Home healthcare via Wireless Biomedical Sensor Network,” RFM 2008.
5. IEEE 802.15 Standard Group Web Site. Available from:<http://www.ieee802.org/15/>
6. M. A. Ismail and F. A. Rashid, “Smart and Cool Home in Malaysia,” Advanced Materials Research Vol. 224 (2011) pp 115-119.
7. D. M. Kumar, “Healthcare Monitoring System Using Wireless Sensor Network,” Int. J. Advanced Networking and Applications 1497, Volume:04 Issue:01 Pages:1497-1500 (2012)
8. A. Monacchi, “An early warning system for ambient assisted living,” Master of Science in Computer Science, School of Computer Science, Reykjavík University, June 2012.
9. Y. J. Chang, C. H. Chen, L. F. Lin, R. P. Han, W. T. Huang, and G. C. Lee, “Wireless Sensor Networks for Vital Signs Monitoring: Application in a Nursing Home,” Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, Volume 2012, Article ID 685107, 12 pages, doi:10.1155/2012/685107.
10. H. Sun, V. D. Florio, N. Gui and C. Blondia, “Promises and Challenges of Ambient Assisted Living Systems,” 2009 Sixth International Conference on Information Technology.
11. N. Nasser, L. Karim, and T. Taleb, “Dynamic Multilevel Priority Packet Scheduling Scheme for Wireless Sensor Network,” IEEE transactions on wireless communications, vol. 12, no. 4, AprilL 2013.
12. C. D. Nugent, L. Galway, L. Chen, M. P. Donnelly, S. I. McClean, S. Zhang, B. W. Scotney and G. Parr. “Managing Sensor Data in Ambient Assisted Living,” Journal of Computing Science and Engineering, vol. 5, no. 3, September 2011, pp. 237-245.
13. S. Kim, R. McLain, “Dust Networks”, MBA290/ENG 298 – Introduction to MOT, 2003.
14. Z. H. Khan, D. G. Catalot, J. M. Thiriet, “ Hierarchical Wireless Network Architecture for Distributed Applications,” The Fifth International Conference on Wireless and Mobile Communications, ICWMC 2009, Cannes, La Bocca : France (2009).
15. C. Gwilliams, A. Preece, A. Hardisty, Benoit Goossens, Laurentius N. Ambu, “Local and global knowledge to improve the quality of sensed data”, International Journal of Digital Information and Wireless Communications (IJDIWC), vol.2, no. 2, April 2012, pp. 164-180.
16. Ouadoudi Zytoune, Driss Aboutajdine, “An optimized energy aware chaining technique for data gathering in Wireless Sensor Network”, International Journal of

Digital Information and Wireless Communication (IJDIWC), vol. 3, no. 2, 2013, pp. 41-49.
17. Sandor Szilasi, Benjamin Babjak, Akos Ledeczi, Peter Volgyesi, "Towards a versatile wireless platform for low-power applications", International Journal of Digital Information and Wireless Communication (IJDIWC), vol.1 no. 2, 2011, pp. 401-414.
18. Sabri Khssibi, Hanen Idoudi, Adrien Van Den Bossche, Thierry Val, Leila Azzouz Saidane, "Presentation and analysis of a new technology for low-power wireless sensor network", International Journal of Digital Information and Wireless Communication (IJDIWC), vol. 3 no. 1, 2013, pp. 75-86.

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 124-142
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

A Genetic Algorithm Analysis towards Optimization solutions

Mujahid Tabassum and Kuruvilla Mathew
Swinburne University of Technology, Sarawak Campus, Jalan Simpang Tiga, 93350, Kuching, Sarawak, Malaysia
mtabassum@swinburne.edu.my and kmathew@swinburne.edu.my

ABSTRACT

In today's world, an optimal and intelligent problem solving approaches are required in every field, regardless of simple or complex problems. Researches and developers are trying to make machines and software's more efficient and intelligent. This is where the Artificial Intelligence plays its role in developing efficient and optimal searching algorithm solutions. Genetic algorithm is one of most pervasive and advanced developed heuristic search technique in AI. Genetic algorithm (GA) is developed to find the most optimized solution for a given problem based on inheritance, mutation, selection and some other techniques. It was proved that genetic algorithms are the most powerful unbiased optimization techniques for sampling a large solution space. In this paper, we have used GA for the image optimization and Knapsack Problems, which are commonly found in a real world scenario. Furthermore, a research based on a tool that uses Genetic Algorithm, called the GA Playground is done to demonstrate the capability of solving the Knapsack Problem with the fitness function and a case study on how images can be reproduced using the optimal parameters. Lastly, a few methods such as the Hash Table and the Taguchi Method are suggested to improve the performance of the Genetic Algorithm.

KEYWORDS

Artificial Intelligence, Genetic Algorithms; Optimization; GA operators; Metaheuristics; JNetic; GA Playground; Knapsack Problems; Travelling Salesman Problem; Taguchi Method;

1 INTRODUCTION

A genetic algorithm (GA) is an artificial intelligence search metaheuristic that is derived from the process of biological organism evolution. Genetic algorithms were described in University of Michigan 1960s and 1970s by John Holland [1].

Genetic algorithms are a subset of the evolutionary algorithm set. The design of evolutionary algorithms is based on processes which occur in living organisms in nature, for example, inheritance, mutation, selection and crossover.

GA's are uniquely distinguished by having a parallel-population based search with stochastic selection of many individual solutions, stochastic crossover and mutation. Many other search methods have some of these elements, but only GA's have this particular combination [2]. Genetic algorithms are used various fields of biology, biotechnology, computer science, engineering, economics, chemistry, manufacturing, mathematics, medicine and pharmacology and other fields and have numerous advantages and disadvantages.

In the event the reader finds difficulty understanding biological concepts, an appendix relating to the explanation of those concepts has been included. However, the reader is expected to be familiar with main concepts and keywords used in the field of artificial intelligence.

The GA requires a fitness statistics or function since it emulates the concept of solution evolution by stochastically developing generations of solution populations. This algorithm is applicable to search for the solution of high degree of complexity that often involves attributes that are large, non-linear and discrete in nature. However, GA does not guarantee that optimality can be found but the results are usually close to the global optimum. Since the solutions are considered to be of probabilistic nature, they are not contained by local optima [3].

In this paper real world problems are been included to evaluate the application and performance of genetic algorithm. The purpose of these case studies is to find the parametric values of the GA that provides the most optimum solution (maximum generation, optimum population, optimum crossover rate, optimum mutation rate) and helps to understand the effectiveness of this algorithm and the methods we could use to improve the performance if searching for the optimum solution.

2 RELATED WORKS

Genetic algorithm have quite a number of advantages which makes it as one of the most preferable and widespread search algorithms in Artificial Intelligence. One of the advantages is the capability of solving any optimization problem based on chromosome approach, another remarkably important feature of this algorithm is its capability to handle multiple solution search spaces and solve the given problem in such an environment. Moreover, genetic algorithms are less complex and more straightforward compared to other algorithms. In addition Genetic algorithms are easier to be transferred and applied in different platforms, thereby increasing its flexibility. Many algorithms in AI are complete and optimal in finding a solution; however, these algorithms are only efficient for single objective solutions which mean there is only a single criterion for the solution state. Whereas, looking for solutions with multiple objectives are much more complex. In such an environment each solution will have more than one criteria and every change in one of these criteria will directly affect the others. Genetic algorithm has this capability to generate efficient solutions in such a complex and mathematically sophisticated environments. This process is performed by using the described implementation operators such as crossover and mutation where the new solutions will inherit from more than one previous chromosome. As a result, as the search space is expanded down the hierarchy each chromosome will be actually the outcome of merging different genes of the previous most optimal solutions. This will subsequently result in finding the final optimal solution which is derived from merging all of the previous generations' optimal solutions (at their own generation time) which is the final desirable outcome in a multiple objective environment [4].

The performance and popularity of GA has brought it as rational choice for industries. The key factors to lead the genetic algorithm applications to success are to have a meaningful fitness evaluation and effective genetic algorithm representation. The elegance and simplicity of the genetic algorithms and its high performance to find the solutions for difficult high-dimensional problems are the reasons that make genetic algorithm become so attractive and widely used. Under some situations, the usefulness and efficiency of genetic algorithms will be enlarged. For example, when mathematical analysis is not available, the failures of traditional search methods, the complexity of the search space and the difficulty to encode the domain knowledge to narrow the search space. During these situations, the use of genetic algorithm will provide efficient and reliable helps to solve the problems. The ability to handle arbitrary kinds of objectives and constraints is one of the advantages of genetic algorithm approach. These things are handled as weighted components of the fitness function and to adapt the genetic algorithm scheduler to particular requirements easily. As a result, the genetic algorithms have been widely used in the world for modeling and problem solving. It can be applied to many technology problem, scientific, engineering problems, entertainment, business and etc. There are a lot of applications that used the genetic algorithm which are robotics, engineering design, automotive design and etc. These three applications will be elaborated to show that how genetic algorithm works on it [5].

2.1 Genetic Algorithm Applications

The flexibility of genetic algorithm has resulted in manipulation of this field in various applications over different industries. Some of these applications are discussed as follows [6]:

2.1.1 Evolvable hardware applications

This field is based on manipulation of GA to produce electronic. The Genetic algorithm models exploit stochastic operators to automatically derive new configurations based on the old configurations. As the model keeps evolving while running in its environment context, finally the desirable configuration which is required by the designer will be reached. An example for such a reconfigurable model can be a robot with capability of manipulating built-in GA to regenerate its configuration after some breakdown due to environmental issues such electromagnetic wave which can cause malfunction in its normal configuration. Moreover, such a robot will be able to alter its configuration to a newer version if it meets a situation where it is requires more functionalities to perform its tasks.

2.1.2 Robotics

As it is clear robotics necessitate designers and engineers to experiment and figure out all the required aspects of a robot such as hardware infrastructure and corresponding software architecture to develop a

comprehensive and efficient robot. For any new task all the mentioned activities are required to be performed again to design a new robot which suits the new objectives. By manipulating Genetic algorithm many of these extra designs requirements can be eliminated. Genetic algorithm will provide this capability to automatically generate a collection of optimal designs which can be used for specific tasks and activities. This approach can even be expanded and result in the generation of robots which can perform several tasks and process more comprehensive applications. Another aspect of using GA in robotics is in navigation process. GA provides optimized solutions for navigation process by which the robot can reach to its required destination without being lost or hitting other objects in the environment. In navigation algorithm each chromosome represents a series of path nodes where each node is a gene. Each gene has an x and y value and a Boolean value which represents whether the next node reachable form the current node or no. By using this approach robot will not only be able to always find a way to the target without hitting the objects but also it will be able to find the most optimal path.

2.1.3 Engineering design

Designing a new engineering model is a complex and time consuming process, but designing an optimal model which uses the minimum resources to deliver the maximum output is even much complex. Such a task requires great deal of effort and experience to be completed perfectly. This is where one more time the functionality of Genetic algorithm comes into action. GA can be integrated into computer based engineering design applications. By following such a strategy the application will be able to analyze different aspect of engineering design principles when generating a new design for a given problem. This approach in addition to providing the required design will also assist the designers to identify the frailties and possible failure points of the design. Such an approach is currently being used in many engineering industries such as aerospace, civil, automotive, robotics, electronics, mechatronics etc. These are a small subset of the fields where Genetic algorithm is currently being used to improve the outcome to the fullest. There are many more fields such as telecommunication routing, trip and shipment routing, gaming, encryption and code breaking, chemical analysis, finance strategies, marketing strategies, which also uses the GA. In summary, Genetic algorithm has obtained a great role in modern world's technological and scientific fields and this is on the increase.

2.1.4 Data Encryption

In the field of cryptology, Genetic Algorithm is being used to produce a new advance encryption by using the operations of Crossover and Mutation. Cryptosystem is a group of algorithm that contains secret keys for the encoding of information or messages into cipher text and decodes them back into their original state. The above figure shows the model proposed by Shannon for a secret key system. A new symmetrical block ciphering system called ICIGA (Improved Cryptology Inspired by Genetic Algorithms) enables the generation of a unique session key in a random process. The users are able to determine the size of the blocks and key lengths during the start of ciphering. The ICIGA is in fact created from an enhancement made upon the system GIC (Genetic algorithms Inspired Cryptography). The ICIGA operates based on the length of secret key being used by the user by splitting the plaintext into parts of the same size. In the ciphering process, the first part is split into block of equal size and being used for the generation of the secret key. The secret key being created will then be used for ciphering other parts of the message. The ciphering process also involves the genetic operations of crossover and mutation by doing a left shit to every block that is processed [7].

2.1.5 Computer Gaming

In the field of gaming, the opponent of the human player is often a form of advance artificial intelligence (AI) that incorporates Genetic Algorithm. Strategies that are being used in the past are being programmed using GA to ensure that the AI is able to learn and improve from the past experience. The learning technique allows the AI to avoid repeating past mistakes, which therefore increase the playability of the game. This enables a more realistic experience to the human player as they are required to change their strategies from time to time. It also helps to avoid the scenario of a situation whereby the human player found a sequence of steps which ultimately leads to success, which means that the game does not pose any challenge anymore. For the GA to work, it requires a method to represent the challenge in terms of the solution and a fitness function for it to decide the quality of an instance. The fitness function first receives a mutated instantiation of an entity and

proceeds to determine its quality. Next, this particular function is customized to fit the problem domain. The fitness function can just be a system timing functions in most cases especially the ones with code optimization. As soon as a genetic representation and fitness function are defined, the GA will instantiate the initial candidates that will then apply the repetitive range of operators made up of selection, crossover and selection to improve the fitness value of the candidates [8].

3 TRADITIONAL METHODS VS GENETIC ALGORITHM

After analyzing the above genetic algorithm examples, we can understand that the genetic algorithm have much more differences from most of the traditional optimization search methods. Instead of working with the variables, genetic algorithm will rather work with the string-coding of variables. Its advantage is that even though the function might be continuous, the coding will still discretizes the search space. Besides that, a discrete or discontinuous function can be controlled with no extra burden because genetic algorithms only require function values at various discrete points. Therefore, genetic algorithm can be applied to solve many problems. Genetic algorithms operators make use of the resemblances in string-structures to make an effective search are also one of the advantages. The genetic algorithm work with a population of points unlike the traditional method which work with single point, it is the most striking differences between them. The expected genetic algorithm solution might be the global solution because more than one string had been processed. Genetic algorithm emphasizes the good information that are found previously by using reproductive operator and propagated adaptively through mutation operator and crossover operator while the traditional method doesn't utilize the obtained information efficiently. By using population based search algorithm, genetic algorithm will reduce the workload to apply the same algorithm many times because multiple optimal solutions can be captured easily in the population. Besides that, there are still other advantages which are GAs use probabilistic transition rules instead of using deterministic rules, work with coding of solution instead of the solution themselves, work with population of solutions instead of single solution and so on. All these advantages can be used to maximize the function and can only obtained by applying genetic algorithm to solve real life problem which the traditional optimization methods cannot [5].

4 CLASSIFICATIONS OF GENETIC ALGORITHMS

A metaheuristic is a procedure which aims to find an acceptable solution in very complex optimization and search problems. [2]

In comparison to other heuristics, metaheuristics make usage of low-level heuristic or search algorithms. Therefore, metaheuristics use concrete heuristics or algorithms which are more abstract.

In comparison to optimization algorithms and iterative methods, retrieved solution is dependent on the set of random variables generated.

Compared to optimization algorithms and iterative methods, metaheuristics do not guarantee that a globally optimal solution can be found on some class of problems. Many metaheuristics implement some form of stochastic optimization, so that the solution found is dependent on the set of random variables generated.

Metaheuristics such as genetic algorithms are characterized by the following attributes [4]:

- Metaheuristics guide search process with defined strategies. Although randomized, metaheuristic algorithms such as GA are not random searches. They exploit prior information to direct the search into a region of better performance within the search space.

- Their goal is to efficiently search the state space to find near-optimal solutions. Thus metaheuristic searches include simple local search algorithms and complex learning processes.

- Metaheuristic algorithms are approximate and usually non-deterministic.

- Metaheuristic algorithms are not problem-specific. They make few assumptions about the optimization or search problem being solved, and therefore they can be used for a variety of problems.

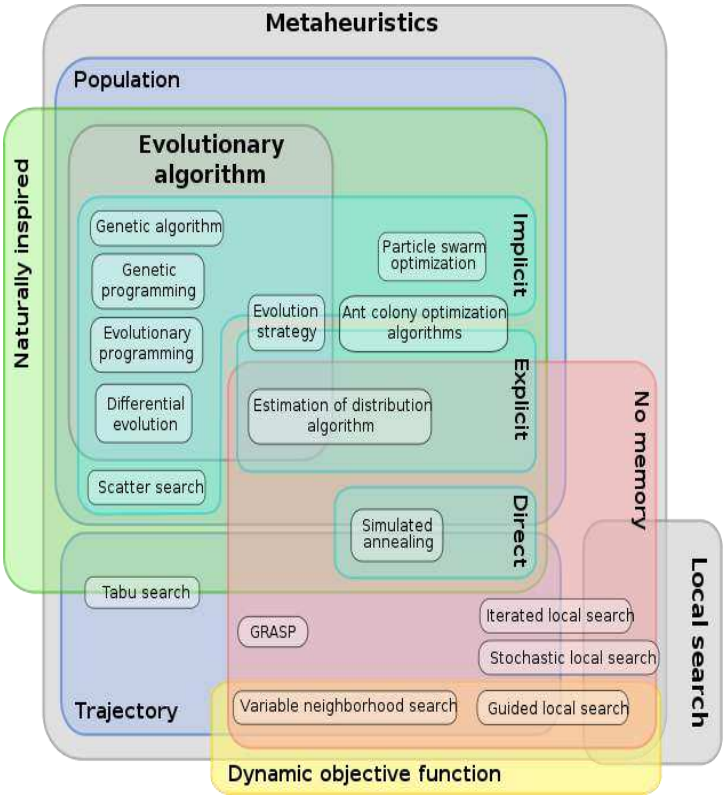


Figure. 1. Metaheuristics classification [9]

GA's are parallel, population-based searches which have a learning component. GA uses stochastic solutions of individual solutions, stochastic crossover and mutation.

4.1 Representation of Genetic Algorithms Using Bit Strings

To represent chromosomes in a genetic algorithm population, strings of bits are normally used, although there are many other existing methods currently in their infancy and under development such as fix-point binary representation, order-based representation, embedded list, variable element list and LISP S-expression.

Every position (locus) in the string (chromosome) has two possible values (alleles): 0 and 1. A solution (chromosome) is composed of several genes (variables). A chromosome represents a solution in the search space of potential solutions.

A population contains a set of chromosomes which are composed of genes. In other words, the solution space contains a set of strings which are composed of bits.

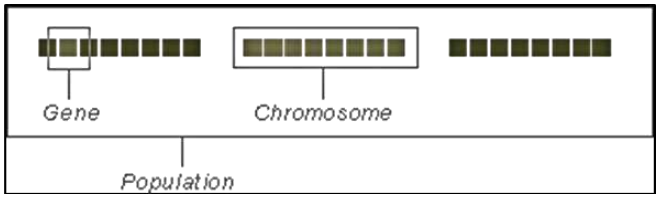


Figure. 2. Population, Chromosomes and Genes [10]

4.2 Implementation Overview

The initial population is first generated at random. Then, the GA goes through the following operators:

4.2.1 Selection operator

The principle here is "survival of the fittest". This operator chooses chromosomes in the population for reproduction. Based on an evaluation function, the 'better' chromosomes are identified and will be more likely to be selected for reproduction. This fitness operator's implementation is dependent on the specific problem at hand.

The 'better' chromosomes may be determined by an objective function which applies uniformly to all chromosomes or by a subjective function, where some rules may not be applied uniformly.

Generally, the probability of selection is proportional to the fitness. It is possible that the best chromosome does not get selected in one run of the GA. However, if the GA is run many times, the probability of selection will converge towards its mathematical expected value.

4.2.2 Crossover operator

This process is closely related to the reproduction of haploid (single-chromosome) organisms or by the fusion of a sperm cell with an ovum to produce a diploid zygote. This operator chooses a position in the string at random. Then, the subsequences before and after that position are exchanged between the strings. The result is the creation of two children strings, i.e. chromosomes.

Given two strings which represent chromosomes, e.g. 11000 and 00111, a crossover operator will randomly select a position, e.g. second position. The resulting children of the crossover will be 11111 and 00000.

Parents: **11|000** and **00|111**
Children: **11111** and **00000**

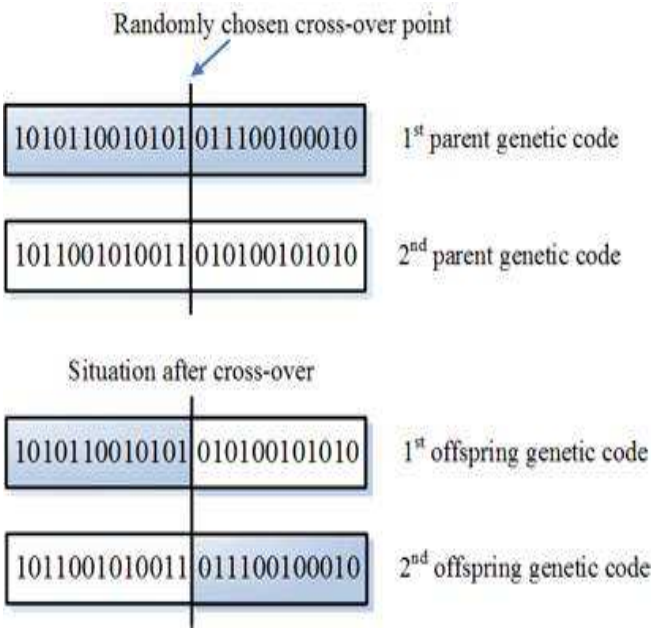


Figure. 3. Genetic Code of the parents and offspring before and after the crossover [11]

Multi-point crossovers are simply crossovers with more than 1 position where crossover will occur.

