EdHat International Research Conference on Information Communication Technology (ICT)

"Research in ICT opens the gate to the modern world"

EdHat INTERNATIONAL
CONFERENCE
23RD NOVEMBER 2018

PARTNER UNIVERSITIES

bucks new university

LINCOLN UNIVERSITY COLLEGE

UWU Uva Wellassa University

AMERICAN INTERNATIONAL UNIVERSITY PRAESIDIUM 1994 BANGLADESH

www.edhatinternational.co.uk

EdHat INTERNATIONAL

CONFERENCE

23RD NOVEMBER 2018

Editor in Chief
Prof Rohana P Mahaliyanaarachchi

Publisher
EdHat International, UK

Organiser
IDM Nations Campus, Sri lanka

All papers are presented in the form in which it was delivered by the authors.
The Organiser is not responsible for statements advanced in papers or
spelling and grammar irregularities.

**EdHat INTERNATIONAL RESEARCH CONFERENCE ON
INFORMATION COMMUNICATION TECHNOLOGY (ICT)**
23RD NOVEMBER 2018, COLOMBO, SRI LANKA

ORGANISING COMMITTEE

Chairperson:

Prof (Dr) Rohana P Mahaliyanaarachchi
(EdHat International, UK)

Members:

Prof. Dayantha Wijeyesekera
(IDM Nations Campus, Sri Lanka)

Janagan Vinayagamoorthy
(EdHat International, UK)

Dr. Carmen Z. Lamagna
(American International University Bangladesh)

Dr Sainey Faye
Buckinghamshire New University)

Dr. M.M.S.N. Premetilake
(Uva Wellassa University)

Dr. Neel Wasantha
(Lincoln University Malaysia)

Editorial Foreword

Prof (Dr) Rohana P Mahaliyanaarachchi
Conference Chairperson and Editor in Chief

Through over eleven years of international educational contribution of cross-national achievement, EdHat International, UK has contributed substantially to the development of higher education and research in number of fields such as IT, Computer Science, Management, Law, Sports Science, etc. The aim of the EdHat International Research Conference on Information Communication Technology (IRCICT) is to provide an international recognized academic forum for researchers to present their research and scholarly knowledge. Because of its international scope, issues can be examined in both a comparative and global context. The main objectives of the IRCICT are to:

- stimulate the exchange of ideas among researchers of different research orientations;

- stimulate creative dialogue among scholars and researchers;

- promote openness to new approaches of data analysis;

- facilitate the evaluation of different research methodologies;

- respect all viewpoints that may contribute to breaking new grounds in ICT and related area research;

- provide greater understanding of the important role that education plays in the development of nations and in shaping individuals.

The EdHat International Research conferences will take place regularly and continuously on different themes at host institutions around the world, also provides an important opportunity for researchers to meet after the period of study participation is over.

International conferences are great opportunities not only for researchers and scientists, but also for experts, policy makers, stakeholders and students. Here are six reasons for you to attend International conferences in your field. International conferences are great opportunities not only for researchers and scientists, but also for experts, policy makers, stakeholders and students. Here are six reasons for you to attend International conferences in your field. They are learning, discussion, presentation, visiting a new place, networking and academic reputation.

By attending this international conference, you get the chance to listen to different points of view and learn new ideas and trends in ICT and related area. It provides you with new techniques, new types of research approaches, data that is yet to be published, and investigators that you may not have heard of.

Therefore, I expect this conference will be a worthwhile and useful one for all the presenters, participants as well as partner universities and organisers.

I must thank to all four partner universities, IDM Nations campus as the key organizer of the event, organising committee, research paper presenters, reviewers, participants and members of the conference office for their great contribution in making this conference a successful one.

TABLE OF CONTENTS

CONTENTS

CONFERENCE CHAIR

Prof (Dr) Rohana P Mahaliyanaarachchi
Conference Chairperson and Editor in Chief

Prof Rohana P Mahaliyannarachchi is the former Vice Chancellor of the Sabaragamuwa University of Sri Lanka. He is the Senior Professor (Chair) of Agri Business Management of the Department of Agri Business Management of the Faculty of Agricultural Sciences of the Sabaragamuwa University of Sri Lanka.

He had participated as a quality assurance reviewer for more than 25 subject reviews during the first round of QA reviews in the university system in Sri Lanka. Further, he has given his contribution as a QA reviewer in Library reviews, Postgraduate Institutions and Graduate faculties reviews in number of Universities in Sri Lanka
.
Most important, he has given his expertise as an Institutional QA reviewer for Institutional reviews in 6 universities in Sri Lanka in 10 occasions. He chaired the review team in four occasions. Apart from Sri Lanka, he has contributed his expertise as a QA reviewer in the Universities in Bangladesh. Prof Mahaliyannarachchi has involved in preparation of code of practices and other QA documents in Sri Lanka and he worked as senior consultant of Quality Assurance of EdHat International, UK that is a UK based educational body during his sabbatical period.

Further, he has worked as a member and chair of the reviewing panel of recognizing degree programmes for private higher education institutes appointed by Ministry of Higher Education, Sri Lanka. He was a member of the panel of preparing the Higher Education Policy of Sri Lanka during 2008/2009.

He has published 7 books and there are more than 30 Research papers which are published in reputed referred journals and more than 40 conference papers presented in national and international research forums under his name.

INVITED SPEAKERS

Chief Guest

Professor K.K.C.K Perera
Vice Chancellor - University of Moratuwa
Bandaranayake Mawatha,
Moratuwa 10400
Sri Lanka

Kapila Perera is a Professor in the Department of Mechanical Engineering who joined the University staff in 1995, also served as a Lieutenant Commander of the regular force of the Sri Lanka Navy prior to joining the Department. He obtained a Masters Degree from University of California Davis in 1989 in Mechanical Engineering, followed by Ph.D. in Mechanical Engineering in 1993 from the same University. Prof. Perera's research interests include Turbulent Heat Transfer and Combustion Modeling, Energy Sources, Renewable Energy Technologies and Co-generation.

Keynote Speaker

Dr. Carmen Z. Lamagna
American International University, Bangladesh (AIUB)
408/1, Kuratoli,
Dhaka 1229,
Bangladesh

She is the current vice-chancellor of American International University Bangladesh (AIUB), appointed by the President of Bangladesh. She served as the president of Association of Universities of Asia and the Pacific (AUAP) and treasurer of International Association of University Presidents (IAUP).

Lamagna earned her bachelor's in Chemical Engineering from Adamson University, Manila in 1978. She got three master's degrees from Philippine Normal University (1990), Rizal Technological University (1994) and University of the East (1995). In 2003, she obtained her Ph.D. in business administration from California Coast University.

Lamagna was a faculty member of Perpetual Help University (1978–1979) and AMA Computer College (1982–1994). She served as the project director of AIUB during 1994 - 1996 and as an assistant vice president of business development in AMA Computer College in 1996. She was appointed as the vice-chancellor of AIUB in 1997. She became the first female VC of a university in Bangladesh.

SESSION GUEST SPEAKERS

Mr. Alan Clarke
EdHat Chief Academic Verifier
EdHat International,UK
Unit 1, 324 Stafford Road,
Croydon,
Surrey,
CRO 4NH,
United Kingdom

Dr. Carmen Z. Lamagna
American International University, Bangladesh (AIUB)
408/1, Kuratoli,
Dhaka 1229,
Bangladesh

Dr Sainey Faye
Principal Lecturer & Course Leader for BSc Accounting
and Finance
Bucks Business School
Buckinghamshire New University
High Wycombe Campus,
Queen Alexandra Road,
High Wycombe
Buckinghamshire HP11 2JZ
United Kingkom

Dr. M.M.S.N. Premetilake
Head of the Department of Computer Science and
Technology,
Uva Wellassa University,
Badulla,
90 000,
Sri Lanka

SESSION GUEST SPEAKERS

Dr. Jayalath Bandara Ekanayake
Senior Lecturer
Sabaragamuwa University of Sri Lanka,
P.O. Box 02,
Belihuloya, 70140,
Sri Lanka

Dr. Shanmuganathan Vasanthapriyan
Senior Lecturer
Sabaragamuwa University of Sri Lanka,
P.O. Box 02,
Belihuloya, 70140,
Sri Lanka

CONFERENCE
PAPERS

A DIGITAL GREEN BANKING FRAMEWORK: A BACKGROUND ANALYSIS FOR FINGERPRINT APPROACHES

N. Lojenaa[1], Ruwan D. Nawarathna[2], S. S. Suthaharan[1]

[1]Department of Physical Science, Vavuniya Campus of the University of Jaffna, Vavuniya, Sri Lanka

[2]Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka

INTRODUCTION AND OBJECTIVES

The fundamental activity of the bank is the money transaction. To perform such transaction, the banking service providers, expect a valid and accurate signature from the customer to verify them. The customer could step forward to make their necessary actions with the placement of signature only.

Even though the signature is an essential factor in Sri Lankan banks, the customers and employees are facing several problems in banking activities. Moreover, the customers change their signature frequently, children and teenagers also can change their signatures until they established a unique one and elders who got the nerve and eye disorder, unable to place their signature properly.

Therefore, the digital fingerprint can be implemented in a banking transaction to make more comfortable services. The fraudulent withdrawal, unauthorized transactions, online financial forgeries, and other offenses can be easily tracked and reported to the relevant legislative actions through this approach. This is because fingerprint is the most secure and powerful biometric element to identify unique customer's identity and validate electronically [1].

RESEARCH METHODS

The initial step is to get the fingerprint as an essential input data into the bank database to store as a unique and primary record. Fingerprint scanners are the input device which senses the pattern of a fingerprint after touching the finger on the surface of a fingerprint scanner. Fingerprint scanners are programmed with a pattern matching algorithm. Generally, a fingerprint scanner captures the fingerprint as an image format, then convert ridges and valleys of the fingerprint region in the image into a binary format (black for ridges and white for valleys) [2].

A banking staff can enter information of the customer along with scanned fingerprint. Once a customer gives his/her fingerprint, then they can gain other facilities without any additional procedures and information. When a customer needs to withdraw money, they can feed their fingerprint using scanner instead of filling forms and place signature. Thereafter the system can verify placed fingerprint with stored fingerprint from the banking database using dedicated pattern matching algorithm for the confirmation of a customer. Further, the fingerprint scanner can be placed with ATM for verification of a customer instead of PIN number and possible to control unauthorized ATM access when its necessary. This may help to block accessing the stolen or lost ATM cards.

This study analyses the possibilities of implementing the fingerprint system in the banking transactions. The customers from several departments were interviewed, and the data was collected through a structured questionnaire according to the guidance of banking administrators. The practical difficulties of manual signature in daily banking practices

for customers and employees have been observed. Further, for the purpose of performance comparison between a number of ATM transaction and manual transaction of selected dates of months' especially salary and pension dates were collected. Further, the existing literature and frameworks were studied to propose a better fingerprint based digital transaction mechanism for the banks. The fingerprint algorithm should have higher throughput, computational efficiency and accuracy as the bank must maintain the quality of service and fast performance to cater a wide variety of customers.

RESULTS AND DISCUSSIONS

According to the survey with banking stakeholders, more than 80% of the bank customers and employees expect the technological advancement of green banking. More than 86% of customers got a knowledge regarding the fingerprint system and nearly 48% of customer knows that already fingerprint system is adopted in several foreign banks. Customers and employees are interested in practice with the fingerprint system rather than a manual signature system. Sri Lankan banks are highly busy on 10th, 20th and 25th dates in a month due to the pension, teacher's salary and other salaries respectively. During these peek days minimum 400 and maximum 700 transactions available in town branches, 350 to 400 transactions are possible in the out of city branches via the counters of the banks. ATM and CDM consist maximum 900 and minimum 150 number of transactions. The number of transactions through the counters depends on the time that the customers spend in the banks. Employee's service time per customer also increase based on the delay of transactions. If the banks introduce a fingerprint system, then the bank can increase the number of transactions approximately by double. Further, the customers also can get faster and accurate service.

Customers can place their fingerprint to withdraw money. If the placed fingerprint is correct, then the system allows to continue further operations regarding withdrawal otherwise it tries up to the maximum level of the attempt, if the attempt is exceeded, then the system stops the transaction and identify the person as not an owner of the account or a perpetrator immediately.

Further, the digital fingerprint framework allows to identify bank frauds easily, detect perpetrator, and unauthorized transactions on time by bank officers before any complaint from customers. This system can be reducing the internal audit works and minimize the time to tally the account at the end of the day.

Based on the literature, there are efficient algorithms available to match the accurate fingerprint from the banking database. Nearly seven algorithm gives more than 90% of accuracy for fingerprint matching algorithm. Especially, "A correlation-based fingerprint verification system", "Pores, Difference of Gaussian filtering, pore – valley descriptor", and "Redundant discrete wavelet transform, Dezert– Smarandache theory" gives 98%, 98%, and 97.98% of accuracy respectively [3].

Some of the governments and non-government sectors introduce fingerprint system to register the attendance of employees. Most of the other developed and developing countries adopt fingerprint technology in the banking sector while doing mobile and online transactions. ATM also installed with fingerprint system as a password protector [4]. Therefore, adopting a fingerprint system to Sri Lankan banks are an effective and efficient strategy to provide better and secure customer services and protect both customers and employees from banking fraud.

CONCLUSIONS

Banking Sectors are getting the manual signature of customers for their basic requirements to fulfill the banking activities; this is a burden procedure to customers as well as employees. This study proposed to convert the traditional signature system into digital fingerprint system in banking sector. More than 80% of general public got a knowledge regarding fingerprint technology and they are eager to accept new technology in the Sri Lankan banking sector. Efficient fingerprint matching algorithms are available with 98% accuracy for match fingerprints. This system increases efficiency and reduces the time consuming of banking activity. It can be used to identify unauthorized activity and banking fraudulent access immediately and accurately. It helps to change over banking framework into digital green banking. Protect unsafety transactions and reduce manual, paper-based works in banks. The banks can attract more customers, and customers would feel secured while dealing their transactions with banks.

REFERENCES

1. Israa M. Alsaadi. (2015). Physiological Biometric Authentication Systems, Advantages, Disadvantages and Future Development: A Review, International Journal of Scientific & Technology Research, 4(12), 285-289

2. Annapoorani. D & Caroline Viola Stella Mery. M, (2014). A Survey Based on Fingerprint Recognition – Minutiae. International Journal of Science and Research (IJSR), 3(10), 607-611

3. Shoba Dyre, & C.P. Sumathi. (2016). A survey on various approaches to fingerprint matching for personal verification and identification, International Journal of Computer Science & Engineering Survey (IJCSES), 7(4), 1-17

4. Priyanka Mahajan, et al. (2016). Secured Internet Banking Using Fingerprint Authentication, International Journal of Innovative Research in Computer and Communication Engineering, 4(3), 6199-6208

EdHat INTERNATIONAL
CONFERENCE
23RD NOVEMBER 2018

METHOD TO DEVELOP AUTOMATIC VEHICLE NUMBER PLATE RECOGNITION SYSTEM

P.M.C.G.Wickramarathnaa, R.L.Dangallaa

aDepartment of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

Keywords : Bounding box, Optical Character Recognition, Localization, Template matching

INTRODUCTION AND OBJECTIVES

A vehicle number plate is a metal plate contains numbers and characters, fixed on to the car body. It used to identify the vehicle. Within past few years the number of the vehicle in Sri Lanka had been increased. With this it makes difficult to keep track of each vehicle for law enforcement. Many techniques have been used for numbers recognition. Each technique has its own advantages and disadvantages. An effective license plate recognition system needs to deal with light changes, weather condition, car movement, damages to the license plate and environmental noise. Recognition of vehicle numbers are challenging tasks due to various formats are used in Sri Lanka. Sri Lankan vehicle number plates include Sinhala letters, provincial codes and various other symbols. The aim of this study is to find a method to develop automatic number plate recognition system.

RESEARCH METHODOLOGY

This approach is implemented and verified using MATLAB. Basically this method consists of three major steps. They are number plate localization, affine transformation and character segmentation, and character recognition. In Plate localization approach input image is subjected to several morphological operations. For further processing input image is converted in to gray scale. Median filtering is applied for noise removal. Noise removed image is subjected to binarization. Horizontal and vertical edges around the number plate region enhance as the first step in character segmentation. Horizontal and vertical edges which do not belong to number plate area are dilated. To obtain connected components in the number plate area the image is subjected to filling operation. Then the components outside the number plate area are removed. For the identification, properties inside the number plate area were enhanced. Letters and numbers are segmented using bounding box method. Template matching technique is used for character recognition in the number plate. Optical Character Recognition used to compare the individual characters against the alphabet and numbers in the template library.

RESULT AND DISCUSSION

In this study Matlab software tool is used. Because of it provided easy approach for debugging and correction of errors in algorithm (Gohil, 2010). Number plate recognition depend on varies factors such as image quality, illumination, viewpoint and vehicle speed.

Image localization was applied to recognize the vehicle number plate area and it is used to eliminate the background and preserve only the number plate area from the input image. A clear image is obtained to have a good quality feature extraction.

Color image is converted in to gray scale image to preserve memory and to speed up the processing. The aim of this step is to increase and improve the visibility of the image. Gray scale image does not have the hue and saturation information while it has the luminance. Median filtering is applied to edge preserving, remove noise and useless and distracting information. At the end noise is reduced and it produces a good quality set of image data for recognition.

Binaries image have less quantity of information in the image and it guard the relevant data. Character segmentation is applied to distinguish objects from background. Failure character segmentation leads to improper division of the characters or improper merging of characters. The sobel edge detection process applies on the image to reduce the amount of data in an image and to preserve the structural properties for further processing. Edge detection performs the locating sharp discontinuities in an image. Different factors were considered when selecting structuring elements for the processing. They are size and shape of the image. Dilation process is applied to fill holes, sharpen the edges of object in an image, join the broken lines and increase the brightness of an image. Interim advantage of dilation step is noise removal (Bhat & Mehandia, 2014). After filling operation it yield a solid segmented image. Connected components were removed to obtained number plate area. During this step plate components were not affected. Because it is surrounded by black background (Anupa H & L, 2016). Obtaining number plate area is the most sensitive and difficult step. Distance, angle of the camera, environmental changes, dirt or damages to the license plate and diverse location of the license plate on cars are the condition which affects on this step. Bounding box created over each character and number plate, each character and number presented on number plate, each character and number is separated out for recognition of number plate (Singh & Sharma, 2015).

In this study template matching is used as its classification based on the individual pixels. An image is compared with predefined images, which are referred to as templates. Optical character recognition is used to recognize an optically processed printed number plate which is based on template matching

CONCLUSIONS

Broken number plate, blurry images, number plate not within the legal specification, low resolution of the characters, and poor maintenance of the vehicle plate are some difficulties associated with character recognition. Accuracy of the segmentation can be improved by processing. Structuring element selection is affect on the preprocessing. Font type, noise in image and tilting is affect on the effectiveness of template matching. In recognition stage the characters on the number plate are converted into texts.

REFERENCES

1. Anupa H, & L, U. (2016). RECOGNITION OF MOVING VEHICLE NUMBER PLATE USING MATLAB. INTERNATIONAL JOURNAL OF ADVANCED COMPUTING AND ELECTRONICS TECHNOLOGY, 3(3), 5.

2. Bhat, R., & Mehandia, B. (2014). RECOGNITION OF VEHICLE NUMBER PLATE USING MATLAB. INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING, 2(8), 5.

3. Gohil, N. B. (2010). CAR LICENSE PLATE DETECTION. (MASTER OF SCIENCE In ELECTRICAL AND ELECTRONIC ENGINEERING Master Of Science), CALIFORNIA STATE UNIVERSITY, SACRAMENTO FALL.

