Proceedings of 5th World Conference on Applied Science Engineering and Technology

(WCASET – 17)

Bangkok, Thailand

11th-12th October' 17

Institute For Engineering Research and Publication

4A, Girija Apartment, MMDA,

Arumbakkam, Chennai-600106, India

www.iferp.in

Publisher: IFERP Explore

Editorial:

We cordially invite you to attend the ***5th World Conference on Applied Science Engineering and Technology (WCASET - 17)*** which will be held at ***The Tawana, Bangkok, Thailand*** on ***October 11th-12th, 2017***. The main objective of ***WCASET*** is to provide a platform for researchers, engineers, academicians as well as industrial professionals from all over the world to present their research results and development activities in relevant fields of Science, Engineering and Technology. This conference will provide opportunities for the delegates to exchange new ideas and experience face to face, to establish business or research relationship and to find global partners for future collaboration.

These proceedings collect the up-to-date, comprehensive and worldwide state-of-art knowledge on cutting edge development of academia as well as industries. All accepted papers were subjected to strict peer-reviewing by a panel of expert referees. The papers have been selected for these proceedings because of their quality and the relevance to the conference. We hope these proceedings will not only provide the readers a broad overview of the latest research results but also will provide the readers a valuable summary and reference in these fields.

The conference is supported by many universities, research institutes and colleges. Many professors played an important role in the successful holding of the conference, so we would like to take this opportunity to express our sincere gratitude and highest respects to them. They have worked very hard in reviewing papers and making valuable suggestions for the authors to improve their work. We also would like to express our gratitude to the external reviewers, for providing extra help in the review process, and to the authors for contributing their research result to the conference.

Since August 2017, the Organizing Committees have received more than 34 manuscript papers, and the papers cover all the aspects in Electronics, Computer Science, Information Technology, Science Engineering and Technology. Finally, after review, about 15 papers were included to the proceedings of ***WCASET - 2017***.

We would like to extend our appreciation to all participants in the conference for their great contribution to the success of ***WCASET 2017***. We would like to thank the keynote and individual speakers and all participating authors for their hard work and time. We also sincerely appreciate the work by the technical program committee and all reviewers, whose contributions made this conference possible. We would like to extend our thanks to all the referees for their constructive comments on all papers; especially, we would like to thank to organizing committee for their hard work.

Editor-In-Chief
Dr. Nalini Chidambaram
Professor
Bharth University

Acknowledgement

IFERP is hosting the ***5th World Conference on Applied Science Engineering and Technology*** this year in month of October. The main objective of WCASET is to grant the amazing opportunity to learn about groundbreaking developments in modern industry, talk through difficult workplace scenarios with peers who experience the same pain points, and experience enormous growth and development as a professional. There will be no shortage of continuous networking opportunities and informational sessions. The sessions serve as an excellent opportunity to soak up information from widely respected experts. Connecting with fellow professionals and sharing the success stories of your firm is an excellent way to build relations and become known as a thought leader.

I express my hearty gratitude to all my Colleagues, staffs, Professors, reviewers and members of organizing committee for their hearty and dedicated support to make this conference successful. I am also thankful to all our delegates for their pain staking effort to travel such a long distance to attain this conference.

Er. R. B. Satpathy
Secretary
Institute for Engineering Research and Publication(IFERP)

5th World Conference on Applied Science Engineering and Technology – 2017

Keynote Speaker

Dr Kanish Kapoor
Assistant Professor
School of Civil Engineering, Lovely Professional University, India

Biography:

Dr. Kanish Kapoor is an Assistant Professor in the School of Civil Engineering, Lovely Professional University, India. He obtained his Doctorate from Department of Civil Engineering, National Institute of Technology Jalandhar. His research work is focused on the reuse and recycling of Construction and Demolition waste and industrial by-products in cementitious materials. Over last five years he has published many international journal and conference papers. He has been responsible for research projects based on pervious concrete incorporating waste plastics and beneficiated recycled aggregates.

DR. KANISH KAPOOR

CONTENTS

CONTENTS

ORGANIZATION COMMITTEE

1. Koichiro Mashiko
Kuala Lumpur, , Malaysia

2. Prof. Kuan Yew Wong
Universiti Teknologi Malaysia,, Malaysia

3. Dr Nasratun Masngut
Universiti Malaysia Pahang , Malaysia

4. Prof. Hakeun Lee
Korea University, South Korea

5. Afshin Davarpanah,
slamic Azad University, Tehran, Iran

6. ANAND NAYYAR
KCL Institute of Management and Technology, G.T. Road, Jalandhar, Punjab,india

7. Dr Farhad Shahnia
Murdoch University, Perth, Australia

8. Mr.Nick Kostovic
KOSTOVIC ACUPUNTURE OF BIO ELECTRONS' LASER, CORP., USA

9. Dr. Amanullah
The University of Agriculture Peshawar, Pakistan., Pakisthan

10. Dr. Chithirai Pon Selvan
Amity University, United Arab Emirates

11. DR. SASITHARAN NAGAPAN
Universiti Tun Hussein Onn Malaysia,, Malaysia

12. Dr. Vimi Jham
Dubai, United Arab Emirates

13. Dr. Vishal Datt. V. Kohir.
K B N COLLEGE OF ENGINEERING KALABURAGI, India

14. Farhan bt Mohd Said
Universiti Malaysia, Malaysia

15. Jun Haslinda Shariffuddin
Universiti Malaysia Pahang, Malaysia

An Overview of Big Data Analytics: Comparison of various Analytics and Visualization Tools

[1] Miss. Avanti Laad, [2] Mr. Bhushan Wanhere, [3] Mr. Nikhil Jawade, [4] Miss. Mansi Choudhary
[1][2][3][4] Dept of CSE, SSGMCE, Shegaon, Maharashtra, India
[1] laadavanti@gmail.com, [2] bhushanwanhere95@gmail.com, [3] nikhil.jawade@gmail.com,
[4] choudhary3296mansi@gmail.com

***Abstract*:-- Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale. The primary purpose of this paper is to provide an in-depth analysis of different platforms available for performing big data analytics. In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Various analytics and visualization tools are discussed in the paper that are analytics tools such as R, Knime, Weka, and Rapid Miner also the visualization tools such as Tableau, Qlikview, Spotfire, MS BI Stack.**

Index Terms: Big data, Analytics,Visualization.

I. INTRODUCTION

Recently, data generation at the scale of terabytes, petabytes and exabytes have become commonplace in many scientific and commercial domains. Streaming data, social media content, electronic medical records, astronomy surveys, genomic and proteomic studies and similar areas generate data at a scale that is becoming increasingly difficult to manage using traditional database technologies. Big Data is an umbrella term used for such massive collections of data. Besides volume, analytics has to face challenges like heterogeneity, timeliness, and complexity, velocity and privacy issues. Due to the rapid growth of data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets. Furthermore, decision makers need to be able to gain valuable insights from such varied and rapidly changing data, ranging from daily transactions to customer interactions and social network data [1]. Such value can be provided using big data analytics, which is the application of advanced analytics techniques on big data. As the technology that helps an organization to break down data silos and analyze data improves, business can be transformed in all sorts of ways. Today's advances in analyzing big data allow researchers to decode human DNA in minutes, predict where terrorists plan to attack, determine which gene is mostly likely to be responsible for certain diseases and, of course, which ads you are most likely to respond to on Facebook[2]. The other challenge is to fuse data sources - of same type or multi- modal and perform joint analytics on the integrated data. The scale of such data poses significant challenges for analytics, going beyond what can be supported by conventional data, storage and retrieval models. To analyze such a large volume of data, Big Data analytics is typically performed using specialized software tools and applications for predictive analytics, data mining, text mining, and forecasting and data optimization [2]. Collectively these processes are separate but highly integrated functions of high-performance analytics. Using Big Data tools and software enables an organization to process extremely large volumes of data that a business has collected to determine which data is relevant and can be analyzed to drive better business decisions in the future.

Notably, the business area getting the most attention relates to increasing efficiency and optimizing operations. Specifically, 62 percent of respondents said that they use big data analytics to improve speed and reduce complexity. There are various tools which can be used for data analytics and visualization that makes the decision making easy and efficient. This paper aims to analyze some of the different analytics methods and tools which can be applied to big data, as well as the opportunities provided by the application of big data analytics in various decision domains.

II. BIG DATA

Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating and information privacy[2].

Sources of Sources of Big data:

1.Public Data:
Public data includes data that is publicly available like data generated by government sectors, weather data, research data, open source data and other data which is freely available to the public.

2. Transactional Data:
Applications used by various enterprises perform transactions like Mobile Applications, Web Applications and many more. In order to support the transactions of these type, there are one or more relational databases which works at backend. This type of data is structured and it is referred to as Transactional Data.

3. Social Media:
Huge amount of data is being generated on social networks like Twitter, LinkedIn, Face book, etc. Thus social media has to capture and manage unstructured[2].

4. Enterprise Data:
Huge amount of data comes from enterprises in different formats such as files, Word documents, emails, spreadsheets, PowerPoint presentations, HTML pages, pdf files, XMLs, etc.
This data which is spread across the organization in different formats is referred to as Enterprise Data.