4.2.3 Mutation operator

This process mimics diversity within a population and helps to prevent premature convergence. This is because mutation and selection alone create hill-climbing algorithms. This operator randomly changes one or some bits in the string representation of a chromosome. For example, a string 10101010 could be mutated in its first position to give 00101010. The probability of mutation and the position of the mutation are controlled by the mutator operator.

Before mutation: 10101010
After mutation: 00101010

4.2.4 Inversion operator

This operator changes the order of genes between two randomly selected points in a chromosome. This fourth operator is rarely used. Its advantages, if any, are not well established. It has been suggested by some artificial intelligence developers that an improvement of search is achieved by such an operator. However, we have noted that all references used for this report indicate that there is no mathematical and empirical evidence of such improvement.

4.3 Effect of Genetic Operators

Using selection alone will have the tendency to generate a population with the fittest chromosomes. Without crossover and mutation, the solution found will be much less optimal.

Using selection and crossover operators only, the GA will converge on a local maxima, which is less optimal than what would otherwise be achieved using all three operators. Using mutation alone makes the GA go on a randomized search through the search space. Using selection and mutation alone will result in the creation of a hill climbing algorithm.

4.4 Pseudo Code for Simple Genetic Algorithm

Assuming that the problem that been clearly defined and candidate solutions have been generated and are represented using X bits, the following steps describe what a genetic algorithm will do, in the following order:

1) Randomly generate n X-bit strings
2) Evaluate fitness of each string in the population using evaluation function f(x)
3) While n offspring have not been created, do
   (a) Select two parent chromosomes based on probability of selection. The higher the fitness, the higher the selection probability.
   (b) Given a crossover and multi-point crossover probability, the crossover operator creates two offspring. In the event no crossover takes place, then the two offspring are identical replicas of their respective parents.
   (c) The two offspring are mutated at each locus using the mutator operator, given a defined mutation probability. The resulting children chromosomes are inserted in the population
   (d) If n is odd, a discard function will have a higher probability of discarding a less fit chromosome
4) The current population is replaced with the new one.

5) Go back to step 2 until fit enough solution is found or until X number of iterations have occurred.

Please note that every time the algorithm goes back to step 2, a new generation is created.

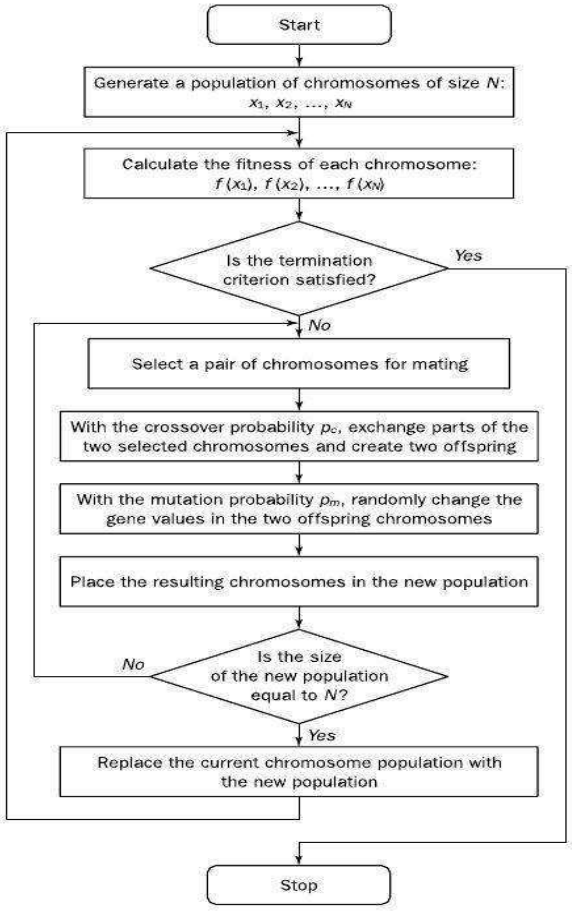


Figure. 4. A basic genetic algorithm [12]

4.5 Genetic Algorithm Parameters

The fundamental parameters of any genetic algorithm are discussed in the following sub-sections.

4.5.1 Cross-over probability

Cross-over points are randomly selected within the length of the chromosomes.

4.5.2 Number of cross-over points

In general, if the chromosome is long, then multiple cross-over points are preferable.

4.5.3 Mutation probability

This refers to the probability of any individual gene which a chromosome to change value from 0 to 1 or vice versa.

4.5.4 The maximum number of iterations

Acknowledging the fact that computational resources are limited, genetic algorithms must have a defined maximum number of generations or a point to indicate when to stop iterating. A maximum number of iterations is especially useful if we have poor knowledge of the ending criteria to indicate the end of iterations.

4.5.5 The value of the termination criterion

If we know how the end criteria should approximately be like, then this parameter is preferred over defining a maximum number of iterations. Examples of termination criteria are:

1. Time Elapsed since fitness increase remains under a certain value

2. If the population diversity falls below a specific value.

4.5.6 The population (generation) size

The greater the number of local maxima, the greater the population size should be. In some cases, the number of local maxima is not known. Therefore the ideal approach would be to have a larger population sample.

5 CONCEPTUAL ADVANTAGE OF GA

Using program evolution to solve computational problems has numerous conceptual applications [13]:

1. To allow a program to adapt and perform according to minimum standards in a changing environment, e.g. robot control in a changing environment or computer interfaces that need to adapt to the idiosyncrasies of an individual user.

2. To allow a program to create innovations on their own, for example, to create a new algorithm to solve a problem. This approach is especially relevant given the fact that intelligent systems are much too complex to program from a ‘top-down’

approach. Therefore an apparently better route to artificial intelligence is through a 'bottom-up' approach where programmers encode the simple rules, and complex behaviors (artificial intelligence) are derived from those rules.

3. Other typical search optimization techniques like linear programming, iterations, simple heuristic functions, depth first search and breadth first searches would involve too much time and space complexity for the solution to be economically viable. By searching over a large set of feasible solutions, metaheuristics can often find good solutions with less computational effort. Therefore metaheuristics like GA are useful for optimization problems.

6 TYPES OF INITIALIZATION OF PROCEDURES

Initialization is actually a very crucial part of Genetic Algorithm, because bad initialization can cause the time it takes to achieve the goal to increase and in the worst case scenario bad initialization can even prevent the algorithm from achieving the best possible solution to the goal. Therefore there are many types of initialization procedures being developed to give developers alternatives and to prevent bad initialization [14].

6.1 Bit String-Uniform Procedure

Firstly is the most commonly used procedure which is the Bit String-Uniform Procedure. Almost all Genetic Algorithm uses the Bit string-Uniform procedure which works by assigning value 0 or 1 with probability of 0.5 to every bit. The argument favoring that procedure is its uniformity in the binary space: using the Bit String-Uniform procedure, all points of 0 and 1 have equal probability of 0.5 to be drawn. Moreover, the bit-wise diversity is maximal, too: at every bit position, there will be as many 0's as 1's. Consider now the point of view of the number of 1's in each bit string. Denote #1(b) the number of 1's in bit string b. The probability distribution of the sum of independent random Boolean variables having probability p to draw 1 follows the well-known binomial law: P [1 = k] = C k n p k (1 p) n.

For large values of n (e.g. n < 30), the binomial law can be approximated by a Gaussian law of mean np and variance np(1- p). Hence for the Bit string-Uniform initialization procedure (p = 0:5), #1 lies in a very restricted range: For instance, for n = 900, 99%.Of all bit strings generated by the standard Uniform procedure have a number of 1's in [411; 489]. Hence, from the point of view of the number of 1's in the bit- strings, the Bit String-Uniform initialization procedure is certainly not uniform.

6.2 Uniform Covering Procedure

Another initialization procedure is theUniform Covering Procedure. The Uniform Covering procedure is designed for each bit string, select a density of 1's, uniformly in 0's and 1's then choose each bit to be one with that specific probability. The Uniform Covering procedure of course leads to a uniform probability distribution of #1. But, on the other hand, the probability for a bit at a given position l in a random bit string to be 1 is 0.5 and the expectation of the total number of 1's in the population is exactly 1/2 of the total number of bits, the Uniform Covering procedure still achieves maximal bit-wise diversity. But let us now consider another alternative point of view, considering the size of sequences of identical bits that are likely to occur in a bit string. When using the Bit string Uniform procedure, sequences of k 1's happen at a given locus with probability 1 1/2 k. With the Uniform Covering procedure, this probability is 1= (k + 1) (= R 1 0 r k dr). Depending on the pre-supposed regularity of the solution (in term of sequences of identical bits), it might be useful to further bias the Uniform Covering procedure to favor the emergence of homogeneous sequences.

6.3 Homogeneous Block Procedure

Besides that, there is another procedure called Homogeneous Block Procedure. The basic idea of the Homogeneous Block initialization procedure is to start from a bit string initialized to a default value for example 0, and to gradually add homogeneous blocks of the other value 1. The critical issue then becomes the number of such blocks, and the way their characteristics are randomly drawn. In the same vein as for the Homogeneous Block procedure, the number of blocks is adjusted depending on a uniformly drawn target density of 1's. Blocks are added to an initial bit string of all 1's, until that target density value is reached. The block characteristics are chosen as follows: A local length L is randomly chosen, uniformly in [0; L max], where L max is user-defined (typically n=10). The center c and the length l of each

block are chosen uniformly, respectively in [0; n] and [0; L]. Finally, to avoid too long loops in the case of high target density and small local length L, a maximum number of blocks is authorized (typically n). Due to the tight control over the density of 1's, the bit-wise diversity should be close to optimal. However, because of the limit on the number of blocks, that procedure is slightly biased toward bit strings having more 0's than 1's. A way to counter-balance that bias is to draw half of the bit strings using 0 as the default value, and the other half by reversing 0 and 1.

6.4 Case-base Initialization

Unlike the above procedures the case-based initialization initializes genetic algorithms in changing environments which is part of an ongoing investigation of machine learning techniques for sequential decision problems. This procedure is a novel combination of two methods of machine learning (genetic algorithms and case-based approaches) to perform learning in a changing environment in which we can monitor and record the changes. Anytime learning with case-based initialization shows a consistent improvement over anytime learning without case-based initialization. Case-based initialization allows the system to automatically bias the search of the genetic algorithm toward relevant areas of the search space. Little time is lost attaining a similar level of learning as in the previous same cases, and then improving on that performance.

One advantage of the method is that the population of the genetic algorithm is seeded without user intervention. Another advantage is that by using good individuals from several different past cases as well as exploratory strategies, default strategies and members of the previous population, our method tends to provide more diversity, guarding against premature convergence.

7 TYPES OF TERMINATION METHODS

Termination is the process where the genetic algorithm decides whether to continue searching or to stop. Each of the enabled termination criteria is checked after each generation to see if it is time to stop. Different types of termination methods include the following [15]:

7.1 Generation Number

The first termination method here is called Generation Number; it works by stopping the evolution the algorithm reached the maximum number of evolution specified by the user. Furthermore the Generation Number is a constantly active termination method is.

7.2 Evolution Time

Secondly Evolution Time is another termination method; it works by stopping the evolution when maximum evolution time specified by the user is exceeded. One flaw of this method is that by default, it will not stop the evolution until the evolution of the current generation is completed even if the maximum evolution time has been exceeded. However this is not a huge flaw as it can be easily solve by allowing the evolution to be stopped during a generation.

7.3 Fitness Threshold

Thirdly Fitness Threshold is another termination method. This method works for both maximize fitness objective and minimize fitness objective. For maximize fitness this termination method will stop the evolution when the fitness of the current population exceeds the fitness threshold specified by the user, while stopping the evolution when the fitness of the current population is lower than the user specified fitness threshold if the objective is to minimize the fitness.

7.4 Fitness Convergence

The fourth termination method is called Fitness Convergence, which stops the evolution when the fitness is considered converged. To determine if the fitness is converged, two filters of different length are used to smooth over the generation obtaining the best fitness and if the best fitness obtained by the long filter and the best fitness obtained by the short filter is less apart than a user specified percentage the fitness is considered converged and the evolution will be stopped.

7.5 Population Convergence

The fifth termination method is called the Population Convergence, which terminates the evolution when the population is considered converged. The population is converged when the average fitness of the population and the best fitness of the population is less apart than a user specified percentage.

7.6 Gene Convergence

The sixth termination method is Gene Convergence, which terminates the evolution when a percentage of genes specified by the user in a chromosome are

converged. A gene is considered as converged when the average value of that gene across all of the chromosomes in the current population is less than a user-specified percentage away from the maximum gene value across the chromosomes.

7.7 Hyper Volume Convergence

Another termination method would be Hyper Volume Convergence, which is a termination method for Multi-Objective Optimization in which the optimal solution set is measured by Hyper Volume of the Pareto-ptimal Front. Evolution will be terminated when the Hyper Volume is stabilized.

7.8 Greedy Search & Back Elimination

Last but not least would be the Greedy Search and Black Elimination termination method, which is for Attribute Selection. When using Greedy Search algorithm, the evolution stops when the fitness does not improve when a single input is added to the previous input set. However when Back Elimination is used, the evolution stops when the fitness becomes worse when a single input is removed from the previous input collection. A variation of the Greedy Search and Back Elimination is also implemented that keeps an elite pool with the size defined as Elite Size. When every solution in the elite pool are checked using above criteria, the search is stopped.

8 TYPES OF OPTIMIZATION PROBLEMS

8.1 Knapsack Problems

The situations; when there are number of items have to pick up into a knapsack. The goal is to optimized item picked be most valuable, which calculate from total value of the item with maximum weight per item to be picked such as problems occur in logistic industrial for efficient transport of product [16].

8.2 Facility Problems

The situations; when a series of facility needed to be placed at the most strategic placement. It is to optimized areas include minimal cost for user, place multi-facilities where existing different type of facilities to minimize the total cost and most strategic area for attraction in order to maximize market opportunity. It is mostly used for shopping complex or city public facilities built [16].

8.3 Traveling Salesman Problem (TSP)

One of most common GA approach could be apply to solve this problem. It is optimized when there are number of nodes which needed for the salesman to travel with minimum cost of distance. These kinds of problem occur in circuit design, routing for transportation and others [17].

8.4 Bin Packing Problem

This problem is derived from terminology of Knapsack problem. It is to optimize when a set of object with different characteristic (volume, weight and other measurement) packed into to numbers of bins, GA helps to minimize the number of bins used. Problem like data packed into a server has to be as minimal as possible; if first bin of space is not fully packed yet, it would be efficient to be able store in data into the remaining [118].

8.5 Multi-modal (global) Optimization Problems

It is to optimize the task where a minimal parameter is received (constraints) but still able to obtain multiple (global or local) best results. It can ease problem if one solution is not useful even through choosing the rest of the results is still at its best estimated result [16].

9 CASE STUDIES USING GENTIC ALGORITM TOOLS PLAYGROUND AND JNETIC

Genetic Algorithm is a method to solve problem and optimize problem at the same time. However, it is impossible to implement without any help from software engineer thus there are existing tools to make this happen. Open source application usually come with limited features while paid application can be custom made depend on what outcome and whether it can be achieve via using Genetic Algorithm.

In this paper we have used two tools such as Genetic Algorithm Playground, which programmed to solve minor problems only, all the problem are listed above types of optimization problem section. Those problems are commonly experienced in real world situation. The application is running cross-platform localized and internet; the application can be run by using applet also known as java plug-in for in computing to make it

convenient to run along the internet, while the application is program using Java programming language, thus it can also be run as a standalone application too if proper installation is adapt. For the testing purpose, due to constraint of sources found, applet is used to run the application through Internet [18].

JNetic is software designed by Steve Bergen. It is used to create stylized images from existing ones, by means of a GA [9]. JNetic makes use of a base image, and tries to replicate it as best it can using a specified number of shapes and colors, all supplied by the user. We already know how the 'fittest' image looks like, all that needs to be done is to evolve a population to get as close as it can to that image. The user can specify the basic GA parameters before running the program. This also includes other settings such as shapes, colors and sizes. However, those other settings will not be varied in this Case study because they are not the focus of this case study is only on GA. The only settings of interest are how changes in GA parameters affect the reproduction of the original image.

9.1 Genetic Algorithm Playground

9.1.1 Problem Description

A knapsack traveller has pick up to valuable item that sufficient needs item for the traveller to travel. The problem the traveller faced is there are only 50 objects can be store with a single knapsack, while these objects contain different variable of weights and values. The goal is to take the greatest valuable object as conceivable [18].

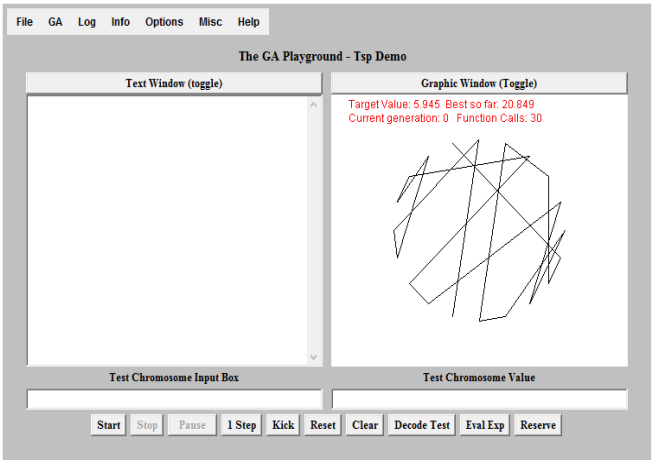


Figure. 5. GA Playground – TSP Demo

9.1.2 Basic Operation of Genetic Algorithm

Reproduction: Duplicate the potential solution of chromosome

Fitness: Calculate each Fitness of all the chromosome

Repeating the following steps

- **Selection:** Define fitness function by choosing two parent
- **Crossover:** Swapping parent to create new children
- **Mutation:** Alter the value of the potential solution

Exit when satisfactory solution is found

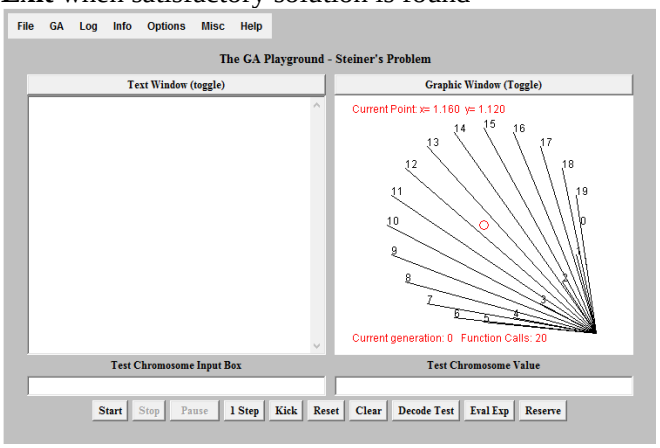


Figure. 6. GA Playground – Steiner's Problem

9.1.3 Optimization and Planning

Those object will be group into set, each set contains a total value and weight cost. The total value of the item can be accumulate from each item included in a set. Item will be compare with different item combination to obtain the highest value. At the same time, total value will be subjected to the total weight, weight is determine where total amount of weight should not exceed maximum weight supported of the knapsack. Thus the outcome must be total value as higher as possible and not exceeding the maximum weight of knapsack.

Let Xi as item, Pi as value and Wi as weight, the collection of cost is than subject to the weight where weight should not be less than maximum weight [18].

$$\text{sum}(x(i)*p(i)) \text{ subject to } \text{sum}(x(i)*w(i)) <= W$$

9.1.4 Application Mode

Parameters can be modified respectively to user's preference, the file must be save locally and it only allow ASCII key. The fitness and function value can be change in population file, this allow the mapped new

chromosome (solution) or when the population type are different. It is optional because the application will generate the default function [18].

In this example, there will not be any use of modification because it must be run locally while program are run through applet.

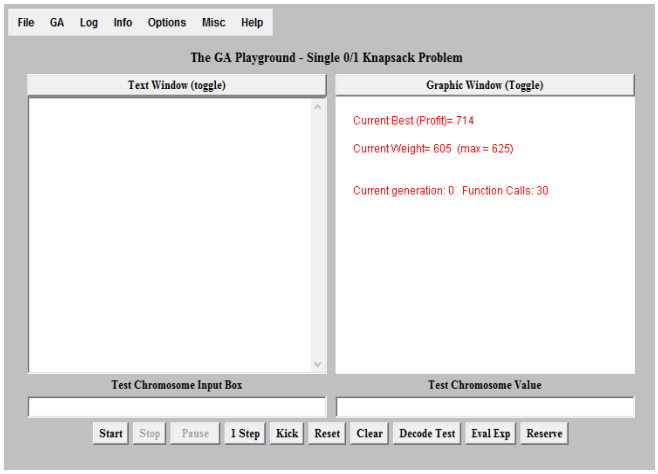


Figure. 7. GA Playground – Application in Process

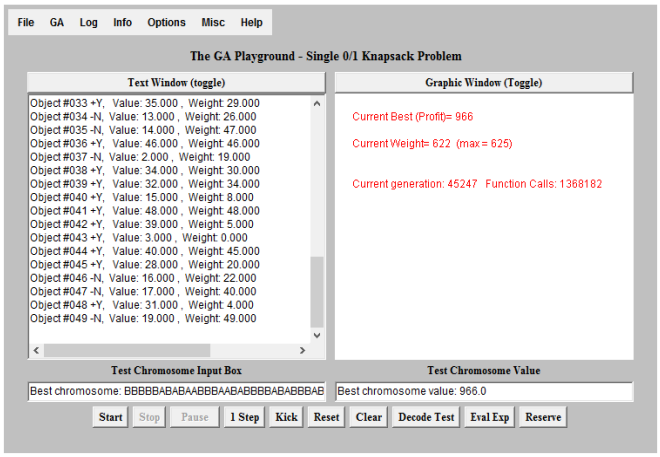


Figure. 8. GA Playground – Application in Process

6.1.4 Results and Finding

Results are attached in Appendix.