4. Singh, R. R., & Sharma, R. (2015). Automatic License Plate Recognition. IJLTEMAS, 4(9), 16.

COMPARING TRADITIONAL SECURITY MECHANISMS WITH SECURITY AND PRIVACY IN BIG DATA

K.L Mahaliyanaarachchi*, S. Ahangama,
Faculty of Information Technology, University of Moratuwa, Moratuwa, Sri Lanka

INTRODUCTION AND OBJECTIVES

The term big data is used to describe the large amount of data which are structured or unstructured. Nowadays big data has become part of modern science technology and businesses. These kind of data are produced from emails, audios, videos, online transactions, posts, images, medical records, search interrogations, science applications, and Social network interfaces [1, 2]. Therefore it is difficult to manage, capture, store, share and filter these kind of huge big data set, because of the massive size and the complexity (Non-structured form, in real time, or streaming) [1,3]. This paper examines and studies on security and privacy mechanism and solution in big data with regards to the architecture or available algorithm. In addition this will identify distinguish between traditional security mechanism and big data approaches.

RESEARCH METHODS

This paper is a review paper and discuss and in section II presents the characteristics of big data and big data approaches. Section III contains the big data privacy and security approaches. This section further categorizes topics under hadoop security, cloud security, monitoring and auditing, key management and anonymization. Section IV is the discussion of this review paper.

RESULTS AND DISCUSSION

To store large volume of data (terabyte and petabyte), Hadoop uses a bock structured distributed file system. It is called as Hadoop distributed file system.

When dealing with big data traditional solutions are not sufficient to make sure privacy and security of big data [13]. Access permissions, encryption schemes, transport layer security and firewalls can be broken; data can be unknown, anonymizes data also can be re-identified [6]. Because of above reasons advanced technologies and techniques are used to protect big data. In this paper security and privacy issues for big data have categorized under four titles. They are Hadoop security, Cloud security, monitoring and auditing, key management and anonymization.

Hadoop was not originally developed for security and it is a distributed process framework [13]. That means Hadoop operates in trustable environment because it does not have high security mechanisms. Security mechanisms and precautions have been started to be developed after when Hadoop become more popular.

Today storing data in clouds becomes a main problem. In a secure way to handle and share big data the service providers must take some precautions [16]. Encryption, decryption, authentication and compression are methods used to secure big data in clouds. Authentication is used to authorize person to access in to the system by using email and password. Encryption method is used to keep data in more secure way. If someone attack or some problem happens to the server that encrypted data have been convert to decrypted data by using secret key [11]. Keeping three backup servers will reduce the vulnerability of losing the data.

Because of the massive increase of data usage and generation, big data needs high privacy and security mechanism for data storing, gathering, processing and transferring. Therefore this study was conducted to discuss about the security and privacy mechanism of big data related to algorithm and techniques. Because of the rapid increase of data generating and processing, in future big data security, privacy & safety will be a huge problem. The amount of data is doubled every year since 2011 and after 2015, it will more than that.

Today NoSQL becomes one of major part of big data, which is more capable of handling unstructured big data. Apache spark increases dramatically the speed of data processing, which is higher than Hadoop and due to this reason most of the companies are now shifting to apache spark. Hadoop projects also get matured and because of the trend that growing in Hadoop, it becomes the core part of big data.

Therefore new big data security technologies, techniques and solutions have to be developed or existing big data security mechanism has be improved. This study will be helpful to understand big data and its security approaches to develop better systems, structures, tools and solutions.

CONCLUSIONS

Big data needs extra privacy and security mechanism for gathering, storing, analyzing and transferring data. This review paper has discussed the characteristics of big data and its approaches. Further, it has discussed the privacy and security approaches with regarding to its algorithms and techniques. Hadoop security Trust mechanism and Random encryption mechanism are discussed here with relate to their algorithms and techniques. The big data security mechanism which related to cloud system has been discussed with their algorithms and also discussed the big data security mechanism under the topics of monitoring & Auditing, Key management & Anonymization. The important of big data security mechanism over traditional mechanism also discussed under above mentioned topics.

REFERENCES

Elisa Bertino "Big data- Security and Privacy" IEEE Interbational Congress on Big Data, 2015

Duygu Sinanc Terzi, Rmazan Terzi and Seref Sagirohlu "A Survey on Security and Privacy Issues in Big Data", ICITST-2015

V. G. Korat, A. P. Deshmukh, K. S. Pamu. Introduction to Hadoop Distributed File System. Int. J. of Engineering Innovations and Research, 1(2): 230-236, March 2012

Pradeep Adluru, Srikari Sindhaoori Datla and Xiaowen Zhang. "Hadoop Eco System for Big Data Security and Privacy", 978-1-4577-1343-9, 2015

ENERGY EFFICIENT VM CONSOLIDATION IN CLOUD COMPUTING ENVIRONMENTS

Jayathilake.T.M.C.S[1] , Jayasena.K.P.N[2]
Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya

Key Words: Energy Consumption, GA, SLAV, VMC, VM Migration

INTRODUCTION AND OBJECTIVES

Virtual Machine Consolidation (VMC) is an emerging solution used in cloud data centers. It is a technique which is performed to maintain minimal number of active hosts by migrating and consolidating the Virtual Machines (VM) into a reduced number of Physical Machines (PM) to conserve energy. Cloud providers have the obligation to provide a suitable Quality of Service (QoS) dealing with energy performance trade off. There are only few algorithms which may produce a better solution within a reasonable amount of time. The VMC policy should change according to the situation. The main contribution of this research is to analyze in the area of VMC in cloud platform and to propose efficient VMC methods using Genetic Algorithm (GA) for better energy consumption, SLA violation with optimum resource utilization. Evolutionary algorithms are used to solve dynamic population based problems and we address a cloud environment where VMs are dynamically migrating.

RESEARCH METHODS

For this research Cloudsim simulation tool have been used with Planetlab data gaining scalable, repeatable and dependable cloud experience for experiments. Cloudsim provides energy aware simulations and dynamic workload application services. CoMon monitoring system provides data from Planetlab nodes which are distributed worldwide.

Table 1 Cloudsim configuration

Number of hosts	800
Number of VMs	1052
Total simulation time	1 day
Date	03/03/2011
Mean	12.31%
Standard deviation	17.09%
Median	6%

The VMC process consists of following steps; host overload detection / host under load detection, VM selection and VM placement. Median Absolute Deviation (MAD), Interquartile Range (IQR), Local Regression (LR), Robust Local Regression (LRR) and Static CPU Utilization Threshold (THR) are the five schemes used to estimate the dynamic CPU utilization threshold for overload host detection. Minimum Migration Time (MMT), Random Selection (RS) and Maximum Correlation (MC) are the three policies used for VM Selection. In this research we are using Genetic Algorithm as the VM Placement algorithm. GA imitates the natural evolution. It is the biological evolution of chromosomes. Better solutions are obtained by recombining with each other based on the idea of the fittest survival. It has the following steps; encoding, chromosome generation, selection, crossover, mutation, eletism and termination. The total number of VMs to be places is represented in number of genes.

RESULTS AND DISCUSSION

In combination of the host overload detection algorithms and VM selection algorithms the research studies 15 Planetlab cloud data based experimental results. Each combination encompasses with the evolutionary algorithm; GA as the VM placement algorithm. MAD behaves better with distributions without a mean or a variance. It is more resilient to outliers in a dataset. MAD is the absolute deviation from the data's median. IQR is a robust statistic and it is a statistical dispersion which can be used to set an adaptive upper utilization threshold. IQR is also known as midspread or middle fifty. It is equal to the first and third quartiles. LR fits simple models to localized subsets of data to make a curve that approximates the data. A tricube weight function is used here to assign neighborhood weights. LRR is an estimation method bisquare to the least squares method for fitting a parametric family. THR dynamically adapts the value of the threshold according to the statistical analysis of workload history. Each host overload detection algorithm combines with 3 VM selection algorithms. MMT selects the VMs which needs a minimum migration time to migrate from the VMs at the list of VMs in the host in terms of resources such as the usage of RAM within a unit supplementary bandwidth available for a host. RS is also called as Random Choice Policy, selects VMs in a random order from a uniformly distributed set of VMs for the migration. GA imitates the natural evolution. It is the biological evolution of chromosomes. A better solution is obtained by recombining with each other based on the idea of the fittest survival. The chromosomes are followed by hosts and genes by VMs. The performance of each combination are measured using five metrics. Energy consumption calculates the total power consumption of the physical machines.
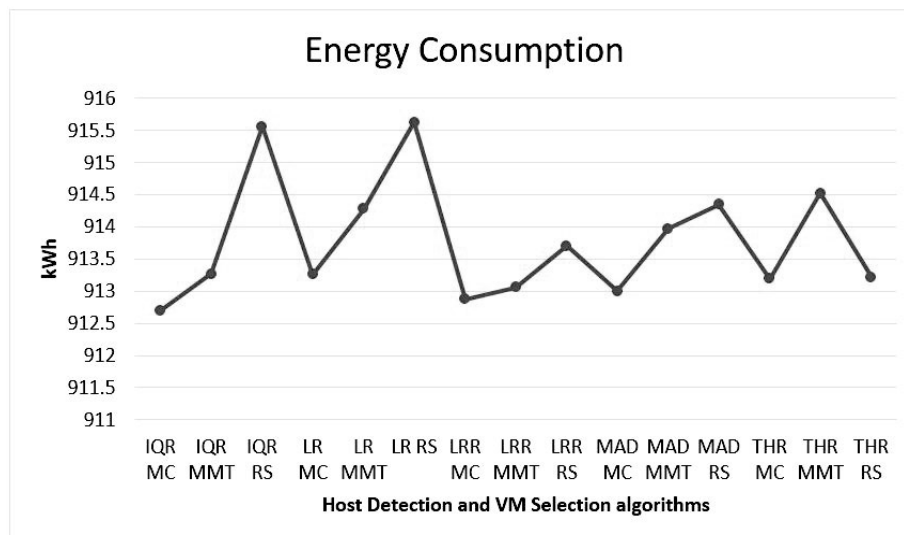

Figure 1 Energy consumption
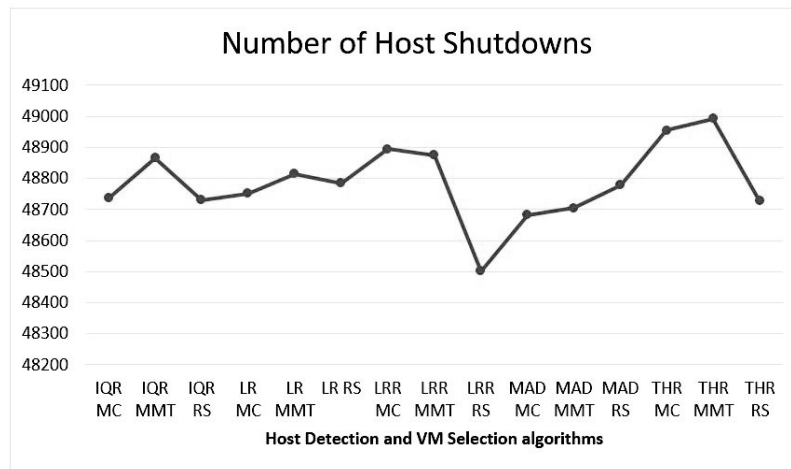
The number of host shutdowns are considered.

Figure 2 Number of host shutdowns

This collects data for the number of Host shutdowns in a unit time and number of VM migrations in a unit time. VM migration consists of four steps. Selecting the PM which is overloaded or under loaded, selecting VMs from the server, selecting a PM to place VMs and transferring VMs to destination server.
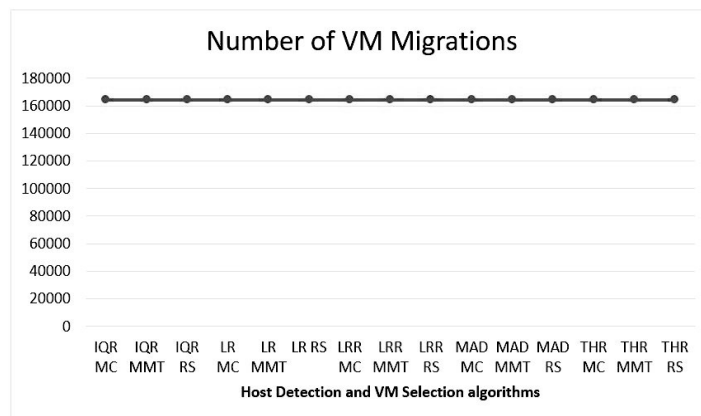

Figure 3 Number of VM migrations

SLA violation is calculated related to SLA violation time per active host and performance degradation due to migrations.
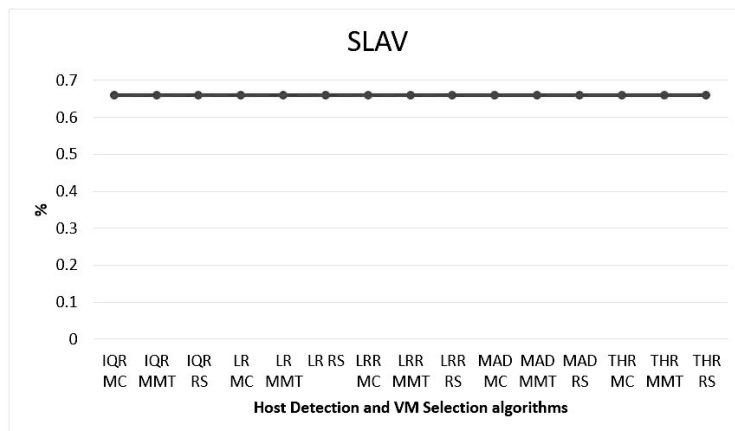

Figure 4 SLA Violation

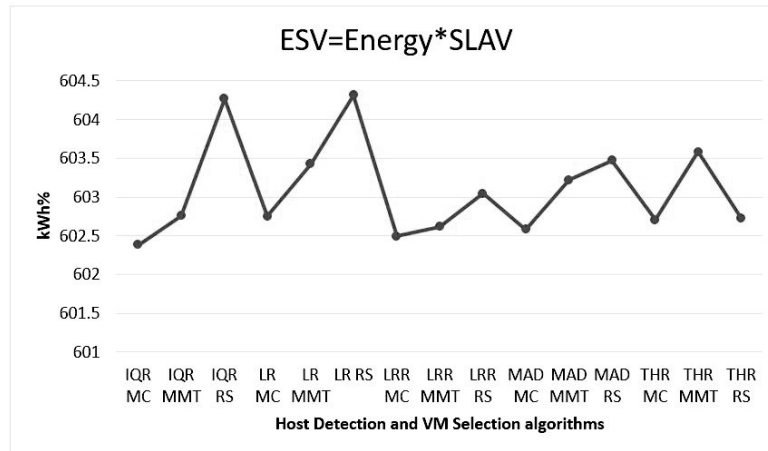Energy SLA violation is calculated related to energy consumption and SLAV.

Figure 5 Energy SLA Violation

CONCLUSIONS

The aspect of the research suggests that the Genetic Algorithm can be adopted to solve Virtual Machine Consolidation problem according to the graphically illustrated simulation results. The proposed model achieves better power consumption, SLAV for optimal resource utilization shown under five metrics energy consumption, SLA violation, energy SLA Violation, number of host shutdowns and number of VM migrations.

REFERENCES

1. Beloglazov, Anton Buyya, R. (2012). Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers. Concurrency Computation Practice and Experience, 24(13), 1397–1420. https://doi.org/10.1002/cpe.1867

2. Chowdhury, M. R., Mahmud, M. R., & Rahman, R. M. (2015). Implementation and performance analysis of various VM placement strategies in CloudSim. Journal of Cloud Computing. https://doi.org/10.1186/s13677-015-0045-5

3. Theja, P. R., & Khadar Babu, S. K. (2015). An evolutionary computing based energy efficient VM consolidation scheme for optimal resource utilization and QoS assurance. Indian Journal of Science and Technology, 8(26). https://doi.org/10.17485/ijst/2015/v8i26/77179

4. Theja, P. R., & Khadar Babu, S. K. (2016). Evolutionary computing based on QoS oriented energy efficient VM consolidation scheme for large scale cloud data centers. Cybernetics and Information Technologies, 16(2), 97–112. https://doi.org/10.1515/cait-2016-0023

NATURE INSPIRED APPROACH FOR EFFICIENCY AWARE RESOURCE SCHEDULING IN CLOUD COMPUTING

Abeywardana.V.P.I.S[1], Jayathilake.T.M.C.S[2], Jayasena.K.P.N[3], Vasanthapriyan.S[4]
Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya

Key words : Algorithms, Evolutionary, Load Balancing, Nature, Resource Scheduling

INTRODUCTION AND OBJECTIVES

Cloud computing has become a trending topic on these days. Because of its high performance and on-demand computing which is an Internet-based computing that provides shared processing resources and data to computers and devices on demand. Cloud computing service vendors can't provide this kind of high performing computation power by maintaining dedicated resources for each application which are hosted in the data center. To provide this kind of a high performing service cloud vendors use several kinds of techniques. Among those techniques most critical processes of the cloud system are task scheduling and load balancing as it affects the performance of whole cloud system. This research is conducted to implement efficient resource management method for cloud computing. To complete this task objectives were defined as this.

I.   Identify the most performing load balancing algorithm form static and nature inspired evolutionary load balancing algorithms.

II. Identify the most performing task scheduling algorithm from static and nature inspired evolutionary task scheduling algorithms.

III. Improve selected algorithm by minimizing response time, makespan and data center processing time to achieve highest performance of the cloud systems.

METHODOLOGY

In this research, various load balancing and task scheduling algorithms have been studied and simulated them using Cloudsim and Cloud Analyst tools to identify best performing algorithms among those algorithms. Cloud Analyst is a tool that made on top of the Cloudsim. The speciality of this tool is the GUI. We don't need to manually upload resource sets, user base, data center specification. Also, we can easily get the result set and present it in a helpful way. Those algorithms are Throttled, Active Monitoring, Round Robin, Particle Swarm Optimisation, Ant Colony Optimisation and Bee Colony Optimisation. Within this research period Two experiments were conducted to check the performance of load balancing algorithms and task scheduling algorithms. First experiment was done to check the overall processing time of algorithms and variance of data center processing time of different task scheduling algorithms over different number of data centers. Second experiment was done to check the variance of makespan of different task scheduling algorithms over different number of tasks. And data centers with same specifications were used during both experiments.

RESULTS AND DISCUSSION

As discussed, Load balancing focuses on keeping the available resources equally busy and avoid overload of one machine with many tasks. There are two kinds of load balancing algorithms which are Static and dynamic load balancing algorithms. Load balancing can be considered as a part of scheduling where time is not considered. In cloud environments resource machines are the bottom part or the core part of the cloud and on top of that cloud providers

make different number of different sized virtual machines. When user started a task it automatically assigned to a virtual machine. But it is not much easy because task size and available resource amount is not a fixed number. Goal of static load balancer is to reduce the overall execution time of a concurrent program while minimizing communication delays. A general disadvantage of all static schemes is that the final selection of a host for process allocation is made when the process is created. In dynamic load balancing algorithms allocated resource can be changed according to the traffic it gets. Within this research period several static and dynamic load balancing algorithms were compared. To compare load balancing algorithms, we used Cloud Analyst as the simulation tool. We simulated performance of several load balancing algorithms under constant set of workloads. Brief of the final results set have been shown in the figure below.