5. Activity Generated data:
Data that has been generated by machines that surpasses the data volume generated by humans. These include data from various machines like images from medical devices, data from sensors, surveillance videos, satellites data and data from mobile towers. These types of data are referred to as Activity Generated data.

III. BIGDATA ANALYTICS

Data analysis, also known as analysis of data or data analytics, is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decisionmaking. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains.[3]

Process of Data Analysis:
There are several phases that can be distinguished, described below:

A. Data requirements:
The data necessary as inputs to the analysis are specified based upon the requirements of those directing the analysis The general type of entity upon which the data will be collected is referred to as an experimental unit (e.g., a person or population of people).[3]

B. Data collection:
Data is collected from a variety of sources. The requirements may be communicated by analysts to custodians of the data, such as information technology personnel within an organization. The data may also be collected from sensors in the environment, such as traffic cameras, satellites, recording devices, etc. [3].

C. Data Processing:
Data initially obtained must be processed or organized for analysis. For instance, these may involve placing data into rows and columns in a table format (i.e., structured data) for further analysis, such as within a spreadsheet or statistical software.[3]

D. Data cleaning:
Once processed and organized, the data may be incomplete, contain duplicates, or contain errors [3]. The need for data cleaning will arise from problems in the way that data is entered and stored. Data cleaning is the process of preventing and correcting these errors [4].

E. Exploratory Data Analytics:
Once the data is cleaned, it can be analyzed. Analysts may apply a variety of techniques referred to as exploratory data analysis to begin understanding the messages contained in the data.[5][6] The process of exploration may result in additional data cleaning or additional requests for data, so these activities may be iterative in nature. Data visualization may also be used to examine the data in graphical format, to obtain additional insight regarding the messages within the data [3].

F. Communication:
Once the data is analyzed, it may be reported in many formats to the users of the analysis to support their requirements. When determining how to communicate the results, the analyst may consider data visualization techniques to help clearly and efficiently communicate the message to the audience [3].

G. Data Visualization:
Data visualization or data visualization is viewed by many disciplines as a modern equivalent of visual communication. It involves the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information"[7].

IV. ANALYTICS TOOLS

A. Knime:

KNIME the Konstanz Information Miner is an open source data analytics, reporting and integration platform. KNIME integrates various components for machine learning and data mining through its modular data pipelining concept. KNIME has been used in pharmaceutical research, [2] but is also used in other areas like CRM customer data analysis, business intelligence and financial data analysis.

B. RapidMiner:

RapidMiner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization.

C. R:

R is an open source programming language and software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing.[10] The R language is widely used among statisticians and data miners for developing statistical software[10] and data analysis. Polls, surveys of data miners, and studies of scholarly literature databases show that R's popularity has increased substantially in recent years [12].

D. Weka:

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Tool	R	Rapid Miner	Weka	Knime
Programming Language	R interpreted language	Java	Java	Java
Price	Open Source	Free Community Edition Commercial Enterprise Edition	Open Source	Open Source Commercial Extensions
License	GNU General Public License	AGPL (Community Edition) Closed (Enterprise Edition)	GNU General Public License	GNU General Public License
Visualization	Bar Chart , Line, Pie Chart, Histogram, Box	Bar chart, Line Bubble, Deviation, Density, Survey plots, Pie chart, Histogram, Box	Bar Chart ,Line, ,Histogram	Bar chart, Line Bubble, Density, Pie chart, Histogram, Box
Platforms	All Except min,jre	All	All	All

Fig 1: Comparison of Data Mining Tools

V. VISUALIZATION TOOLS

Data visualization is a modern branch of descriptive statistics. It involves the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information. Some of the tools are:

A. Tableau

This software adopts a very different mental model as compared to using programming to produce data analysis. Think about the first GUI that made computers public friendly, suddenly the product has been repositioned. "Pretty Graphs" are useless if they just look pretty and tell you nothing. But sometimes making data look pretty and digestible also makes it understood to the average person.

B. Spotfire

Spotfire is a smart, secure, governed, enterprise-class analytics platform with built-in data wrangling that delivers AI-driven, visual, geo, and streaming analytics. Whether you are beginning your analytics journey by building a simple dashboard, or working on generating deeper insights with a hyper-rich interactive analytic app, Spotfire supports you with numerous tools and techniques.

C. Qlikview

QlikView is a leading Business Discovery Platform. It is very powerful in visually analyzing the relationships between data. It does in-memory data processing and stores the data in the report itself that it creates. It can read data from numerous sources including files and relational databases [8].

D. Microsoft BI Stack

The Microsoft BI stack provides all the tools you need to build, manage and use a BI solution. SQL Server 2008 is the foundation of the stack as the data platform hosting the data mart or data warehouse. SQL Server 2008 includes three BI components: Integration Services (SSIS), Analysis Services (SSAS), and Reporting Services (SSRS).

Criteria	Spotfire	Qlikview	Tableau	MS BI Stack
Implementation Speed	Good	High	Good	Average
Scalability	Unlimited	Limited by RAM	Very Good	Good
Pricing	High	Above Average	High	Average
Licensing/support cost	High	High	High	Average
Interactive Visualization	Very Good	Excellent	Very Good	As good as excel

Fig 2: Comparison of Visualization Tools

CONCLUSION

This paper conducts a comparison between four analytics tools against various criteria namely: platform, visualization, and programming language. The paper also conducts a comparison between four visualization tools that are Tableau, Spotfire, Qlikview, and MS BI Stack. RapidMiner seems to support wider range visualization types. Nevertheless, R is holding the second place in popularity behind RapidMiner, which is leading the market. In terms of classifiersapplicability, we conclude that WEKA was the best tool to run the selected classifiers followed by R, RapidMiner, and finally KNIME respectively. In the context of visualization tools Qlikview is the fastest to implement, Microsoft is the price leader, Spotfire is best at scalability and Tableau is very good at visualization.

ACKNOWLEDGMENT

We would like to take this opportunity to express our heartfelt thanks to our college Shri Sant Gajanan Maharaj College Of Engineering, Shegaon, M.H. India and all our faculty members of Computer Science Department for their encouragement and whole hearted cooperation. We would like to thank Asst. Prof. V.S. Mahalle for his esteemed guidance and encouragement. His suggestions broaden our vision and guided us to succeed in this work. Our greatest thanks is to all who wished us success especially our parents, friends as their moral support and care always encourages us.

REFERENCES

[1] .Pratyusha Gudavalli, Reshma Pothuri ,Rama Devi Gunnam “Survey on Importance and Tools Used: Big Data” Volume 6, Issue 5, May 2016 ISSN: 2277 128X

[2]. http : // www .webopedia .com / TERM / B / big _ data _ analytics.html

[3]. O'Neil, Cathy and, Schutt, Rachel (2013). Doing Data Science. O'Reilly. ISBN 978-1-449-35865-5

[4]. Clean Data in CRM: The Key to Generate Sales-Ready Leads and Boost Your Revenue Pool Retrieved 29th July, 2016

[5]. Stephen Few-Perceptual Edge-Selecting the Right Graph For Your Message-September 2004

[6]. Behrens-Principles and Procedures of Exploratory Data Analysis-American Psychological Association-1997

[7]. Michael Friendly (2008). "Milestones in the history of thematic cartography, statistical graphics, and data visualization"

[8]. JVP makes 4,000% return on QlikTech Archived December 3, 2013, at theWayback Machine

[9]. How to choose between tableau and spotfire analytics for your company by Butler Analytics,August 2017, Whitepaper

[10]. Vance, Ashlee "Data Analysts Captivated by R's Power". New York Times. 2009-01-06

[11]. David Smith (2012); R Tops Data Mining Software Poll, Java Developers Journal, May 31, 2012.

[12]. Karl Rexer, Heather Allen, & Paul Gearan (2011); 2011 Data Miner Survey Summary, presented at Predictive Analytics World, Oct. 2011.

[13]. Robert A. Muenchen (2012). "The Popularity of Data Analysis Software".

[14]. Tippmann, Sylvia (29 December 2014). "Programming tools: Adventures with R". Nature. 517: 109–110.doi:10.1038/517109a.

[15]. Samiya Khan Kashish Ara Shakil and Mansaf Alam: “Cloud-Based Big Data Analytics – A Survey of Current Research And Future Directions”

[16]. Ahmad Al-Khoder, Hazar Harmouch “Evaluating four of the most popular Open Source and Free Data Mining Tools ” IJASR International Journal of Academic Scientific Research ISSN: 2272-6446 Volume 3, Issue 1 (February - March), PP 13-23

[17]. A. Wahbeh, Q. Al-Radaideh, M. Al-Kabi and E. Al-Shawakfa, "A Comparison Study between Data Mining Tools over some Classification Methods," International Journal of Advanced Computer Science and Applications (IJACSA), Special Issue on Artificial Intelligence, vol. 2.

[18]. Beisken, S.; Meinl, T.; Wiswedel, B.; De Figueiredo, L. F.; Berthold, M.; Steinbeck, C. (2013). "KNIME-CDK: Workflow-driven Cheminformatics". BMC Bioinformatics. 14: 257. PMC 3765822 Freely accessible. PMID 24103053. doi:10.1186/1471-2105-14-257.