The knapsack has a maximum capacity of 625 and needed to obtain as higher as possible but not exceed the maximum weight. We uses 0/1 knapsack solution to maximize the total profit in this area. 0/1 means item is either entirely store or totally leave all item not store any item.

GA is an artificial intelligent uses heuristic search algorithm, the idea is derived from genetic finding. It can find all most usable solution for a set of parameters. Thus, answers are being shown in chromosome, which represent a structure of DNA. The chromosome obtain is: <BBBBBABABAABBBAABABBBBABABBBABB BBBAABABBBBBBBBAABA>. The DNA is then being encoded into readable value which is 966, therefore chromosome is representing the content of the problem and each of the DNA has individual data.

Firstly, the knapsack capacity has been stack into 605 weights and with a profit of 714. After applied Genetic Algorithm (GA) to the problem, surprisingly it had maximized the profit from 605 to 966. In addition, the weight did not exceed the maximum capacity, the weight still have 3 left. The weight is added in exchange to pick up more valuable object while it is still at the same amount of weight.

9.1.5 Limitation of the GA playground

- GA playground can be very slow in process when it is solving problem which is much complex and large number of parameters.
- GA playground need manual observation and configuration because it has the potential of malfunction, never ending load when result and condition coalition. In knapsack problem, weight found is at 3 number away but never stop looping because there are no more parameter to load up 3 number to maximum weight.
- The GA playground is in opening stage, thus the number of problem can be solve is very minimal.
- The knapsack problem can only choose minimum of 50 objects into a single knapsack and cannot be configure when in applet mode.

9.1.6 Advantages

- GA allows to optimized huge variety of problems, for example, TSP can eventually link to circuit design, scheduling, and delivery problems.
- GA is a problem solving mechanism and is able to generate multiple results that are most appropriate.
- Bad parameter initially does not affect the end result, moreover GA will discard those bad parameter.
- GA can solve problem with more complex and finite number of parameter, it uses fitness function.
- GA can relate easily to existing real world situation instead of just genes encoding.

9.1.7 Disadvantages

- GA does not always come with global optimum all the time especially when overall solution has various population.
- A tool is dependent on computers speed and only real time application can produce quick respond time.
- To a non-professional, the return encoded result maybe not be able to understand in application to the user's problem, because GA does not simulation the instructions but rather it shows the encoded chromosome as solution values (not applicable to everyone to use).
- Every time GA provides different results which only allow situation that tolerate with trial and failure results
- GA is complex to be able to apply problem or to be understand.

9.2 JNetic

9.2.1 Selection of Source Image

The image in figure 5 is the SUTS logo, and we will be using this for the experiments.

Figure. 9. SUTS Logo

9.2.2 Choosing Geometric Primitives

The smaller the "pixel" used, the better the final image solution would be. However, the computation resource required to create a solution image which is even grossly similar is extremely high for an average laptop. Therefore, we used large rectangles to represent pixels.

9.2.3 Choosing a Color Scheme

By default the color model of the geometric primitives will follow a full RGB color scheme, which includes all 16 million colors with alpha transparency, i.e., which have varying degrees of transparency and opacity. A range of transparency allows the creation of graphics that have smoother edges.

Transparency has been disabled due to high resource usage. Therefore, all primitives will have 256 of each red, green and blue value, rendered over a white background.

9.2.4 Setting GA Parameters

The success of the final image is extremely dependent on the GA parameters used. The recommended number of generations was set to 500, population to 150, crossover rate to 100%, mutation rate to 0.1% and tournament size (number of individuals needed to fill a tournament during selection) to 4.

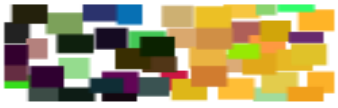
Image correctness is ranges from 0 to 1, 1 being an exact replica of the original image. Image correctness is measured automatically by the built-in fitness monitor by the program.

Important: Because of the randomness of a GA, every run is likely to produce a slightly different final image

9.2.5 Changing the Number of Generations

Population = 200,
Crossover rate = 100%,
Mutation = 1%,
Tournament size = 4

TABLE I. NO OF GENERATIONS

Value	Average Image Correctness (5 d.p)	Image
100	0.45238	
200	0.54667	
500	0.71397	
1000	0.79594	

1500	0.81322	

The GA appeared to reach convergence at about 1500 generations, using current parameters.

9.2.6 Changing Population

Max generation =200, Crossover rate = 100%, Mutation = 1%, Tournament size = 4

A lower population dramatically decreases computational time. The tradeoff however, is a far below optimum solution. The most optimum population was 100. After this point, there was a 'diminishing returns' on the ratio at an increasing rate.

TABLE II. POPULATION

Value	Average Image correctness (5 d.p)	Computation time/s	Ratio (correctness/time)	Image
50	0.58707	0.077	7.624	
100	0.69151	0.139	4.974	
150	0.71944	0.199	3.615	<negligible difference>
200	0.72498	0.245	2.959	<negligible difference>
250	0.72589	0.304	2.388	<negligible difference>

9.2.7 Changing Crossover Rate

Max generation = 200, Population = 200, Mutation = 1%, Tournament size = 4

TABLE III. CHANGING CROSSOVER RATE

Value	Average Image Correctness (5 d.p.)	Image
10	0.69028	
30	0.64933	
50	0.62929	
70	0.65025	
90	0.69936	
100	0.64544	

From the above table, the best crossover rate is about 90%. The value of 100% yielded diminishing returns.

9.2.8 Changing Mutation Rate

Max generation = 200, Population = 200, crossover rate = 100%, tournament size = 4

TABLE IV. CHANGING MUTATION RATE

Value	Average Image Correctness (5 d.p.)	Image
0.1	0.66306	
0.2	0.70309	<negligible difference>
0.4	0.72663	<negligible difference>
0.6	0.73970	<negligible difference>
0.8	0.7399	<negligible difference>
1.0	0.74140	
1.2	0.73669	<negligible difference>
1.5	0.71198	<negligible

		difference>
2.0	0.67763	

From the above table, it is observed that the best mutation rate is about 1.0 %.

After 1200 generations, population size =100, crossover rate = 90%, mutation rate = 1.0% and using large "pixels" of 10 by 30 to represent rectangles, the best possible result is shown in figure 6.

It is possible to achieve near 100% replica if the pixel sizes used were much smaller. However, the population sizes and generations must be increased and this demands days of computational power on a laptop.

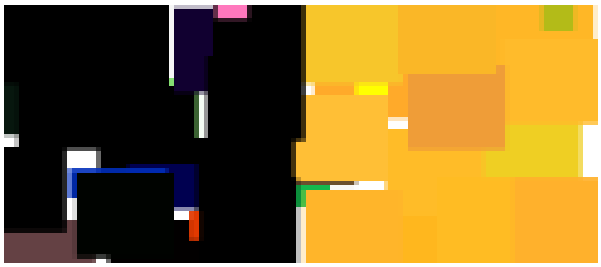


Figure. 10. FINAL SUTS Logo

A progress chart labels the best, worst and average fitness over time. Once the average and best finesses reach near the same value, the GA has converged, and cross-over will be near-useless. This made a compelling case to increase mutation rate to maintain diversity.

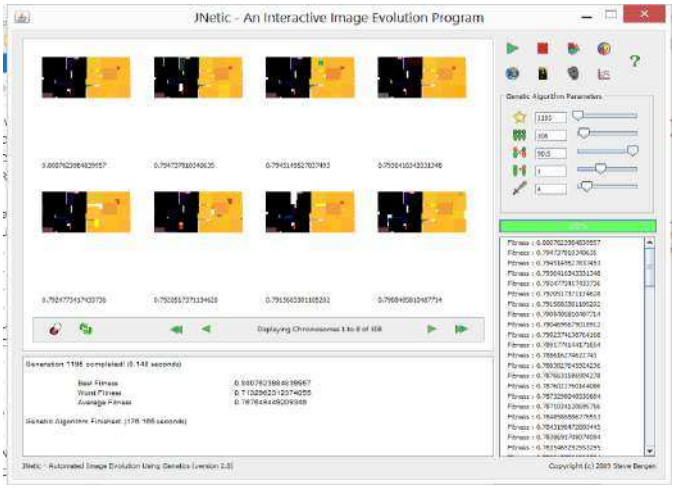


Figure. 11. JNetic Software

10 ACKNOWLEDGEMENT

We are thanks to Sailesh Choytooa, Caleb Lai, Gary Lim and Nelson Ting for their contribution to this work.

11 CONCLUSION

In this paper, we have implemented two case studies to find the parametric values of the GA that provides the most optimum solutions. In conclusion, Genetic Algorithm (GA) is the best application to solve various common problems using fitness function. Genetic Algorithm is an exhaustive approach which can be applied in Artificial intelligence field to find optimal solution in complex search spaces. It is a heuristic search algorithm that will exploit the historical information for best solution. The genetic algorithm is well suited for high complexity problems without any known sophisticated solution techniques like the combinatorial optimization problems. By applying the comprehensive functionalities of Genetic Algorithm in real world scientific and industrial fields, it will not only be ensured that many of the existing problems will be resolved but also there will be a promising future towards development of agents which can perform such a task efficiently and effectively without human intervention. Besides that, the application also optimize problem through generate multiple most appropriate solution. Most of the operation are taken in calculate the fitness by repeating recalculate in order to obtain multiple solution with the most accurate result. Thus, it has taken more time for processing problems. In order to improve the performance, these are the methods should be apply:

- Hash Table

In this method, there are 3 steps should be taken to improve the performance of GA. First, create function that receives single parameter the size of core data structure. Secondly, implement put method to takes two parameters which are the key and the element. Those key are referral for the element. Lastly, get method, which can receive one parameters of key and it can return two values. Those two values include whether the element found and the element [19].

- Taguchi Method(TM)

This method is to select the ones which are useful to select as a result and let go redundancy of calculation;

it helps to reduce the number of conducts calculation. The result selection should covers these 3 priority: (1) Factor of interest, (2) the level of interest and (3) desired cost limitation [20].

- Binary Tree Algorithm

Binary tree algorithm is very similar to Hash table approach where put and get method are implemented. The only difference between Hash algorithm is it uses binary tree method to store data structure while Hash algorithm uses array. Binary tree method is by branching when key is greater than the current value using binary format [19].

- Keep Last Generation Algorithm

This method has been derived from Hash algorithm, the difference in this method is, its stored previous fitness value and able to reuse those values as comparison and such. By doing so it able to reduce the size of hash table than the Hash Table in Hash algorithm where it keeps grows as large as possible [19].

In order to obtain the maximum improvement of GA, these method are recommended to be implement, as from prove it does significantly improve the performance of simple GA approach.

12 REFERENCES

1. E. Eiben (1994). "Genetic algorithms with multi-parent recombination". PPSN III: Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature: 78–87. ISBN 3-540-58484-6.
2. M. Melanie (1999), "An Introduction to Genetic Algorithms", Cambridge, MA: MIT Press. ISBN 9780585030944.
3. Marijke,K 2002, "Genetic Algorithms, An overview", http://www.meteck.org/gaover.html.
4. Prof. Dr. R. Safaric & Dr. A. Rojko, "Genetic code of the parents and the offspring before and after the cross-over", Maribor December 2006, http://www.ro.feri.uni-mb.si/predmeti/int_reg/Predavanja/Eng/3.Genetic%20algorithm/_04.html.
5. Tom, v.M n.d., "Genetic Algorithm", http://www.civil.iitb.ac.in/tvm/2701_dga/2701-ga-notes/gadoc/
6. n, "Real World Uses of Genetic Algorithms", brainz learn something, http://brainz.org/15-real-world-applications-genetic-algorithms/
7. Ankita Agarwal, "Secret Key Encryption algorithm using GA", http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.17.661&rep=rep1&type=pdf , Vol 2, Issue 4, April 2012. ISSN: 2277128X
8. Lucas, S., and Kendell, G. 2006, "Evolutionary computation and games" In IEEE Comput Intell Mag. February, 10–18. IEEE.
9. T. Wong & H. Wong, "Introduction to Genetic Algorithms", Department of Computing Imperial College of Science and Technology and Medicine UK, http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol1/hmw/article1.html.
10. Prof. Dr. R. Safaric & Dr. A. Rojko, "Genetic code of the parents and the offspring before and after the cross-over", Maribor December 2006, http://www.ro.feri.uni-mb.si/predmeti/int_reg/Predavanja/Eng/3.Genetic%20algorithm/_04.html.
11. E. Schultz, J. Mellander & C. Endorf (2008), "Intrusion Detection & Prevention - A basic genetic algorithm", [image online], http://my.opera.com/blu3c4t/blog/show.dml/2636486.

12. S. Bergen, "JNetic Textures", http://www.cosc.brocku.ca/~bross/JNeticTextures/.
13. Varma & Nathan Erhardt, "Artifical Intelligence from a philosophical and biological perspective", http://biology.kenyon.edu/slonc/bio3/AI/index.html.
14. Leila Kallel & Marc Schoenauer, "Alternative Random Initialization in Genetic Algorithm", ICGA. 1997.
15. Termination Methods n.d., NeuroDimension Inc., http://www.nd.com/genetic/termination.html
16. Silvano Martello & Paolo Toth, "Knapsack Problems: Algorithms and Computer Implementations", DEIS, University of Bologna, John Wiley & Sons, New York.
17. Obitko, M n.d., IV, "Genetic Algorithm", University of Applied Sciences, http://www.obitko.com/tutorials/genetic-algorithms/ga-basic-description.php .
18. Dolan, A n.d., "A single knapsack problem with 50 objects", http://www.aridolan.com/ga/gaa/Knapsack01.aspx
19. Povinelli, RJ n.d., Improving Computational Performance of Genetic Algorithms: "A Comparison of Techniques", Marquette University, http://povinelli.eece.mu.edu/publications/papers/gecco2000b.pdf.
20. Anagun, AS & Sarac, T 2006, "Optimization of Performance of Genetic Algorithm for 0-1 Knapsack Problems Using Taguchi Method", Eskishir Osmangazi University, http://mmf2.ogu.edu.tr/sanagun/Docs/dok4.pdf
21. Mujahid Tabassum, Kuruvilla Mathew and Sailesh Choyooa, "A Genetic Algorithm Approach towards Image Optimization", ICIEIS2013, SDIWC, ISBN: 978-09891305-2-3.
22. Anil,K & Ghose,M.K n.d., "Information Security using Genetic Algorithm and Chaos, http://www.skoch.in/images/stories/security_paper_knowledge/ Information%20Security%20using%20Genetic%20Algorithm%20and%20Chaos.pdf.
23. Baldi, MM, Crainic, TG, Perboli, G & Tadei, R 2011, The Generalized Bin Packing Problem, Cirrelt, https://www.cirrelt.ca/DocumentsTravail/CIRRELT-2011-39.pdf.
24. Dolan, A n.d., *A general GA toolkit implemented in Java, for experimenting with genetic algorithms and handling optimization problems*, http://www.aridolan.com/ga/gaa/gaa.html#DefinitionFiles.
25. Drezner, T, Drezner, A & Salhi, S 2002, *Solving the Multiple competitive facilities location problem*, Elsevier, http://www.ceet.niu.edu/faculty/ghrayeb/IENG576s04/papers/SA/Solving%20the%20multiple%20competitive%20facilities%20location%20problem.pdf.
26. Genetic Algorithm n.d., http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/tcw2/report.html#TSP.
27. Ramsey, LC & Grefenstette, JJ n.d., Case-Based Initialization of Genetic Algorithms, citeseerx edu. , http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.17.661&rep=rep1&type=pdf.
28. Ronkkonen, J 2009, *Continuos Multimodal Global Optimization with Differential Evolution-based Methods*, Lappenranta University of Technology, https://www.doria.fi/bitstream/handle/10024/50498/isbn%209789522148520.pdf.
29. *Genetic Algorithms To Constraint Satisfaction Problems* n.d., University of Portsmouth, http://www.economicsnetwork.ac.uk/cheer/ch13_1/ch13_1p16.htm.

Appendix: GA Playground – results:

Single 0/1 Knapsack Problem - Solution:		
Chromosome: <BBBBBABABAABBBAABABBBBABABBBABBBBBAABABBBBBBBBBAABA>		
Value: 966.000		
Object#000 +Y	Value: 49.000	Weight: 21.000
Object#001 +Y	Value: 23.000	Weight: 1.000
Object#002 +Y	Value: 21.000	Weight: 4.000
Object#003 +Y	Value: 37.000	Weight: 2.000
Object#004 +Y	Value: 28.000	Weight: 28.000
Object#005 +Y	Value: 27.000	Weight: 47.000
Object#006 +Y	Value: 21.000	Weight: 2.000
Object#007 +Y	Value: 6.000	Weight: 23.000
Object#008 +Y	Value: 38.000	Weight: 12.000
Object#009 +Y	Value: 7.000	Weight: 12.000
Object#010 +Y	Value: 11.000	Weight: 45.000
Object#011 +Y	Value: 0.000	Weight: 0.000
Object#012 +Y	Value: 23.000	Weight: 29.000
Object#013 +Y	Value: 27.000	Weight: 36.000
Object#014 +Y	Value: 15.000	Weight: 43.000
Object#015 +Y	Value: 0.000	Weight: 0.000
Object#016 +Y	Value: 17.000	Weight: 1.000
Object#017 +Y	Value: 17.000	Weight: 38.000
Object#018 +Y	Value: 5.000	Weight: 4.000
Object#019 +Y	Value: 41.000	Weight: 33.000
Object#020 +Y	Value: 32.000	Weight: 18.000
Object#021 +Y	Value: 32.000	Weight: 3.000
Object#022 +Y	Value: 21.000	Weight: 38.000
Object#023 +Y	Value: 28.000	Weight: 11.000
Object#024 +Y	Value: 0.000	Weight: 35.000
Object#025 +Y	Value: 49.000	Weight: 13.000
Object#026 +Y	Value: 30.000	Weight: 29.000
Object#027 +Y	Value: 5.000	Weight: 3.000
Object#028 +Y	Value: 11.000	Weight: 32.000
Object#029 +Y	Value: 41.000	Weight: 26.000
Object#030 +Y	Value: 28.000	Weight: 21.000
Object#031 +Y	Value: 42.000	Weight: 47.000
Object#032 +Y	Value: 22.000	Weight: 9.000
Object#033 +Y	Value: 35.000	Weight: 29.000
Object#034 +Y	Value: 13.000	Weight: 26.000
Object#035 +Y	Value: 14.000	Weight: 47.000

Object#036 +Y	Value: 46.000	Weight: 46.000
Object#037 +Y	Value: 2.000	Weight: 19.000
Object#038 +Y	Value: 34.000	Weight: 30.000
Object#039 +Y	Value: 32.000	Weight: 34.000
Object#040 +Y	Value: 15.000	Weight: 8.000
Object#041 +Y	Value: 48.000	Weight: 48.000
Object#042 +Y	Value: 39.000	Weight: 5.000
Object#043 +Y	Value: 3.000	Weight: 0.000
Object#044 +Y	Value: 40.000	Weight: 45.000
Object#045 +Y	Value: 28.000	Weight: 20.000
Object#046 +Y	Value: 16.000	Weight: 22.000
Object#047 +Y	Value: 17.000	Weight: 40.000
Object#048 +Y	Value: 31.000	Weight: 4.000
Object#049 +Y	Value: 19.000	Weight: 49.000

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 143-154
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Experimental Comparison of Uninformed and Heuristic AI Algorithms for N Puzzle and 8 Queen Puzzle Solution

Kuruvilla Mathew and Mujahid Tabassum
Swinburne University of Technology(Sarawak Campus), Jalan Simpang Tiga, 93350, Kuching, Malaysia
kmathew@swinburne.edu.my and mtabassum@swinburne.edu.my

ABSTRACT

This paper compares the performance of popular AI techniques, namely the Breadth First Search, Depth First Search, A* Search, Greedy Best First Search and the Hill Climbing Search in approaching the solution of a N-Puzzle of size 8, on a 3x3 matrix board and for the solution of the classic 8 Queen Puzzle. It looks at the complexity of each algorithm as it tries to approaches the solution in order to evaluate the operation of each technique and identify the better functioning one in various cases. The N Puzzle and the 8 Queen is used as the test scenario. An application was created to implement each of the algorithms to extract results for each of the cases. The paper also depicts the extent each algorithm goes through while processing the solution and hence helps to clarify the specific cases in which a technique may be preferred over another.

KEYWORDS

Artificial Intelligence; N Puzzle Solution; 8 Queens Puzzle; Uninformed and Heuristic AI Techniques;

1 INTRODUCTION

Artificial Intelligence (AI) attempts replicating the human ways of reasoning in computing. As a full replication may not be approachable at once due to is magnitude and complexity, research now targets commercialisable aspects of AI towards providing "intelligent" assistive services to the human users [1]. Decision making in this paradigm involves evaluating a number of alternatives in different spatial configurations, environments and circumstances and to find better of the alternatives. It also involves decision making even when an ideal alternative is not derivable. This paper is therefore limited to comparing the implementations of the popular AI algorithms, namely Breadth First Search, Depth First Search, A*, Best First Search and Hill climbing algorithms for solving a sliding n-puzzle in an attempt to look at the better efficient of the algorithms for this case. To solve the sliding n puzzle problem, one moves a set of square tiles arranged randomly in a square board to arrive at a pre-determined order. The board has only one blank square and each tile can only move to the blank space adjacent to itself. Our aim in this paper is to apply the AI approaches to the case and compare their performances in the problem solving.

This paper is organized as follows. Section II looks at related work in the areas of research. Section III outlines the problem and our approach towards the solution. Section IV details the experiment and results whereas section 5 presents the summary and conclusion based on results observed in section IV.