Table 1: Load balancing simulation result

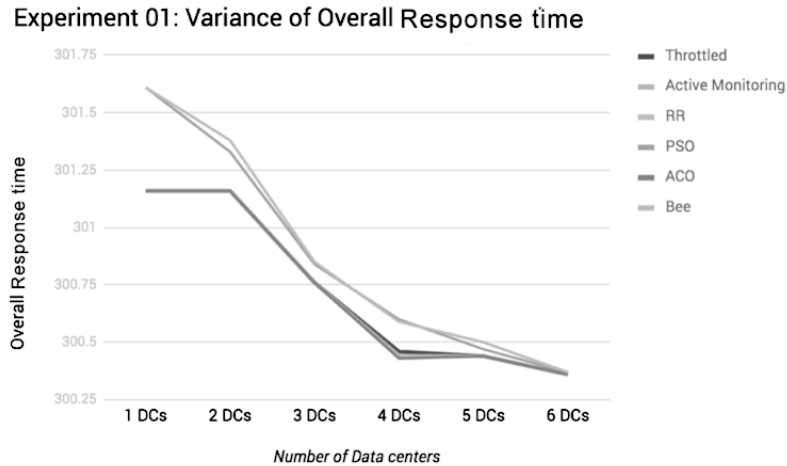|	Overall Response time						Data center processing time					
	Throttled load balancing algorithm	Active Monitoring load balancing technique	Round Robin load balancing algorithm	PSO	ACO	Bee	Throttled load balancing algorithm	Active Monitoring load balancing technique	Round Robin load balancing algorithm	PSO	ACO	Bee
1 Data Center with 100Vms	301.16	301.16	301.16	301.61	301.16	301.61	1.45	1.45	1.45	1.89	1.45	1.89
2 DCs with 50Vms	301.16	301.16	301.16	301.33	301.16	301.38	1.45	1.45	1.45	1.62	1.45	1.66
3 DCs with 35,35,30 Vms	300.76	300.76	300.76	300.84	300.76	300.85	1.05	1.05	1.04	1.12	1.05	1.13
4 DCs with 25Vms	300.46	300.44	300.44	300.60	300.43	300.59	0.74	0.73	0.72	0.88	0.71	0.88
5 DCs with 20Vms	300.44	300.44	300.44	300.47	300.44	300.50	0.73	0.73	0.73	0.75	0.73	0.79
6 DCs with (4 DCs with 15Vms and 2 with 20 Vms)	300.36	300.36	300.36	300.37	300.36	300.37	0.65	0.65	0.65	0.66	0.65	0.66
Mean	300.723	300.72	300.72	300.87	300.718	300.883	1.012	1.01	1.007	1.153	1.007	1.168
Median	300.61	300.6	300.6	300.72	300.6	300.72	0.895	0.89	0.885	1	0.89	1.005
Standard Deviation	0.333	0.335	0.335	0.455	0.337	0.460	0.334	0.335	0.337	0.455	0.338	0.456

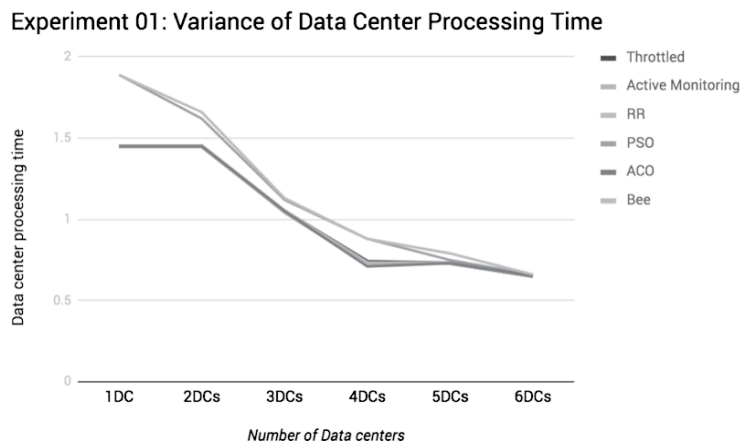Figure 1 Variance of Overall response time

Figure 2 Variance of Data Center processing time

According to the simulation results of the figure 1, all algorithms show a gradual decrement of response time and data center processing time when data center count is increasing. According to the results, ACO perform better than all other static and dynamic algorithms which we have simulated. In the same way figure 2 shows gradual decrement of data center processing time when data centre count is increasing.

Table 2 : Task scheduling simulation result

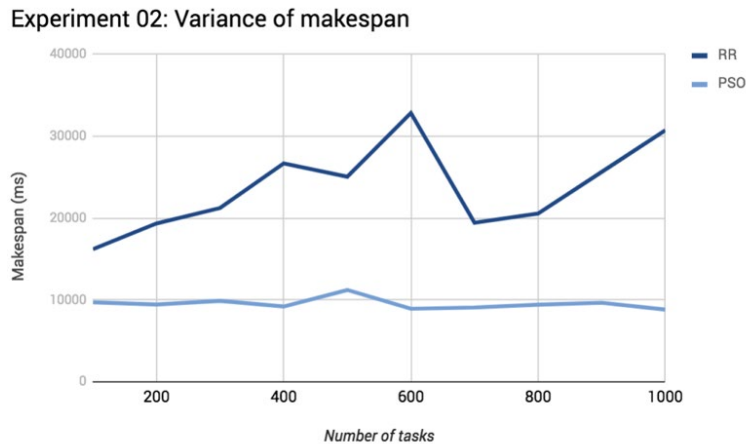|	Makespan	
Number of Tasks	RR load balancing algorithm	PSO
100	16203.30	9712.09
200	19348.80	9449.76
300	21242.61	9893.72
400	26692.27	9197.88
500	25059.25	11220.01
600	32858.90	8926.42
700	19440.50	9080.97
800	20566.86	9420.00
900	25646.25	9661.48
1000	30743.96	8827.63
Mean	23780.27	9538.996
Median	23150.93	9434.88
Standard Deviation	5071.091034	650.0894587

Figure 3 Variance of makespan

Task scheduling is a major and critical area of the resource management. Graph of the figure 3 has drawn According to the simulation results that produced by the CloudSim. In this research, we have executed Round Robin algorithm and Particle Swarm Optimization algorithm with different number of tasks and checked the variance of the makespan of those algorithms. According to this graph it clearly shows the performance gap between static and nature inspired evolutionary algorithms. Throughout this simulation process Particle Swarm Optimization algorithm performed better than Round Robin algorithm and Particle Swarm Optimization algorithm perfectly performed even when the task count is increasing.

CONCLUSIONS

According to the experiments which have done within this research nature inspired evolutionary algorithms showed high performance than other types of algorithms. According to the final results of load balancing and task scheduling simulations these results can be proposed: Ant Colony Optimisation algorithm for load balancing in the cloud platform for higher performance and Particle Swarm Optimisation algorithm based task scheduling technique can be used to minimize makespan of cloud platform.

REFERENCES:

1. Singh, S., & Chana, I. (2016). A Survey on Resource Scheduling in Cloud Computing: Issues and Challenges. Journal Of Grid Computing, 14(2), 217-264. doi: 10.1007/s10723-015-9359-2

2. Madni, S., Latiff, M., Coulibaly, Y., & Abdulhamid, S. (2016). Resource scheduling for infrastructure as a service (IaaS) in cloud computing: Challenges and opportunities. Journal of Network And Computer Applications, 68, 173-200. doi: 10.1016/j.jnca.2016.04.016

3. S. Singh and I. Chana. (2014) QRSF: QoS-aware resource scheduling framework in cloud computing. The Journal of Supercomputing, vol. 71, no. 1, pp. 241-292.

4. Kalra, M., & Singh, S. (2015). A review of metaheuristic scheduling techniques in cloud computing. Egyptian Informatics Journal, 16(3), 275-295. doi: 10.1016/j.eij.2015.07.001

DATA ANALYTICS IN FOG COMPUTING USING DOCKER AND KUBERNETES

H.M. Buddhika Priyabhashana. K.P.N. Jayasena
Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka.

Abstract - Nowadays Internet-of-Things (IoT) applications produce a large amount of data and require a powerful analytical process. These data should transmit to data centers for analytics. This kind of data may affect the networks, overload server memory. We are going to implement an architecture with the pre-analysis process which reduces data before sending them to data centers. We extend the existing fog computing platform to a more efficient with the help of Docker and Kubernetes. We implement the analytics applications in fog stations other than implementing them in the central server. As results: (i) Suggest extended application for Fog Nodes, (ii) Analyze and extract the result from data which generate from multiple Fog Devices, (iii) Application of analysis phase and monitoring phase of Fog Stations. (iv) Applicability of this experiment.

INTRODUCTION AND OBJECTIVES

In the current world Internet-of-Things (IoT) applications are very popular. When the number of IoT devices is getting higher, the amount of data is also getting larger. To handle, those data, require powerful analytical approaches for getting expected results because IoT devices usually have limited resources. IoT applications transmit data to resource-rich data centers for analytics. However, it may congest networks because of significant amount data, overload data centers, and increase security vulnerabilities. In this case, focus on implementing an analytical platform to generate a result before sending them to the data centers. We launch an analytical application using container management technologies among the devices without sending everything to the data centers and collaborating several IoT devices and pre-processing the data before transmitting them over the Internet. As a result, expect pre-processing data before sending them to the main server and reduce the size of data before storing them. We use Docker for dynamic deployment and Kubernetes for efficient management.

RESEARCH METHODS

As a solution, the proposed architecture is implemented using end devices called Raspberry Pi and add some sensors to the device. Sensors could be any type. Mainly focus on the combination of multiple devices together and called fog devices. After that pass them to the analytical point called fog station. We use Docker and Kubernetes for container management in fog station. Then analyze data, get a result and send only the result to the central server.
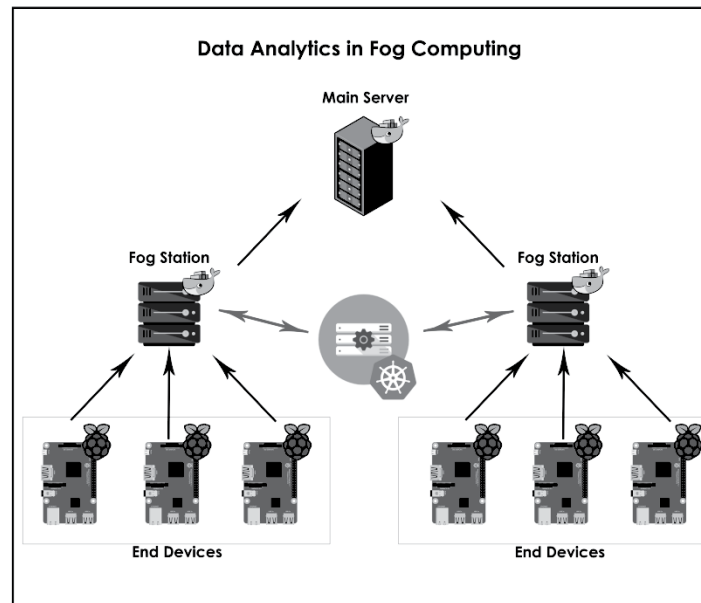
Figure 1. Overview of the structure.

By using this way can reduce data amount that has to store in the main server.

- End devices (Fog devices) – IoT devices that collect data via sensors. Sensors could be any type. When collecting data from fog devices, then pass them to a mini server called "Fog station."

- Kubernetes – Docker containers are running on heterogeneous devices at different locations, we need a tool to manage and monitor them, such as Kubernetes.

- Docker – In this case, hard to install all the libraries on every fog device. Therefore, use Docker containers to package the operators and required libraries into Docker images.

- Fog Station – Fog station is a server that responsible for multiple fog devices and manage them. Fog stations are an extended version of fog nodes. In this case, we introduce Fog station as a solution.
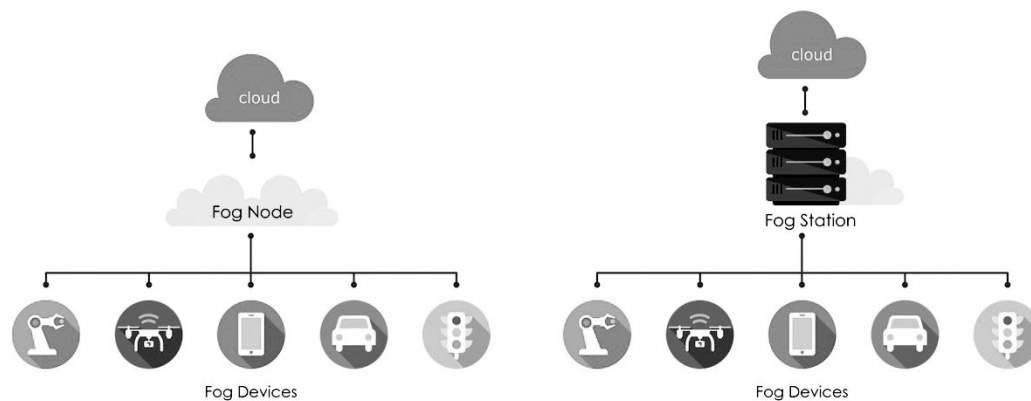

Figure 2. Application of Fog Node and Fog Station.

RESULTS AND DISCUSSION

There are many advantages could achieve from this system. There are many sensors in fog devices; generated data pass to the fog stations. By using Docker, we could install all the libraries that require for fog devices without any issues. Docker containers, package the operators and required libraries into Docker images.

These devices generate a processed data, send directly to the fog station. In fog station, there are several programs to handle different kind of data that received from different kind of data sources. The importance of the different programs, there are many fog devices connected to a single fog station such as motion detection, cameras, videos, gas sensors. So they have their analytical processes for each data source.

If there is a Fog node, the main issue is when functioning same devices such as weather data monitoring devices; then they are trying to access the application at Fog node for an analytical part in same time. We should have a method to avoiding this kind of conflicts. That is why we put the mini server with server capabilities called Fog station. It can handle this issue, and it could manage a different kind of devices also at the same time. When analyzing data using fog station send the result to the central server for future use. That is the way to reduce data amount that generated from each device. When considering this method, it is more efficient than the previous method. Because in the old method, data that generated in every device send them to the main server via Fog hub. However, by using this new method can analyze data and get a result from them before sending to the main server. As a result, can reduce data amount and avoid storing unnecessary data inside the main server and can save the memory of the main server.

Next important point is Fog station monitoring. We achieve this using Kubernetes. As mention above Docker containers running on multiple devices at different locations, we use Kubernetes to manage and monitor them. By using Kubernetes, it allows to monitor the performance of Fog stations such as memory, processing power. Then we can have a clear idea about processes inside the Fog stations and create a clear view of them. It will help to monitor performance also.

Following Graphs shows expected to result from the system. These graphs do not base on actual values. Only use graphs to simulate the outcomes.
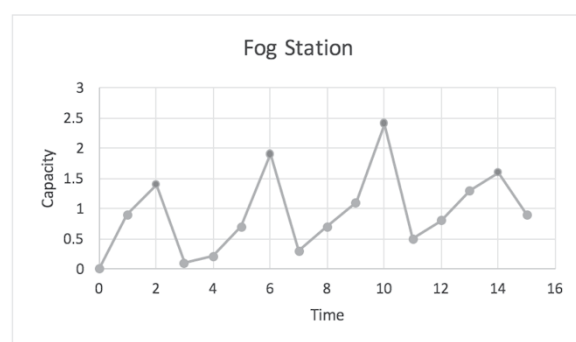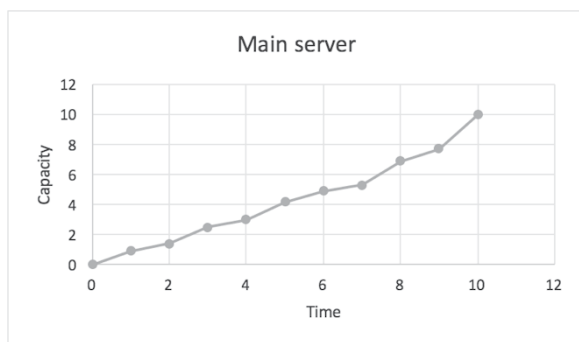

Figure 3 - Time vs Capacity related to main server　　　Figure 4 - Time vs Capacity related to Fog Station

Figure 3 shows the expected diagram for Time vs. Data Capacity in the server without using Fog Stations. The capacity continuously growing with time is shown in the figure. It will be a massive bulk to the server.

Figure 4 shows the expected Time vs Capacity graph inside the Fog Station server. Blue markers describe cut-points of the server for process data. Then send the result to the main server and remove processed data from Fog Station and keep the result inside Fog station.

Figure 5 shows the expected Time vs Capacity graph for the main server with and without Fog Station. When comparing these two graphs, it will be shown how fog stations help to save the capacity of the main server.
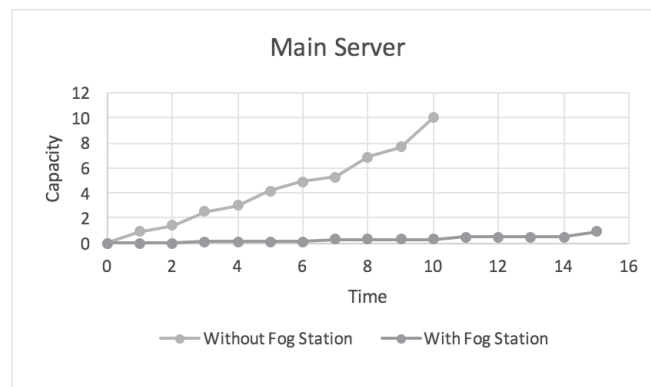
Figure 3 - Time vs Capacity related to main server (with and without Fog Station)

CONCLUSIONS

In this research, we propose an architecture for handle fog devices and reduce a large amount of data that generate by fog devices before storing in the main server. Mainly we focus on implement fog stations instead of fog nodes. Next, main part is handle data coming from different sources and analyze them using fog stations and get the result for making predictions or future use. This architecture covers data flow from fog devices to the main server. In this case, we adopt open-source projects such as Docker and Kubernetes. We used Kubernetes for analyzing and monitor the performance of fog stations.

REFERENCES

Hong, H., Tsai, P., & Hsu, C. (2016). Dynamic module deployment in a fog computing platform. Retrieved from https://ieeexplore.ieee.org/document/7830701/

Hong, H., Tsai, P., & Hsu, C. (2017). Distributed Analytics in Fog Computing Platforms Using TensorFlow and Kubernetes. Retrieved from https://ieeexplore.ieee.org/document/8094194/

Hong, K., Lillethun, D., Ramachandran, U., Koldehofe, B., & (2013). Mobile fog: a programming model for large-scale applications on the internet of things. Retrieved from https://conferences.sigcomm.org/sigcomm/2013/papers/mcc/p15.pdf

Wu, D., Arkhipov, I., Kim, M., Talcott, L., Regan, C., & McCann, A. (2016). ADDSEN: adaptive data processing and dissemination for drone swarms in urban sensing. Retrieved from https://ieeexplore.ieee.org/document/7497529/

HOW CAN BUSINESS UTILISE BUSINESS INTELLIGENCE FOR COMPETITIVE ADVANTAGES IN RETAIL & CONSUMER GOODS INDUSTRY

Dr. Madhawa S. Ranawake,

INTRODUCTION AND OBJECTIVES

Retailers and consumer goods companies face tough competition and businesses earn very thin margins. Almost every company mainly focus on improving on profit margins and how they can improve quality on products and services. Big data has become a hot topic in the consumer retail industry as it addresses one of the most critical issues challenging retail today.