[19]. Singh and Reddy, ‖A Survey on platforms for big data Analytics‖ Journal of Big Data 2014

[20]. Blog post: Thoran Rodrigues in Big Data Analytics, titled 10 emerging technologies for Big Data‖, December 4, 2012 —

WannaCry Ransomware Attack: Aproach and Analysis

[1] Mr. Bhushan Wanhere, [2] Miss. Avanti Laad, [3] Miss Neha Mahalle, [4] Prof.V.S. Mahalle
[1][2][3][4] Dept. of CSE, SSGMCE, Shegaon, Maharashtra, India
[1] bhushanwanhere95@gmail.com, [2] laadavanti@gmail.com, [3] mnehamahalle@gmail.com,,
[4] vsmahalle@gmail.com

***Abstract*:-- Earlier, "Cryptography" is used for information processing and communications, it helps people to store information securely and it is used very frequently for private communications but Cryptovirology is the study of applications of cryptography to build the malicious software (malware). It is an investigation, how modern cryptographic tools and paradigms can be used to strengthen, develop and improve new malware attacks. This paper presents the idea of cryptovirology which really an opposite side of the cryptography. On the contrary of cryptography, it is offensive in nature. It tells how modern cryptographic paradigms and tools can be used to strengthen, improve, and develop new malicious software attacks. The world has experienced a massive global ransomware cyber-attack known as "WannaCrypt" or "WannaCry". Hundreds of thousands computers worldwide have been hit and affected more than 150 countries. This paper clears the process of "how the Ransomware virus can damage the victim's system, based on Extortion mechanism" with showing the live scenario that causes the loss of access to information, loss of control and loss of money also. Here we discuss and explain some of the latest events regarding to this cryproviral extortion like Eternal Blue and DoublePulsar. This paper also suggests some of the countermeasures, mechanisms to defend and prevent such attacks. And can be able to provide some of the safe guards initially.**

Index Terms: ransomware; encryption; decryption;Eternal Blue;DoublePulsar;wannacry;

I. INTRODUCTION

Ransomware is a type of malicious software from cryptovirology that threatens to publish the victim's data or perpetually block access to it unless a ransom is paid. While some simple ransomware may lock the system in a way which is not difficult for a knowledgeable person to reverse, more advanced malware uses a technique called cryptoviral extortion, in which it encrypts the victim's files, making them inaccessible, and demands a ransom payment to decrypt them[1][2][3][4]. Ransomware is a type of malicious software from cryptovirology that threatens to publish the victim's data or perpetually block access to it unless a ransom is paid. While some simple ransomware may lock the system in a way which is not difficult for a knowledgeable person to reverse, more advanced malware uses a technique called cryptoviral extortion, in which it encrypts the victim's files, making them inaccessible, and demands a ransom payment to decrypt them. The world has experienced a massive global ransomware cyber-attack known as "WannaCrypt" or "WannaCry" (Ransom: Win32/ WannaCrypt) since Friday, May 12 2017. Hundreds of thousands computers worldwide have been hit and affected more than 150 countries [15]. WannaCry is far more dangerous than other common ransomware types

because of its ability to spread itself across an organization's network by exploiting a critical vulnerability in Windows computers, which was patched by Microsoft in March 2017 (MS17-010).

The malware has the capability to scan heavily over TCP port 445 (Server Message Block/SMB), spreading similar to a worm, compromising hosts, encrypting files stored on them then demanding a ransom payment in the form of Bitcoin. It is important to note that this is not a threat that simply scans internal ranges to identify where to spread, it is also capable of spreading based on vulnerabilities it finds in other externally facing hosts across the internet.

II. BACKGROUND

Cryptography: Cryptography deals with the actual securing of digital data. It refers to the design of mechanisms based on mathematical algorithms that provide fundamental information security services. You can think of cryptography as the establishment of a large toolkit containing different techniques in security applications. There are mainly two aspects in cryptography algorithms and the key used [6]. The important encryption techniques are symmetric, asymmetric, hash function algorithms.

Cryptovirology: Cryptovirology is the study of the applications of cryptography for implementation of malicious software [6]. It is an area that employs public key cryptography to mount attacks on computer systems, showing that cryptography has also "negative" usage. It's

the combination of virus science and cryptography which created Cryptovirology.

Kleptography: "Kleptography" is the study of stealing the information more securely. A kleptographic attack is an attack in which a malware designer deploys an asymmetric backdoor. There is an explicit distinction between confidentiality of the messages and awareness of the attack is taking place in this attack. A kleptographic attack requires a private key known only to the attacker in order to use the backdoor. The reverse-engineer cannot use it because if reverse engineering detects the key generation code then he or she will get the idea that attack is under process [6].

Cryptanalysis: Cryptography is the science of making the secret unintelligible and cryptanalysis is the science of retrieving the secret from that unintelligible data for the further use [6]. It involves the study of cryptographic mechanism with the intention to break them. The cryptographic process results in the cipher text for transmission or storage. Cryptanalysis is also used during the design of the new cryptographic techniques to test their security strengths.

III. METHODOLOGY

"Crypto-viral extortion, which uses public key cryptography, is a denial of resources attack. It is a three-round protocol that is carried out by an attacker against a victim. The attack is carried out via a crypto-virus that uses a hybrid cryptosystem to encrypt host data while deleting or overwriting the original data in the process [1]."

The purpose of this whole is to allow the attacker to indirectly run code of his own Trojan without being blamed for installing it.

The three-round protocol is interesting. It consists of the following:

1. Virus author-> Victim
2. Victim->Virus author
3. Virus Author -> Victim

Algorithm:

1. Start
2. Generate RSA-2048 key pair.
3. Placed Key pair within virus.
4. Deploy virus on victim's machine.
5. if virus successfully deployed
    a. Then establishes connection with virus author using TOR onion servers.
    b. Generates key K & initialization vector IV.
    c. Encrypts host data using different AES-128-CBC key per file.
    d. The AES key is encrypted using the infection specific RSA key pair.
    e. Encrypted text is held for RANSOM.
    f. Virus notifies the attack & state demand.
6. Checks for IP address of the infected machine and same subnet are scanned for additional vulnerable machines connected to via port 445 TCP
    a) If Connection established then data containing the exploit payload is transferred to the connected pc. Go to step (1)
    b) Else try again for deployment.
7. If victim want data,
    a) Then pay ransom & transmit CT to virus author.
    b) Virus author decrypts CT.
    c) Send key & IV to victim decryptor application.
    Else
    Data can never be decrypted.
    Else
    Try again for deployment.
8. Stop

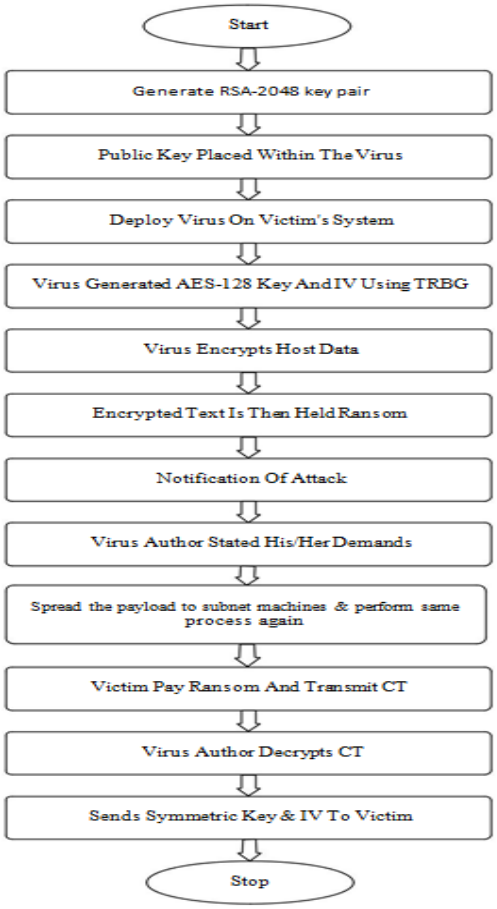


Fig. 1: Flowchart of extortion Approach

IV. WANNACRY EXPLOITS

TSB published several leaks containing hacking tools from the NSA. Specifically, these exploits and vulnerabilities targeted • enterprise firewalls, anti-virus products, and Microsoft products.

A. Eternal Blue

Eternal Blue exploits a vulnerability in Microsoft's implementation of the Server Message Block (SMB) protocol [9]. The vulnerability exists because the SMB version 1 (SMBv1) server in various versions of Microsoft Windows mishandles specially crafted packets from remote attackers, allowing them to execute arbitrary code on the target computer.

Eternal Blue proceeds as follows for the exploitation:

•Sends an SMB Echo request to the targeted machine
•Sets up the exploit for the target architecture
•Attempts exploit
•If successful exploitation occurs.
•Pings the backdoor to get an SMB reply

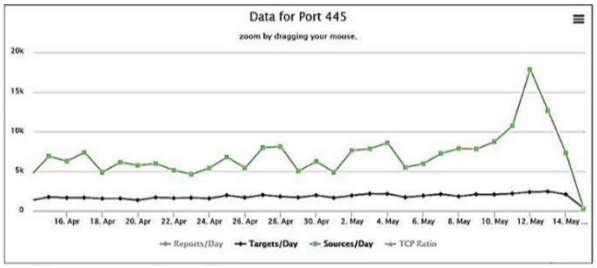


Fig 2: Port 445 Hits on Infected System [26]

B. DoublePulsar

WannaCry made use of DOUBLEPULSAR which is a persistent backdoor that is generally used to access and execute code on previously compromised systems. This allows for the installation and activation of additional software, such as malware. This backdoor is typically installed following successful exploitation of SMB vulnerabilities addressed as part of Microsoft Security Bulletin MS17-010[15]. This backdoor is associated with an offensive exploitation framework that was released as part of the Shadow Brokers cache that was recently released to the public. It runs in kernel mode, which grants cybercriminals a high level of control .These connections allow an attacker to establish a Ring 0 level connection via Server Message Block via TCP port 445[20].