2 RELATED WORKS

Brooks discussed about intelligence without perception [1] in a case where intelligence is not about individual sub-system decompositions but about parallel activity decomposers that interact directly with the world, with notions of peripheral and central systems fading away.

Drogoul and Debreuil presented a distributed approach in solving the N Puzzle [2], an approach based on decomposition of a problem into independent sub-goals, which is in turn decomposed into agents that satisfy the sub-goals. This approach is used to demonstrate emergent solutions and for solving for very large N Puzzles.

Kumar et. all presented summary of results from a parallel best first search of state-space graphs [3]. The paper looks at several formulations of A* Best First

algorithm and discussed how certain searches are better or lesser suited for some search problems.

Blai Bonet and Héctor Geffner study a family of heuristic planners [4], applying them in the context of Hill Climbing and Best First search algorithms and tested on a number of domains to analyse the best planners and the reasons why they do well.

Korf, R. E. discussed a study on Depth First search as asymptotically optimal exponential tree searches [5]. They discuss that the Depth First iterative-deepening algorithm as capable of finding optimal solution for randomly generated 15 puzzle.

Korf, R. E. presented a Linear Best First Algorithm [6], exploring nodes in the best first order, and expands fewer nodes. This works on the sliding puzzle with reduced computation time, but with a penalty on solution cost.

Russel. S and Norvig. P explored the detailed concepts of AI using intelligent agents and multi-agent systems search algorithms etc. in the distributed problem solving approach in AI in their book[7]. The book gives good insight towards the modern approaches in AI.

3 THE PROBLEM

3.1 The N Puzzle

The sliding puzzle is a simple but challenging case for demonstrating artificial intelligence concepts. It involves having a set dimension of puzzle space (usually 3x3 for an 8-puzzle) and denoting the dimensions as N being the columns and M being the rows. In the puzzle space, there is a random arrangement of cells/blocks, with one empty space that enables adjacent cells/block to slide into them. Since the pieces are square blocks, only the top, bottom, left and right blocks adjacent to the empty space may slide into its place. The cells can either be numbered or printed with a fragment of the whole picture that the rearranged puzzle should show[2].

3.2 The 8 Queen Problem

The classic 8 Queen puzzle is a very similar puzzle as the N puzzle. This involves an 8 x 8 game board, commonly used in the game of chess. The challenge is to place 8 queens of a chess game so that no queen is facing another, i.e. not in the same line, either diagonally or directly. The initial state of the puzzle starts with an empty board and the goal state has all the 8 queens placed on the board, not facing each other, as shown in the figure.

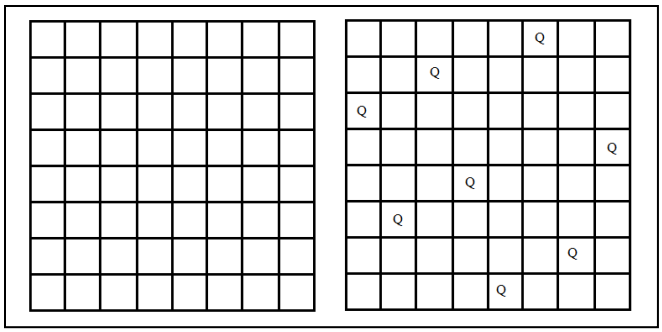


Figure. 1. The 8 Puzzle, initial and goal state

3.3 The Problem Formulation

Problem formulation is done by analysing the environment of the puzzle and deriving its characteristics using PEAS, as in Table 1.

3.4 Approaching the Solution

3.4.1 N Puzzle

For solution searching, it would be most useful to distil the possible arrangements of tiles as individual States. Thus, each State shows a possible combination of tile positions within the given puzzle space. The collection of all possible States is called the State Space. With the increase of N or M of the puzzle, the size of the State Space shall increase exponentially.

In every state, the empty space position determines which States can be transitioned to. For instance, when the empty space is in the middle of a 3x3 puzzle, tiles at the Top, Bottom, Left or Right can move into it. But if the empty space is at the top left corner, only the right or bottom tiles can slide into it.

Thus, after each slide, a new State is transitioned into. If puzzle is to begin with an Initial State of tile arrangements, then its subsequent transitions into other States can be represented by a Graph. An example of this can be seen in Figure 2. A search attempt will need to begin with an Initial State and a Goal State to

achieve. As puzzle traversal can often pass through the same state at different intervals. We will consider the instances of decisions as nodes. By aligning the node arrangements to start from the Initial Node to possible routes leading to the Goal nodes, a search tree is formed, as we see in Figure 3. The algorithms explored in this paper will traverse the Search Tree in different ways to find the Goal State from the Initial State.

TABLE I. ENVIRONMENT ANALYSIS – N PUZZLE

Sl No	Environment Characteristics of Puzzle	
		Description
1.	Performance	Arrangement of tiles/cells/blocks in the whole puzzle space. Main performance gauge is from the least number of moves to solve the puzzle.
2.	Environment	Puzzle space determined by N (columns) and M (rows), always with a single empty space for tiles to slide into. Numbers range from 1 to (N*M)-1. Initial state arrangements must be derived from Goal state arrangement or else there will not be possible solutions.
3.	Actuators	Tiles are moved into the empty space, either from Top, Bottom, Left or Right of the empty space.
4.	Sensors	Fully software, so the agent will have full view of the puzzle space.

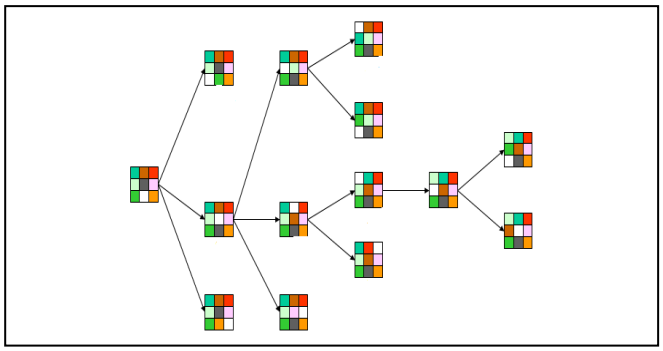


Figure. 2. Puzzle transition graph for a 3x3 N puzzle

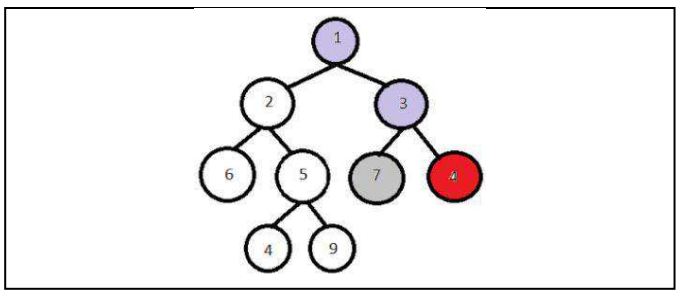


Figure. 3. The Search Tree

3.4.2 The 8 Queen Puzzle

The 8 queen is very similar to the N puzzle, the solution space defined by the 8x8 game board on which queens can be placed. Queens can move vertically, horizontally or diagonally, both in the forward and backward directions and the objective is to place all the 8 queens on the board so that they don't face each other. Unlike the N puzzle, the state space in this puzzle is limited to the 8x8 board and its combinations. A solution state will start from the initial state and transition through one of the possible states and will be considered complete when it reaches the goal state. An example search tree is shown in figure 4. The algorithms explored in this paper will traverse the Search Tree in different ways to find the Goal State from the Initial State.

TABLE II. ENVIRONMENT ANALYSIS – 8 QUEEN PUZZLE

Sl No	Environment Characteristics of Puzzle	
		Description
5.	Performance	Arrangement of 8 queens in an 8x8 puzzle space of an 8x8 game board. Main performance gauge is from the least number of moves to solve the puzzle.
6.	Environment	Puzzle space determined by the 8x8 game board. Initial state is an empty board and final state is the board with 8 queens placed, none facing another.
7.	Actuators	Each queen is placed, one after the other on a square on the board in a position that does not face any other.
8.	Sensors	Fully software, so the agent will have full view of the puzzle space.

3.5 The Ecosystem

This paper attempts to demonstrate the implementation of Breadth First Search, Depth First Search, A*, Best First Search and Hill Climbing algorithms for solving a sizeable sliding puzzle and for solving an 8 queen puzzle. An Object Oriented Approach is applied for the N Puzzle in order to modularize the application so that different search algorithms (encapsulated in different classes) can be made to work with the same interface. For the N puzzle, the input parameters and output results are to be presented in a Text File. We will also assume that the input parameters co ntain only valid configurations, implying that the Initial State is solvable, only if it can be done in

reverse, to reach the initial state starting from the Goal State. Otherwise, the N puzzle is unsolvable. The 8 Queen

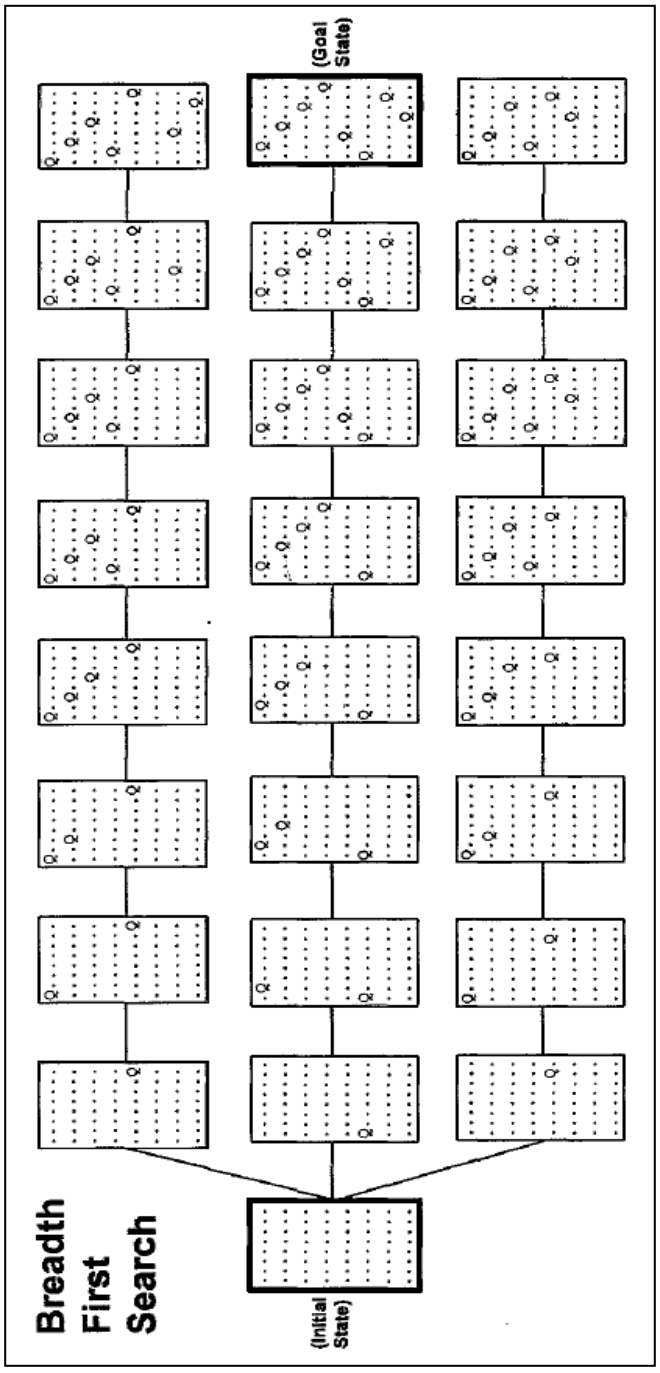


Figure. 4. Breadth First Search – 8 Queens

3.6 System Design Overview

The flow chart in Figure 4 illustrates the top-level design of the application used to run the tests. The Search Algorithms will have their own unique flowcharts, which will be detailed during their own specific sections of this paper.

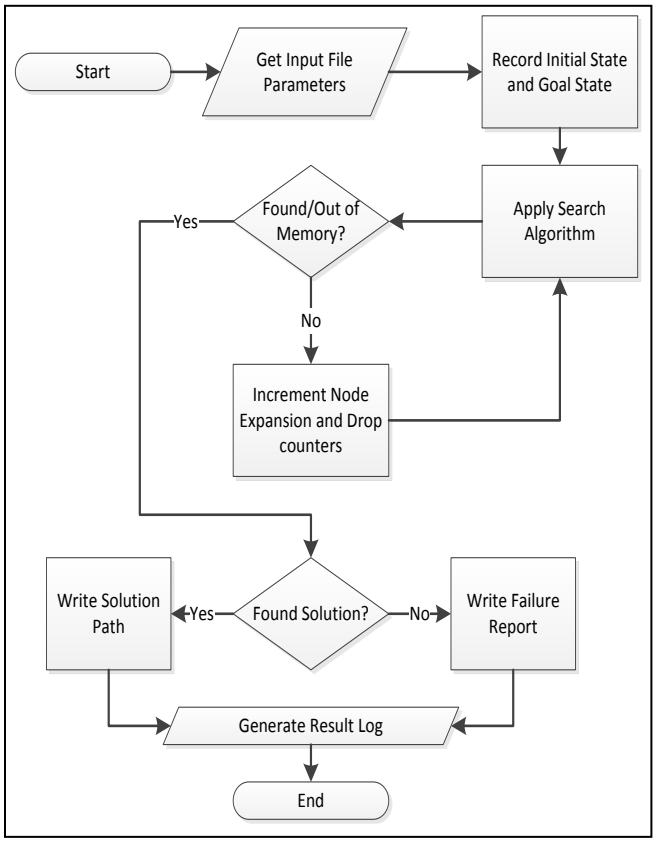


Figure. 5. Flow chart test application application

A separate file reading and writing facility will help in parsing the input parameters as well as outputting the process logs as a new results file. This design enables the different search methods to use the same interface for extracting parameters, processing and generating the results log. The application was developed in C# using the Visual Studio .NET 2013 IDE.

4 The Algorithm Implementations and Experiments

There are 5 different search algorithms that are explored in this paper, namely the Breadth First Search, Depth First Search, A*, Best First Search and Hill Climbing. The first two are Uninformed techniques while the others are Heuristics based. The first part will discuss the N Puzzle solution and the second part will discuss 8 Queen Puzzle solution. The theory, software implementation and results of each of

the AI techniques in consideration are explained in the following sections.

4.1 N Puzzle

4.1.1 Breadth First Search

The Breadth First Search is an uninformed type algorithm, so it does not start will full knowledge of the entire State Space. Instead, it builds its own memory of the State Space by remembering all the explored nodes that it passes through. In addition, the only way for the BFS to know when to stop is by finally arriving at a node that has the same state as the Goal Node. The BFS traverses the Search Tree by uncovering the frontier one level at a time. The sequence is illustrated in Figure 5.

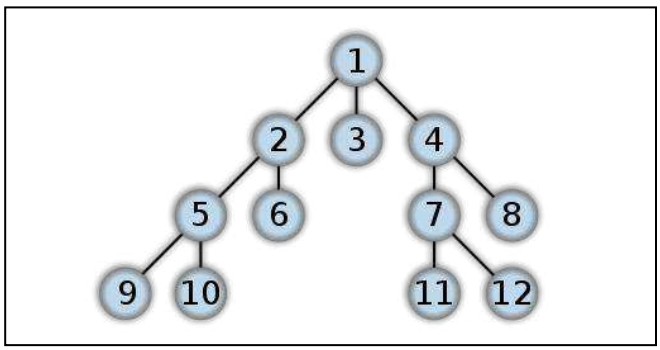


Figure. 6. Breadth First Tree Traversal – N puzzle

The algorithm strategy can be explained plainly in the following steps:

Step 1: Begin by making the Root Node as the Active Node.
Step 2: Add the Active Node into the Solutions List.
Step 3: Check if the Active Node is the Goal Node.
Step 4: If it is the Goal Node, then go to Final Step.
Step 5: If not, derive all unexplored successors and add them into the Queue.
Step 6: If the queue is empty, go to Final Step.
Step 7: Else, take the next element on the Queue as Active Node, and repeat Step 2.

Final Step:
Trim all Solution List nodes that do not lie between the Goal Node and the Root Node.

Examining the steps, it can be concluded that the Queue used to arrange the order of nodes examined, determines the horizontal-first motion of expansion (First In, First Out). The configuration input parameters are maintained for all 5 algorithm tests to ensure a level testing condition. An excerpt from the results (shown as the following) shows that for a 4 move solution, it has expanded 23 nodes. The reason for the 22 drops was that the Successor trimming also removes the state that it the active node is transitioning from. This shows that while BFS can find the shortest path to the Goal State, it has to expand quite a lot of nodes and remember them in the process.

Result Summary:

Nodes Expanded = 23
Nodes Dropped = 22
Solution Length = 4

4.1.2 Depth First Search

The Depth First Search algorithm shares the similar mode for Goal identification and Search Tree traversal, but not in its motion of expansion. Instead of going Horizontal first, the DFS method chooses a branch and expands it continuously until it reaches a dead end. If a goal has not been found, it will backtrack to the next available alternate branch and repeat the process. This motion is illustrated in Figure 6.

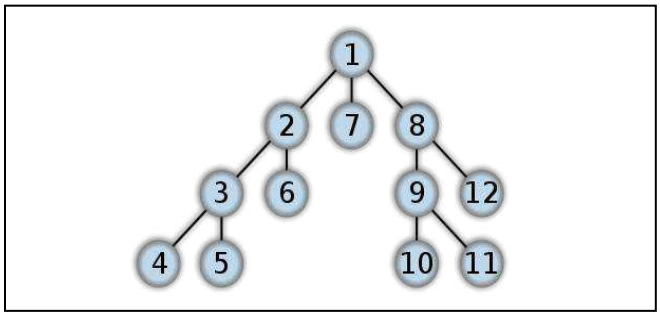


Figure. 7. Depth First Search Tree Travesal

As mentioned in the BFS strategy, the motion of expansion is influenced by the type of data structure used. Instead of First-In-First-Out, DFS will use a Stack (First-In, Last Out). The rest of the algorithm is exactly the same as BFS. But pure DFS has a high possibility of infinite depth, so it will need a depth limit to force the agent to backtrack. This will transform the DFS into Depth-Limited Search.

As the difference between DFS and BFS is only the frontier data structure, the programmatic flow chart for DFS is exactly the same as BFS, except that the Frontier is now a Stack where BFS uses a queue. While the Depth Limited Search do not consume as much memory as BFS when it expanded 4743 but dropped 4832 nodes, it takes variable amount of time to search, depending on whether it found a branch with the goal at the end. Also, the DFS is not a complete method because it can get lost expanding a branch without reaching an end or goal. In this experiment, it took much longer than BFS with 4743 nodes expanded and the solution is inefficient with 4730 moves. The summary result, using the same configuration as the BFS test is as the follows:

Nodes Expanded = 4743
Nodes Dropped = 4832
Solution Length = 4730

4.1.3 A* Search

The A* Search is a Heuristics-based Informed search algorithm. The term 'Informed' refers to the use of State Space awareness in the form of the Heuristics Function. All informed search methods rely on knowing how much a particular node have in common with the Goal Node. Put simply, Heuristics H(n) = Difference (Goal State,Current State). The higher H(n) is, the less ideal the alternative.

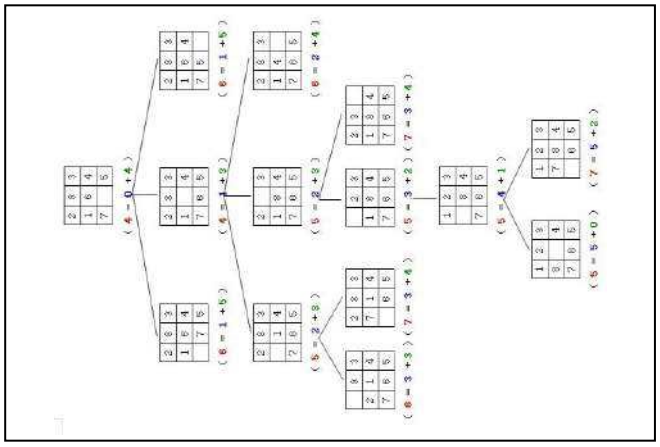


Figure. 8. A* traversal of the 8-Puzzle search tree

A heuristics search algorithm will basically use the Breadth First Search's horizontal motion of expansion, but will only expand leaves that are most 'ideal', i.e. having the smallest H(n). The A* algorithm takes this further by incorporating the Path Cost into the Heuristics function for every node. So now, not only are nodes chosen according how closely they resemble the Goal State, but also how far they are from the Root Node. An example of the A* traversal can be seen as Figure 7. The Blue numbers are the Path Costs and the Green numbers are the Heuristics function.

The overlying structure of the A* search algorithm is the BFS, but significant changes has been made to the generation of successors and queue management which involves the heuristics and path cost calculations. These details are reflected in the flowchart shown in Figure 8.

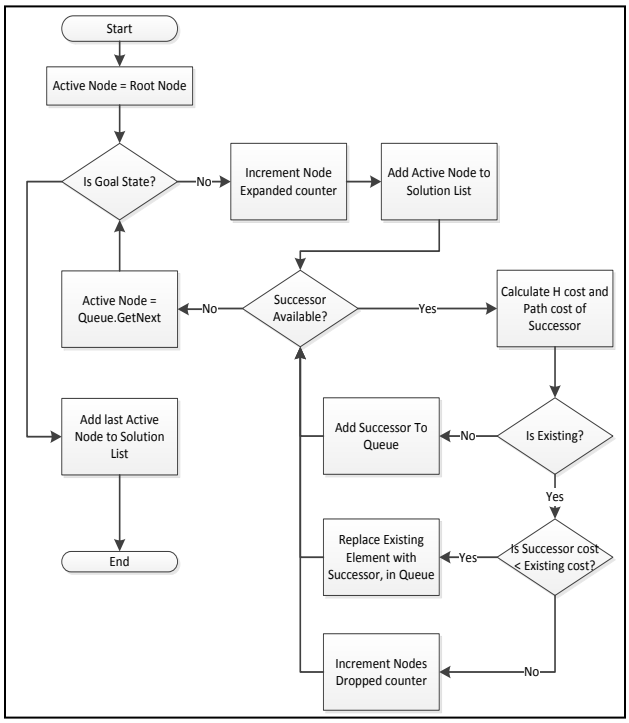


Figure. 9. A* Search Algorithm Implementation

It is observed that the A* algorithm managed to find the best solution like BFS in 4 steps, but has managed to do so with only 4 nodes being expanded. Also, it has managed this with only generating 9 nodes and dropping 3, presumably due to them being too far different from the Goal State. It is more memory efficient compared to DFS and is faster than BFS but requires more computation power as each node is not only evaluated for it matching the Goal State, but also how much difference between them and how far down the tree it has gone.