The world of data is changing and retail and consumer goods companies are challenged by the increasing scale, complexity and velocity of data accumulation. Its evident that retail &consumer goods industry use big data for competitive advantage. Even though it's spread in every corner of the industry still lots of retail & consumer goods companies are trying to depend on an in-house data which accumulates data originated from internal systems. Which means those organisations are not utilizing big data technologies, where lots of companies are losing to the competition or not reaching the full potential benefits of easily available data for further processing.

In this study the main areas in focus are, to understand what technologies available for big data analysis, what are the inhibitors in adopting the big data technology, key benefits of using the technology and possible methods of adopting and achieving competitive advantage through implementing the technology.

RESEARCH METHODS

This research being carried out as a literature survey with inductive approach to identify the GAPs of current knowledge in implementing Big Data in retail industry. Those GAPs being analysed in order to find out root causes where those summarized to get into conclusion on how to rectify those. Various quantitative and qualitative analysis techniques were used to further validate the outcome conclusion. As this research was highly depend on published credible literature, both technology base research papers and business vertical focus research papers being selected.

RESULTS AND DISCUSSION

Literature review was conducted to explore the technology availability for big data. To further validate the data analysis outcome, analysis was carried out to check the vendor solution offering in big data business intelligence solution provides and what are the base technologies they employ. With the analysis of both outcomes, Hadoop can be identified as the key technology available. According to Michele and Ambiga (2012), even many technologies identified in big data technologies Apache Hadoop is the main database technology used by many instances. Chodorow (2013), describes MongoBD is another database technology available in big data practice and that comes under NoSQL technology.

When analysing on challenges faced in the retail and consumer industry in big data implementation. Gantz&Reinsel (2012) describe Real-time security monitoring and privacy has always been a challenge, given the volume and velocity of data streams. In retail and consumer industry more private data is being gathered in in-house running systems, for an example with the Point of Sale (POS systems). In the analysis practice, the information security and confidentiality governance is a challenge.

As cultural challenge LaValle, et al. (2014), describes the adoption barriers in managerial level with technology being a cultural challenge rather than a technological challenge. According to Nrusimhamand Mohd(1013, p 614) further explain scale of data which growing in massive scale and store that data and compute power to analysed that data in timely manner need very powerful compute power. This required a massive infrastructure investment other than your existing data warehouse system in the retail and consumer goods industry.

As another challenge Nrusimhamand Mohd (2013, p 614) describes big data is to automatically generate the right metadata (data about data) to provide details of how data being recorded and describe what data has recorded and measured, converting the raw data and storing that in an in house databases becomes a key challenge.

Katina and Keith (2013), describes big data produce people interests on what they like, how they consume it, people thinking patterns and many other behavioural patterns of the people. Analysis emerged "with use of big data the potential growth for the retail and consumer business will increase by 0.5% in year 2020 and lead the profit margin by 60 present those who use big data for the maximum." By providing the value added services like loyalty and membership benefits. Big data will support provide information on what types on program people are looking and the feedback on the existing services.

When considering on ways in which big data can adopt in business intelligence, running the data analytics in cloud based environment is much more effective and efficient for retail and consumer goods companies as it scale for the data growing in a massive scale.

Based on the discussion on implementation model in retail and consumer goods companies needs to have following features, Streamline on Meta data population and stranded across the databases, Cloud base bid data analysis mechanisms for retail and consumer goods companies, Hybrid cloud database big data solution, based on those findings below framework being formulated to develop set of guidelines.
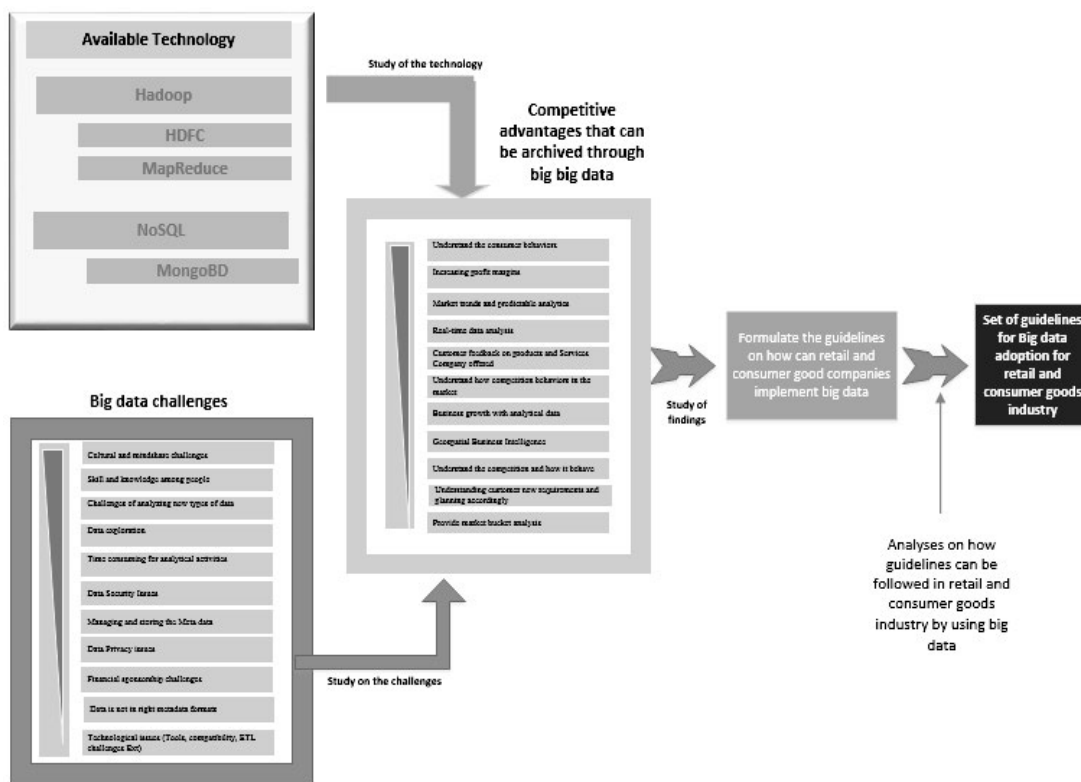

Figure 1: Framework on how to implement Big Data in order to achieve competitive advantage (Source: Author, 2018)

CONCLUSION

As findings shows the highest impact coming from people and its culture based on the analysis, retail and consumer goods companies need to develop people skill development workshops on technology impact for business success targeting business users and all the senior management of the organization, including CIO to IT administrators need to keep exploring technological update, providing necessary trainings for the different IT personals need to be continually formulated. As unstructured data and machine generated data going to continually grow, to avoid these challenges retail and consumer goods companies need to embrace cloud based big data solutions, which will provide the technological ability to store and integrate structured data with unstructured data in a scalable manner. Hybrid cloud solutions will provide the ability to retain the confidential data within the company premises and public data in the public cloud and integrate both data for analytical purposes. With hybrid cloud Hadoop based big data solutions, companies can easy adopt the could base big data solution while keeping initial investments low and addressing security, privacy and scalability challenges.

REFERENCES

Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., &Widom, J. (2011). Challenges and Opportunities with Big Data 2011-1.

Agrawal, D., Das, S., & El Abbadi, A. (2011, March). Big data and cloud computing: current state and future opportunities. In Proceedings of the 14th International Conference on Extending Database Technology (pp. 530-533). ACM.

Davenport, T. H., Barth, P., & Bean, R. (2012). How 'big data' is different. MIT Sloan Management Review, 54(1), 22-24.

PR, N 2014, 'Global Big Data Market in the Retail Sector 2014-2018', PR Newswire US, 4 April, Regional Business News, EBSCOhost, viewed 8 January 2014.

EdHat INTERNATIONAL
CONFERENCE
23RD NOVEMBER 2018

FIND NEW UNSUPERVISED ALGORITHM FOR FREELANCER EMPLOYEES RANKING

W.G.S. Lakshitha[1], R.L. Dangalla[2]
[1,2]Department of Computing and Information Systems
Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka.

INTRODUCTION AND OBJECTIVES

A freelancer or freelance worker is a term commonly used for a person who is self-employed and is not necessarily committed to a particular employer long-term. Such as these marketplaces are growing fast. The number of U.S. freelancers hit 57.3 million in 2017 and the thousands of businesses around the world looking for new and creative ways to empower their extended workforce.

Today, freelancer employees are ranked using the number of the projects that he has completed. Such as some sites are using user reviews to ranking their employers.

It is not the best way to solutions for freelancers ranking. There are some problems.

- Top-ranked completes the project more than the allocated budget.
- Rules of the top-ranked.
- Consumed time might be longer than the others.
- Freelancers are asking for payments before any work has been completed.

This novel concept introduces a new unsupervised learning algorithm for freelancer ranking problems.

Objectives:
- Use different parameters to ranking.
- Find the most suitable employees for a project.governance is a challenge.

RESEARCH METHODS

This study is suggested new parameters to develop a new freelancer ranking algorithm. There are match skills point (MSP), completed project percentage (P), on budget and time value. This study selected 5 parameters to freelancers ranking.

Number of completed projects (n)
Calculate the $\sqrt{n}$ by using n.
The n represents the total completed projects value.

Match Skills Point (MSP)
The match skills point represent some employer is how much suitable for the selected project.

$$MSP = \frac{\text{Number of employer has skills}}{\text{Number of skills need to the project}}$$

Completed Projects Percentage (P)

$$P = \frac{\text{Number of completed projects}}{\text{Number of Bid won project}}$$

P value is represented employer completed projects percentage.

On Budget

This study considered the freelancer employer is finishing some projects within the on budget or not. Some project finish under minimum budget than allocated budget. Its benefits go to the customers.

$$B_{min} = \frac{\text{Completed projects under minimum budget}}{\text{Number of total completed projects}} \times 100\%$$

$$B_{max} = \frac{\text{Completed projects under maximum budget}}{\text{Number of total completed projects}} \times 100\%$$

$$B_{dif} = B_{min} - B_{max}$$

Time Value

The time value is represented who have completed the projects in the allocated time. Some project allocate time is one day and some projects are allocated a few months.

$$T_{min} = \frac{\text{Completed number of projects before allocated time}}{\text{Number of total completed projects}} \times 100\%$$

$$T_{max} = \frac{\text{Completed number of projects after allocated time}}{\text{Number of total completed projects}} \times 100\%$$

$$T_{dif} = T_{min} - T_{max}$$

Finally, calculate F_{RP} value by using this novel algorithm.

$$F_{RP} = MSP \left(\sqrt{n} . P + \frac{1}{4} B_{dif} . \log_{10} T_{dif} \right)$$

F_{RP} = Freelancer employer Ranking Point

MSP = Match Skills Point

n = Number of completed projects

P = Completed Projects Percentage

$B_{dif} = B_{min} - B_{max}$

$T_{dif} = T_{min} - T_{max}$

RESULTS AND DISCUSSION

Enterprises, workers, and customers all benefit from a transparent system that values and promotes integrity. Today freelancer represents 35% of the US workforce and 16.1% represent the European Union.

Parameter study was how to find a new unsupervised algorithm for freelancer employee ranking by using different parameters. Also, it is selected some important parameters to rank, like as a number of completed projects, Project Match Skills, Completed Project Percentage and well on budget. Match Skills Point (MSP) is zero it represents the consider freelancer employees don't have suitable skills for the project. Some freelancer employer get Max Value for MSP (max value = 1.0). It describes employer is 100% suitable for the project.

The project needs to estimate the time duration and effort for each task of the project. This research is considered the selected employer is the number of projects completed in allocated time. This study isn't considered the time duration in the projects because time is depending on the projects. Tmax minimum value is zero because some consider employer complete all projects are before allocated time. Tmin is represented the before completed number of projects allocated time.

Employees after bids a project this suggest novel algorithm run the site backend automatically and will select the best 50 bids in among the bids. All freelancer sites are considering the employee's skills to select the best employer for the project. The employer must have to fulfill all skills to grant the project. It helps to complete the project with success. This study proposes an unsupervised algorithm for employees ranking in freelancer sites. The proposed algorithm automatically runs in the freelancer site backend. Algorithm ranking is a technique that helps to select the best one among all dataset. In this study used training dataset and testing dataset for train novel freelancer ranking algorithm. Both training and testing study obtained more than 95% accuracy of the novel algorithm.

CONCLUSIONS

Freelancer customer loss their time, cost and quality of the products. So it must have a better ranking unsupervised learning algorithm to rank the employees to give better service to the customers and the good chance of the freelancer employees. Nowadays freelancer sites only consider a number of completed projects to ranking. This study helps to select the best 50 employees to do their projects. The future works are to use more parameters to rank and occurrence of the increase the accuracy of the freelancers ranking algorithm.

REFERENCES

A zkaUmair, (2017). Individual Work Behavior in Online Labor Markets: Temporality and Job Satisfaction[researchgate] https://www.researchgate.net/publication/319285202

Marios Kokkodis, Panagiotis Papadimitriou, Panagiotis G. Ipeirotis, (2015). Hiring Behavior Models for Online Labor Markets [ACM Digital Library] https://dl.acm.org/citation.cfm?doid=2684822.2685299

S.V. Semenikhin, L.A. Denisova, (2011). Learning To Rank Based on Modified Genetic Algorithm[ieeexplore]https://ieeexplore.ieee.org/document/7819080/

Shihchieh Chou, Chinyi Cheng, Szujui Huang, (2009). A ranking algorithm for query expansion based on the term's appearing probability in the single document [emeraldinsight]https://emeraldinsight.com/doi/full/10.1108/14684521111128014

EdHat INTERNATIONAL
CONFERENCE
23RD NOVEMBER 2018

INTRODUCTORY PROGRAMMING:FACTORS AFFECTING COURSE OUTCOME OF NOVICE STUDENTS IN SRI LANKAN UNIVERSITIES

Dahanayake W. C. S.*, Dhanapala W. W. G. D. S*
*Department of Computing & Information Systems, Faculty of Applied Sciences,
Sabaragamuwa University of Sri Lanka, Belihuloya.

Keywords: Introductory programming, novice, course outcome, comfort level, linear regression

INTRODUCTION AND OBJECTIVES

Computer programming is an integral part of the software engineering and its related education. The demand for computing professionals has led to increasing attention for computer programming education. The failure rates in introductory programming courses are a universal problem as supported by many research studies. The first year students are suffering from the pressure to start higher education and all problems are related to an adjustment to university education. In this study, identify the factors it affects the outcomes of the introductory programming courses in Sri Lanka. These include factors that are related to programming behavior and psychological and cognitive characteristics and takes steps towards understanding their relationship with programming aptitude and performance. (Sarpong, Arthur, & Owusu, 2013) There are some interesting findings which will have implications for teaching and learning. Identification of the relationship to these factors will help educators to make the best decisions on how to improve computer curriculum and programs and help students become better programmers (Philip, Renumol V. G., & Gopeekrishnan R, 2013).

RESEARCH METHODS

This study is performed as a quantitative research and the measuring instrument as a questionnaire which is distributed among the university students in Sri Lanka who are following computer programming as the major. In total, 270 students enrolled in a first-year programming module voluntarily participated. The sample was composed of 48.52% male students and 51.48% female students.The respondents were basically from Sabaragamuwa University (35.19%), Uva Wellassa University (21.85%), Rajarata University (20.74%) and the rest of 22.22% respondents were from other universities in Sri Lanka.
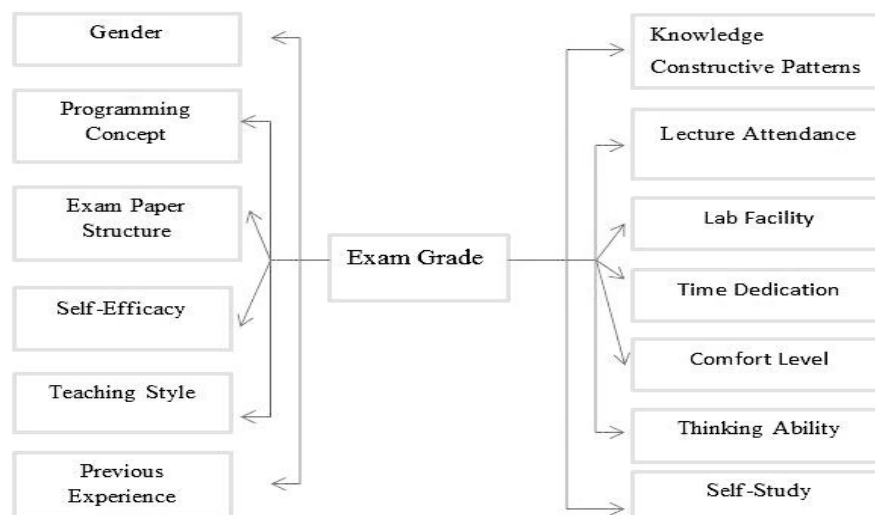

Figure 1: Conceptual framework

Considering the dependent and independent variables identified in the conceptual framework, thirteen hypotheses have been taken into the account. For example, H11 refers to course outcome of novice students have an involvement with gender. The statistical analysis was carried out via SPSS Version 21. Before the actual distribution of the questionnaire, a reliability test was carried out. As per the pilot test, the alpha value for each scale measured at .80 or higher which indicates to be in a good level to continue the study. Analyses were carried out in two stages to test the hypothesis (Byrne & Gerry Lyons, 2014). First, we generated descriptive statistics and tests of statistical significance for each index. Second, we used correlation and linear regression analysis to determine the relationships among the variables; with the particular goal of understanding the greatest predictors (Qian & James D. Lehman, 2016)

RESULTS AND DISCUSSION

Table 1: Theoretical results

		Frequency	Percent
Valid	A	19	7.0
	A-	16	5.9
	A+	13	4.8
	B	43	15.9
	B-	24	8.9
	B+	33	12.2
	C	34	12.6
	C-	18	6.7
	C+	54	20.0
	D	2	.7
	D+	12	4.4
	E	2	.7
	Total	270	100.0

Table 2: Practical Results

		Frequency	Percent
Valid	A	32	11.9
	A-	25	9.3
	A+	13	4.8
	B	47	17.4
	B-	38	14.1
	B+	37	13.7
	C	30	11.1
	C-	15	5.6
	C+	30	11.1
	D	2	.7
	E	1	.4
	Total	270	100.0

Source: SPSS

According to the results in table 1, only 29.9% of respondents have scored at least B+ grade in first theoretical exam. According to the table 2, 39.7% of the students have scored at least B+ in their first practical exam. Compared with the theoretical results lower grades percentage has reduced in students' practical results. So we can come to conclusion students' practical knowledge is greater than theoretical knowledge. We can conclude the students' exam results are not satisfied at this level but the failure rate is relatively less comparing with other countries as mentioned in literature. (0.7% in theoretical exam and 0.4% in practical exam)

Table 3: Difficulty level in programming concepts

Descriptive Statistics

	N	Mean	Std. Deviation
Abstraction	270	1.82	1.062
Interfaces	270	1.87	1.082
Encapsulation	270	1.92	1.065
Polymorphism	270	1.92	1.062
Static_Members	270	1.94	1.046
Structures	270	1.97	1.075
Inheritance	270	1.99	1.084
Pointers	270	2.03	.977
Array	270	2.04	.957
Method	270	2.04	1.048
Classes_Objects	270	2.05	.931
Loop	270	2.06	1.065
Conditional_Statement	270	3.07	.973
Operators	270	3.07	.977
Variable	270	3.09	.920
Data_Type	270	3.14	.859

Source: SPSS

According to the table 3, this situation is very crucial to understand as one bigger problem of students does not seem to be understanding of the basic concepts but rather learning to apply them. The topic, 'abstraction' in programming was identified as the most difficult topic to learn by novice students. Immediate feedback during conceptual matters and detailed explanation of less understood aspects could probably help many students. Programming includes several dynamic concepts that many times are taught through static means. According to the Pearson correlation coefficient, previous programming experience variable indicates to have the highest coefficient($r=0.337$ $p<0.001$) which spotlights that there is a positive relationship. The lowest correlation coefficient value ($r=-0.178$ $p<0.01$) is given by the gender variable. According to the student's point of view, paper structure variable is not significant to the course outcome. In regression analysis, the predictor set one with self-efficacy, previous experiences, difficulty level in concepts, comfort level, lecture attendance and self-study were the best model to predict introductory course outcome in this study. According to the coefficient table, the regression coefficients beta values of all the independent variables are statistically significant. As the results of the regression analysis, the highest magnitude represents the previous experience variable having a coefficient beta value of 0.215 suggesting that there is a positive relationship between previous experience and overall GPA variable. According to the model summary the model explains 61% ($R2=0.611$) of course outcome and performance in introductory programming. In predictor set two, regression coefficients beta values of gender, teaching style, knowledge constructive patterns, lab facility and time dedication variables are statistically significant.