V. TECHNICAL ANALYSIS OF THE ATTACK

A. Distribution of the WannaCry Ransomware

Once the malware is on a system, its worm capability will try to spread further through SMB. After initializing the functionality used by the worm, two threads are created. The first thread scans hosts on the LAN. The second thread gets created 128 times and scans hosts on the wider Internet. The scanning thread tries to connect to port 445, and if so creates a new thread to try to exploit the system using the ETERNALBLUE SMB vulnerability (MS17-010). If the exploitation attempts take over 10 minutes, then the exploitation thread is stopped [15].

B. Encryption

When a computer becomes infected with WanaDecrypt0r, the installer will extract an embedded file into the same folder that the installer is located in. This embedded resource is a password- protected zip folder that contains a variety of files that are used by and executed by WanaCrypt0r.
Following are the key encryption approaches used in WannaCry Ransomware:

1. Each infection generates a new RSA-2048 key pair.
2. The public key is exported as blob and saved to "00000000.pky".
3. The private key is encrypted with the ransomware public key and saved as "00000000.eky" [20].
4. Each file is encrypted using AES-128-CBC, with a unique AES key per file.
5. Each AES key is generated CryptGenRandom.The AES key is encrypted using the infection specific RSA key pair.
6. The RSA public key used to encrypt the infection specific RSA private key is embedded inside the DLL and owned by the ransomware authors [20].

C. Kill-Switch And Kill-Mutex

A kill switch is an event that is used to stop a program from continuing to execute. In the case of WannaCry, the kill switch is a domain name that the Worm component of WannaCry connects to when it starts. When the WannaCry worm was released on March 12th, the kill switch domain was set to:
www[.]iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea[.]com
The malware stops if it finds the following domain exists:
www[.]iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea[.]com

It is to be noted that organizations that use proxies will not benefit from the kill-switch, unless it is a transparent proxy. The malware is not proxy-aware, so it will not be able to connect to the kill-switch domain and thus the malware will not be stopped. The malware tries to create a Mutex named MsWinZonesCacheCounterMutexA [15]. If it exists already, the encryption phase will not be done.

VI. IMPACT OF THE ATTACK

Ransomware not only targets home users; businesses can also become infected with ransomware, leading to negative consequences, including:

• Temporary or permanent loss of sensitive or proprietary information.
• Disruption to regular operations.
•Financial losses incurred to restore systems and files.
•Potential harm to an organization's reputation.

Paying the ransom does not guarantee the encrypted files will be released; it only guarantees that the malicious actors receive the victim's money, and in some cases, their banking information. In addition, decrypting files does not mean the malware infection itself has been removed [15].

VII. DETECTION OF THE ATTACK BY ANTI-VIRUS

Microsoft Anti-Malware products detect the present version of this WannaCry ransomware as Ransom: Win32.WannaCrypt from definition version 1.243.291.0[15].Various anti-virus software detect the malware as:

- Ransom.Wannacry
- Ransom.CryptXXX
- Trojan.Gen.8!Cloud
- Trojan.Gen.2

VIII. RECOMMENDATIONS FOR NETWORK PROTECTION

- Apply the patch (MS17-010). If the patch cannot be applied, consider:
- Disabling SMBv1
- Blocking all versions of SMB at the network boundary by blocking TCP port 445
- Ensure anti-virus software is up-to-date.
- Implement a data back-up and recovery plan to maintain copies of sensitive or proprietary data in a separate and secure location.

CONCLUSION

So we found that Crypto-virology has offensive growth over years. We had observed a numerous propositions by security agencies like CERT to notify the existence of crypto-viral extortion attacks and their countermeasures. But its existence in today's developing modern world showing its dominance over past. We must see this as being disregarded fundamentally to security stimulus as it is the key demonstration of reactive security is prevailing over the proactive security. We believe Wanna Cry attack is the tip of the iceberg which is warning of a widespread attack which will securely steal information being unnoticed to the user. It seems as an indication of such bulk attacks, which are already illustrated in literature. These attacks are destined to be overlooked unless and until a large scale world attack will be publicized.

ACKNOWLEDGMENT

We are thankful to Shri Sant Gajanan Maharaj College Of Engineering, Shegaon, M.H. India and all our faculty members of Computer Science Department for their encouragement and whole hearted cooperation. I would like to thank Asst. Prof. V.S. Mahalle for his guidance and valuable suggestions time to time. We are lucky to have god like Parents, they provided me very excellent moral support and Friends for their encouragement during this work.

REFERENCES

[1] Young, A.; M. Yung (1996). Cryptovirology: extortion-based security threats and countermeasures. IEEE Symposium on Security and Privacy. pp. 129–140. ISBN 0-8186-7417-2. doi:10.1109/SECPRI.1996.502676.

[2] Jack Schofield (28 July 2016). "How can I remove a ransomware infection?". The Guardian. - 28 July 2016.

[3] Michael Mimoso (28 March 2016). "Petya Ransomware Master File Table Encryption". threatpost.com. -28 July 2016.

[4] Justin Luna "Mamba ransomware encrypts your hard drive, manipulates the boot process". Neowin. September 21, 2016

[5] https://www.us-cert.gov/ncas/alerts/TA17-132A

[6] Cryptovirology.com FAQ: http://www.cryptovirology.com/

[7] Fox-Brewster, Thomas. "Equation = NSA? Researchers Uncloak Huge 'American Cyber Arsenal'". Forbes.

[8] "Latest Shadow Brokers dump – owning SWIFT Alliance Access, Cisco and Windows". Medium. 14 April 2017. - 15 April 2017.

[9] The legacy code behind WannaCry – the skeleton in the closet". scademy.com/blog/ -18 April 2017.

[10] "Microsoft Security Bulletin MS17-010 – Critical". TechNet. Microsoft. 13 May 2017.

[11] "WannaCry Ransomware Attack Hits Victims With Microsoft SMB Exploit". eWeek. 13 May 2017.

[12] "WannaCry: BSI ruft Betroffene auf, Infektionen zu melden" (in German). heise online. 14 May 2017.

[13] "Microsoft issues 'highly unusual' Windows XP patch to prevent massive ransomware attack". The Verge. Vox Media.

[14] MSRC Team. "Customer Guidance for WannaCrypt attacks". Microsoft.

[15] The Wannacry Ransomware"-White Paper, The Computer Emergency Response Team Of Mauritius (Cert-Mu) May 2017

[16] Cameron, Dell "Today's Massive Ransomware Attack Was Mostly Preventable; Here's How To Avoid It". Gizmodo - 13 May 2017

[17] "How One Simple Trick Just Put Out That Huge Ransomware Fire". Forbes. 24 April 2017.

[18] "Player 3 Has Entered the Game: Say Hello to 'WannaCry'". blog.talosintelligence.com

[19] "Almost all WannaCry victims were running Windows 7". The Verge. Vox-Media.https://arstechnica.com/information-technology /2017/04/ 10000-windows-computers-may-be-infected-by-advanced-nsa-backdoor/

[20] "WannaCry|WannaDecrypt0r NSA-Cyberweapon-Powered Ransomware Worm" https://gist.github.com/rain-1/ 989428fa5504f378b993ee6efbc0b 168/

[21] Kanwalinderjit K Gagneja. "Knowing the Ransomware and Building Defense Against it – Specific to HealthCare Institutes, Availability, Reliability and Security (ARES), 2016 11th International Conference 31 Aug.-2 Sept. 2016

[22] Shivale Saurabh Anandrao," Cryptovirology: Virus Approach"

International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.4, July 2011

[23] "Wannacry"- Ransomware Attack - Technical Intelligence Analysis Report-EYGM May 2017

[24] Young, A. And Yung, M. Malicious Cryptography— Exposing Crypto-virology ,Wiley, 2004. (Book)

[25] Bleepingcomputer.Com

"https://www.bleepingcomputer.com/news/security/wana-decryptor-wanacrypt0r-technicalnose-dive/ "

[26] https://isc.sans.edu/port.html?port=445

Analysis Of Microbial Inactivation Performance On Ballast Water Treatment System Prototipe Using Combination Of Active Carbon And UV Radiaton

[1] P. Trika, [2] S. Maya, [3] F. H. Nur
[1] Department of Marine Engineering, Institut Teknologi Sepuluh Nopember, Surabaya-60111, Indonesia
[2] Department of Biology, Institut Teknologi Sepuluh Nopember, Surabaya-60111, Indonesia
[3] Department of Marine Engineering, Institut Teknologi Sepuluh Nopember, Surabaya-60111, Indonesia
[1] trika@its.ac.id, [2] maya@bio.its.ac.id, [3] harisnf50@gmail.com