The summary of the test results, using the same configuration script for BFS and DFS, are as the follows:

Nodes Generated=9
Nodes Expanded = 4
Nodes Dropped = 3
Solution Length = 4

4.1.4 Best First Search

The Best First Search is also a type of informed search method, in that it relies on the Heuristics Function for its working. Unlike the A* method, which requires higher computational capacity, the Best First Search only compares the heuristic value of nodes, ignoring the path cost. Theoretically, this will cut down the required computations of the A* by half, in expense of the possibility of having endless loops or very expensive total path cost of solutions. Other than that, it operates over a Breadth First Search's motion of expansion.

The Best First Search is also known as the Greedy Best First Search. When it expands nodes and gets a list of possible successors that were not explored before, it will derive the heuristics value of each of the successors and pick the best one to expand. The other leaves will be unexplored. It is not optimized, as the cheapest path may involve going through heuristically suboptimal nodes but yet have less path cost. Other than this, the Greedy Best First Search has the same mode of operation as the A*.

The Greedy Best First Search only differs from the A* Search at the point of determining the fate of possible Successors. Instead of taking into account calculations of path cost, it directly decides on which leaf to expand by using only the H function. In this particular instance, its findings and performance match the A* test, because the solution only spans 4 steps. On longer solutions and more complex Search Trees, the same cannot be expected of the Greedy BFS search as there is a higher possibility of it getting stuck on a lost branch with initially low H values. By supplying the same configurations as the previous tests, the following results are observed:

Nodes Generated=9
Nodes Expanded = 4
Nodes Dropped = 3
Solution Length = 4

4.1.5 Hill Climbing Search

Similar to the Greedy Best First Search, the Hill Climbing method borrows from the A* Search but simplifies it even further. Where the Greedy search only bases its decisions on the Heuristics Function, Hill Climbing works in the same way but totally disregards memory of explored nodes. Therefore, it travels down the Search Tree by selecting the successor with the cheapest heuristics value, without retaining memory of explored states.

This will ensure that the heuristics technique functions with minimal use of memory, least computation possible but still retain the advantage of an informed method of solution finding. The downside of Hill Climbing is that due to the absence of memory, resulting in the possibility of repeating the same states and getting stuck in some state of local maxima.

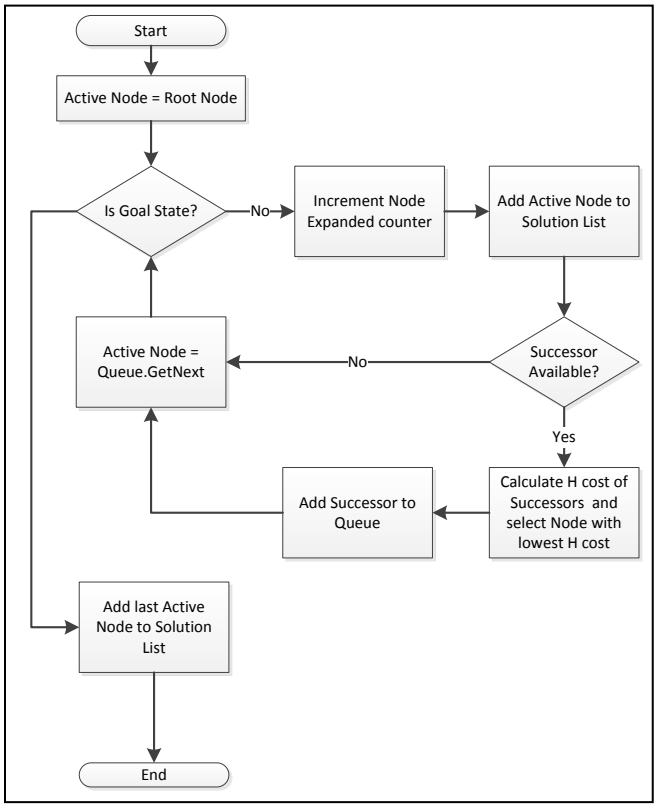


Figure. 10. Hill Climbing Search Algorithm

While the changes between the Greedy BFS and A* are only in the selection of successors, Hill Climbing

greatly simplifies the Successor generation and selection process, as illustrated in the flow chart shown in Figure 9.

From the results, it seems that the Hill Climbing technique manages to find the solution in the same time as Greedy BFS and A*, but dropped more nodes than the other two. Also, it shares even more risk of failure than Greedy search in cases of longer solutions or more complex problems. By supplying the same configurations script, the following results are observed:

Nodes Generated=5
Nodes Expanded = 4
Nodes Dropped = 7
Solution Length = 4

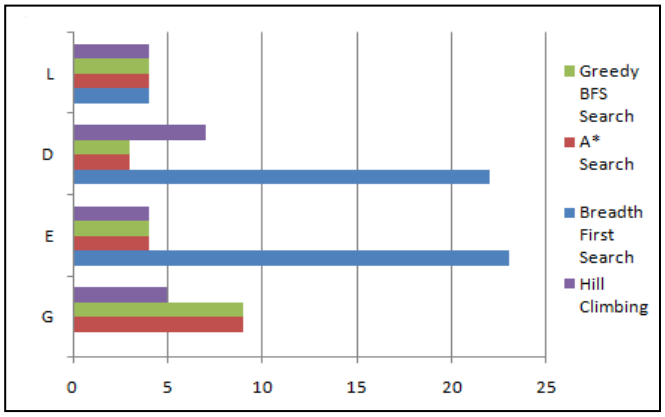


Figure. 11. Comparisson of the Search Algorithm Observations

4.1.6 Conclusion and Future work

A summary of the comparison of the 5 AI techniques are as observed in table II. The findings listed in the table are plotted in a comparison chart in figure 10. The values of Depth first search has been avoided in the chart since the value is well out of range.

TABLE III. SUMMARY OF RESULTS

Sl No	Summary of AI Algorithm Results on the N Puzzle				
	Algorithm	*G*	*E*	*D*	*L*
1	Breadth First Search	NA	23	22	4
2	Depth First Search	NA	4743	4832	4730
3	A* Search	9	4	3	4
4	Greedy BFS Search	9	4	3	4
5	Hill Climbing	5	4	7	4

[a] G: Nodes Generated, E: Nodes Expanded, D: Nodes Deleted, L: Solution Length

The following conclusions can be made based on the results as we observe in table II and figure 10.

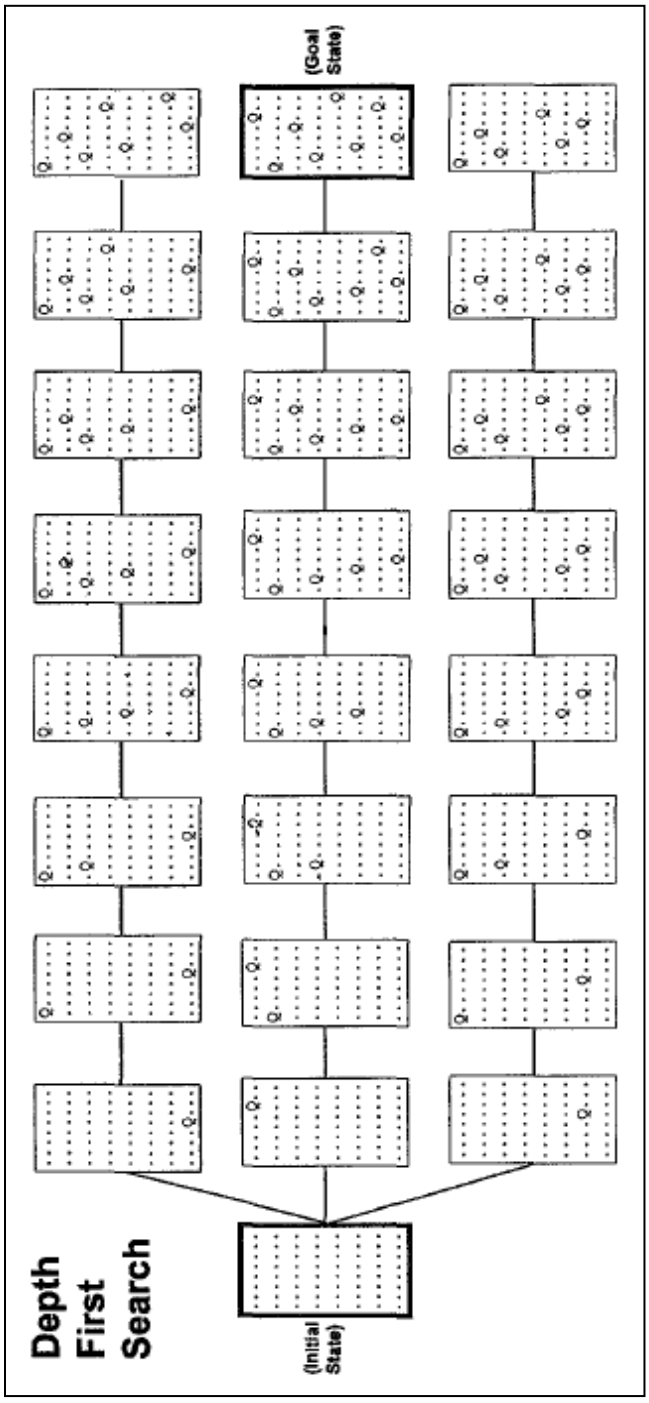


Figure. 12. Depth First Search – 8 Queen Puzzle

- **Breadth First Search**: More complete and concise uninformed technique that manages to find the best solution using minimal

computation but suffers from intensive memory use. BFS will be most recommended for small dimension sliding puzzles, but for the expandable N*M type with exponentially scaling complexity, BFS becomes less efficient and can run out of memory before completing long solutions.

- **Depth First Search**: Explores branches individually before backtracking. It is actually good for long solutions, but only if it is lucky enough to start on a branch with a possible Goal State. The sliding puzzle has single definite solutions, so it will not be optimal to use DFS. A complete search method is preferable.
- **A* Search**: By utilizing both heuristics values and path costs, the A* search manages to find the shorted solutions with moderate memory and time performance. A lighter method will be preferable if it was a small puzzle. However, this dynamically up-scaling puzzle will benefit more from the A*'s complete and heuristic approach.
- **Greedy BFS**: The Best First Search also uses the A*'s heuristic method, but negates the path cost. For this particular problem, the Best First Search is quite effective, matching the performance of the A* search on shorter solutions. For longer solutions, the A* is safer but requires more computations.
- **Hill Climbing**: This is the most memory efficient heuristics approach but has high risks of failure due to local maxima issue and disregard of memory, especially for moderate to long solutions. This is not recommended for the sliding puzzle problem as there is a high chance for the failure conditions to present themselves.

In conclusion, the best approaches to apply to this dynamically scaling sliding puzzle will either be the A* or the Greed Best First Search. Greedy BFS is more memory efficient and matches the performance of A* for shorter solutions. For longer and more complex solutions, the A* is the best choice, in order to avoid possibilities of getting a suboptimal solution due to the Greedy BFS's disregard for path cost computations.

4.2 The 8 Queen Puzzle

Additional study on the 5 AI techniques was carried out to generate the solution for the classic 8 Queen Puzzle. Since the algorithms have already been explained in the sections above, this section will focus on the application on the 8 Queen Puzzle. The experiments will compare the convergence of the algorithms on the 8 Queens puzzle to extend the results of the AI techniques on the solution for the N puzzle

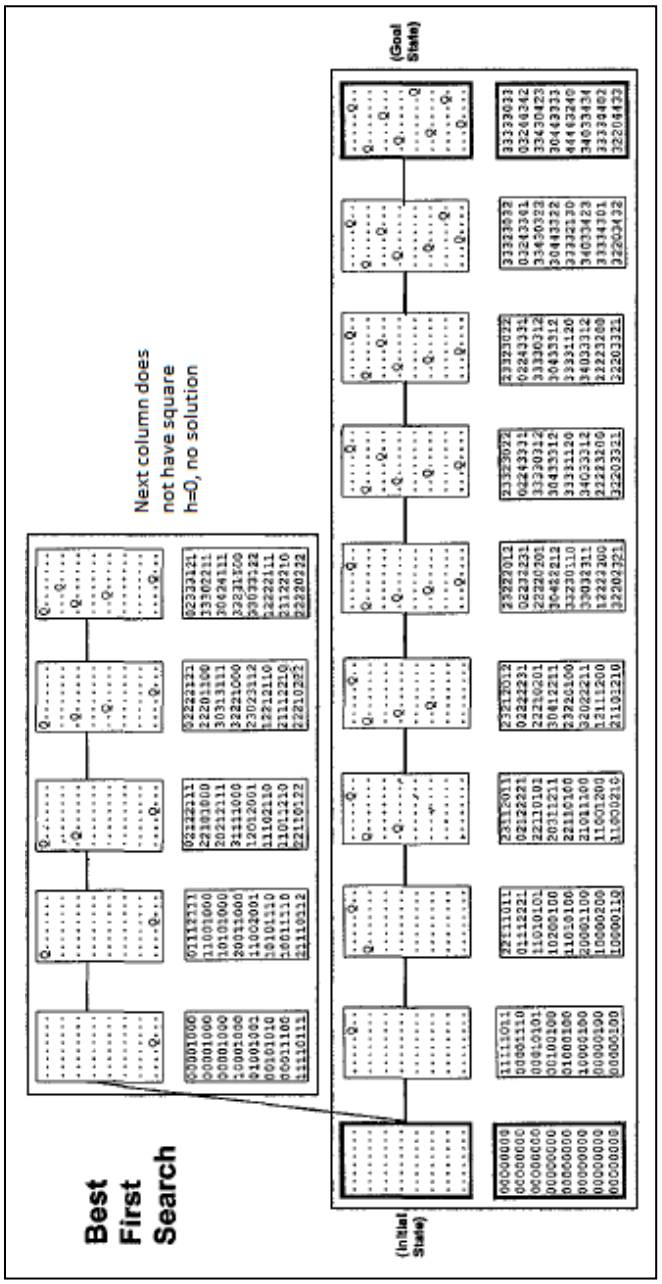


Figure. 13. Best First Search – 8 Queen Puzzle

4.2.1 Breadth First Search (BFS)

Breadth first search's initial state is an empty board. After the initial step, the next step would be to place

the first queen. This is done row by row. For example, the first initial queen will be placed at row 1 column 1. The next initial queen is placed at row 1 column 2 etc. The algorithm then checks each square row by row to see if there's any squares which do not crash with the initial queen placed or any other queens on the board, whether upwards, downwards, left, right or diagonally. When that square is discovered, a queen will be placed there and the algorithm continues to search for such squares until all 8 queens are placed. If 8 queens fail to be placed on the board, the board will be cleared and a new initial queen will be placed. The goal state is reached when 8 queens have been placed on the board without any queen facing any of the other queens on the board.

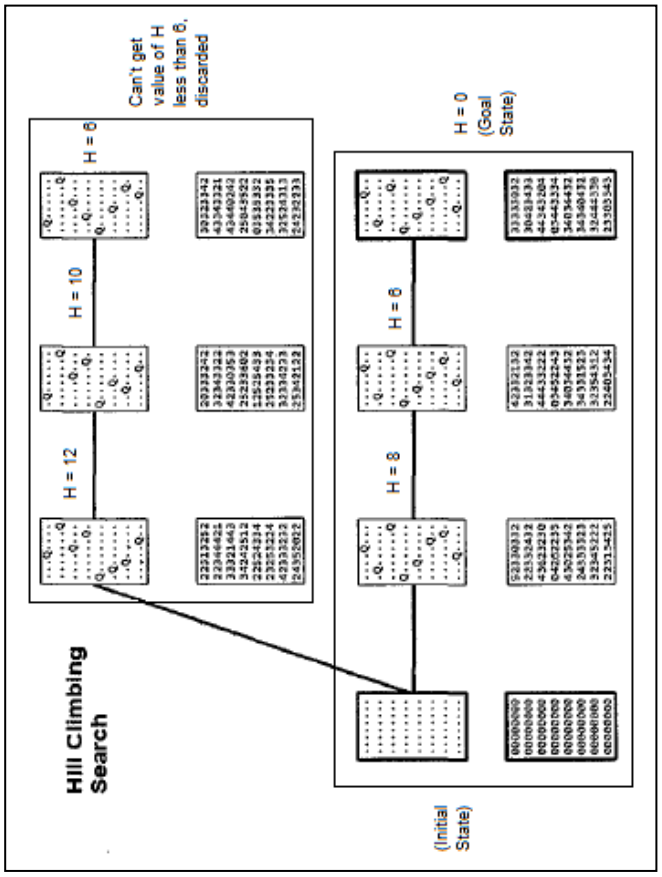


Figure. 1. Hill Climbing Search – 8 Queen Puzzle

In the diagram for Breadth first search (figure 4), there are three examples given ach with a different initial queen. From the diagram shown below, when the queen is placed on row 6 column 1(the middle), board with 8 queens has been successfully generated. In other cases, queens failed to be placed on the board because there were no more open squares to place the remaining queens without facing any of the existing queens on the board.

4.2.2 Depth First Search (DFS)

Depth first search is similar to breadth first search. The only difference is that when placing the initial queen, in goes column by column. And also, the search for squares with no crashes is done column by column.

In the diagram for Depth first search (figure 12), there are three examples given ach with a different initial queen. For the experiment, the depth first search produces a board with eight queen when the initial queen is placed on row 1, column 6.

4.2.3 Best First Search

For best first search, the *h(n)* will be the total number of crashes for the board. In this algorithm, the crash value for each square is calculated. For example, if there are no queens that can attach that square, it will be 0. If there are two queens, one above it and one to the left of it, the crash value for that square will be 2.

The best first search will insert an initial queen. Then, it will check to see if there are any squares containing 0 crash value for each column. If there is a square containing 0 crash value (h=O), it will insert queen in that square and move on to the next column. However, if after checking all the squares in that column and there are no squares containing 0 crash value, the algorithm will reset the board and start anew with a new initial queen position. Therefore, the algorithm can obtain 8 queens by making sure that the value remains as O(lowest possible value) by having no crashes between each queens at all. Figure 13 shows the success path where there are no crashes between queens as well as a fail path. As shown, in column six of the latest state, there are no squares with h=O and choosing a square with crash value higher than 0 will cause the board's number of crashes to increase. Therefore, the board is discarded.

4.2.4 Hill Climbing Search

Hill climbing search for this 8-Queen puzzle, in order to reach the goal state where h = 0, it will continue to loop to find moves in the direction of decreasing *h(n)*.

It will terminates when there is no lower *h(n)* than the previous ones. By looking at the diagram (Figure 14), the hill climbing search will randomly generate 8 random placement of the queen in the 8x8 board after the initial state it will then calculate the *h(n)* and then during the next state it will swap the Q position column by column in search of the *h(n)* that is lesser than the previous *h(n)* until it reach *h=0*.

Hence, based on the evaluation function *f(n)=h(n)*, so the results will be *f(n)*1=0. The board will terminate if there is no *h(n)* that is less than the previous *h(n)*. When board is clear, a new random placement of the queens is placed again and the process is repeated until it reaches the goal state.

4.2.5 A* Search

The goal state for every a-Queen solution is *h(n)* is equal to 0. Thus, for A* search, the algorithm used can generate 8 random placement of the queen in the 8x8 board without any conflict between the Queens. As shown in the diagram below, once the A*search found the first solution that has h = 0, it will choose that as the goal state even though there are other solution available.

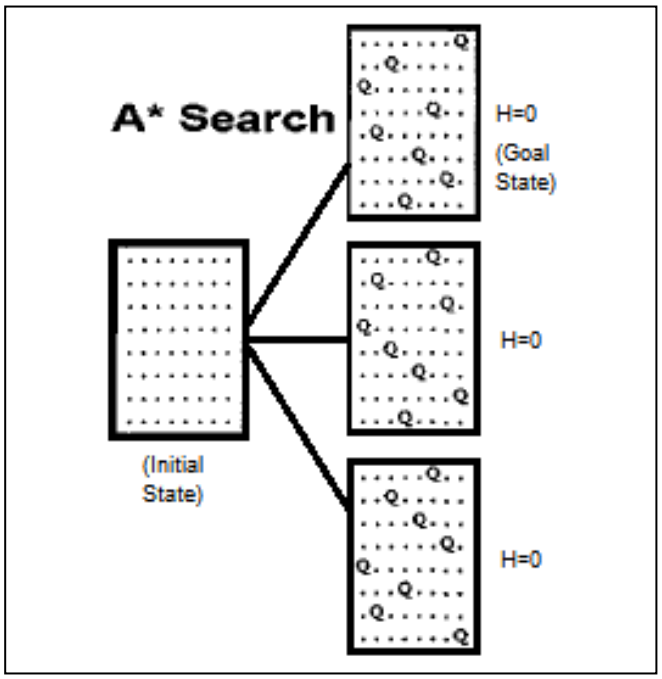


Figure. 1. A* Search – 8 Queen Puzzle

According to the evaluation function, *f(n)=g(n)+h(n)*. *h(n)* is the number of conflicts from goal state and g(n) is the number of conflicts from initial state. So at the goal state, when *h(n)=0*, and since A* search places every Q without any conflict between1each queen, hence *g(n)=0*. Based on the evaluation function, *f(n)* will be 0.