Table 4: Predictor set one

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.658[a]	.611	.295	.60178

a. Predictors: (Constant), Self_study, Lec_attendance, Comfort_level, Self_Efficacy, PConcept_difficulty, Previous_experience

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.215	.335		3.631	.000
	Self_Efficacy	.162	.033	.197	1.849	.046
	Previous_experience	.331	.087	.215	3.816	.000
	Comfort_level	.290	.082	.185	3.532	.000
	PConcept_difficulty	-.206	.067	-.170	-3.073	.002
	Lec_attendance	.222	.053	.210	3.172	.000
	Self_study	.363	.105	.189	3.439	.001

a. Dependent Variable: Overall_GPA

Source: SPSS

Table 5: Predictor set two

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.238	.232		5.331	.000
	Gender	.204	.083	.143	2.471	.014
	KnowledgeCon_pattern	.203	.093	.135	2.194	.029
	Teaching_style	.198	.088	.136	2.249	.025
	Time_dedication	.257	.091	.164	2.834	.005
	Lab_facility	.120	.044	.160	2.731	.007

a. Dependent Variable: Overall_GPA

Source: SPSS

CONCLUSION

A predictive combination of factors was found to be a comfort level on the module, lecture attendance, previous experiences, self-efficacy, self-study and the competency level in programming concepts accounting for 61% of the variance in course outcome of introductory programming. Our study provided further evidence on the importance of gender, teaching style, knowledge constructive patterns, lab facility and time dedication as predictors of introductory programming performance. The study recommends including computer programming as a subject in secondary education since the strongest relationship exists between previous experiences and programming performances. It is verified that problems relating to high concept areas of programming must be addressed. To enhance teaching and learning of computer programming courses, we recommend to teachers to adapt more than one teaching style and make students feel comfortable in the classroom. Students should participate in lectures and practical sessions in order to have higher grades. The study also found that the factor, exam paper structure does not significant with course outcome of introductory programming according to students' perception.

REFERENCES:

Byrne, P., & Gerry Lyons. (2014). The Effect of Student Attributes on Success in Programming. Research Gate, 1-6.

Philip, M., Renumol V. G., & Gopeekrishnan R. (2013). A Pragmatic Approach to Develop Computational Thinking Skills in Novices in Computing Education. 2013 IEEE International Conference in MOOC, Innovation and Technology in Education (MITE), 2(0), 1-7.

Qian, Y., & James D. Lehman. (2016). Correlates of Success in Introductory Programming: A Study with Middle School Students. Journal of Education and Learning, 5(2), 1-11.

Sarpong, K. A.-M., Arthur, J. K., & Owusu, P. Y. (2013). Causes of Failure of Students in Computer Programming Courses: The Teacher – Learner Perspective. International Journal of Computer Applications, 77(12), 1-6.

AN EXPERT SYSTEM FOR TOOTH REMOVAL - A CASE ON TOOTH EXTRACTION FORCEPS

K. Banujan[1]# and S. Vasanthapriyan[1]
1 Department of Computing & Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

INTRODUCTION AND OBJECTIVES

Even though tooth extraction is one of the common surgical procedure in the dental field, it needs an extensive knowledge and practical experiences when handling the dental extraction equipment. Otherwise, it will be more complex or even in a worse case, it may cause damages to patients' mouth area (Malden, 2001). So, it is very important to have a sound knowledge of the instruments to be used, especially on extraction forceps. Sometimes different terminologies are used to express the same concept. Due to the unstructured, incomplete, general nature and varied formats of the information, the knowledge is not reaching everybody (Walisadeera, Ginige, & Wikramanayake, 2015).

Ontologies are widely used for various purposes such as knowledge management, e-commerce, natural language processing, intelligent integration of information and semantic web (Vasanthapriyan, Tian, Zhao, Xiong, & Xiang, 2017). The aim of this work is to contribute to an improvement in the management and usage of dental extraction forceps in hospitals by developing an ontology-driven solution that organizes and describes clearly related knowledge.

RESEARCH METHOD

As a first step, we gathered the contextual information from various sources of related field experts from Sri Lankan universities, Sri Lankan hospitals by using formal and informal interviews and discussions, research articles, online articles, books, newspapers, radio, and television. By analyzing the information gathered, we identified the information needed by the dentists and their assistants in the form of questions for our study.

After identifying the needs of information, we used grounded theory for data collection. Two dentists with extensive knowledge of dental (mainly on extraction) and an expert on ontology engineering took part. We created Competency Questions (CQs) which will work as requirement's specification of the dental extraction forceps ontology. If CQs contains all the necessary and sufficient axioms that correctly answer the CQs, it is possible to know whether an ontology was created correctly (Vasanthapriyan, Tian, & Xiang, 2017).

Then, we used Grüninger and Fox's methodology (Grüninger & Fox, 1995) for our work as it publishes a formal approach for designing the ontology and also it provides a framework for evaluating the developed ontology. The ontology was implemented by using the Protégé-OWL Ontology Editor 5.1. Finally, developed ontology was evaluated and validated using by experts in the field ontology, by using inbuilt FaCT++ 1.6.5 reasoner and online ontology evaluator called OOPS! (http://oops.linkeddata.es/) (Poveda-Villalón, Suárez-Figueroa, & Gómez-Pérez, 2012).

RESULTS AND DISCUSSION

Tooth extraction is one of the common surgical procedure in the field of dental, which mainly depends on the knowledge and experience of the dentists. Therefore in this research, great importance is given to knowledge for dental extraction forceps, and the potential benefits of managing dental extraction forceps knowledge, an ontological approach to representing the necessary dental extraction forceps knowledge within the dentists' context was developed.

Designing this type of ontology is not a simple task, because we need to gain vast domain knowledge. In this paper, we have explained how we designed and developed the ontology to organize domain knowledge by meeting particular access requirements effectively. Using this approach, dental extraction forceps ontology to include information needs to be identified for dental extraction activities to be designed.

First, we gathered the needed data from the dental experts. Competency questions were used to determine the scope of the ontology and used to identify the contents of the ontology because contextualized information fulfils the expressiveness and reasoning requirements of the software testing ontology. Grüninger and Fox's methodology focuses on building ontology based on first-order logic (FOL) by providing strong semantics. The associative relationships are to identify the concepts and relationships with meaningful relations and to define the relationships and their inverse relationships. Finally, modelled them into ontology using Protégé-OWL Ontology Editor 5.1. Description Logic (DL) query was used to query the competency questions. The DL query answered all the CQs listed in the beginning.

The validation and evaluation have been done separately. We validated the ontology in terms of accuracy and quality by using the FaCT++ reasoner which is and inbuilt tool in Protégé-OWL Ontology Editor 5.1 and by using web-based tool OOPS!. We evaluated the ontology with the help of ontology expert by examining the deficiencies of the artifacts we used. Based on Ontology experts' responses, comments, and suggestions the ontology was redeveloped.

The developed ontology will be used in the development of knowledge management (KM) portal which will be used in knowledge dissemination process. In the KM portal, the authorized persons (i.e. doctors) can share their experiences.

CONCLUSIONS

Identification of the suitable extraction forceps for the given case is resolved by developing a domain ontology on dental extraction forceps. Designing this type of ontology is not a simple task because we need to gain vast domain knowledge. This research presents dental extraction forceps ontology to represent dental extraction forceps domain knowledge which includes dental extraction forceps concepts, properties, and their relationships. We believe our dental extraction forceps ontology can support other hospitals to improve the sharing of knowledge and learning practices.

REFERENCES

Grüninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies.

Malden, N. (2001). Surgical forceps techniques. Dental update, 28(1), 41-44. Poveda-Villalón, M., Suárez-Figueroa, M., & Gómez-Pérez, A. (2012). Validating ontologies with oops! Knowledge Engineering and Knowledge Management, 267-281.

Vasanthapriyan, S., Tian, J., & Xiang, J. (2017). An Ontology-Based Knowledge Framework for Software Testing. Paper presented at the International Symposium on Knowledge and Systems Sciences.

Vasanthapriyan, S., Tian, J., Zhao, D., Xiong, S., & Xiang, J. (2017). An ontology-based knowledge management system for software testing. Paper presented at the The Twenty-Ninth International Conference on Software Engineering and Knowledge Engineering (SEKE).

Walisadeera, A. I., Ginige, A., & Wikramanayake, G. N. (2015). User centered ontology for Sri Lankan farmers. Ecological Informatics, 26, 140-150.

WHY KEY EMPLOYEES QUIT THEIR JOBS

Aruna Sanjeewa[1], Jayalath Ekanayake[2]
Sabaragamuwa University of Sri Lanka, Belihuloya

[2] Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya

INTRODUCTION AND OBJECTIVES

The objective of this study is to explore the factors, which affect the employees to leave their organizations.

Employees are the key for growth and productivity of any organization. Satisfied employees tend to work harder and smarter, which in turn can boost a company's bottom line. Retaining such key employees is a challenge for an organization. Employees leave or change their firms due to various reasons. Identifying such reasons is an important task. Several studies [1][2][3] have explored this topic using different methodologies. However we use modern machine-learning algorithms to analyze and identify the complex relationship among the attributes of the employee behavior historical dataset.

The outcome of this study shows that the satisfaction level of an employee is a key factor for quitting the job and hence, the top management should always monitor the satisfaction level and should conduct programs to enhance the satisfaction of employees.

RESEARCH METHOD

The dataset was collected from the repository https://www.kaggle.com/giripujar/hranalytics. It contains 14999 instances, which 3571 records belong to the employees already left from the company and 11428 records belong to the employees currently working in the company. Table 2.1 shows the description of the feature set selected for the experiment. The dataset was cleaned and then correlation analysis was conducted to identify the correlated features.

Table 2.1 - Feature description of the dataset

Attribute	Description
Satisfaction level	Employee satisfaction level. Ranges between 0 and 1.
Last evaluation	Grade got in last evaluation. Ranges between 0 and 1.
Number of projects	Number of projects currently working on.
Average monthly hours	The average number of hours worked for a month.
Time spent at company	Number of years the employee has been working for the company.
Work accident	Whether the employee has met with an accident during the working hours (Yes-1, No-0).
Promotion at last 5 years	Whether the employee has got promoted (Yes-1, No-0).
Salary level	Salary column convert to the numeric values (Low-1, Medium-2, High-3).
Left (Class variable)	Whether the employee has left or not (Working-0, Left-1).

J48, Support Vector Machine, and K-Nearest-Neighbors algorithms were trained using the above dataset to predict the class variable Left.

RESULTS AND DISCUSSION

Correlation analyze
According to the Figure 3.1, satisfaction level of the employee is highly correlated with class variable Left and further the correlation is negative. Moreover, number of projects carried out by an employee is positively correlated with the average number of working hours per month, which is an obvious fact as the employees are assigned more projects they have to work more hours.
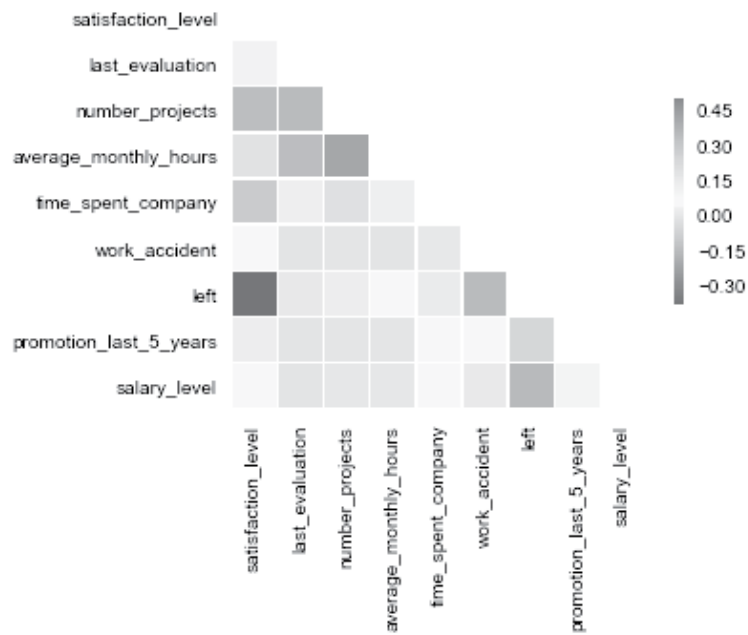
Figure 3.1 - Correlation coefficients represented in a Heat-map

The promotion of the employees positively correlated with the number of projects that they undertake, which is an interesting fact. This fact encourages employees to work on more projects.

A decision tree was trained on top of the above dataset using the J48 algorithm [4], which is implemented in WEKA data mining tool to predict the class variable Left.
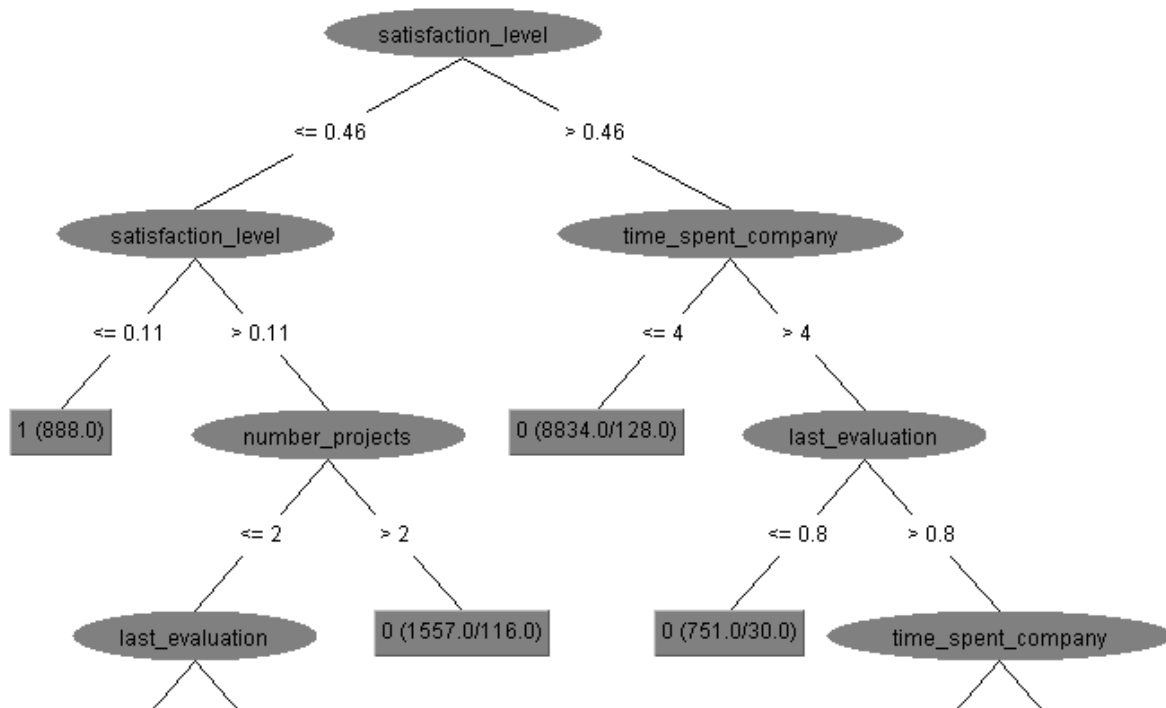

Figure 3.2 – Decision tree generated using WEKA J48 learning algorithm

Figure 3.2 depicts the first four levels of the decision tree. The topmost decision node in a tree, which corresponds to the best predictor called root node. In the above tree the root node is satisfaction level of the employees and hence, the satisfaction level is the best predictor of the class variable Left. This fact is further confirmed by the correlation analysis, which revealed that there is a strong correlation between the class variable Left and the satisfaction level of the employees.

The working period of an employee is in the second level of the decision tree. The interesting factor is that the employees having less experiences (less than 4 years) are willing to continue on their current job.

The grade that the employee got at their last evaluation is also an influential factor for the decision whether to continue or quit from the job. According to the above decision tree if the employees are promoted into lower grades then there is a possibility to continue their job.

Support Vector Machine and K-nearest neighbor algorithms were also trained using the above dataset to predict the class variable Left. The prediction accuracy of these two algorithms is comparative lower than the decision tree. Also, these algorithms can only predict the class of a given instance, but cannot explain the reasons for the predictions as they are not human readable.

CONCLUSION AND RECOMMENDATION

The objective of this project is to explore the reasons why the key employees quit from an organization. The dataset was explored using three modern machinelearning algorithms.

The satisfactory level of an employee is highly influencing for the decision to quit from the organization and hence, the top management should take appropriate actions to create a friendly working environment within the organization. Moreover, the service period of an organization is also an impact on the decision to quit. It is also revealed that the less experience workers are willing to continue their jobs, however, when they become experience they tend to change their organization.

REFERENCES

1. Holmes, P., Chapman, T., & Baghurst, T. (2013). Employee Job Embeddedness: Why People Stay. International Journal of Business Management & Economic Research, 4(5).

2. Mitchell, T. R., Holtom, B. C., Lee, T. W., Sablynski, C. J., & Erez, M. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. Academy of management journal, 44(6), 1102-1121.

3. William Lee, T., Burch, T. C., & Mitchell, T. R. (2014). The story of why we stay: A review of job embeddedness. Annu. Rev. Organ. Psychol. Organ. Behav., 1(1), 199-216.

4. Quinlan, John R. (1992). Learning with continuous classes. In 5th Australian joint conference on artificial intelligence, vol. 92, pp. 343-348.

SINGLE AND MIXED LINE RATE VARIATION FOR DEDICATED PROTECTION IN ELASTIC OPTICAL NETWORKS

Satkunarajah Suthaharan, Rukshani Puvanendran, and Vanuja Chandrakumar
Vavuniya Campus of the University of Jaffna

Keywords: elastic optical networks; dedicated protection; single-line rate; spectrum efficiency

INTRODUCTION AND OBJECTIVES

WDM is a multiplexing technique of data transmission in which it divides the huge transmission bandwidth available on a fiber into several non-overlapping wavelength channels and enables data transmission over these channels simultaneously [1, 2]. WDM systems enable to increase the capacity of the network without laying additional fiber. In our research study, we have focused on survivability approach, specifically, dedicated protection method in WDM optical networks that received much attention in the research community nowadays. Elastic optical (or flexgrid) networks (EONs) have recently been introduced to use the frequency spectrum more efficiently [3]. 12.5 GHz fine granular frequency slots or flexible grids are used for setting up lightpaths instead of using 50 GHz or 100 GHz fixed grid spacing in EON [4]. Our objective in this paper is to investigate variation of the spectrum efficiency of traditional shared protection and elastic optical networks using MLR in various scenarios.