Abstract:-- **This research is aimed to analyzed the ballast water treatment system prototipe performance in inactivating pathogenic microbes in ballast water. The concept that used to analyzed the prototipe performance is by performing ballast water treatment using ballast water treatment system prototipe and then carrying out the quantitative analysis test of the microbial content in samples that have been processed using two different methods. In the treatment stage ballast water was pumped with capacity of 5 lpm, 10 lpm and 20 lpm into active carbon filter and UV reactor. In the filtration process, ballast water was filtered by active carbon that loaded into a filter house that has diameter of 6.5 cm and depth of 20 cm . In the UV reactor ballast water was illuminated by UV-C lamp with maximum dose of 16,58 mW/cm2. After the treatment process has done, then the ballast water sample was analyzed using Total Plate Count (TPC) method with steril sea water diluent and TPC method with sterile aquades solution diluent. From these observations it found that when ballast water sampel was analyzed using TPC method with steril sea water diluent obtained inactivation result until 99,80%, while when it was analysed using TPC method with steril aquades solution diluent obtained inactivation result until 100%.**

Index Terms: ballast water treatment, active carbon, uv-c, total plate count

I. INTRODUCTION

Ballast water is very important to keep the safety operation of ship. Ballast water is used to control trim, draft, stability and tension on ships hull caused by adverse ocean conditions or as a result of changes in cargo weight [1]. Besides giving positive impact for ship, ballast water can make major threat for environment, publict health and economy. This problem is due to the spread of patogenic microbes form one region to another through ballast water. If the environmental condition where the patogenic microbes release is suitable for their living, these species will be survive and can reproduce sigficantly than become Invasive Alien Species (IAS) or Harmful Aquatic Organism and Pathogens (HAOP) that can make the extinction of native species, effects on local or regional biodiversity, effects on coastal industries that use water extraction, effects on public health and impacts on local economies based on fisheries [2].

To solve that problem, in 2004 International Maritime Organization (IMO) establish Ballast Water Management Convention that requires all ships in international traffic manage their ballast water and sediments to a certain standard according to a ship's specific ballast water management plan. Additionally, every ship must meet the performance standard of ballast water exchange which state that when ships conduct ballast water management shall discharge less than 10 living organisms of more than or equal to 50 micrometers per 1 m3 and for microorganisms that hold between 10 and 50 micrometers, only 10 microorganisms can be removed every 1 millimeter. As for the type of microbe, for vibrio cholerae must be less than 1 cfu per 100 ml, echerichia coli less than 250 cfu per 100 ml and intestinal entercocci less than 100 cfu.

To fulfill the requirements from IMO Ballast Water Convension several ballast water treatment methods were developed. Such as in [3] there was a research to understand the potential technology of UV radiation that used to inactivate microorganisms in ship's ballast water. In their study the UV dosage used for Chorella inactivation was given in the range of 0-540 mJ/cm2. From their research they got a result that state that the greater the UV dose given in water treatment the greater the rate of bacterial inactivation.That statement was demonstrated in the inactivation results which is when the amount of UV dose given in inactivation was 36 mJ/cm2 the bacterial inactivation reached 97.6%, whereas when UV doses increased up to 90 mJ/cm2 the bacterial inactivation increase to be 99.6%.

Besides using UV radiation, there was also the use of carbon materials in ballast water treatment especially for

water filter. Such as in [4] there was a research that focused on the analysis on the performance of activated carbon that used as a water filter media. In this study was used two pieces of carbon that have different granular sizes. The first carbon with the GAC-A code has a granular size of approximately 1.5 mm while the second carbon with the GAC-B code has granular size of approximately 2.3 mm. The pore distribution of the GAC-B carbon was randomly distributed, while in GAC-A was spread evenly. Both carbons were used to filter tap water that has turbidity level of 0.79 NTU and well water that have turbidity of 1.06 NTU. From the results obtained, both carbons can provide good treatment results in reducting the turbidity value. The GAC-A carbon filter resulted turbidity reduction value at tap water about 70.8% and well water about 68.86%, while the GAC-B carbon filter reduce turbidity values in tap water about 48.10% and well water about 66.03 %.

Referring to the positive results of the used of activated carbon and UV radiation as medium for treating ballast water, in [5] there was a research that continue the study on the use of these materials for ballast water treatment. In that studythe active carbon and UV were combined become a ballast water treatment system. In that study ballast water with capasity of 5 lpm, 10 lpm and 20 lpm was drining into active carbon filter and UV reactor. In filtration process, ballast water was filtered by active carbon that loaded into a filter house that has diameter of 6.5 cm and depth of 20 cm. In UV reactor ballast water was illuminated by UV-C with maximum dose of 16.58 mW/cm2. Ballast water that has been processed using prototype then analyzed using TPC method with sterile sea water diluent. The results obtained from this analysis show inactivation values that tend to be inconsistent. Based on [5], the inconsistency of the results of that analysis was hypothesized to be caused by several causes. The first hypothesis said that the inconsistency caused by a prototype performance factor that can not work optimally in providing treatment to seawater so there were still a lot of living microbes. The second hypothesis said that the inconsistency caused by the contamination of sea water microbes that came from the medium diluent that have not died even though the sea water has been sterilized before. The third hypothesis said that the inconsistency caused by microbes that grow significantly when implanted in agar medium because in agar medium they got an ideal environment that has salinity, ph, water content and other factor that support the microbial growth.

To prove the truth of these hypotheses in this study conducted ballast water treatment experiment using prototipe that has developed in [5] and also carried out quantitative analysis using TPC method with aquades solution diluent to avoid contamination from foreign microbes.

Finally in this study shown the comparison of the ballast water treatment system inactivation performance that obtain from the observation of microbial inactivation results using TPC method with sterile seawater diluent and TPC method with aquades solution diluent.

II. PROTOTIPE OF BALLAST WATER TREATMENT SYSTEM

A. Prototipe Developement

The prototipe that has been built in [5] is used active carbon and UV-C light as it's main component to treat ballast water. Active carbon in this prototipe is used as a filter. Active carbon is the derivative products of a carbon. Active carbon almost like a sponge and has many tiny microscopic pores that soak up water. All these little micropores create a huge inner-surface area. Activated carbon is produced specifically to achieve a very big internal surface (between 500 - 1500 m2/g) [6]. This big internal surface makes active carbon ideal for adsorption. Adsorption is a process where a solid is used for removing a soluble substance from the water [6]. Active carbon comes in two variations: Powder Activated Carbon (PAC) and Granular Activated Carbon (GAC). The GAC version is mostly used in water treatment [7]. The active carbon that used in this study has precision of filtering up to 10 microns.

UV-C in this ballast water treatment system is used as componen to inactivate patogenic microbes. It used to inactivate patogenic microbes because it's light can be absorbed by proteins, ribonucleic acid (RNA) and deoxyribonucleic acid (DNA), and can lead to cell mutations and/or cell death, therefore it is effective in inactivating pathogens [8]. UV-C lamp that used as a component of UV reactor has the same specifications as the UV-C lamp used in research that has been implemented by [5], which it has power of 30 watt with length of 57 cm and amounted to two pieces and both of these lamps mounted on a UV reactor that has a length of 53 cm and a diameter of 9.7 cm.

B. Prototipe Working Scheme And Design

The prototipe working scheme in this study is still the same as the prototipe working scheme that has been developed in [5]. Bellow is the prototipe working scheme of ballast water treatment

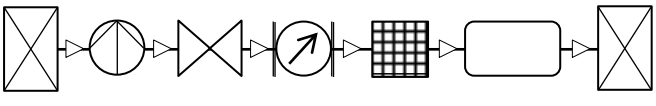


Fig.1 Ballast water treatment working scheme [5]

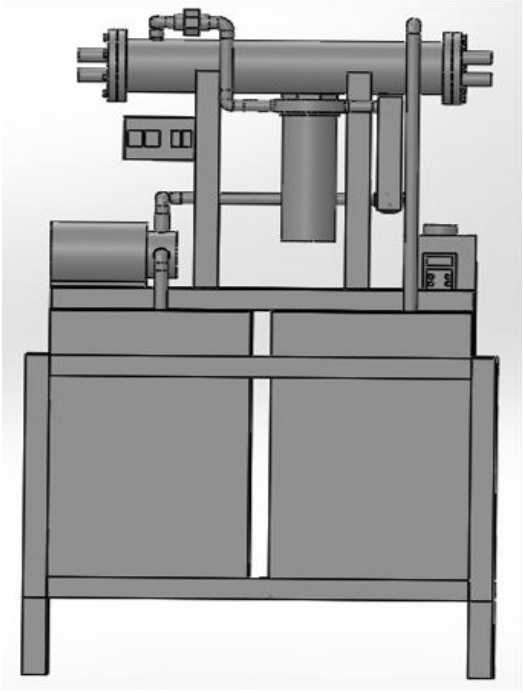


Fig. 2 Ballast water treatment design front view [5]

Where :
1 = existing water tank
2 = water pump
3 = regulating valve
4 = flow meter
5 = filter
6 = UV reactor
7 = processed water tanks
8 = voltage regulator

Based on the working diagram above, ballast water treatment prototipe consists of tank number one which serves as a reservoir for seawater in existing condition. Water in tank number one pumped into active carbon filter with predetermined capasity. The capasity of ballast water in the prototipe was regulated using regulating valve located on the discharge of the pump. When adjusting ballast water capasity with regulating valve also carried out monitoring on flow meter which located between regulating valve and flow meter. This monitoring was conducted to determine the amount of water capasity in the system whether it has been in accordance with the variation of the predetermined capasity. After getting out from flow meter, ballast water flown into filter to lead filtering process using crumb rubber and carbon filter. Than ballast water flown into UV reactor to undergo microbial inactivation process using UV-C lamp. The dose of UV-C light radiation on this reactor varied by regulate the voltage of UV-C lamp using an electrical regulator. Finally after getting out from UV reactor, ballast water accommodated in tank number two. From this tank was taken ballast water water samples to be analyze the amount of microbes that contained in it.