4.2.6 Conclusion and Summary

After exploring the 5 algorithms to solve the 8-Queen puzzle, we have come to the conclusion that A* is the best search algorithm for the 8 Queen Solution. Breadth First Search, Depth First Search and Best First Search can only generate one board which same each time it is run. Hill Climbing on the other hand, attempts to swap the position of the queens on the board and may discard many boards before it discover correct solution. A* algorithm is able to reach the goal state of 8 queens by positioning 8 queens on the board so that there is only one queen per column and one queen per row where the total crashes between queens is 0. In addition to this, it will guarantee to produce a board which solves the 8-Queen puzzle. The board generated can also be different each time the A* algorithm is used and need not be static like those generated using BFS, DFS and Best First Search. The summary of the result is as shown in Table IV.

TABLE IV. ENVIRONMENT ANALYSIS – 8 QUEEN PUZZLE

Comparison between algorithms					
	BFS	**DFS**	**Best First**	**Hill Climbing**	**A***
Min Cost by	-	-	No of crashes of placed queens	No of crashes of placed queens	No of crashes of placed queens
Back-tracking	No	Yes	Yes	No	Yes
Complete	Yes	No	No	No	Yes
Time	$O(b^{d+1})$	$O(b^m)$	$O(b^m)$	$O(\infty)$	Exponential with path length
Space	$O(b^{d+1})$	$O(b^m)$	$O(b^m)$	O(b)	Stores all nodes
Optimal	Yes	No	No	No	Yes

b stands for branching factor
d stands for depth
m stands for maximum path length

Summary, Conclusion and Future Work

As we have seen in the studies of on the N Puzzle as well as the 8 Queens puzzle, the A* is seen to perform

best, with its heuristics and faster convergence at cheaper cost and therefore conclude that the A* search is best suited for solving issues of this nature. In both cases, A* gave the best results and also has the possibilities to be applied towards issues of growing dimensionalities.

Further studies can be carried out to measure the rate of convergence and do a parallel comparison for both puzzles. Also, further improvement can be implemented in the form of morphing algorithm that can opt to repeat the search in A* if the initial attempt with Greedy BFS turns out unusually long solution. Also, a fail-safe function to switch from A* to Greedy BFS can be implemented, to kick in during events of critically reduced memory conditions.

5 Acknowledgements

We acknowledge the contributions of Mark Tee Kit Tsun Carmella Sim Lee Yoong, Carmen and Lo Xiang for the contributions in coding for the algorithm implementations.

6 References

[1] Brooks, R. A., “Intelligence without representation” Artificial Intelligence, Volume 47, Issues 1–3, January 1991, Pages 139-159, ISSN 0004-3702

[2] Drogoul, A., and Dubreuil, C. "A distributed approach to n-puzzle solving." Proceedings of the Distributed Artificial Intelligence Workshop. 1993.

[3] Kumar. V. Ramesh, K. and Rao, V. N. "Parallel Best-First Search of State-Space Graphs: A Summary of Results." AAAI. Vol. 88. 1988.

[4] Bonet, B. and Geffner, H. “Planning as heuristic search,” Artificial Intelligence, Volume 129, Issues 1–2, June 2001, Pages 5-33, ISSN 0004-3702

[5] Korf, R. E. "Depth-first iterative-deepening: An optimal admissible tree search." Artificial intelligence 27.1 (1985): 97-109.

[6] Korf, R. E. “Linear-Space Best-First Search,” Artificial Intelligence, Volume 62, Issue 1, July 1993, Pages 41-78, ISSN 0004-3702

[7] Russell, S. and Norvig, P. “Artificial Intelligence: A Modern Approach Author: Stuart Russell, Peter Norvig, Publisher: Prentice Hall Pa.” (2009): 1152

International Journal of Digital Information and Wireless Communications (IJDIWC) 4(1): 155-168
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Performance Evaluation of Adaptive LDPC Coded Modulation Cooperative Wireless Communication System with Best-Relay Selection

Ahmed S. Mohamed, Mohammed Abd-Elnaby, and Sami A. El-dolil
Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt,
Ahmed.abouhiba@telecomegypt.com, moh_naby@yahoo.com.com, msel_dolil@yahoo.com

ABSTRACT

In this paper the performance of a cooperative wireless communication system based on combined best relay selection (BRS) and adaptive LDPC coded modulation (ACM) scheme is investigated. These investigations are focused on evaluating the performance of the proposed cooperative wireless communication system over independent non-identical Rayleigh fading channels in terms of bit-error rate (BER) using MATLAB® computer simulations and comparing the system performance with ACM direct transmission and ACM cooperative with single relay. The simulations results show that the proposed cooperative scheme achieves lower signal-to-noise ratio (SNR) values for desired bit-error rate (BER) and high spectral efficiency as compared to ACM direct transmission and ACM cooperative with single relay.

KEYWORDS

Cooperative communication, LDPC, ACM, best-relay selection, single relay cooperative, Rayleigh fading

I. INTRODUCTION

Transmission over wireless channels suffers from random fluctuations in signal level known as fading and from co-channel interference [1]. Diversity is a powerful technique to mitigate fading and improve robustness to interference. In classical diversity techniques, the data signal is conveyed to the receiver over multiple (ideally) independently fading signal paths (in time/frequency/space) [2]. Appropriate combining at the receiver realizes diversity gain, thereby improving link reliability. There are several approaches to implement diversity in a wireless transmission. Multiple antennas can be used to achieve diversity. But multiple antennas are not always available or the destination is just too far away to get good signal quality. Recently, cooperative communications for wireless networks have gained much interest due to its ability to mitigate fading in wireless networks through achieving spatial diversity, while resolving the difficulties of installing multiple antennas on small communication terminals. The basic idea in cooperative communication is that in addition to the direct transmission from the transmitter to the receiver, there can be other nodes, which can be used to enhance the diversity by relaying the source signal to its destination, hence forming a virtual multiple-input multiple-output (MIMO) system. In a cooperative communication system, users act as information sources as well as relays. There are two main cooperative methods: amplify-and-forward (AAF) (non-regenerative relays) and decode-and-forward (DAF) (regenerative relays) methods. In the AAF method, the relay receives a noisy version of the signal transmitted by the source and then amplifies its received signal and re-transmits it to the destination [3]. In the DAF method, the relay decodes the noisy version of the signal transmitted by the source and then re-encodes and re-transmits it to the destination.

In cooperative networks with multiple relays, cooperative diversity protocols can be generally categorized into fixed and adaptive relaying protocols. Relay selection is attractive because of its high performance, efficient use of power and bandwidth resources, and simplicity [4]. The best-relay selection scheme for cooperative networks has been introduced in [5] and they called it opportunistic relaying. According to opportunistic relaying, a single relay among a set of N relay nodes is selected, depending on which relay

provides for the "best" end-to-end path between source and destination. The authors in [5] showed that this scheme has the same diversity order as the cooperative diversity using space-time-coding [6] in terms of the outage probability for both decode-and-forward and amplify-and-forward schemes. However, this important result was given using semi-analytical asymptotic analysis at high SNR only (without deriving a closed-form expression for the outage probability).Many different schemes for single relay selection have been proposed in [5], [7]–[9].

The authors in [10] analyzed the adaptive DAF relaying technique where among N relays that can participate, only k relays ($k < N$), with good channels to the source decode and forward (retransmit) the source information to the destination. The authors in [10] proved that increasing the number of potentially participating relays, N, does not always decrease the outage probability. To improve the outage probability performance, the authors in [11] suggested that only the best relay among the decoding group, $Ð$ will send another copy of the source signal to the destination. Hence, the total number of channel (or time slots) needed is reduced from $k+1$ to two only. For this proposed scheme, the authors in [11] derived the high-SNR outage probability approximation and they showed that it outperforms distributed space-time codes for networks with more than three relaying nodes. The authors in [12] presented a performance analysis for the AAF cooperative communications with relay selection and derived a closed-form expressions for the average SER performance for BPSK, M-PSK, and M-QAM signals, also derived a closed-form expression for the outage probability, and an analytical expression for the average end to- end SNR gain obtained from relay selection. The authors in [13] evaluated the performance of coded modulation scheme based on LDPC codes considered direct transmission over additive white Gaussian noise (AWGN) and flat Rayleigh fading channels. The authors in [13] presented a simple adaptive LDPC-coded modulation scheme for direct transmission over flat slowly-varying Rayleigh fading channels. In this scheme, six combinations of encoding and modulation pairs are employed for frame by frame adaptation and the spectral efficiency varies between 0.5 and 5.0 bits/symbol/Hz during data transmission. Their Simulation results show the power and spectral efficiency of coded modulation scheme based on LDPC codes, also their results show that the adaptive LDPC-coded modulation has the benefit of offering better spectral efficiency while maintaining an acceptable error performance.

The authors in [14] proposed an energy efficient cooperative MIMO technique where LDPC code is used as an error correcting code. Their simulations results show that the cooperative scheme outperform SISO scheme in the presence of LDPC code.

In this paper, we propose a cooperative wireless communication system based on combined best relay selection and adaptive LDPC coded modulation (ACM) scheme. Our investigations are focused on studying the performance of the proposed system over independent non-identical Rayleigh fading channels. The main objective of this scheme is to achieve higher spectral efficiency while guaranteeing the same diversity order as that of the conventional cooperative scheme.

To improve spectral efficiency, we introduced adaptive LDPC coded modulation (ACM) at source node which provides multiple coded modulation transmission schemes (CMSs), where each scheme is specified by one of the M-ary quadrature amplitude modulation (M-QAM) and LDPC code pair. The source node selects a CMS for transmission and adapt the transmit power on frame-by-frame basis based on the instantaneous signal-to-noise (SNR) between the source and the best selected relay and SNR adaption thresholds, respectively. To contrast the performance of the proposed cooperative communication system we compare its performance with ACM-direct transmission and ACM-cooperative system with single relay. We have employed MATLAB® to write a computer program designed for simulation of the proposed cooperative system to allow various parameters of the system to be varied and tested.

The rest of this paper is organized as follows. In section II, we introduce a system model of the proposed cooperative wireless communication system includes multi-node relay-selection

cooperative diversity model with adaptive LDPC coded modulation transceiver structure. In section III, the best relay selection scheme and adaptive LDPC coded modulation (ACM) are introduced. In section IV, error control codes (ECC) and low-density parity-check codes (LDPC) are introduced. The simulation model and MATLAB® simulation results of the proposed cooperative wireless communication system is presented in Section V, followed by conclusions in Section VI.

II. SYSTEM MODEL

We consider a wireless communication system as shown in Figure 1, with one source node (*S*), one destination node (*D*), and *N* half duplex relay nodes R_1, R_2, ……, R_N which are randomly distributed between source and destination. Let $h_{S,D}$, h_{S,R_i} and $h_{R_i,D}$, *i=1,2,…,N* denote The *S-D*, *S-*R_i and R_i*-D* channel coefficients, respectively. We assume that the channel coefficients $h_{S,D}$, h_{S,R_i} and $h_{R_i,D}$ are mutually independent zero-mean, complex Gaussian random variables (flat Rayleigh fading) with variances $\delta^2_{S,D}$, δ^2_{S,R_i} and $\delta^2_{R_i,D}$ respectively. The additive white Gaussian noise (AWGN) terms of all links are modeled as zero-mean complex Gaussian random variables with variance N_0 .We will assume a special case where the source and relay nodes transmit with equal power P.

Assuming that the relay nodes use in-band half duplex relaying, where the relaying is done over two time slots, in the 1st time slot the source node transmits the data to both the destination node and a set of N-relay nodes. The received signals $y_{S,D}$ and y_{S,R_i} at the destination and the relay, respectively, can be written as

$$y_{S,D} = \sqrt{p}\, h_{S,D}\, x + n_{S,D} \quad (1)$$

$$y_{S,R_i} = \sqrt{p}\, h_{S,R_i}\, x + n_{S,R_i} \quad (2)$$

where *p* is the transmitted power at the source, *x* is the transmitted information symbol, $n_{S,D}$, and n_{S,R_i} are additive noise at the receiving nodes.

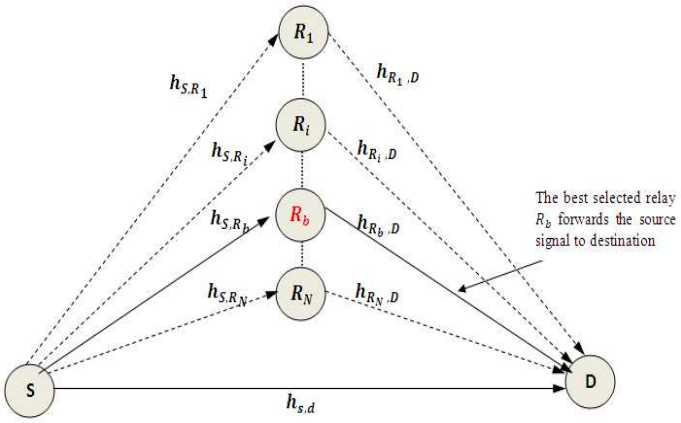


Figure 1. Multi-node cooperative communication system with best relay selection

In a cooperative diversity scheme, if the signal-to-noise ratio of a signal received at any relay of the N-relay nodes exceeds a certain threshold, the relay decodes the received signal and has the ability to forward the decoded information to the destination node. On the other hand, if the channel between the source and the relay suffers a severe fading such that the signal-to-noise ratio falls below the threshold, the relay remains idle. All relay nodes which has the ability to fully decode the source data will belong to a decoding group Ð

In the 2nd time slot, the best selected relay from the decoding group *Ð* re-encodes and retransmits the source data to the destination. At the destination node the signals of the direct link and the best cooperative link can be combined using appropriate combing technique i.e., MRC technique [15, 16]. The received signal at the destination from the best selected relay, can be written as

$$y_{R_b,D} = Q\,(y_{S,R_b})\, h_{R_b,D} + n_{R_b,D} \quad (3)$$

where the function Q (·) denotes the ACM cooperative protocol implemented at the selected best relay node[17,18], R_b , $h_{R_b,D}$ is the channel coefficient from the selected best relay to destination and $n_{R_b,D}$ is an additive noise.

Since our proposed cooperative wireless communication system based on combine best relay selection and ACM scheme, Figure 2 shows

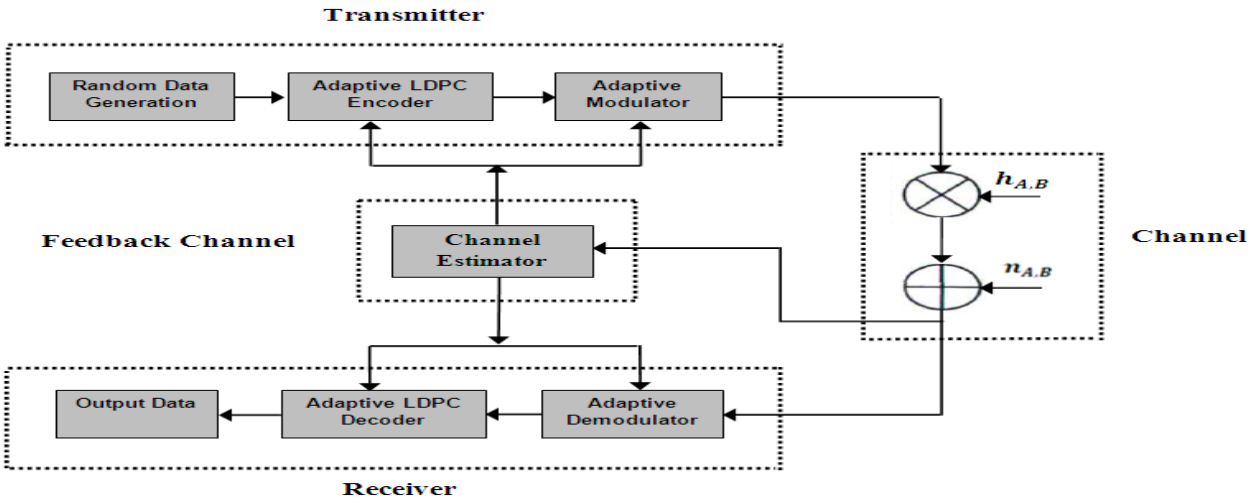


Figure 2. Transceiver structure with adaptive LDPC coded modulation at source node

a simplified block diagram of ACM transceiver structure [13], Where $h_{A,B}$ and $n_{A,B}$ is the channel coefficient and the additive white Gaussian noise (AWGN) of a link from *A* to *B*, the terms *A* and *B* can be either source (*S*) and destination (*D*) or source (*S*) and Relay (R_i).

III. BEST RELAY SELECTION AND ADAPTIVE LDPC CODED MODULATION (ACM)

In best relay selection (BRS) scheme, a single relay among a set of *N* available relay nodes is selected depending on which relay provides for the "best" end-to-end path between source and destination. Let $\gamma_{h_{S,D}} \triangleq |h_{S,D}|^2 \frac{E_s}{N_0}$ is the instantaneous SNR between the source and destination, $\gamma_{h_{S,R_i}} \triangleq |h_{S,R_i}|^2 \frac{E_s}{N_0}$ is the instantaneous SNR between the source and relay R_i and $\gamma_{h_{R_i,D}} \triangleq |h_{R_i,D}|^2 \frac{E_s}{N_0}$ is the instantaneous SNR between relay R_i and destination. $\gamma_{h_{S,D}}$, $\gamma_{h_{S,R_i}}$, and $\gamma_{h_{R_i,D}}$ are exponentially distributed with parameters $1/\bar{\gamma}_{h_{S,D}}$, $1/\bar{\gamma}_{h_{S,R_i}}$, and $1/\bar{\gamma}_{h_{R_i,D}}$, respectively, where $\bar{\gamma}_{h_{S,D}} \triangleq E[\gamma_{h_{S,D}}] \triangleq \delta_{S,D}^2 \frac{E_s}{N_0}$, $\bar{\gamma}_{h_{S,R_i}} \triangleq E[\gamma_{h_{S,R_i}}] \triangleq \delta_{S,R_i}^2 \frac{E_s}{N_0}$, and $\bar{\gamma}_{h_{R_i,D}} \triangleq E[\gamma_{h_{R_i,D}}] \triangleq \delta_{R_i,D}^2 \frac{E_s}{N_0}$, and E [•] denotes expectation[19].

We assume that the channel state information (CSI) are available at the source and relay nodes. For a given required spectral efficiency $\bar{\eta}$ the decoding group $Đ$ is given by

$$Đ = \left\{ k \in N, |h_{S,R_k}|^2 \geq \bar{\eta}, k < N \right\} \quad (4)$$

Where $\bar{\eta}$ is given by

$$\bar{\eta} \triangleq \frac{2^{2\eta}-1}{E_s/N_0}.$$

The relay selection algorithm selects the best relay R_b from the decoding group $Đ$ such that [20]

$$b \triangleq argmax_{i=1,2,\ldots,N} \min\left\{\gamma_{h_{S,R_i}}, \gamma_{h_{R_i,D}}\right\}. \quad (5)$$

The mutual information of the proposed cooperative wireless communication system with combined BRS and AMC is given by [21]

$$I_{BRS} = \begin{cases} \frac{1}{2}\log\left(1+2\frac{E_s}{N_0}|h_{S,D}|^2\right), & |h_{S,D}|^2 \geq \eta_D \,\&\, |h_{S,R_i}|^2 < \bar{\eta} \\ \frac{1}{2}\log\left(1+\frac{E_s}{N_0}|h_{R_b,D}|^2\right), & |h_{S,D}|^2 < \eta_D \,\&\, |h_{S,R_i}|^2 \geq \bar{\eta} \\ \frac{1}{2}\log\left(1+\frac{E_s}{N_0}|h_{R_b,D}|^2+\frac{E_s}{N_0}|h_{S,D}|^2\right), & |h_{S,D}|^2 \geq \eta_D \,\&\, |h_{S,R_i}|^2 \geq \bar{\eta} \\ \textit{The Source is Idle}, & |h_{S,D}|^2 < \eta_D \,\&\, |h_{S,R_i}|^2 < \bar{\eta} \end{cases} \quad (6)$$

where $\eta_D \triangleq \frac{2^{\eta}-1}{E_s/N_0}$

At the source node, the adaptive LDPC coded modulation (ACM) provides multiple coded modulation transmission schemes (CMSs), where

each scheme is specified by one of the *M*-ary quadrature amplitude modulation (M-QAM) and LDPC code pair. The source node selects a CMS for transmission and adapt the transmit power on frame-by-frame basis based on the CSI feedback from the receiver. To select the appropriate CMS for ACM wireless communication system, we need to know the SNR thresholds. We assume that *L* CMSs are candidates for the ACM system. The source node decides which CMS should be used at the start of each frame transmission according to a given set of SNR thresholds. For a certain level of BER, we define adaption SNR thresholds set, $\lambda \in \{\lambda_1, \lambda_2, \dots, \lambda_{L-1}\}$ for the instantaneous SNR for every transmitted frame. Thus, each of the *L* candidate CMS is assigned to operate in a particular SNR region. When the threshold λ falls within a given SNR region, $\lambda_l \leq \lambda < \lambda_{l+1}$ where $l \in \{1,2,..,L-1\}$, the associated CSI is sent back to the source node to adapt its corresponding CMS. The corresponding spectral efficiency of each candidate CMS is denoted as η_l and $\eta_1 < \eta_2 < \cdots < \eta_l$.

The instantaneous SNR for each frame of a link from *A* to *B* is given by

$$\left(\gamma_{h_{A,B}}\right)_w = \left(\left|h_{A,B}\right|^2\right)_w (E_s/N_o) \quad (7)$$

where *w* is the index of the transmitted frame.