RESEARCH METHODS

The configuration for both WDM and elastic optical networks scenario, a recently proposed switch architecture [5] which achieve faster performance has been used in our investigation. The switch uses components such as variable optical splitters (VOSs) and combiners, bandwidth variable transponders (BVTs), flexible wavelength selective switches (Flex WSS). Further, less power consumption caused by the components can be achieved in the switch architecture. Therefore, to illustrate shared protection in elastic optical networks for MLR, we use switch architecture in the NSF network topology. A single wavelength carries a higher data rate than a single frequency slot carries. Therefore, when considering MLR in both wavelengths and frequency slots vary from SLR which has the unique data rate throughout the entire connections establishment. Further, to implement limiting the number of working links, we propose an algorithm. This algorithm consists of setting up primary and backup paths and their wavelengths or frequency slots. Source, destination, and the data rate are the inputs. This is used to accommodate lightpaths and to calculate the spectrum efficiency for MLR.

RESULTS AND DISCUSSION

We simulate the traditional dedicated protection approach in elastic optical networks. To perform the experiment we use NSFNET (14 nodes and 21 bidirectional links) topology for our study. We consider 352 frequency slots each of which consists of 12.5 GHz spacing. Various data rates such as 100 Gbps, 400 Gbps, 1 Tbps are considered in both mixed-line rate and single-line rate with their appropriate bandwidths and are followed the uniform distribution. Request arrival process follows Poisson distribution and holding time of requests follow exponential distribution with unit mean. Traffic requests arrive in dynamic network environment. Source node and destination node of each request follow uniform distribution. We assume the guard band in between two frequency slots is 12.5 GHz. Each experiment is simulated with various request arrivals in order to compute the total amount of bit rate and bandwidth used. In this study, we consider dedicated protection in EON particularly, for MLR. This is because to measure the spectrum efficiency caused by both MLR and SLR in various scenarios. We select the traffic bit rate ranges from 20 Tbps to 100Tbps for all comparisons that are used to calculate spectrum efficiency. Such that they provide approximately the same spectrum efficiency. This helps us to find and compare the relative impact of the performance on different scenarios. Our performance study is

A FRAMEWORK FOR MYSQL PRE-PROCESSOR WITH INTERACTIVE USER INTERFACES

Krishnamoorthy Santhanakrishnan and Vijayanathan Senthooran
Department of Physical Science, Vavuniya Campus of the University of Jaffna

ABSTRACT

The error messages displayed after compilation of MySQL query shows the MySQL error code, SQLSTATE value and text string in command prompt. The displayed error messages will lead an inconvenience to users who are willing to assess the full details of the errors. The proposed framework presents the design of pre-processor with attractive interactive user interfaces to reduce the ambiguity of the error messages generated after the compilation of MySQL query. The output of the proposed pre-processor gives the packed error details when an error occurs during the compilation of query and it shows clear error messages to the users instead of giving existing warning message in the command prompt. This will facilitate the users who are not an expert in MySQL and they easily tackle the type error message and where they occurred. The proposed pre-processor also provides a lot of interactive interfaces for handling databases and tables stored within the MySQL server as supporting elements for creation of queries generates Interactive text editor with intellisence facilities for users to minimize their careless or typing errors. Finally, the prototype of this pre-processor supports different types of queries with an expected error details and reduces the ambiguity of the warning messages displayed after the compilation of MySQL query.

INTRODUCTION AND OBJECTIVES

The MySQL Server is a very fast and a reliable Relational Database Management System and it is mostly used in the development of web based and windows based applications as it is open source and free software. Because of the SQL engine only gives warnings rather than giving clear error messages after compilation of queries, an undesirable feature of MySQL for entry developers is the ambiguity of error messages due to the complex syntax displayed in the command prompt. There are some or few of compilers for MySQL with user interfaces. They do not have their own error messaging system for compilation. Instead, when compilation is done they only extract error messages from MySQL compiler and just display them to the user without giving any further explanations. To simply understand the error messages produced by MySQL, it is proposed a pre-processor with attractive user interfaces that outputs the error messages which easily realize the syntax of the error message. The proposed pre-processor stores the all the information relating to the compilation of a query. As MySQL server runs on batch mode, the proposed pre-processor uses batch files to interconnect user interfaces with MySQL server. The objective of the proposed study is to develop a pre-processor with attractive user interfaces that displays its own error messages by analyzing the compiling query.

PROPOSED METHODOLOGY

In the proposed system, all the data are stored and defined within the MySQL Server. Therefore, the system should actually have the interconnection facility between these interfaces and MySQL Server. Moreover, the queries are compiled with errors through the analysis of the data which are internally stored within the MySQL Server as the C#.NET 2008 has some capabilities to deal with the interconnection with MySQL Server. It uses a COM object that handles all the functionalities relating to the connection between C#.NET and MySQL Server. The proposed system uses batch files to interconnect C#.NET and MySQL Server. The class StreamWriter is used for writing any files at system run time and helps us to write batch files. Moreover, C#.NET is capable of executing batch files by using the

classes ProcessInfo and Process which are used to write and execute batch files at runtime. The system must first be connected with MySQL Server at the starting stage of the system. When connecting the MySQL Server and the system with proper user authentication by giving the host which may be running on local host or remote server, user name and password, one batch file is generated with these request and then it is executed. Here, it is no need to produce the output file for this connection. The only thing is to make sure whether the server is successfully connected or not. But, in contrast, the situation is different after the connection is established successfully with MySQL Server. In most of the times, When the need of handling databases, tables and other information stored within the server, the information that are requested from the system must be provided with the output file. The following flowchart simply depicts the methodology in the designing of MySQL pre-processor.
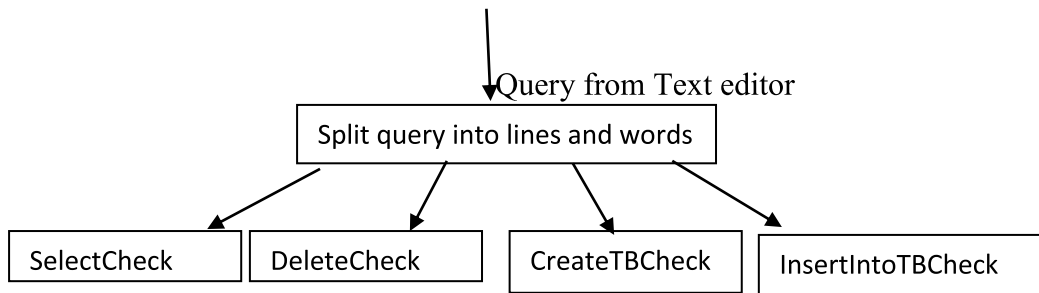
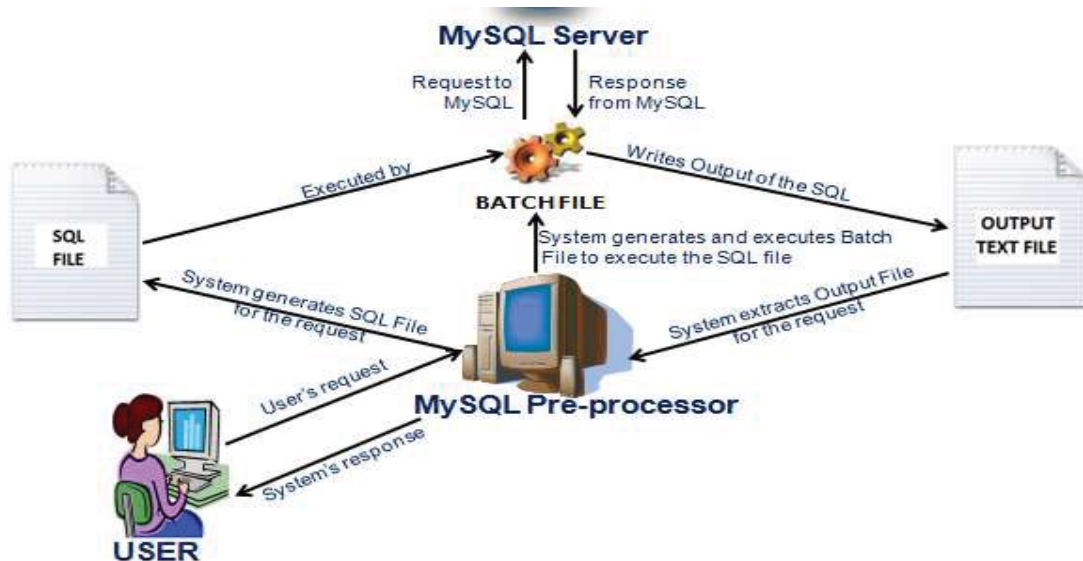Pre-processing of query

Figure 1

System Functionality

Figure 2

RESULTS AND DISCUSSION

Compilation of query, SelectCheck, BracketCheck and RealTimeSelectCheck methods are implemented using C# according to the methodology with attractive user interfaces. The results of the existing MySQL warning message is compared with the error message generated by proposed pre-processor and their features are discussed based on the ambiguity of users. The proposed pre-processor provides sequence of user interfaces for connecting server, viewing databases, help and syntax and IntelliSense facilities. In the connecting server window, it is specified whether the MySQL Server is running on local host or on a remote server and the IP Address for the remote server, user name and password for the server. Further, it also identifies the directory location of MySQL bin and path for storing batch files, query files and output text files which are needed for the system's operations. In the viewing database window, databases, tables and fields are displayed as tree view in the system and the proposed system also provides the data grid view and design view of a given table. These interfaces definitely give proper guidance for creation and compilation with minimum effort for users. While typing query, the help and syntax window loads and displays the help and syntax for that query to the user. These also helps user to minimize the errors caused by lack of knowledge about syntax of that query. The pre-processor provides the IntelliSense facilities. The IntelliSense facilities also minimize the typing and careless error. As soon as the keyword of the query is typed, all possible values next to that keyword are displayed in the context menu and they can be added into the query by clicking on the context menu. In the SelectCheck algorithm, only three clauses such as SELECT, FROM and WHERE are checked. Other clauses GROUP BY, ORDER BY and HAVING can easily be checked by slightly modifying the SelectCheck algorithm. For checking the attributes and their types, methods isAttribute ( ) and getTypeofAttrib ( ) can be used. When the comparison is made with the use of functions, the method TypeCheck( ) can also be used to make sure whether type of attribute and type of the attribute value for corresponding attribute of the function are matched. The sample queries with mismatching of select, from, where, delete, brackets and semicolons are compiled with the proposed pre-processor as well as MySQL. The pre-processor specifies the error that indicates which data type the attribute value should have and the clause where the error is occurred. It is compared with the warning message for the same query compiled in MySQL. Based on the error messages displayed in two different ways, the proposed pre-processor presents the simple and syntax clear error messages compared with existing MySQL warning messages for entry level MySQL users.

CONCLUSIONS

The proposed pre-processor with interactive user interfaces is quite and completely successful. The designing and implementation of the pre-processor has been focused on giving a clear and meaningful error messages instead of giving unclear and ambiguous error messages and warnings which are typically given by any other MySQL Compiler when compilation of MySQL queries is done. Hence, the error message that is deliberately displayed during the compilation process is fairly simple and easy to understand for the entry level users who are working with MySQL manipulation. The errors with attributes and attribute values are clearly shown by this system rather than just giving a warning. The brackets that are placed within the queries are effectively checked. When the system shows error messages on bracket checking, error message includes which type of brackets that may be opened brackets or closed brackets are missed or incorrectly placed within the query. A lot of interactive interfaces with the information, which are needed for the creation and compilation of queries such as IntelliSense facilities, data grid view and design view of the table, and help with syntax of the query that is being written, are given for helping user to manipulate the query by minimizing typing and careless mistakes with a minimum effort.

REFERENCES

1. Amlanjyoti Saikia, S. J. (March 2015). Comparative Performance Analysis of MySQL and SQL Server Relational Database Management Systems in Windows Environment. International Journal of Advanced Research in Computer and Communication , Vol. 4, Issue 3, 160-164.

2. Dipina Damodaran, S. S. (Vol. 5, No. 2, April 2016). PERFORMANCE EVALUATION OF MYSQL AND MONGODB DATABASES. International Journal on Cybernetics & Informatics (IJCI) , 387- 394.

3. Letkowski, J. (2014). Doing database design with MySQL. Journal of Technology Research, Volume-6, pages-15.

4. Wille, W. (2013, 09 08). https://www.codeproject.com/Articles/627073/MySQL-Fullltext-Start-for-a-Simple-Search-engine-u. Retrieved from https://www.codeproject.com.

LIGHTWEIGHT, CROSS PLATFORM AND WEB BROWSER BASED SYSTEM APPLICATION DEVELOPMENT SERVER RUNTIME WITH SDK - NEUTRALINOJS

S.J.M.D.P.Samarakoon, M.A.S. Suranga, C.Hettiarachchi
Uva Wellassa University, Sri Lanka

INTRODUCTION AND OBJECTIVES

When building applications they have to be able to run on any platform since the targeted audience is not using a single operating system. Therefore achieving cross application development was important. Nowadays cross-platform application development is popular with Javascript. Most popular and leading frameworks are Electron and Nw.js (NodeWebkit). These solutions bundle Chromium browser embedded view and Nodejs runtime into a package. Therefore the bundled application is large although a simple application is developed and tends to take more RAM usage since it uses both Chromium and Nodejs.

From this research an alternative method to develop cross-platform applications without above said dependencies will be introduced.

The main objectives of this research are,

•   Creating an environment for developing and executing cross platform browser-based applications without any dependency along with native functionalities (specially for lightweight applications).

•   Smaller bundled application output for multiple platforms.

RESEARCH METHODS

NeutralinoJs is a framework to develop cross platform applications with native functionalities and it allows to run applications inside web browsers.

Application development with Electron and NWjs requires Nodejs and hundreds of Javascript libraries to be installed and special data files for the bundling will be downloaded at deployment and also the application output is large due to embedded chromium browser and Node runtime.

NeutralinoJs will give developers portable lightweight SDK for cross platform application development and it reduces the application output to a smaller size by replacing embedded chromium with user's default browser and Node Js runtime with lightweight multithreaded API server.

PLATFORM COMPONENTS

Neutralino Server
Platform dependent runtime that process kernel function requests (JSON) from user desired port. This program is implemented using C++ using multithreaded socket programming approach on top of open source projects as per below.

Neutralino SDK
This is Javascript library for developers who are going to build applications with NeutralinoJs. All kernel function js aliases are defined inside this library. Eventually the source code is compressed for faster loading by webpack.

Neutralino Docs
The API documentation along with the tutorials regarding NeutrolinoJS is added here. The documentation website is built using open source solution docsify.

RESULTS AND DISCUSSION

This platform releases the use of Chromium browser in ordinary platforms and replaces it with any browser to reduce the package size and also the use of Nodejs can be released with a lightweight application server. The light weighted application building is achieved by NeutralinoJS.

Eventually this application will prove any system core function can be accessed, executed and results can be captured to/via application SDK. Following core functions will be added for the initial version.

Module	Functions
Neutralino.os	Get environment variable
	Run system command and get output
Neutralino.settings	Get settings
Neutralino.computer	Get RAM usage details
Neutralino.filesystem	Creates directory
	Removes directory
	Writes a file
	Reads a file
	Removes a file

Table 1 Modules and functions

Solution architecture
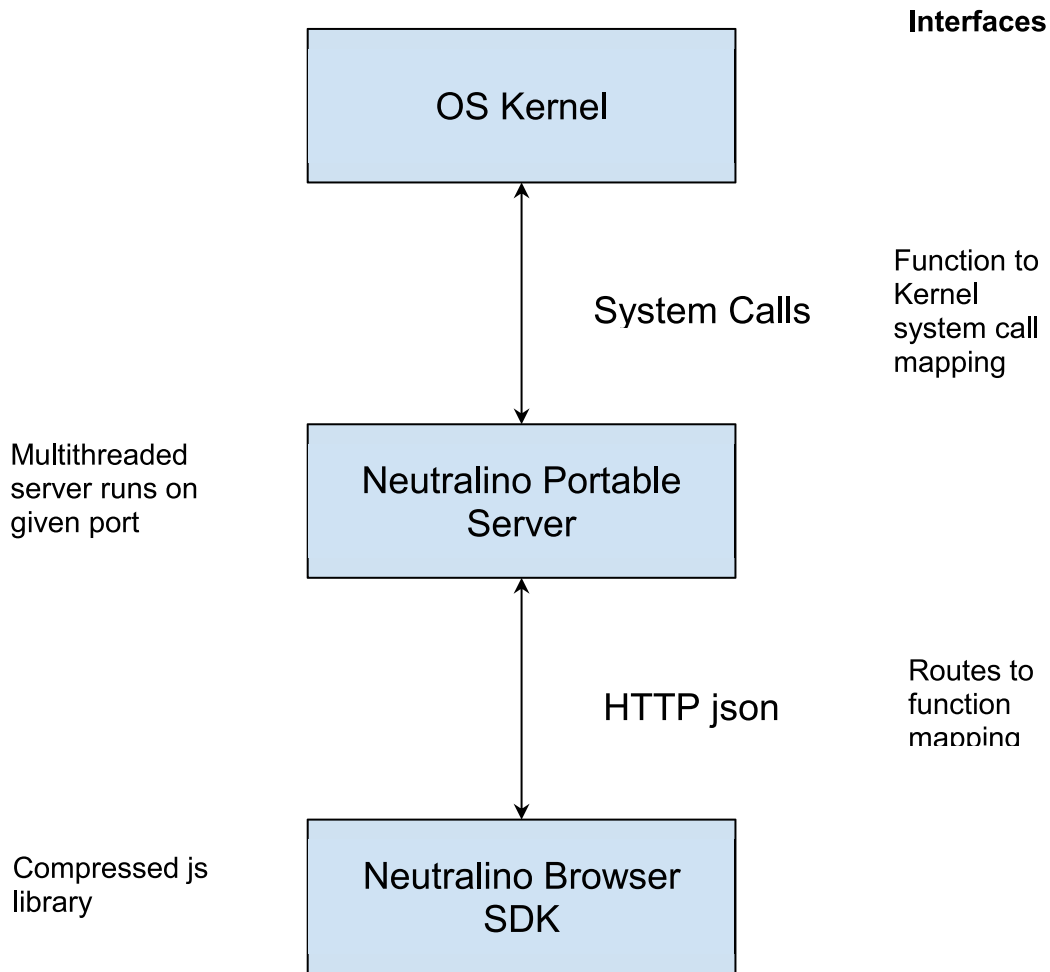
Figure 1 Solution Architecture

DEVELOPMENT WORKFLOW

Electron and Nw.js	NeutralinoJs
Install Nodejs	Download Portable SDK
Install library and automatic installation of dependencies	Develop the application
	Done
Develop the application	
Download build files for platform	
Build application	
Large bundled output	**Lightweight output**

Figure 2 Development Workflow

COMPARISON OF THE SOLUTIONS

NeutralnoJS grants the possibility of developing cross-platform applications with native functionalities (controlling system core and infrastructure) that can be executed without any dependency in user's browser on a socket created by a lightweight application server.

Through this research developing a framework to create lightweight cross-platform applications without any dependency with smaller bundled application size is achieved.

	Bundled Application Size (compressed size)		
	Electron	Nwjs	NeutralinoJS
For **Windows** Environment	66.2MB	61MB	~700kB
For **Linux** Environment	55.2MB	74MB	~150kB

Figure 3 Comparison of the solutions

NeutralinoJS provides the best solution in building applications which are web browser based and cross platform and by reducing their application size significantly.