C. Inaktivation Eksperiment

In the inaktivation experiment, seawater pumped into ballast water treatment system prototype with capasity of 5 lpm, 10 lpm dan 20 lpm. In addition, the UV dose was also varied to determine the relatioanship between the amount of ballast water capasity and UV doses that required for pathogenic microbe inactivation in seawater. The UV dose that given for water treatment were 7.10 mW/cm2, 14.20 mW/cm2 dan 16.58 mW/cm2. In the last step carried out quantitative analysis of pathogenic microbial contents in ballast water samples to know the amount of microbial content in proessed ballast water sample

D. Quantitative Analysis Of Pathogenic Microbe

The quantitative analysis of pathogenic microbial contents in ballast water samples that have been processed using ballast water treatment prototipe was performed to determine the performance of ballast water treatment system prototype in inactivating pathogenic microbes. The quantitative analysis of microbial content was done by using Total Plate Count (TPC) method with steril aquades diluent. Steril aquades solution was used as Na diluent in TPC method to meet the protocol of bacteria growing medium on this method and also as a comparison of quantitative analysis of pathogenic microbes that have been done in [5] using TPC method with sterile seawater diluent.

There were several steps carried out in quantitative analysis using TPC method, such as the making of bacterial growth media, the making of dilution solution and the sterilization of the equipments to be used in testing and sterilization of the test. In the preparation of bacteria growing media, the medium that often used is the Plate Count Gel. This medium dissolved as much as 17.5 grams into 1000 ml of sterile aquades then heated until boiled so that the media dissolve perfectly in the aquades. During the boiling process was also done stirring to the media so that no sedimentation occurs. After the media has completely dissolved then the media was sterilized using autoclave for 15 minutes at a temperature of 121 degrees Celsius or pressure at 1.5 ATM. After the preparation of the growth medium has been completed, the next step was preparing the diluent solution. The diluent solutions that can be used include phosphate buffer solution and pepton water [9]. The last Preparation that performed before carrying out quantitative tests using TPC method was the sterilization of testing equipment.

After the preparation was ready, then conducted the quantitative analysis of pathogenic microbes with was initiated by inserting a 0.1 ml seawater sample into a reaction tube containing 9.9 ml of sterile distilled water

and homogenized with a vortex. This tube is then called dilution 10^{-2}. Then from dilution 10^{-2} was taken as much as 0.1 ml and transferred into a reaction tube containing 9.9 ml of sterile aquadest and homogenized again. The tube then referred to as 10^{-4} dilution. Dilution continues until dilution 10^{-8} is obtained. From 10^{-4} to 10^{-8} dilution were taken as much as 100 µL / 0.1 ml of sample solution using micropipette and transferred into empty petri dish then added nutrient agar (Na) liquid medium with temperature ranging from 48°C - 50°C. Samples and media were homogenised by rotating to form a pattern of eight (8). Each dilution was poured twice then incubated at 35°C for 24 hours. When the incubation time has reached 48 hours the bacteria contained in the cup will grow so it can be calculated against the number of bacteria. Calculation of the number of bacterial colonies in the cup using a standard called "Standard Plate Count". Based on this standard, the selected and calculated cup is the cup that contain colony amounts of from 25 to 250. The cup having the number of colonies in the range is then calculated using the formula $N = \frac{\sum C}{[(1 \times n_1)+(0{,}1 \times n_2) \times d]}$ [10], where

N : Number of colonies per ml or g of product
$\sum C$: Sum of all colonies on all planets counted
n_1 : Number of plates in first dilution counted
n_2 : Number of plates in second dilution counted
d : Dilution from which the first counts were obtained

III. RESULT AND DISCUSSION

A.Microbial Content In Sea Water Under Existing Conditions

The analysis of microbial content in seawater at existing condition was done in [5] using Total Plate Count (TPC) method with mendium bacteria grower in the form of Sodium agar (Na). In that analysis, water that used as sample was taken from Kenjeran sea Surabaya. After being observed by TPC method in Kenjeran sea water sample contain of 1.31 x 105 cfu of pathogenic microbes.

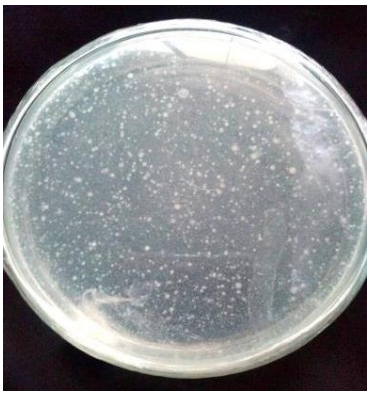


Fig.3 Microbes in Kenjeran sea water sample [5]

B.Quantitative Analysis of pathogenic microbe using TPC Method With Steril Aquades Diluent

From the observations of pathogenic microbes that have been done using TPC method with steril aquades diluent obtained results as in Table 1. The results of the ballast water treatment in Table 1 obtained results that all of the samples do not contain microbe. With this result can be conclude that the prototype can inactivate 100% of the microbes contained in sea water samples using minimum UV dose of 7.10 mW/cm2 and maximum water discharge of 20 lpm.

Table 1.
Table of microbial quantitative analysis using

Eksperiment number	Debit (lpm)	UV dose (mW/cm2)	Number of alive microbes (Cfu/ml)
1	5	7.10	0
2	10	7.10	0
3	20	7.10	0
4	5	14.20	0
5	10	14.20	0
6	20	14.20	0
7	5	16.58	0
8	10	16.58	0
9	20	16.58	0

C. Comparative study of the quantitative analysis results of pathogenic microbes using TPC method with sterile aquades solution diluent and TPC method with sterile sea water diluent.

This chapter describes the comparative study of the quantitative analysis results of pathogenic microbes using TPC method with sterile aquades solutions diluent with quantitative analysis results of pathogenic microbes using TPC method with sterile seawater diluent which was taken from the results of the research in [5].

The results of the microbial inactivation that obtained in [5] to be inconsistent. The inconsistency that inactivation can be seen from the comparison of UV dosage with the number of living microbes in medium growth as in Table 2 below.

Table 2.
Table of microbial quantitative analysis using tpc and sterile sea diluent [5]

Eksperiment number	Debit (lpm)	UV dose (mW/cm2)	Number of alive microbes A (Cfu/ml)
1	5	7.10	9.6×10^4
2	10	7.10	1.5×10^5
3	20	7.10	8.3×10^6
4	5	14.20	0
5	10	14.20	0
6	20	14.20	0
7	5	16.58	7.2×10^4
8	10	16.58	2.6×10^2
9	20	16.58	7.5×10^6

In the first three experiments in Table 2 which the sea water with UV dose of 7.10 mW/cm^2 obtained the greatest inactivation results in sample number one with the number of dead microbes are 3.5×10^4 Cfu or 26.71% of the microbes from water samples under the existing conditions. In the next three experiments the dose of UV radiation was added until become 14.20 mW/cm^2. With this UV dose ballast water treatment prototipe results 100% inactivation of the amount of microbes from water samples under the existing conditions. In the last three experiments the UV dose was increased to 16.58 mW/cm^2. With the addition of UV doses this should be increase the performance of inactivation microbe, but the results obtained in this experiment showed the opposite results, which with the addition of UV dose makes the inactivation performance decrease and only result the most significant inactivation of 1.3×10^5 Cfu or 99.80%.

The inconsistency of the quantitative analysis result obtained in the research that has been done in [5] was hypothesized to be caused by several causes that have been explained in introduction section. Based on the results obtained in this study that shown in Table 1 it can be seen that there is a significant difference with the results obtained in [5] that shown in Table 2. The inactivation results that obtained in this study are better and more consistent than the results obtained in penelirtian [5]. The factors that are believed to be the cause of inaccurate inactivation results are the hypotheses generated in [5] as explained in introduction

From the hypotheses given in [5], the hypothesis stating that inconsistency was caused by the contamination of the microbial sea water derived from the medium diluent that has not died even after it has been sterilized is the most likely to happen. This hypothesis is considered the possibility because based on the results of quantitative analysis that has been done by using TPC method with aquades solution that results consistent inactivation value, whereas when tested by using TPC method with sterile seawater diluent conduct inconsistent results or the result of the inactivation is not linear with the large of the water capasity and UV doses that given, which it may be caused by contamination from microbes derived from seawater that used as diluents. In addition, the hypothesis of microbial contamination from seawater was supported by the third hypothesis which states that the medium is an ideal environment for microbial growth because it has salinity, ph, water content and other supporting factors that support microbial growth. With this ideal growth medium, it able to make microbes that contaminate the grower medium will

multiply significantly so that will make inaccurate results in microbial quantitative analysis from comparative study that has been done on the results of water treatment research by using a combination of carbon filter and UV radiation that analyzed by TPC method with sterile sea water diluent and TPC method with sterile aquades solution obtained conclusion that the value of inactivation from TPC with aquades diluent has better accuracy than the value obtained from quantitative analysis using TPC with sea water diluent. Because of that, the inactivation result from TPC method with aquades solution diluent is used to represent the ballast water treatment system prototype inactivation performance.