As mentioned before, when $0 \leq \left(\gamma_{h_{A,B}}\right)_w \leq \lambda_1$, the first candidate CMS is employed during the w^{th} transmitted frame. The $(l+1)^{th}$ candidate CMS is employed during w^{th} transmitted frame, when $\left(\gamma_{h_{A,B}}\right)_w$ satisfies the following inequality

$$\lambda_l \leq \left(\gamma_{\mathrm{h_{A,B}}}\right)_{\mathrm{w}} \leq \lambda_{l+1} \quad l = (1,2,..,L-1) \quad (8)$$

We can deduce the inequality for instant fading amplitude, $\left|h_{A,B}\right|_w$ when $(l+1)^{\mathrm{th}}$ candidate CMS is employed [22],

$$\sqrt{\frac{\lambda_l}{E_s/N_o}} \leq \left|\mathrm{h_{A,B}}\right|_{\mathrm{w}} < \sqrt{\frac{\lambda_{l+1}}{E_s/N_o}} \quad l = (1,2,..,L-1) \quad (9)$$

The adaptation thresholds for the fading amplitude are then determined by the relationship

$\sqrt{\frac{\lambda_l}{E_s/N_o}} = v_l$, thus

$$v_l \leq \left|\mathrm{h_{A,B}}\right|_{\mathrm{w}} < v_{l+1} \quad l = (1,2,\dots,L-1) \quad (10)$$

Since our proposed cooperative system is based on ACM, the main advantage of ACM is that it makes a good use of time varying nature of wireless channel and hence improves spectral efficiency while keeping the performance at an acceptable level. Thus, due to its adaptive nature, the spectral efficiency of the proposed cooperative system is varied as a function of the instantaneous SNR. The average spectral efficiency (ASE) is defined as the average number of information bits transmitted per symbol duration. For the ACM scheme there are *L* CMSs candidates and the corresponding spectral efficiency of each candidate CMS is denoted as η_l and $\eta_1 < \eta_2 < \cdots < \eta_l$. The ASE is defined as [23]:

$$ASE = \sum_{l=1}^{L} \eta_l \,.\, P_l(\alpha) \quad (11)$$

where $P_l(\alpha)$ is the Rayleigh probability distribution for α being in the interval [v_l, v_{l+1}]. The probability density function of the Rayleigh distribution is given by

$$p(\alpha) = 2\alpha.\exp(-\alpha^2) \quad (12)$$

Thus, the ASE is expressed by

$$ASE = \eta_1.P(0 \leq \alpha \leq v_1) + \eta_2.P(v_1 \leq \alpha \leq v_2) + \cdots + \eta_L.P(v_{L-1} \leq \alpha \leq \infty) \quad (13)$$

where

$$P(v_l \leq \alpha \leq v_{l+1}) = \int_{v_l}^{v_{l+1}} p(\alpha)\, d(\alpha) = \int_{v_l}^{v_{l+1}} 2\alpha.\exp(-\alpha^2) d(\alpha) = \exp(-v_l^2) - \exp(-v_{l+1}^2) \quad (14)$$

thus the ASE is given by,

$$ASE = \eta_1.[\exp(-v_0{}^2) - \exp(-v_1{}^2)] + \eta_2.[\exp(-v_1{}^2) - \exp(-v_2{}^2)] + \cdots + \eta_L.[\exp(-v_{L-1}{}^2) - \exp(-v_L{}^2)]$$

$$= \sum_{l=1}^{L} \eta_l.[exp(-v_{l-1}{}^2) - exp(-v_1{}^2)] \quad (15)$$

where $v_0 = 0,\ v_L = \infty$.

IV. ERROR CONTROL CODES (EEC)

In wireless communication systems, errors in data transmission can come from many different sources (i.e., random noise, interference, channel fading and physical defects etc.). These channel errors must be reduced to an acceptable level to ensure reliable data transmission. To combat the errors, we normally use two strategies, either stand-alone or combined. The first one is the automatic repeat request (ARQ). An ARQ system attempts to detect the presence of errors in the received data. If any errors are found, the receiver notifies the transmitter of the existence of errors. The transmitter then resends the data until they are correctly received. In many practical applications retransmission may be difficult or not even feasible at all. For example, it is impossible for any receiver in a real-time broadcasting system to request data to be resent. In this case the second strategy, known as error control codes (ECC) is the only viable solution.

ECC are used for detecting the presence of errors and correcting them. It first adds redundancy (parity bits) to the message to be sent and form a codeword that contains both the message and the redundancy; this process is called encoding and is carried out at the transmitter. It then corrects errors based on the redundancy in a process called decoding that is performed at the receiver [24].

Several different types of ECC exist, but we may loosely categorize them into two families of codes. One is called block codes, which encode and decode data on a block-by-block basis where the data blocks are independent from each other. Block codes include repetition codes, Hamming codes [25], Reed Solomon codes [26], and BCH codes [27]. Consequently block coding is a memoryless operation and can be implemented using combinational logic. In contrast, another family of codes, namely, the convolutional codes [28], works on a continuous data stream, and its encoding and decoding operations depend not only on the current data but also on the previous data. As such, convolutional coding contains memory and has to be implemented using sequential logic.

ECC can also categorized based on decoding algorithms into two families, one is non -iterative decoding algorithms, such as syndrome decoding for block codes or maximum likelihood (ML) nearest codeword decoding for short block codes, algebraic decoding for Reed Solomon and BCH codes, and Viterbi decoding or sequential decoding for convolutional codes. The other family of decoding algorithms is iterative decoding algorithms, such as turbo decoding with component MAP decoders for each component code, and the sum product algorithm (SPA) [29] or its lower complexity approximation, min-sum decoding [30] for low density parity check codes (LDPCs). Since our proposed cooperative wireless communication system based on combined best relay selection and adaptive LDPC coded modulation, LDPC code is considered in this paper as ECC for its superior error correcting capabilities.

A. Low-Density Parity-Check Codes (LDPC)

Low-density parity-check (LDPC) codes, also known as the Gallager codes are a class of linear block error correction codes. LDPC codes were first discovered by Robert Gallager [31, 32] in the early 60s. For some reason, though, they were forgotten and the field lay dormant until the mid-90s when the codes were rediscovered by David MacKay and Radford Neal [33, 34]. Since then, the class of codes has been shown to be remarkably powerful, comparable to the best known codes and performing very close to the theoretical limit of error correcting codes.

More recently, due to their advantages LDPC codes have been proposed for several state-of-the-art wireless standards including IEEE 802.16e wireless MAN [35], IEEE 802.11n wireless LAN [36] and second generation satellites for digital video broadcasting (DVB-S2) [37].

An LDPC code is an (n, k) or (n, w_c, w_r) linear block code whose parity-check matrix H contains only a few 1's in comparison to 0's (i.e., sparse matrix). For an $m \times n$ parity-check matrix (where $m = n - k$), we define two parameters: the column weight w_c which equal to the number of nonzero elements in a column and the row weight w_r which equal to the number of nonzero elements in a row where $w_c << n$ and $w_r << m$.

An LDPC code is regular if w_c is constant for every column and w_r is also constant for every row, and a regular LDPC code will have,

$$m \cdot w_r = n \cdot w_c \quad (16)$$

the coding rate of the regular LDPC is given by,

$$R = 1 - \frac{w_c}{w_r} \quad (17)$$

On the other hand, if H is low in density (sparse) but w_c and w_r are not constant, the code is irregular.

Besides the general expression of LDPC codes as an algebraic matrix, LDPC codes can also be represented by a bipartite Tanner graph, which was proposed by Tanner in 1981 [38].A Tanner graph is a bipartite graph introduced to graphically represent LDPC codes. It consists of nodes and edges. The nodes are grouped into two sets. One set consists of n bit nodes (or variable nodes), and the other of m check nodes (or parity nodes). The creation of such a graph is straightforward: Check node i is connected to bit node j if $h_{i,j}$ of the parity matrix H is a 1. From this we can deduce that there are totally $m \cdot w_r$ (or $n \cdot w_c$) edges in a Tanner graph for a regular LDPC code. Apparently the Tanner graph has a one-to-one correspondence with the parity-check matrix.

The Tanner graph of the H (6, 2, 3) rate 1/3 regular LDPC code is shown Figure 3.

The LDPC codes used in this paper is based on the WiMax 802.16e standard. In the WiMax 802.16e standard, the LDPC codes are a set of systematic linear block codes which are built from Richardson-Urbanke encoding algorithm [39] and a special class of Quasi-Cyclic (QC) LDPC codes [40 , 41].

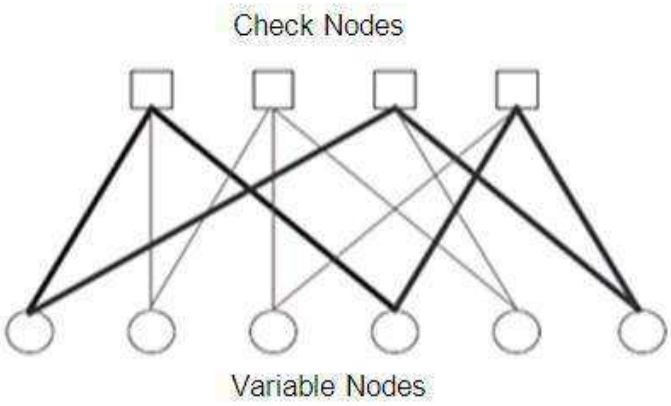


Figure 3. Tanner graph representation of the parity-check matrix in (18), where the bold lines represent 6-cycle

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \quad (18)$$

The LDPC codes adapted by the IEEE 802.16e WiMAX standard [35] support 19 different codeword lengths with four distinct code rates namely, (1/2, 2/3, 3/4, 5/6) and six different code classes (1/2, 2/3 A, 2/3 B, 3/4 A, 3/4 B, 5/6). The parity check matrix of all six code classes consists of 24 columns and (1-R) ×24 rows, with each entry describing a z-by-z sub-matrix which is either a permuted identity matrix or a zero matrix. The first R×24 columns correspond to the systematic information, the second (1-R) ×24 columns for the parity information which have a fixed structure required by the encoder design [42].

The sub-matrices z-by-z has a variable size that ranged from 24×24 to 96×96 with incremental granularity of 4, providing 19 codeword lengths ranging from n = 576 to n = 2304 bit with incremental granularity of 96 bit. Figure 4 shows the generic structure of the parity check matrix used in IEEE 802.16e WiMAX standard, with rate 1/2 and codeword length of n = 2304 bit (i.e., z = 96). In WiMAX LDPC codes the BER performance improved with increasing codeword length n [13], thus in all simulations we select LDPC codes with codeword length n=2034 bit.

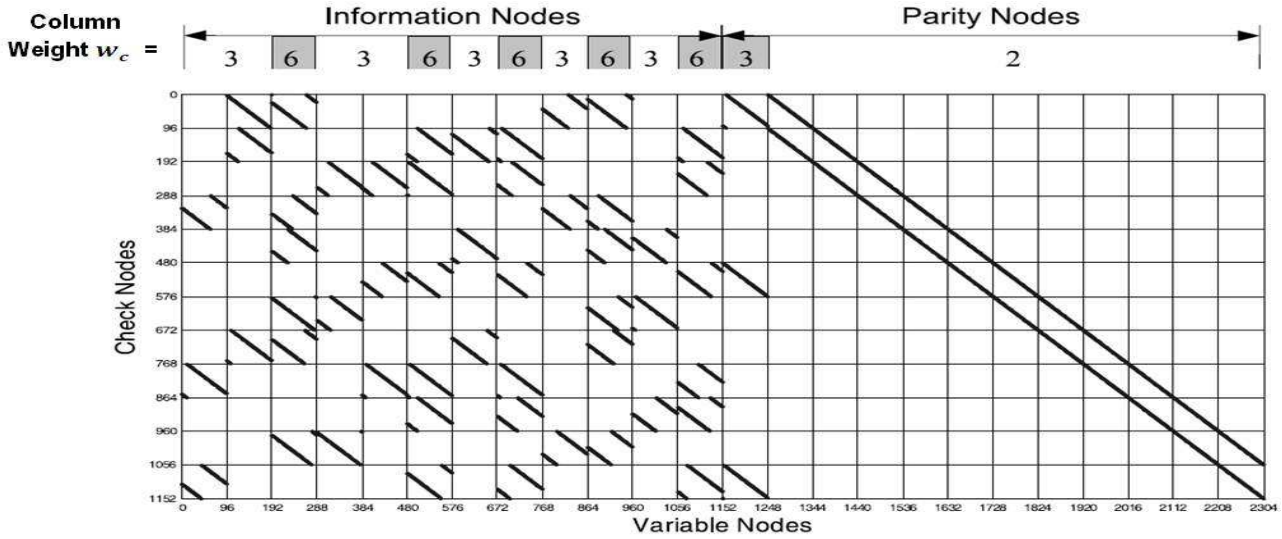


Figure 4. Structure of the parity check matrix *H for* a rate 1/2 WiMAX 802.16e LDPC code with codeword length *n*=2304 bit (*z* = 96), where the bold lines represent elements "1" in *H*

V.SIMULATION MODEL AND RESULTS

In this section we present the simulation model and discuss the simulation results. As we have stated before, our research goal is to evaluate the performance of cooperative wireless communication system with combined best relay selection (BRS) and adaptive LDPC coded modulation (ACM). The system model and block diagram of ACM transceiver structure are shown in Figures 1, 2. We have employed MATLAB® to write a computer program designed for simulation of the proposed system to allow various parameters of the system to be varied and tested. The possible considered parameters and their corresponding values are mentioned in Table 1.

We considered three scenarios for simulations:

1- Direct transmission (without relays), in this case we have only a direct link of distance *d*=1Km from source node (*S*) to destination node (*D*).
2- Cooperative wireless communication system with single relay (*N*=1).
3- Cooperative wireless communication system with best relay selection (BRS), where a single relay among a set of *N* relay nodes is selected, depending on which relay provides for the "best" end-to-end path between source and destination.

Table.1 Simulations parameters and their values

Parameters	Values
Digital Modulation	QPSK, 8QAM, 16QAM, 64QAM, 265QAM
Distance between Source and Destination, *d*	1Km
Number of Relay Nodes, *N*	1, 4, 8
Channels	AWGN, flat Rayleigh fading
LDPC Codes	WiMAX LDPC codes
LDPC Code Rate, r	1/2, 2/3, 3/4, 5/6
Codeword Length, n	2304 bit
Encoding	Richardson-Urbanke algorithm
Decoding	Logarithmic BP algorithm
Maximum number of iterations	30

To evaluate the performance of the proposed ACM scheme, we adopt six different combinations of LDPC coded modulation schemes (CMSs) at which the spectral efficiency can vary from 1 to 6 bits/s/Hz, as shown in Table 2.
Figure 5 shows the BER performance comparison between cooperative wireless communication system with single relay (*N=1*) and cooperative wireless communication system with BRS with

different number of relay nodes *N* between source and destination, considering transmission over Rayleigh fading channels using rate 1/2 LDPC code and QPSK modulation. It can be observed that, in all cases, increasing the number of available relay nodes between source and destination decreasing the BER level for a given value of SNR.

Figure 6 shows the SNR adaptation thresholds of each candidate pair at $\mathrm{BER} = 10^{-2}$ for different scenarios of simulations. The SNR adaptation thresholds at $\mathrm{BER} = 10^{-2}$ are summarized in Table 3.

Table.2 Spectral efficiency of six candidate pairs for the proposed adaptive LDPC coded modulation scheme

CMS	Modulation	Coding rate	Spectral efficiency (bits/s/Hz)	SNR adaptation thresholds
CMS-1	QPSK	1/2	1	$0 \geq SNR \leq \lambda_1$
CMS-2	8-QAM	2/3	2	$\lambda_1 \geq SNR \leq \lambda_2$
CMS-3	16-QAM	3/4	3	$\lambda_2 \geq SNR \leq \lambda_3$
CMS-4	64-QAM	2/3	4	$\lambda_3 \geq SNR \leq \lambda_4$
CMS-5	64-QAM	5/6	5	$\lambda_4 \geq SNR \leq \lambda_5$
CMS-6	256-QAM	3/4	6	$SNR \geq \lambda_5$

Table.3: SNR adaptation thresholds at $\mathrm{BER} = 10^{-2}$ for different scenarios of simulations

SNR Adaptation Threshold(dB)	λ_1	λ_2	λ_3	λ_4	λ_5
Direct Transmission	20	25	28.9	32.5	35
Cooperative Transmission with Single Relay	13.6	19	23.5	28.5	32
Cooperative Transmission with BRS	8.3	14.6	19.5	24.2	28.9

Figure 7 shows the BER performance comparison between direct transmission with different CMSs and with ACM at the source node. Our results show that direct transmission with ACM at the source node achieves lower SNR values for desired BER as compared to direct transmission with different CMSs. We can also observed that the BER curve of direct transmission with ACM is between CMS-3 and CMS-4 and for a $\mathrm{BER} = 10^{-2}$ direct transmission with ACM can have a SNR gain between 3 to 7 dB compared to CMS-5 and CMS-6, respectively and a SNR loss between 3.6 to 6.3 dB compared to CMS-3 and CMS-2, respectively.

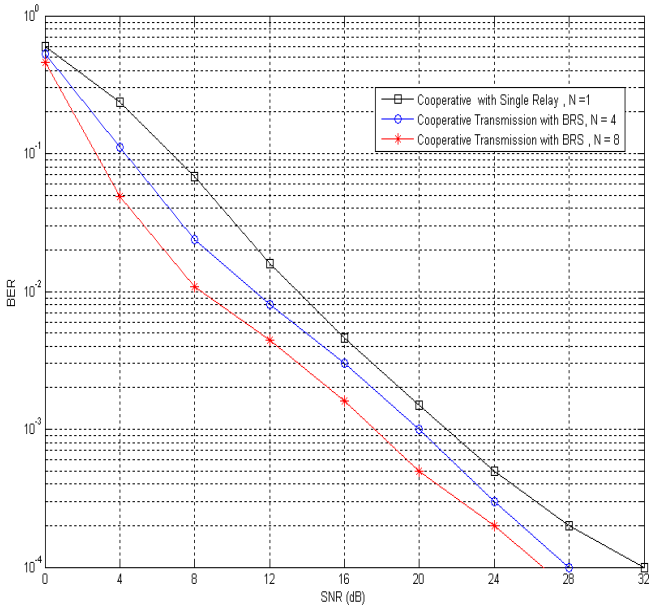


Figure 5. BER versus SNR for different number of relay nodes *N*, considering transmission over Rayleigh fading channels (*r*=1/2, *n*=2304 bit, QPSK, and *d*=1Km)

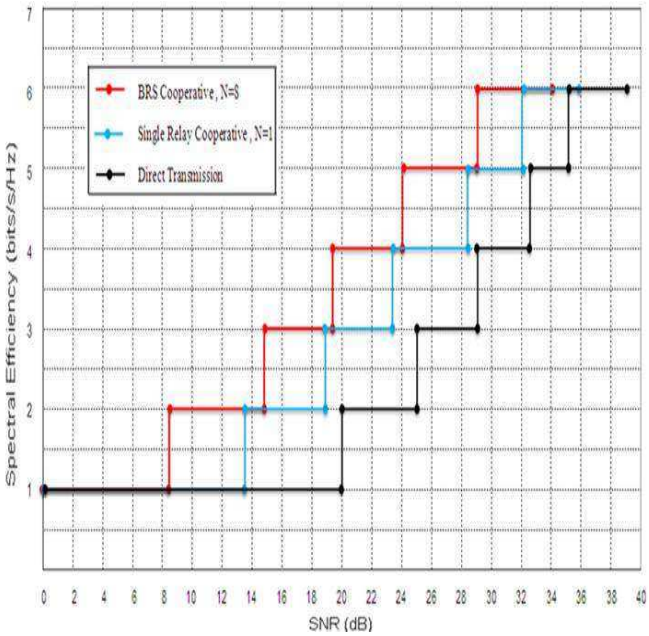


Figure 6. SNR adaptation thresholds of each candidate pair at BER=10^{-2} for different scenarios of simulations

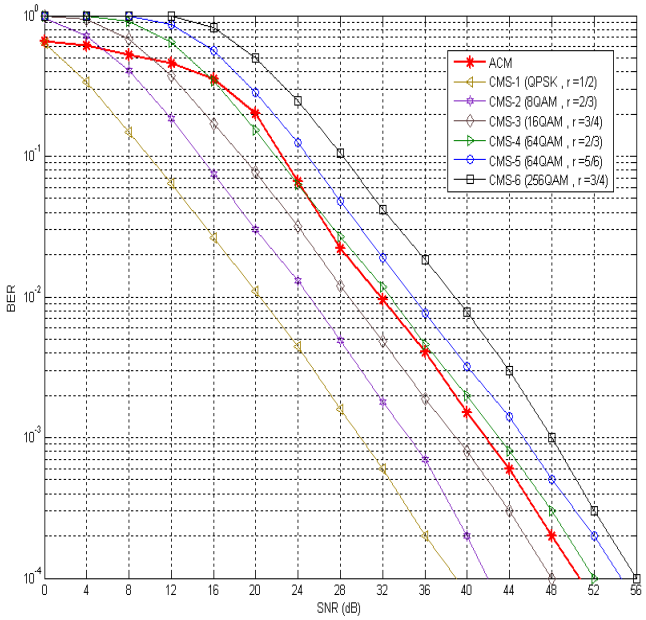


Figure 7. BER performance comparison between direct transmission with different CMSs and with ACM

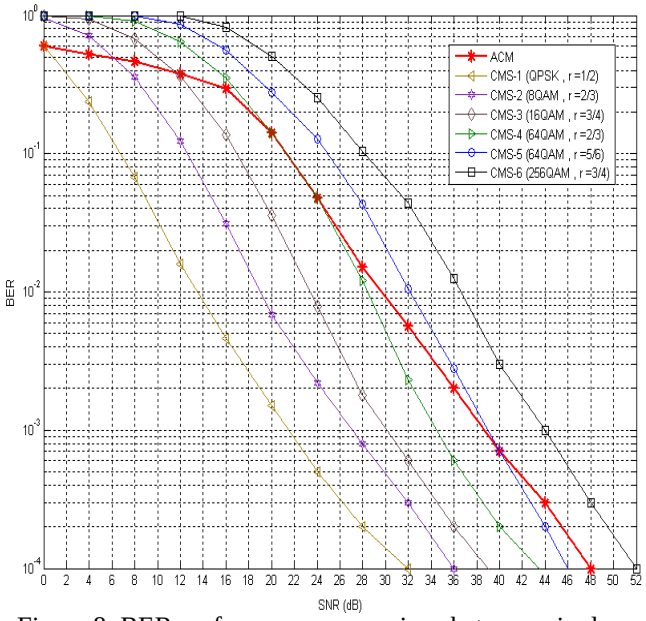


Figure 8. BER performance comparison between single relay cooperative with different CMSs and with ACM, *N*=1

Figure 8 shows the BER performance comparison between single relay cooperative transmission with different CMSs and with ACM at the source node. Our results show that single relay cooperative transmission with ACM at the source node achieves lower SNR values for desired BER as compared to single relay cooperative transmission with different CMSs. We can also observed that the BER curve of single relay cooperative with ACM is between CMS-4 and CMS-5 and for a $BER = 10^{-2}$ single relay cooperative transmission with ACM can have a SNR gain between 2.2 to 6.4 dB compared to CMS-5 and CMS-6, respectively and a SNR loss between .6 to 6 dB compared to CMS-4 and CMS-3, respectively. Our results also show that single relay cooperative transmission with different CMSs and ACM achieves lower SNR values for desired BER as compared to direct transmission.