This framework provides a stable support for both Linux and windows environment.

NeutralinoJS provides an automatic run-time shutdown after inactive session and a basic (token based) authorization method to protect inside Networks.

NeutralinoJS has also gone to give support for hosting NeutralinoJS applications on Cloud platform and addingconfiguration for cloud/ desktop based apps in settings.json.

Results also add a Angularjs template for Neutralinojs, Reactjs template for Neutralinojs and Vuejs template for Neutralinojs.

CONCLUSION

This research will be done by finding an alternative lightweight approach for developing cross platform applications with native functionalities other than electron and Nw.js which bundles heavy components into user code. As per now the concept is implemented as an open source project with clear documentation for developer community with the support from the developer community.

REFERENCES

1. Abeer Alkhars, Wasan Mahmoud (2016) Cross-Platform Desktop Development (JavaFX vs. Electron), Sweden

2. GitHub. (2018). Large file size of small electron app ? · Issue #10861 · electron/electron. [online] Available at: https://github.com/electron/electron/issues/10861 [Accessed 27 Feb. 2018].

3. GitHub. (2018). NW Executable File Size: make a lightweight standalone executable · Issue #3955 · nwjs/nw.js. [online] Available at: https://github.com/nwjs/nw.js/issues/3955 [Accessed 27 Feb. 2018].

4. reddit. (2018). Is there a more lightweight alternative to Electron? • r/webdev. [online] Available at: https://www. reddit.com/r/webdev/comments/6s1al7/is_there_a_more_lightweight_alternative_to/ [Accessed 27 Feb. 2018].

IMPROVING EFFORT ESTIMATION PRACTICES IN SMALL SCALE AGILE SOFTWARE PRODUCTS

J.N.N.Mahisha[1]*, H.M.K.T.Gunawardana[1] and R. M. K. T. Rathnayaka[2]
[1]Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Srilanka.

[2]Department of Physical Sciences and Technology, Sabaragamuwa University of Sri Lanka, Belihuloya, Srilanka.

INTRODUCTION AND OBJECTIVES

Estimation of effort in a software project is a project management activity which is an essential task in software industry. To build a successful project its effort needs to be estimated nearly to the accurate level using a standardized method. Software development effort estimation can predict the exact or most accurate amount of effort in terms of money or man hours that is needed to develop a particular software system. Among various software development practices, agile practices have become a popular software development approach in software industry. Agile software development describes the process where continuous evaluation and reimplementation occurs. The main objective of this study is to identify which factors affect most to increase the accuracy of effort estimation in agile software projects. Since agile is found to be a broad topic this research is mainly focused on Scrum based agile software projects. The objective of exploring the effect of parameters on Effort estimation is simplified along with four main parameters including Lines of Codes, Involvement of team members, Historical Data and Involvement of Expertise.

METHODOLOGY

This research was carried out on two main methods, including a systematic literature review and a survey. Through the systematic literature review 14 different parameters which would affect to the effort estimation process and related effort estimation models were identified. From the identified 14 parameters, 4 commonly and most applicable parameters were selected to build the questionnaire. Based on the findings of systematic literature review, a taxonomy was developed and in order to evaluate and check the accuracy of the taxonomy, a survey questionnaire was carried out among the 50 IT professionals in the industry. Hence, a quantitative research was conducted along a questionnaire distributed among the effort estimation practitioners which resulted many towards the importance of effort estimation in projects.

The survey questionnaire designed for this research title comprises of two main sections including demographic data, which would help in the descriptive analysis while the second section includes questions related to the effort estimation practices, parameters that are used in agile software development process. A five point Likert scale which ranges from Strongly Agree to Strongly Disagree was set up in the second section as a means of facilitating convenience in responding the questions. The questions were designed and prepared considering the respondent's perception as much as possible in order to ensure the convenience and understandability of the questionnaire.

RESULTS AND DISCUSSION

Results obtained through the systematic literature review were put on to the taxonomy by considering their effect on each selected effort estimation model. Statistical Package for the Social Sciences (SPSS) software was used to present statistics for comprehensive understanding of data collected from the questionnaire.

As the first step, a reliability analysis was done to ensure the validity of the questionnaire. Reliability is assessed with the Cronbach's alpha. A value of at least 0.70 was used as the threshold to indicate adequate reliability. Table1 shows the reliability results gained through SPSS for the selected parameters.

Table 1: Results of Reliability test

Variable	Cronbach's Alpha	Accept/Reject/Moderate
Lines of Codes	0.821	Accepted
Involvement of team members	0.646	Accepted
Historical Data	0.750	Accepted
Expertise Involvement	0.770	Accepted

The correlation among constructs was estimated using Pearson Correlation Coefficient. Correlation coefficient quantifies the strength and direction of the linear relationship between two constructs and generally, it ranges between -1 and +1.

Table 2: Correlation Analysis Results

Correlations

		Avg_factor_1	Avg_factor_2	Avg_factor_3	Avg_factor_4
Avg_factor_1	Pearson Correlation	1	.383**	.362**	.552**
	Sig. (2-tailed)		.006	.010	.000
	N	50	50	50	50
Avg_factor_2	Pearson Correlation	.383**	1	.341*	.287*
	Sig. (2-tailed)	.006		.015	.043
	N	50	50	50	50
Avg_factor_3	Pearson Correlation	.362**	.341*	1	.318*
	Sig. (2-tailed)	.010	.015		.024
	N	50	50	50	50
Avg_factor_4	Pearson Correlation	.552**	.287*	.318*	1
	Sig. (2-tailed)	.000	.043	.024	
	N	50	50	50	50

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

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

From the above Table 2, all the factors have gained positive values. This means each factor has a positive relationship with one another. From the table, it is clear that the strongest relationship is shown by the parameter Avg_factor_4 which refers to the factor Expertise Involvement. The second strongest factor is involvement of team members. The proposed taxonomy is shown in table 3.

Table 3 Finalized solution table

	High	Medium	Low
Influence of historical Data	COCOMO, Story point, FPA		Planning poker, T-shirt sizing, wideband Delphi
Involvement of team members in effort estimation	Planning poker, T-shirt sizing, Wideband Delphi		COCOMO, Story point analysis, FPA
Expertise Involvement	Story point analysis, Planning poker, FPA, T-shirt sizing, Wideband Delphi		COCOMO
Dependability directly on Programming language (LOC)	COCOMO	T-shirt sizing	Story point analysis, FPA, Planning poker, Wideband Delphi

When comparing the taxonomy with the statistical results, we can prove that most accurate estimations could be done with the involvement of Expertise in the effort estimation process. Effort estimation models like, Story point analysis, Planning poker, Function point analysis, T-shirt sizing and Wideband Delphi requires expertise in the effort estimation process. Out of the selected six effort estimation models, five models falls under high influence of Expertise in the taxonomy. The next highest influence was from the parameter, Involvement of team members. When compared with the taxonomy, it clearly shows, out of the six selected effort estimation models, three models including Planning poker, T-shirt sizing, Wideband Delphi falls under the highest influence category. Therefore, the accuracy of the taxonomy could be clearly evaluated when comparing it with the statistical analysis results.

CONCLUSION

This research thesis provided an overview of different software development effort estimation methods. The taxonomy presented provides a framework to understand, categorize, assess and effort estimation approaches to support their selection and tailoring for specific purposes. To make more accurate estimations, the crucial factor would be the involvement of expertise, then the involvement of team members, next would be the use of historical data. Also, models like Story point analysis, Planning poker, Function point analysis, Wideband Delphi could be highly accepted to make the most accurate estimations.

REFERENCES

1. Boehm, B. W. (1981). Software engineering economics (Vol. 197): Prentice-hall Englewood Cliffs (NJ).

2. Molokken-Ostvold, K., & Haugen, N. C. (2007). Combining estimates with planning poker--an empirical study. Paper presented at the Software Engineering Conference, 2007. ASWEC 2007. 18th Australian.

3. Muketha, G. (2016). A Review of Agile Software Effort Estimation Methods.

4. Suri, P., & Ranjan, P. (2012). Comparative analysis of software effort estimation techniques. International Journal of Computer Applications (0975–8887), 48(21).

5. Hastings, T. E., & Sajeev, A. S. M. (2001). A vector-based approach to software size measurement and effort estimation. IEEE Transactions on Software Engineering, 27(4), 337-350.

6. Mahnic, V. (2011). A case study on agile estimating and planning using scrum. Elektronika ir Elektrotechnika, 111(5), 123-128.

7. Ungan, E., Çizmeli, N., & Demirörs, O. (2014, August). Comparison of functional size based estimation and story points, based on effort estimation effectiveness in scrum projects. In Software Engineering and Advanced Applications (SEAA), 2014 40th EUROMICRO Conference on (pp. 77-80). IEEE.

8. Maxwell, K., Van Wassenhove, L., & Dutta, S. (1999). Performance evaluation of general and company specific models in software development effort estimation. Management Science, 45(6), 787-803.

9. Jørgensen, M. (2007). Forecasting of software development work effort: Evidence on expert judgement and formal models. International Journal of Forecasting, 23(3), 449-462.

10. Britto, R., Mendes, E., & Börstler, J. (2015, July). An empirical investigation on effort estimation in agile global software development. In Global Software Engineering (ICGSE), 2015 IEEE 10th International Conference on (pp. 38-45). IEEE.

METHOD TO DEVELOP AN IOT BASED SMART AGRICULTURE ANALYZING DEVICE FOR LARGE AGRICULTURAL INDUSTRY

Jayasinghe J.A.S.S. *, Dangalla R L.*
*Department of Computing and Information Systems, Faculty of Applied Sciences,
Sabaragamuwa University of Sri Lanka, Belihuloya

Keywords: IoT, Agriculture, Smart Analyzing, SMS

Agriculture plays the vital role in economy and development of the country. Many people are depend on agriculture and The farmers and corps are the main players of agriculture. Most developing countries exports their Agricultural products to other countries and a grate income to the country. But sometimes farmers get low yields due to various reasons. Unscheduled pesticide usage and unscheduled fertilizing are some problems which are farmers facing. Less awareness about the soil is the main reason for the less harvest. The main solution to this problem is smart agricultural approach by modernizing the traditional methods of measuring the soil, environmental and weather conditions using smart technology. Smart farming systems are expected to play an important role in improving farming activities and cultivating activates. There are some important factors for the growth of the plants and those factors are directly affected to have a good crop cultivation. When considering some areas the Climate and weather changes are directly affected to the plant growth .it is better the famer have an idea or forecast about the future or the current situation of the cultivation. The real need is a smart agricultural device that can accurately measure soil conditions and environmental condition.

The objective of this research was to develope a smart agriculture analyzing device based on Internet of Things (IOT) technology. Because most of people are using smart phones and laptops now a days. And people are looking easy and fast ways to do their works Quick. When a smart device is their it will very helpful. This smart device was used several sensors to read and send some values regarding condition of soil and environment. Those sensor values were sent to a Server using an ADS1115 board and it was connected to ESP8266 (NodeMCU) board with sensors. The provided sensors values stored to the database. Such as soil moisture, soil Ph, humidity, sunlight and temperature.

The client gets a username and secret key for login, and client observes soil and environmental conditions by logging into the system. The observed raw sensors values should be under international stranded level. In one moment by interacting with options of the website and facilitated to get a report with past details. The values are not in the international stranded range user will be notified with a web pop up message and Mobile SMS. This method reduced the number of field observation and manual paper base work. The advantage of this system is simply plug the device, switch on and user can get the data from the web app.

DATA MINING APPROACH TO PREDICT SUITABLE SEASONS TO CULTIVATE CROPS

Senadeera T.L*, Ekanayake E.M.U.W.J.B **

*Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya

**Department of Physical Sciences and Technologies, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya

This project addresses three key prevailing issues in small-scale farming in rural areas of Sri Lanka. Badulla district and Monaragala district were considered for the data collection. The farmers have limited access to markets, lack of knowledge sharing and inadequate technical guidance, consequently farm profits are decreased and farmers are staying away from farming. The proposed web based system links farmers to markets while significantly reducing the operational cost. Also, this system implements a localized Chat room, which farmers can interact with professionals or experts in agriculture domain, peers, agro logistics suppliers as well as consumers or buyers and share their knowledge, views and new ideas using their own languages. Furthermore, the Decision Support System integrated to this system provides timely guidance that helps farmers for well-planned cultivations. Here considered most important factors to cultivate factors as Rainfall, Temperature and Soil type. Collected data from various resources for those important factors and mined it by Simple K means clustering to identify different hidden patterns. Clusters that got from K means clustering and basic important factors about crops that collect are compared with each other and took decisions about crops and when they have to cultivate and where they have to grow. Result from the other research identifying patterns in prices of crops in E Agro Platform compared with the result of this research. Through this research socio-economic state of the small-scale farmers will be uplifted. Also, the farmers will aware to apply mobile and ICT in farming and consequently, the rural communities become more knowledgeable.

TOWARDS BRIDGING THE INFORMATION GAP BETWEEN RURAL PADDY FARMER AND URBAN MARKET

A.K.Upani Padmila[1], Gajhanan Vettivel[2], Chandima H. de Silva[3]

[1]Lecturer in IT cum Programme Coordinator, CINEC Metro Campus, 144/3, S De S Jayasinghe Mawatha, Nugegoda, Sri Lanka.

[2]Network Administrator cum Instructor ,CINEC Metro Campus, 144/3, S De S Jayasinghe Mawatha, Nugegoda, Sri Lanka.

[3]Head, CINEC Metro Campus, 144/3, S De S Jayasinghe Mawatha, Nugegoda, Sri Lanka.

Key Words: Information gap, Farmer dissatisfaction, Paddy trading, IT enabled Business model

INTRODUCTION AND OBJECTIVES

Lack of information is one of the major factors behind most issues of rural farmers in Sri Lanka. At the extreme such issues have even ended up in the affected parties committing suicide. This research is aimed at exploring the nature and the causes with respect to deficiency of information on the part of the Sri Lankan rural farmer, who faced numerous difficulties in selling their produce at a reasonable price. According to the literature review it is shocking to identify that farmers who cultivate traditional rice and organic fruits and vegetables are the mostly negatively affected parties due to the current trading process even though the demand and popularity for these food items have been increased in Sri Lanka. Hence, this study focused on farmers who are engaged in cultivating organic vegetables, fruits and traditional rice. Further, this research proposes an IT based solution to bridge this information gap as a primary effort to reengineer the current trading process which is in favor for intermediaries. A comprehensive study of the relevant stakeholders including farmers, buyers and government institutions linked with paddy and food industry has been conducted to explore the nature of the problem. Having examined both the IT and non-IT solutions adopted by Sri Lanka and in other countries of paddy and a range of other crops, a new business model is proposed and tested on a pilot scale. Accordingly, this research makes a valuable contribution towards addressing the marketing issues of the organic food and traditional paddy industry enriching the rural farmer with the key information resources.

Main Objective:
- The main objective of this research is to fill the information gap between rural paddy (traditional rice) farmer and urban market which will be a solution for the identified question of how to fill the information gap between rural paddy farmer and urban market.

Sub Objectives:
- To conduct a comprehensive background study to identify factors effecting the dissatisfaction of traditional rice farmer
- To introduce an IT solution as a primary step to re-engineer the current business process
- To verify the proposed IT solution

METHODOLOGY

The research is exploratory in nature and it is more into collection of qualitative data. Three broad sections called data collection, proposing the IT based solution and verification of the solution can be identified within the methodology of this study.

Data Collection
Kurunegala Assistant Government Agent (AGA) Division within the Kurunegala district has been selected for the data collection as it consists of the highest number of rural farmers. Sample of rural farmers representing three Grama Niladhari (GN) Divisions within Kurunegala AGA Division which are Bamunugedara, Asswedduma and Kurunegala Town – North East have been taken for the final data collection and analysis. Structured and unstructured interviews for government institutions related with the paddy and agriculture industry and surveys with questionnaires for farmers were used to collect data.

PROPOSING A SOLUTION

Having examined marketing channels in the traditional rice industry along with a comprehensive business process study, the research team has proposed an IT based business model. The IT solution is Android Mobile Application named as CROP-EX which facilitates the buy and sell orders from sellers/traditional rice farmers and buyers.

Verifying the Solution
As an extension of this study, research team has conducted an attitude survey with the selected sample of rural paddy farmers regarding the proposing solution.

RESULTS AND DISCUSSIONS

From a comprehensive background study, the research team has identified there are several dissatisfaction factors affect to the rural paddy farmers' desperate situation as presented in Figure 1. According to the findings among all the other dissatisfaction factors, the information gap between paddy farmer and the paddy industry plays a major role for the issues faced by the rural paddy farmers. Hence, the research group has identified the significance of this information gap and has focused on further research and analysis.
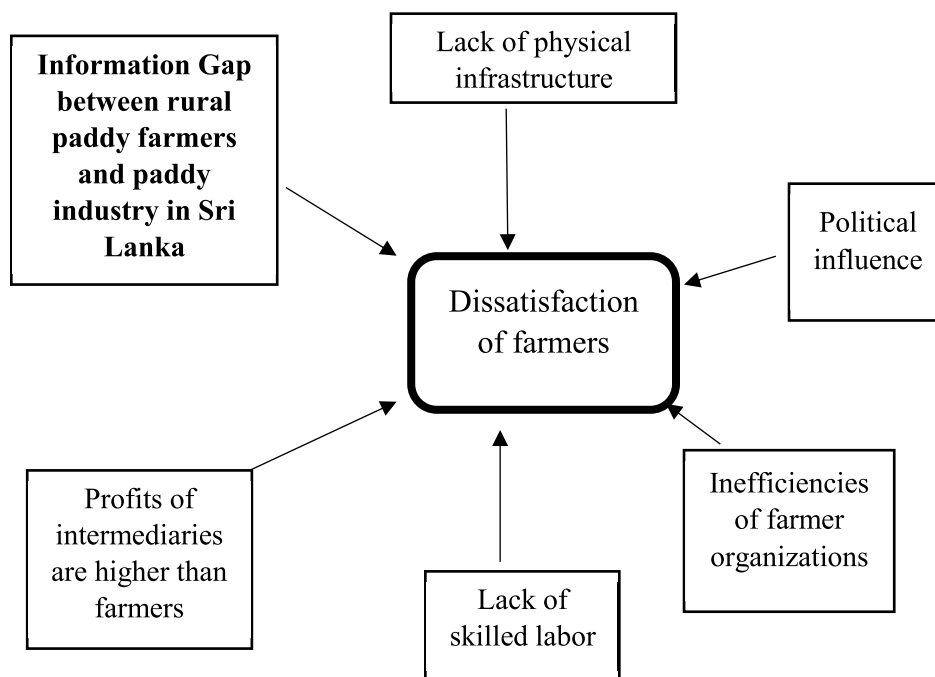