REFERENCES

[1] International Maritime Organisation, "International Converance On Ballast Water Management For Ship," International Maritime Organisation, London, Converance BWM/CONF/36 , 2004.

[2] European Maritime Safety Agency. (2013) European Maritime Safety Agency. [Online]. http://www.emsa.europa.eu

[3] Z.J. Ren et al., "Microorganism Removal from Ballast Water using UV Irradiation," Journal of Residuals Science & Technology, pp. 13(1), 31-35, 2016.

[4] Y. K. Siong, J. Idris, and M. M. Atabaki, "Performance of Activated Carbon in Water Filters," ResearchGate, January 2013.

[5] P. Trika, S. Maya, and F. N. Haris, "Preliminary Study Of The Developement Of Ballast Water Treatment Using Filtration And UV Radiation Method," International Journal of Marine Engineering Innovation and Research, no. to be published, 2017.

[6] Science Buddies. (2015, December) Scientific American. [Online]. https://www.scientificamerican.com

[7] Eddy and Metcalf, Wastewater Engineering: Treatment, Disposal, and Reuse, 3rd ed. Singapore: McGraw-Hill Inc, 1991.

[8] J. R. Bolton, Ultraviolet application handbook. Ontario, Canada: Bolton Photosciences Inc, 1999.

[9] N. M. Dian. (2017, April) Info Laborat. [Online]. http://www.infolaborat.com

[10] M. Larry and P. T James. (2001) Bacteriological Analytical Manual: Aerobic Plate Count. [Online]. https://www.fda.gov

[11] theguardian. (2011, February) theguardian. [Online]. https://www.theguardian.com

Development of Real-Time 2.5 Dimensional Topographical Map for AR System of Excavators

[1] Joo Hwan Park, [2] Do Hyoung Shin (Corresponding Author)
[1] [2] Department of Civil Engineering, Inha University
[1] pjh5208@gmail.com, [2] dshin@inha.ac.kr

***Abstract*:-- The purpose of the present study is to develop a real-time 2.5-dimensional topographical map based on the AR-Navigator system being developed by the authors, so as to provide the visual information required by excavator operators. The development of a real-time 2.5-dimensional topographical map is expected to lay the foundations toward automated construction equipment capable of autonomous operation.**

Index Terms: Earthwork, Excavator, 2.5Dimensional Map, Construction, Automation

I. INTRODUCTION

Numerous efforts are being made from various approaches in the construction industry to enhance efficiency, and in the case of construction equipment, visual information maps have been studied as a means of providing useful information to operators to increase efficiency [1]. Based on this study, the authors are currently developing an AR-Navigator that Augmented Reality (AR)-based guide system for excavators.

For the realization of the AR-Navigator system, a 2.5D topographical map that updates variations in ground elevation with excavation progression is necessary. Therefore the purpose of the present study is to develop the real-time 2.5D topographical map update system required by AR-Navigator. The following description of the technology at hand assumes an excavator with a bucket width of 1m.

II. AR – NAVIGATOR

AR-Navigator is a system that provides visual information to excavator operators through a tablet PC so as to improve the efficiency of work. One of the key features of AR-Navigator, shown in Fig. 1, is the showing of the current ground elevation with green color and as-designed ground elevation with purple color for the end of the bucket with visual AR image, so as to facilitate excavation. To this end, a 2.5D topographical map that updates the current ground elevation for the area being excavated is necessary.

The real-time 2.5D topographical map update system developed for this purpose is a key system comprising AR-navigator, and expresses topographical changes in earthwork sites in real-time, conveying changes in the ground level of the work area to the excavator operator. This is a key feature of AR-Navigator. As for data necessary for use of the real-time 2.5D topographical map update system, these are the as-designed ground level, coordinates of the edge of the bucket, and an initial topographical map produced through a topographical scan. The initial topographical map is produced in point cloud data form by scanning, and is expressed in a data form suitable for the present system through the refining process shown in III.A of the present study. The as-designed ground elevation is obtained from the engineering drawings, while the position of the edge of the bucket is determined using GPS and an inclinometer [2].

ACKNOWLEDGMENT

This research was supported by a grant(17SCIP-B079689-04) from Construction Technology Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

Figure 1. Edge of bucket, as-designed ground elevation, and current ground elevation shown on AR-Navigator

III. REAL-TIME 2.5D TOPOGRAPHICAL MAP

The real-time 2.5D topographical map requires an as-scanned topographical map that shows the ground elevation prior to excavation and which is obtained using a laser scanner, an as-designed topographical map, and real-time excavator bucket coordinates (XB, YB, ZB).

A.2.5D topographical map generation

To generate the initial ground levels for the real-time 2.5D topographical map system, point cloud type as-scanned data from a laser scanner is used. However, use of a point cloud type topographical map in the present system without refinement poses the risk of slowing topographical map updates due to the excessive computation load caused by hundreds of millions of point data.

To resolve this potential issue, the point cloud type topographical map needs to be transformed into mesh data by assigning Z coordinates using a grid comprising X and Y planes, reducing system load.

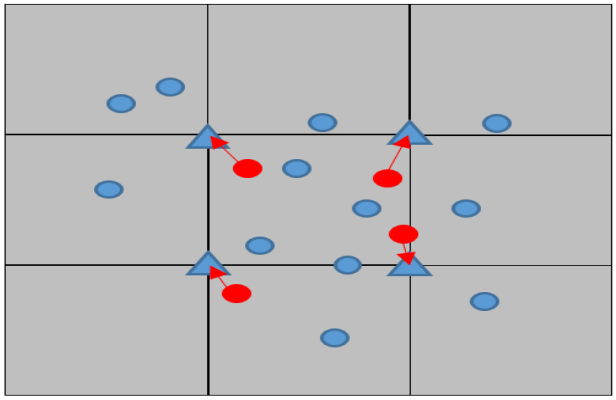


Figure 2. Dense Point Cloud Assignment

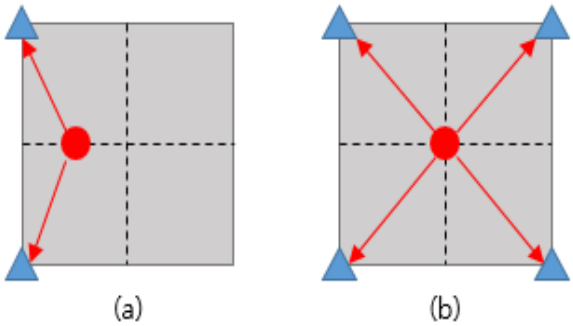


Figure 3 Sparse Point Cloud Assignment

That is, this process can be said to be a process where, due to storage capacity limitations, a high-resolution image is converted into a low-resolution image. The intersections on the grid used to convert point cloud type data to mesh data are referred to as grid points, and the point cloud points are assigned to these grid points. The X and Y coordinates indicate the modified position of the excavator due to turning or movement.

However, simple conversion of point cloud data topographical map data into mesh data to reduce system load may cause a loss in accuracy. The intervals of the grid can be likened to the resolution of an image, as explained in the foregoing, this resolution for grid must be able to guarantee a certain degree of accuracy for the present system. Here, the resolution for X and Y in grid is 0.2m, this resolution having been decided through a trade-off between accuracy and the specifications of the system used for AR-Navigator..

Assigning of mass point cloud data to mesh data involves the entering of Z coordinate values into a grid expressed using only X and Y axes. Here, there are two methods of entering point cloud Z coordinate values into a grid point, depending on whether the point cloud is dense or sparse. In dense areas, when the point cloud is placed to overlap the grid is set as the grid point value, the Z_P value which is the point cloud data closest to a grid point is assigned to each grid point as Z_G, as shown in Fig. 2. The points indicated using circles show the point cloud coordinates (P_p, Y_P, Z_P), while the intersecting points of the squares, including the triangles, indicate the grid point coordinate values (X_G, Y_G, Z_G).

As for areas where the point cloud is sparse, an insufficient number of points causes the situation shown in Fig. 3. If, as in Cases (a) and (b), a point equidistant from two or more adjacent grid points, indicated with triangles, to be assigned the same value. In this manner, a point cloud data type topographical map may be converted into mesh data and used for real-time topographical map updates.

However, prior to converting the point cloud data into mesh data, there may be noise caused during the scanning process by trees, workers, and construction equipment, etc. on the earthwork site, as shown in Fig. 4.

Figure 4. Noise in Scanned Point Cloud

As can be seen in Fig. 4, noise appears at a certain vertical distance from the ground. Unless the earthwork with explosions, the ground excavated is for the most part comprised of sandy soil and clay soil. Sandy soil, where

the particles are weakly bound together, is expressed in the point cloud as a group of consecutive points with little variation among adjacent points, and it is judged that there can be no substantial variation in Z coordinates between adjacent points at any given location. As for clay soil, whereas the particles are more strongly bound than in sandy soil, it is judged that high variation among a group of consecutive points on the point cloud is very low in validity.