Figure 9 shows the BER performance comparison between BRS cooperative transmission with different CMSs and with ACM at the source node. Our results show that BRS cooperative transmission with ACM at the source node achieves lower SNR values for desired BER as compared to BRS cooperative transmission with different CMSs. We can also observed that the BER curve of BRS cooperative with ACM is between CMS-4 and CMS-5 and for a $BER = 10^{-2}$ BRS cooperative transmission with ACM can have a SNR gain between 2.6 to 8 dB compared to CMS-5 and CMS-6, respectively and a SNR loss between 1.6 to 6 dB compared to CMS-4 and CMS-3, respectively. Our results also show that BRS cooperative transmission with different CMSs and ACM achieves lower SNR values for desired BER as compared to direct transmission and single relay cooperative.

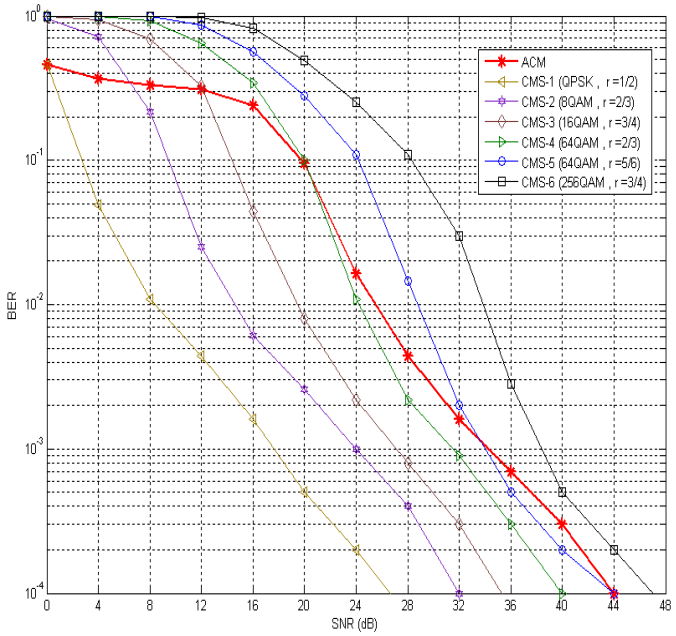


Figure 9. BER performance comparison between BRS cooperative with different CMSs and with ACM, *N*=8

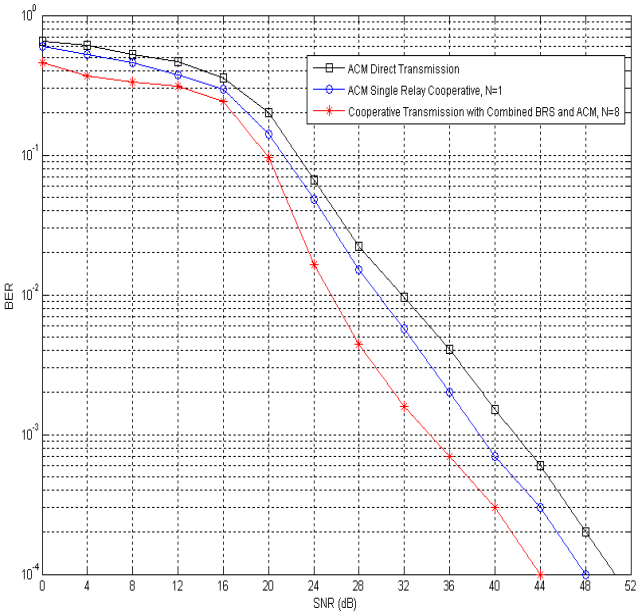


Figure 10. BER performance comparison between ACM direct transmission, ACM cooperative with single relay and Cooperative system with combined BRS and ACM, SNR adaptation threshold at BER $= 10^{-2}$

Figure 10 shows the BER performance comparison between ACM direct transmission, ACM cooperative system with single relay and the proposed cooperative system with combined BRS and ACM. Our results show that cooperative system with combined BRS and ACM achieves lower SNR values for the desired BER as compared to ACM direct transmission and ACM cooperative system with single relay, for BER $= 10^{-2}$ the proposed cooperative system with combined BRS and ACM can have a SNR gain between 4.4 to 6.5 dB compared to ACM cooperative system single relay and ACM direct transmission, respectively. All result for different scenarios of simulations are concluded in Table 4; from the results it is clear that cooperative transmission with BRS outperform all other scenarios at a given BER level for different CMSs. Figure 11 shows the effect of selecting SNR adaptation thresholds at different BER levels ($BER = 10^{-2}, 10^{-3}$) on the performance of the proposed cooperative system with combined BRS and ACM, from the figure it is clear that the BER performance of the proposed cooperative system improved by selecting SNR adaptation thresholds at lower BER level where lower CMSs are used more frequently.

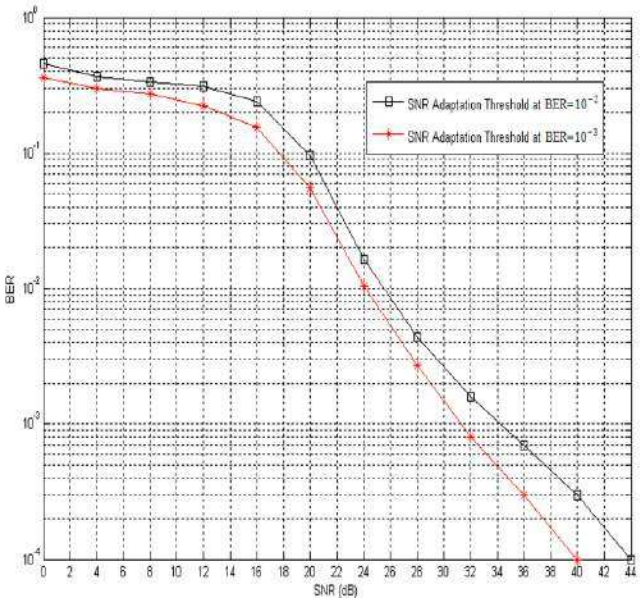


Figure 11. The effect of SNR adaptation threshold on the BER performance of the proposed cooperative system with combined ACM and BRS, *N*=8

Table 4. SNR required at BER level 10^{-2} for different scenarios of simulations

CMSs	Direct Transmission	Cooperative Transmission with Single Relay	Cooperative Transmission with BRS
CMS-1	20	13.6	8.5
CMS-2	25	19	14.6
CMS-3	28.9	23.5	19.5
CMS-4	32.5	28.5	24.2
CMS-5	35	32	28.9
CMS-6	39	36.8	34
ACM	32	29.9	25.5

Figure 12 shows the average spectral efficiency (ASE) of different ACM transmission scenarios, SNR adaptation threshold at BER=10^{-2}. From the figure it is clear that the ASE increases as the SNR increases, also it is clear that the proposed cooperative system with combined BRS and ACM required lower SNR levels to attain the same ASE compared to all other scenarios, and as the number of relay nodes increases the proposed cooperative system achieves lower SNR values for the desired ASE. From the results; we can also observe that, the proposed cooperative system with combined BRS and ACM required SNR value of 25.3 dB to

achieve BER level of 10^{-2} as shown in Figure 9, and the ASE at this value of SNR is 6 bits/s/Hz as shown in Figure 12. For the same BER level of 10^{-2}, the SNR values of CMS-1, CMS-2, CMS-3, CMS-4, CMS-5 and CMS-6 are 8.5, 14.6, 19.5, 24.2, 28.9 and 34dB, respectively with spectral efficiency from 1 to 6 bits/s/Hz. Thus, the proposed cooperative system can outperform CMS-5 and CMS-6. Although it required higher SNR value compared to CMS-1, CMS-2, CMS-3 and CMS-4. Figure 13 shows the effect of selecting SNR adaptation thresholds as different BER levels ($BER = 10^{-2}, 10^{-3}$) on the ASE performance of the proposed cooperative system. From the figure it is clear that the ASE of the proposed cooperative system using higher SNR adaptation threshold at BER level of 10^{-2} have a gain of approximately .5 to 1.25 bits/s/Hz for the same SNR value compared to that with lower SNR adaptation threshold at BER level of 10^{-3}.

VI. Conclusions

In this paper, a new cooperative scheme based on combined best relay selection (BRS) and adaptive LDPC coded modulation (ACM) is investigated. To improve the spectral efficiency of the proposed cooperative system, we introduced ACM at source node which provides multiple coded modulation transmission schemes (CMSs), where each scheme is specified by one of the *M*-ary quadrature amplitude modulation (M-QAM) and LDPC code pair (candidate pair).

To contrast the performance of the proposed cooperative system we compare its performance with ACM-direct transmission and ACM-cooperative system with single relay. The simulation results show that the proposed cooperative scheme achived lower SNR values for the desired BER levels and required lower SNR values to attain the same average specral efficiency as comared to ACM direct transmission and ACM cooperative with single relay.

We studied also the effect of selecting SNR adaptation thresholds at different BER levels ($i.e., BER = 10^{-2}, 10^{-3}$) on the performance of the proposed cooperative scheme, the simulation results show that the BER performance of the proposed cooperative scheme improved by selecting SNR adaptation thresholds at lower BER level where lower CMSs are used more frequently. The results also show that the ASE of the proposed cooperative scheme using SNR adaptation threshold at higher BER level of 10^{-2} have a gain of approximately .5 to 1.25 bits/s/Hz for the same SNR value compared to that with lower SNR adaptation threshold at BER level of 10^{-3}.

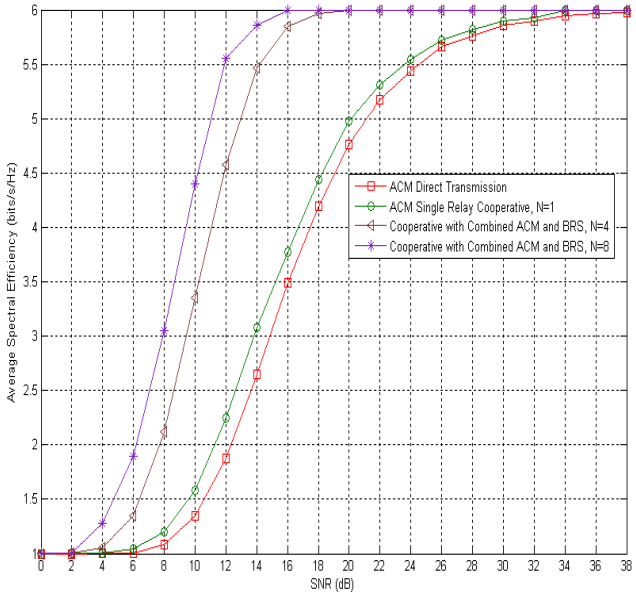


Figure 12. ASE of different transmission scenarios, SNR adaptation threshold at BER $= 10^{-2}$

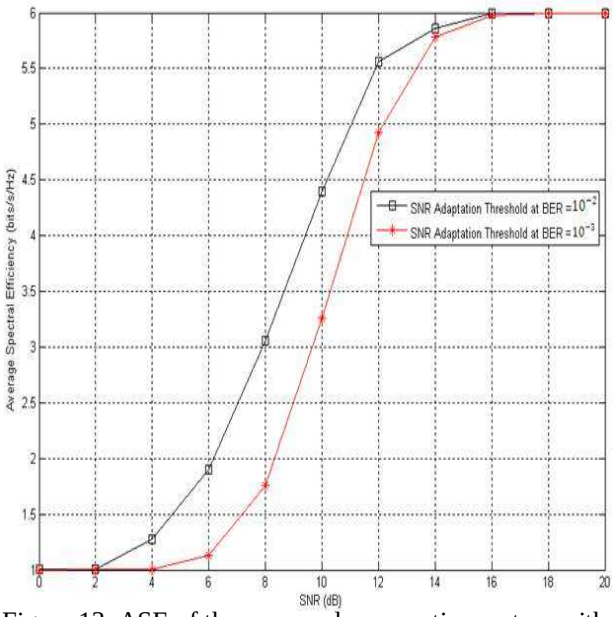


Figure 13. ASE of the proposed cooperative system with combined ACM and BRS, SNR adaptation thresholds at different BER levels ($BER = 10^{-2}, 10^{-3}$), *N*=8

References

[1] M. K. Simon and M.-S. Alouini," Digital Communication over Fading Channels," 2nd edition. New York: Wiley, 2005.
[2] R. U. Nabar, H. Bölcskei, and F. W. Kneubühler,"Fading Relay Channels: Performance Limits and Space–Time Signal Design," IEEE JSAC, vol. 22, no. 6, Aug 2004.
[3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behaviour," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.
[4] Bahrami, Hamid Reza, and Tho Le-Ngoc, "Distributed BLAST with relay selection for multi-antenna AF relaying," IJCS, vol. 24, No.4, pp. 473–482, Apr. 2011
[5] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," IEEE J. Select. Areas Commun, vol. 24, pp. 659–672, Mar. 2006.
[6] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," IEEE Trans. Inf. Theory, vol. 49, pp. 2415–2425, Nov. 2003.
[7] A. Bletsas, H. Shin, and M. Z. Win, "Outage optimality of opportunistic amplify-and-forward relaying," IEEE Communications Letters, vol. 11, pp. 261–263, Mar. 2007.
[8] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," IEEE Trans. Wireless Commun., vol. 6, pp. 3114–3123, Aug. 2007.
[9] D. S. Michalopoulos, G. K. Karagiannidis, T. A. Tsiftsis, and R. K. Mallik, "Distributed transmit antenna selection (DTAS) under performance or energy consumption constraints," IEEE Trans. Wireless Commun., vol. 7, pp. 3168–1173, Apr. 2008.
[10] N. C. Beaulieu and J. Hu, "A closed-form expression for the outage probability of decode-and-forward relaying in dissimilar Rayleigh fading channels," IEEE Commun. Lett., vol. 10, no. 12, pp. 813-815, Dec. 2006.
[11] E. Beres and R. S. Adve, "Selection cooperation in multi-source cooperative networks," IEEE Trans. Wireless Commun., vol. 7, no. 1, pp. 118-127, Jan. 2008.
[12] M. Torabi, W.Ajib and D.Haccoun, "Performance Analysis of Amplify-and-Forward Cooperative Networks with Relay Selection over Rayleigh Fading Channels," Vehicular Technology Conference, 2009, pp.1 – 5, Spring 2009.
[13] R. Yang, "LDPC-coded Modulation for Transmission over AWGN and Flat Rayleigh Fading Channels," Master's thesis, Faculte' Des Sciences Et De Genie' Universite' Laval Que'bec, 2010.
[14] M. R. Islam and Y. S. Han, "Cooperative MIMO Communication at Wireless Sensor Network: An Error Correcting Code Approach," Sensors, 2011.
[15]Guan, Zhang-Jun, Zhang, Wei-Jiong,Zhou, Xi-Lang, "Performance analysis of multi-antenna relay communication systems with MRC," IJCS, vol. 25, No.11, pp. 1505–1512, Nov. 2012
[16]Hang Long, Meiying Wei, Ruchen Duan, Kan Zheng, and Wenbo Wang, "Performance analysis of the unicasting mode in multi-relay systems," IJCS, vol. 24, No.4, pp. 513–525, Apr. 2011
[17] K. J. R. Liu, A. K. Sadek, W. Su, and A. Kwasinski, "Cooperative Communications and Networking," Cambridge University Press, 2008.
[18] Rui, Xianyi, Hou, Jia, Zhou, and Liulei, "Decode-and-forward with full-duplex relaying," IJCS, vol. 25, No.2, pp. 270–275, Feb. 2012
[19] D. S. Michalopoulos, H. A. Suraweera, G. K. Karagiannidis, and R. Schober, "Amplify-and-Forward Relay Selection with Outdated Channel Estimates," IEEE Trans. Communications, vol. 60, No. 5, May. 2012
[20] S. Ikki and M. H. Ahmed, "Performance analysis of cooperative diversity wireless networks over Nakagami-m fading channel," IEEE Communications. Lett., vol. 11, pp. 334–336, Apr. 2007
[21]Q. Deng and A. G. Klein , "Relay selection in cooperative networks with frequency selective fading," EURASIP Journal on Wireless Communications, , Nov. 2011
[22] S. M. Alamouti and S. Kallel, "Adaptive trellis-coded multiple-phase-shift keying for Rayleigh fading channels", IEEE Transactions on Communications, Vol. 42, No. 6, pp. 2305-2314, Jun. 1994.
[23]Y. Zhang, D. Yuan, and H. Zhang, "Adaptive LDPC for Rayleigh fading channel," Computers and Communication, 2004. Proceeding ISCC 2004. 9th International Symposium on, Vol. 2, pp. 651- 656, Jul. 2004.
[24] Y. Jiang, "A Practical Guide to Error-Control Coding Using MATLAB®," Artech House, 2010.
[25] R. Hamming, "Error detecting and error correcting codes," Bell System Technical Journal, Vol. 29, pp. 147 160, Apr. 1950.
[26] I. Reed and G. Solomon ,"Polynomial codes over certain finite fields," Society for Industrial Applied Mathematics Journal, Vol. 8, no. 2, pp. 300-304, Jun. 1960.
[27] Bose, R.C.; Ray Chaudhuri, D.K," On a class of error correcting binary group codes," Inform. Contr. 1960, 3, 68-79.
[28]P. Elias , "Coding for Noisy Channels," IRE Conv. Rec., Part 4, 1955, pp. 37–47.
[29] Viterbi, A. J, "An Intuitive Justification and a Simplified Implementation of the MAP Decoder for Convolutional Codes," *IEEE J. Selected Areas Commun*, Vol. 16, No. 2, February 1998, pp. 260–264.
[30] Robertson, P., E. Villebrun, and P. Hoher, "A Comparison of Optimal and Sub-Optimal *MAP* Decoding Algorithms Operating in the Log Domain," *Proc. Intl. Conf. Commun*.,Seattle, WA, 1995, pp. 1009–1013.
[31] Gallager, R. G, "Low-*Density Parity-Check Codes*,"Cambridge, MA: The MIT Press, 1963.
[32] Gallager, R. G, "Low-Density Parity-Check Codes," IRE Trans. Info. Theory, Vol. IT- 8, January 1962, pp. 21–28.

[33] MacKay, D. J., and R. M. Neal, "Near Shannon Limit Performance of Low Density Parity Check Codes," Electron. Letter, Vol. 32, No. 18, 1996, pp. 1645–1646.
[34] MacKay, D. J, "Good Error-Correcting Codes Based on Very Sparse Matrices," IEEE Trans. Info. Theory, Vol. 45, No. 2, March 1999, pp. 399–431.
[35] IEEE 802.16e, "Air Interface for Fixed and Mobile Broadband Wireless Access System, "IEEE P802.16e/D 12 Draft, Oct. 2005.
[36] IEEE 802.11.n, "Wireless LAN Medium Access Control and Physical Layer specifications: Enhancements Higher Throughput," IEEE P802.16n/D1.0, Mar. 2006.
[37] European Telecommunications Standards Institute (ETSI),"Digital Video Broadcasting (DVB) Second generation framing structure, channel coding and modulation systems for Broadcasting," EN 302 307 VI. 1.1. www.dvb.org.
[38] Tanner, R. M, "A recursive approach to low complexity codes," IEEE Transactions on
Information Theory, vol. IT-27, pp. 533–547, Sept. 1981.
[39] T. J. Richardson and R. L. Urbanke , "Efficient encoding of low-density parity-check codes, " IEEE Transactions on Information Theory, Vol. 47, no. 2, pp. 638-656, Feb. 2001.
[40] S. Myung, K. Yang and J. Kim, "Quasi-cyclic LDPC codes for fast encoding, " IEEE Transactions on Information Theory, Vol. 51, no.8, pp. 2894- 2900, Aug. 2004.
[41] M. P. C. Fossorier, "Quasi-cyclic low-density parity-check codes from circulant permutation matrices, "IEEE Transactions on Information Theory, Vol. 50, no. 8, pp. 1788-1794, Aug. 2004.
[42] T. Brack, M. Alles, F. Kienle, and N. When, "A synthesizable IP Core for WiMax 802.16e LDPC Code Decoding," In Proc. 2006 Personal Indoor and Radio Communications Conference (PIMRC '06), Helsinki, Finland, Sept. 2006.