Fig 1 : High Level Conceptual View

The factors affecting to the identified gap had been categorized under main two sections as shown in Figure 2 and Figure 3 according to the findings derived from the literature review, questionnaire responses and interviews with stakeholders and from the background study as information gap occurs at pre cultivation period and information gap occurs at post cultivation period.
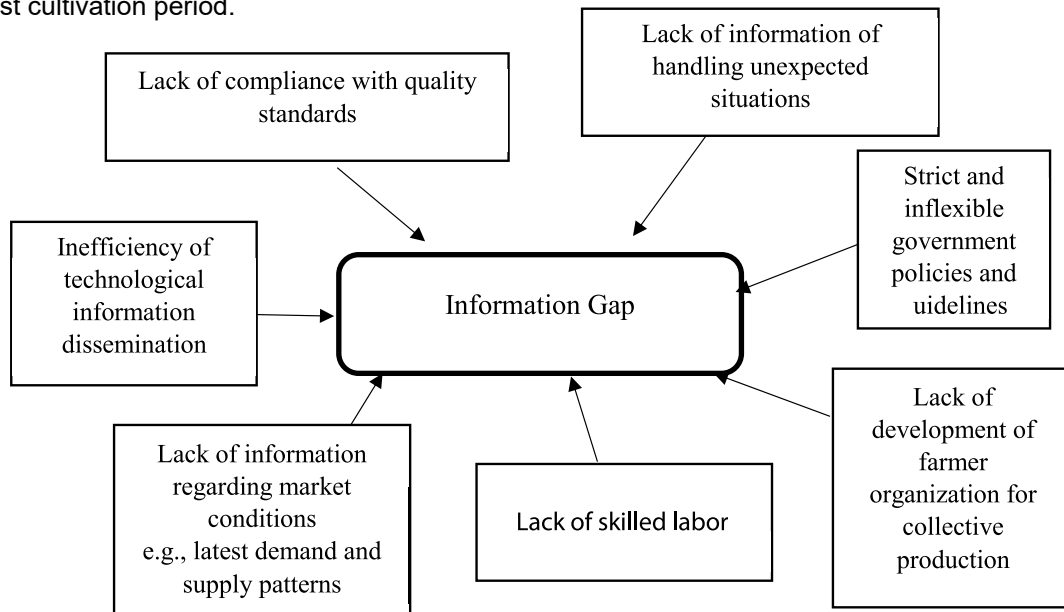

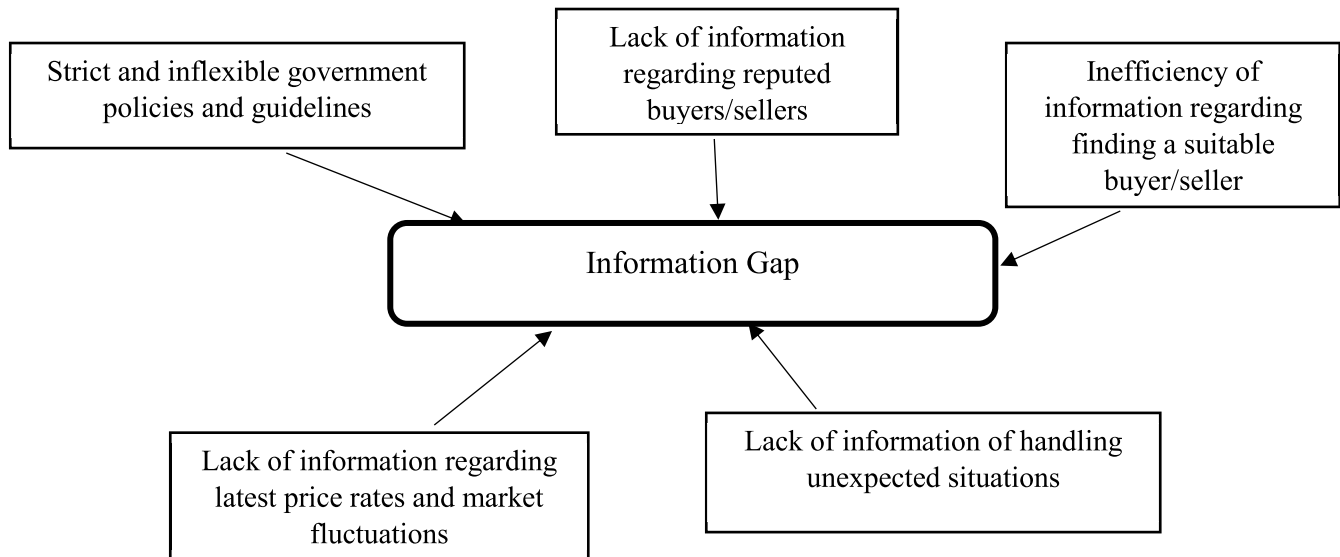Fig 2: Conceptual View of Information Gap occurs in Pre Cultivation Period

Fig 3: Conceptual View of Information Gap occurs in Post Cultivation Period

Questionnaire was comprised of questions to capture the factors identified from the background study which are represented in Figure 2 and Figure 3. The responses of the farmers are categorized mainly into two categories as favourable and less favourable for the proposing business model.

For each question of the distributed questionnaire the project team has calculated the percentages of favourability and less favourability by applying to the below mentioned formulas.
Favourable Response % = (Count of favourable responses per a question) / (Number of farmers) x 100 (1)

Less Favourable Response % = (Count of less favourable responses per a question) / (Number of farmers) x 100 (2)

As per the conceptual views represented by Figure 2 and Figure 3, the analysed data of the questionnaire is applied to the independent variables in order to ensure the existence of the social issue and emphasize the importance of the proposing business model.

Questions in the questionnaire are categorized under each identified factor in two conceptual views which they belong to. Then sum of the favourable response percentage and less favourable response percentage are calculated under each variable and then their average is calculated.

Favourable Response % = (Sum of favourable response percentages per an independent variable)/ (Number of questions per an independent variable) (3)

Less Favourable Response % = (Sum of less favourable response percentages per an independent variable)/ (Number of questions per an independent variable) (4)

Figure 4 elicits the analysis of data on questionnaire results of this study.

Following is the numbering of variables in Figure 4,

1. Inefficiency of technological information dissemination
2. Lack of information of handling unexpected situations
3. Lack of compliance with quality standards
4. Lack of information regarding market conditions

E.g., latest demand and supply patterns

5. Lack of skilled labour

6. Inefficiency of information dissemination regarding finding a suitable buyer/seller

7. Lack of information regarding reputation level of buyer/ seller

8. Lack of information regarding latest price rates and market fluctuations

9. Strict and inflexible government policies and guidelines

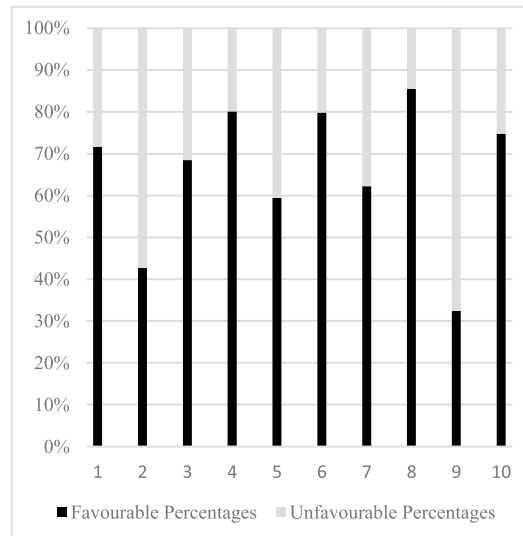10. Lack of development of farmer organization for collective production

Fig 4: Data Analysis Findings

According to the data analysis there are eight supporting variables and only two unsporting variables. Variables which have more than 55.00% of favourable percentage were taken as supported variables and variables which have less than 55.00% of favourable percentage were taken as unsupported variables for the introduced conceptual view and for the proposing solution. According to this criteria identified supported variables are mentioned below.

• Inefficiency of information dissemination regarding finding a suitable buyer/seller

• Lack of information regarding reputation level of buyers/sellers

• Lack of information regarding latest price rates and market fluctuations

• Inefficiency of technological information dissemination

• Lack of compliance with quality standards

• Lack of information regarding market conditions e.g., latest demand and supply patterns

• Lack of development of farmer organization for collective production
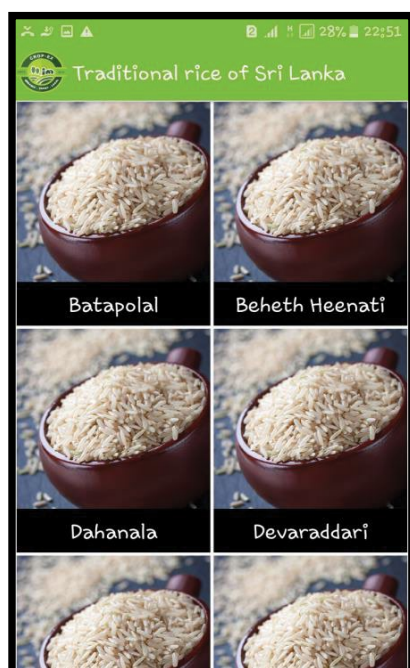
• • Lack of skilled labour

According to the analysis on data presented in Figure 4, the highest favourable percentage is recorded by the independent variable/cause of "lack of information regarding latest price rates and market fluctuations". The second highest favourable percentage is recorded by two variables called "lack of information regarding market conditions" and "inefficiency of information dissemination regarding finding a suitable buyer/seller". This shows that there is a significant market information gap between rural paddy farmer (traditional rice farmer) and the urban market.

Though two variables called "lack of information of handling unexpected situations" and "strict and inflexible government policies and guidelines" have been identified as unsupported variables since they have recorded less than 55% of favourable percentage, there is a possibility to conduct future research on those areas also. The research team has identified that there are future research areas regarding money management issues, enhancing education levels and methods to change attitudes of famers as well as other stakeholders in paddy trading industry. And also when it comes to trading of food items humans need touch and feel sense. Therefore integration of paddy quality checking methods to the proposing solution will be another future research area.

The proposed CROP-EX Android Application can be downloaded to any kind of a smart phone. Buyers can register them in the system and can submit buy orders. Sellers basically farmers have to register through a farmer organization. Farmer organizations can be provided the smart phones. Based on the collective production, the farmer organization can submit a sell order. Buyer can make the payment on the receipt of goods. The rating facility available for both the parties.

As buyers can demand rice for future use, the farmers can plan their cultivation based on the demand. CROP-EX facilitates the customer oriented and demand based traditional rice production which is a prevalent requirement in the organic food industry.
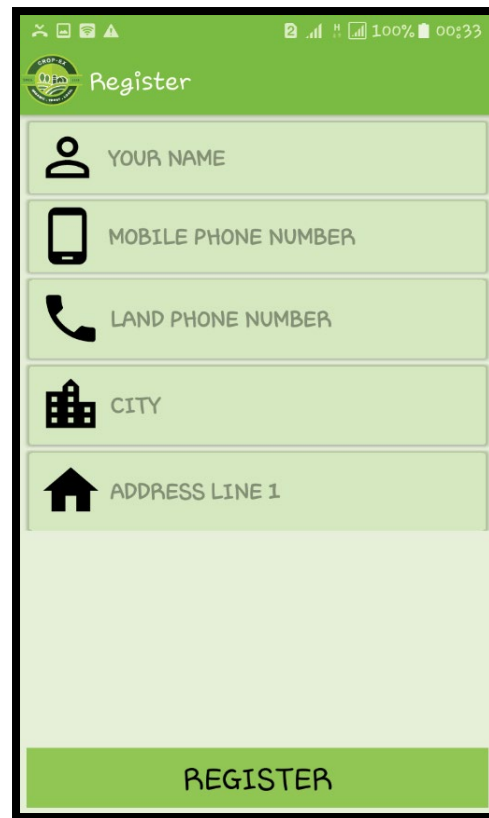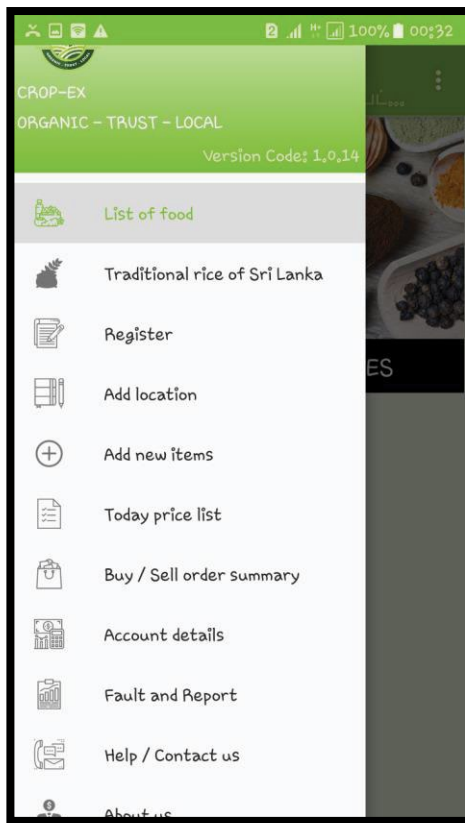
Few of the GUIs of functions of the CROP-EX IT solution are as below.

CONCLUSION

Research findings reveal that Sri Lankan rural farmers faced difficulties when it comes to sell their produce at a reasonable price due to deficiency of information regarding market data. The proposing IT based business model will ease this situation by filling the identified information gap by enabling a method for farmers to obtain market information in an expeditious way. This IT solution mainly consists of a web application and a mobile application which acts as a platform to connect sellers/farmers and buyers directly. It will enable the sellers and buyers to put their sell/buy orders through the system which will completely eliminate the intermediaries. And also this enables the customer oriented and demand driven production where framers do not need to suffer with the excess production.

Since research team has received a positive response from the attitude survey conducted using the selected sample of rural paddy farmers and the discussion with the Ministry of Agriculture regarding the proposing solution there is a possibility to implement this solution within Sri Lanka. This solution can be developed as a web application where users can obtain required information through the internet.

This solution will mitigate irregularities in Sri Lankan paddy and organic food market and it will enable the stakeholders of the market to do their trading activities in a proper and profitable way

REFERENCES

1. Mohamed Ismail Mujahid Hilal and Kaldeen Mohamed Mubarak, "Rice Marketing: Lesson and Driver for Sri Lankan Producers", in the title of Published Research Paper, Department of Management, South Eastern University of Sri Lanka, Oluvil, Sri Lanka.

2. Eastern Provincial Council, "Eastern Develeopment Plan 2012 - 2016", in the title of Published Research Paper, Trincomalee.

3. W.C.Gamawelagedara, Y.M Wickramasinghe and C.A.K Dissanayake, "Impact of Rice Processing Villages on House Hold Income of Rural Farmers in Anuradhapura District", Department of Agricultural systems, Faculty of Agriculture, Rajarata University of Sri Lanka, 2011.

4. Sharmini Kusum Kumara et al., "Viability of Rice Processing at Household Level", in the Title of Published Research Paper, Hector Kobbekaduwa Agrarian Research and Training Institute, Colombo 07, Sri Lanka, 2008.

INDEX OF
AUTHORS

INDEX OF
AUTHORS

PROGRAMME OF SCIENTIFIC SESSIONS

DATE : 23.11.2018
VENUE : SRI LANKA FOUNDATION, COLOMBO 07

SESSION ONE

Venue : Main Auditorium

Session Chair : Dr. Carmen Z. Lamagna

Rapporteur: Mr. Deshanath Kulasinghe

Time: 11:00 am – 1:00 pm

Time	Abstract Title	Authors
11:00 am – 11:30 am	A Digital Green Banking Framework: Replacement of the Manual Signature System with A Finger Print System	Lojenaa Navanesan ,Ruwan D Nawarathna, Shelton Sarath Suthaharan Satkunarajah
11:30 am – 12:00 pm	Method to develop automatic vehicle number plate recognition system	Charith Gayashan Wickramarathna, Ravindra Lakmal Dangalla
12:00 pm – 12:30 pm	Comparing traditional security mechanisms with Security and Privacy in Big Data	Kasun Lakshan Mahaliyanaarachchi, Supunmali Ahangama

SESSION TWO

Venue: Hall 01

Session Chair: Mr. Alan Clarke

Rapporteur: Dr. Madhawa Ranawake

Time: 11:00 am – 1:00 pm

Time	Abstract Title	Authors
11:00 am – 11:30 am	Energy efficient VM consolidation in cloud computing environments	Chathurika Jayathilake, Pubudu Jayasena
11:30 am – 12:00 pm	Nature inspired approach for efficiency aware resource scheduling in cloud computing	Isuru Abeywardana, Pubudu Jayasena, Dr. Shanmuganathan Vasanthapriyan
12:00 pm - 12:30 pm	Data analytics in fog computing using docker and kubernetes	Buddhika Priyabhashana, K.Pubudu Nuwanthika Jayasena

1.00 pm – 2.00pm LUNCH

SESSION THREE

Venue   : Main Auditorium

Session Chair : Dr. M M S N Premathilake

Rapporteur: Ms. Tharaka Sri Kuruparan

Time: 2:00 pm – 3:30 pm

Time	Abstract Title	Authors
2:00 pm – 2:30 pm	How can business utilise business intelligence for competitive advantages in retail & consumer goods industry	Dr. Madhawa Ranawake
2:30 pm – 3:00 pm	Find new unsupervised algorithum for freelancer employees ranking	Sameera Lakshitha, Ravindra Dangalla
3:00 pm – 3:30 pm	Introductory programming: factors affecting course outcome of novice students in Sri Lankan universities	Wasana Dahanayake, Samadara Dhanapala

SESSION FOUR

Venue: Hall 01

Session Chair: Dr. Jayalath Ekanayake

Rapporteur: Ms. Srinivasini Sasitharasarma

Time: 2:00 pm – 3:30 pm

Time	Abstract Title	Authors
2:00 pm – 2:30 pm	An expert system for tooth removal – a case on tooth extraction forceps	Banujan Kuhaneswaran, Dr. Vasanthapriyan Shanmuganathan
2:30 pm – 3:00 pm	Why key employees quit their jobs	Aruna Sanjeewa, Dr. Jayalath Ekanayake
3:00 pm – 3:30 pm	Single – line rate and mixed – line rate variation for dedicated protection in elastic optical networks	Satkunarajah Suthaharan, Rukshani Puvanendran,Vanuja Chandrakumar

SESSION FIVE

Venue   : Main Auditorium

Session Chair : Dr. Sainey Faye

Rapporteur: Ms. Anushika Elangasinghe

Time: 4:00 pm – 5:30 pm

Time	Abstract Title	Authors
4:00 pm – 4:30 pm	A Framework for MySQL Pre-processor with interactive user interfaces	Krishnamoorthy Santhanakrishnan, Vijayanathan Senthooran
4:30 pm – 5:00 pm	Lightweight, Cross platform & Web browser based System Application Development Server Runtime with SDK - NeutralinoJS	Deepal Samarakoon, Shalitha Suranga, Chathumaduri Hettiarachchi
5:00 pm – 5:30 pm	Improving effort estimation practices in small scale agile software products	Nizrin Mahisha, Kalinga Gunawardhana, Kapila Rathnayaka

SESSION SIX

Venue: Hall 01

Session Chair: Dr. Jayalath Ekanayake

Rapporteur: Ms. Piyumi Sirigampala

Time: 4:00 pm – 5:30 pm

Time	Abstract Title	Authors
4:00 pm – 4:30 pm	Method to develop an IoT based smart agriculture analyzing device for large agricultural industry	Saumiyanath Suboda, Ravindra Dangalla
4:30 pm – 5:00 pm	Data Mining Approach to predict Suitable Seasons to Cultivate Crops	Tharindu Senadeera, Dr. Jayalath Ekanayake
5:00 pm – 5:30 pm	Towards Bridging the Information Gap between Rural Paddy Farmer and Urban Market	Upani Padmila, Gajhanan Vettivel, Chandima de Silva

EdHat INTERNATIONAL
CONFERENCE
23RD NOVEMBER 2018

EDhat
INTERNATIONAL

Delivering world-class qualifications

EdHat International UK
Unit 1, 324 Stafford Road, Croydon, Surrey
CRO 4NH, United Kingdom.

Phone : +44 (0) 207 183 4556
Fax   : +44 (0) 208 686 9559

E-mail : info@edhatinternational.co.uk

www.edhatinternational.co.uk