Accordingly, given a group of consecutive points, it is assumed that large variations of 0.5m or greater are outliers caused by construction equipment or workers. To remove such outliers, the Z_p value of a point cloud data point other than the outlier is assigned to the corresponding grid point.

It is judged that such method substantially reduces system load and contributes to enhancing computation speed. The generated 2.5D topographical map is expressed to the excavator operator on a plane comprising X and Y axes, with the Z axis values expressed using color variations.

B. Updating the topographical map

Using the mesh data type topographical map made by assigning point clouds to a grid, the changes in topography caused by earthwork are updated in real-time. Here, an update involves the updating of a Z_G coordinate corresponding to X and Y coordinates on the grid to the Z_B coordinate representing a point that the edge of the excavator bucket passes by. The positions where the coordinates (X, Y) need to be updated during excavation are found based on the movement of the edge of the bucket, as shown in Fig. 5. Also, the shape of the bucket, shown in Fig. 6 with a bucket tooth width of 0.1m, was given consideration. Therefore the movement of the bucket edge was expressed using a bucket edge area, which is defined as the bucket width (1m) times the bucket tooth width (0.1m). At a given point in time, the Z_B coordinates for all points within the bucket edge area are the same. In the case that, for a X_P, Y_P coordinate set within the area, Z_B is less than Z_G, the Z_G of the grid point is updated to Z_B. As excavation work in general is carried out in linear fashion without turning, non-linear movement was not given consideration.

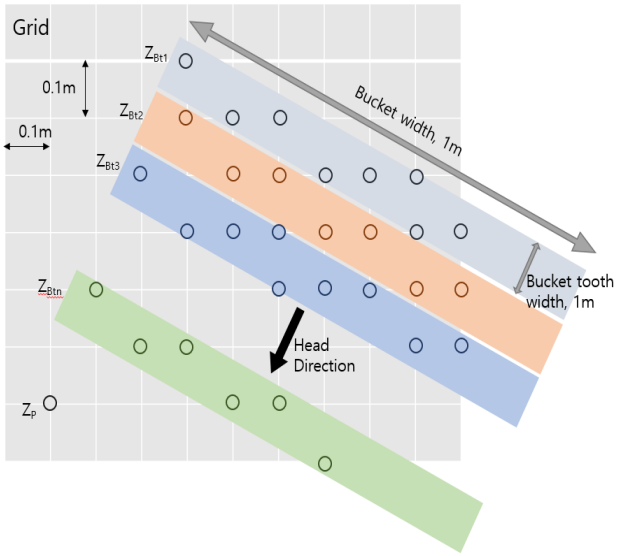


Figure 5. X, Y Coordinate Detection with Bucket Movement

Figure 6. Area of Bucket on Grid

The algorithm for the real-time topographical map update system, which is updated depending on the movement of the bucket edge during excavation, is as follows.

1) Receive the real-time position of the bucket from the GPS and inclinometer in the form of X_B, Y_B, and Z_B coordinates.

2) Find X_G, Y_G on the path of movement of the bucket edge area during excavation.

3) To determine whether or not to update Z_G, compare the Z_G coordinate value for the X_G, Y_G position against the Z_B coordinate value.

4) If Z_B is smaller, update Z_G with the Z_B value.

Real-time updating is defined as using the above-state algorithm, it is judged whether or not to update the Z_G

value of the grid point, and carrying out updating accordingly.

In AR-Navigator, the as-is ground elevation can be determined in real-time through such computations, and it is possible to compare the position of the bucket edge against the as-scanned topographical map to calculate differences in elevation.

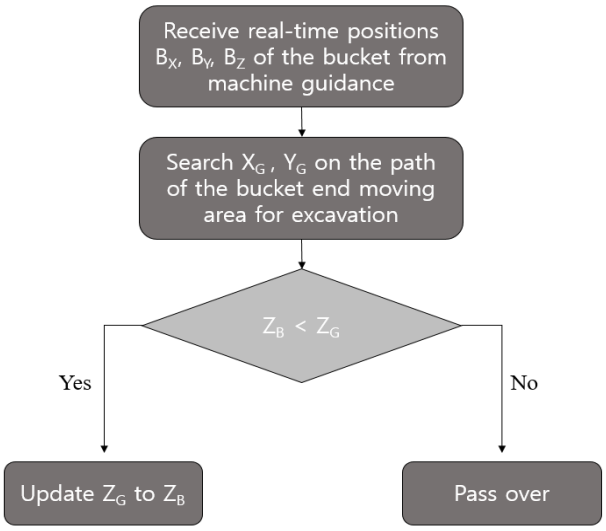


Figure 7. Algorithm for 2.5Dimensional Map Update

VI. CONCLUSION

In the present study, a real-time 2.5D topographical map required for AR-Navigator was developed. Tests as to the usability and accuracy of this real-time 2.5D topographical map update system in an actual construction site are being prepared, and points of improvement will be identified following such tests to further improve the present system. AR-Navigator with real-time 2.5D topographical map provides bucket edge, as-is ground elevation and as-designed ground elevation information in the form of an AR image, and is expected to improve excavator work efficiency. There is sufficient room for further improvement of the present system through future improvements to GPS system accuracy, and the present study is expected to contribute to the automation of construction equipment as well as the mechanical automated control of excavators.

REFERENCES

[1] S. Song, H. Kang, H. Kim, T. Moon and D. Shin, "Task-Visual Information Map to Develop AR Navigators of Construction Equipment." Korean Journal of Construction Engineering and Management, Korea, vol. 17, no.3, pp. 116–124, May 2016.

[2] Lee. K., Park. J., Kang. H. and D. Shin, "Seamless Superimposition Technique of Virtual Objects for AR System of Excavator Based on Image Processing" Korean Journal of Construction Engineering and Management, Korea, vol. 18, no.2, pp. 21–29, March 2017.

Development of Sambiloto (Andrographis paniculata) Leaves as A Source of Eco Accumulator to Help Combating Forest Degradation

Mahtuf Ikhsan
Department of Forest Management, Faculty of Forestry, Bogor Agricultural University, Indonesia
mahtuf123@gmail.com

Abstract:-- **The problems now faced by Bukit Suligi Educational Forest are approximately 70 % of FTC (Forest Training Facility) has been burned and palm oil is illegally planted by the surrounding community. From burned area, rehabilitation is the real choice to preserve Bukit Suligi Educational Forest. According to the problem, a study needs to be done to prove that there are trees in the forest can produce an alternative energy for the community. This is what will change the perception of people who think the value of forest is lower than palm oil plantation. Utilization of sambiloto (Andrographis paniculata) leaves as one of alternative sources that can generate electricity based on the characteristics of sambiloto leaves produce a bitter taste in the tongue entirely. This research is done by getting sambiloto leaves in Bukit Suligi Educational Forest and making eco accumulator model from sambiloto leaves extract in Forestry Vocational School of Pekanbaru. Based on the results, 720 ml eco accumulator model can generate electricity of 3,87 volts. Therefore, sambiloto leaves extract is potential for a new alternative electrical energy. From this research, it can be expected that community around the forest will take an initiative to restore degraded forests with sambiloto plant.**

Index Terms: sambiloto leaves, green battery, rehabilitation.

I. INTRODUCTION

Indonesia's forest wealth is one of the greatest biodiversity content in the world. Its diversity is spread across Indonesia, especially in Kalimantan and Sumatera. Various types of plants have vast benefits that can be used as foods, traditional medicine and renewable energy alternative that have not been optimized as a whole.

Sambiloto plants are often found in forests and seaside where all parts of this plant has its own benefits and content ranging roots, stems, fruit and leaves. One of the lesser known pieces of benefit and usefulness is the sambiloto leaves, which is often wasted in a useless manner without meaningful utilization that has more value and extraordinary benefits. One of the benefits that should be developed besides the utilization of sambiloto leaves as an air pollutant absorber is sambiloto leaves can serve as a source of electrical energy in Indonesia. Utilization of extract from sambiloto leaves can be used as renewable energy to overcome the increasing electricity demand in Indonesia.

The supply of community electricity based on Micro Hydro Power Plant (PLTMH) resources can be set in 200 W per house connection (220 V, 1A). The availability of electrical energy from State Electricity Company (PLN) installed capacity, amounting to 72.85% of energy generated from fossil fuels. The main allocation of rural household using generally is for night time lighting, with consideration during the day most people work. New potential local energy sources can be realized into power plants, among others : solar, wind, and microhydro energy. New and renewable energy especially local potential, needs to be studied and utilized as primary energy source for electricity generation (Widodo 2012).

Based on chemical properties, mahogany leaves, which are alkaline with a bitter taste along with other solutions such as metal lead (Pb), can be utilized as an electrolyte solution in battery cells. Accumulator with strong electrolyte solution is not environmentally friendly and also dangerous. Therefore, it is of interest in reviewing and applying the utilization of wasted sambiloto leaves as an environmentally friendly electrolyte solution to the batteries. It is also a learning experience for inland communities, which are not yet reached by electricity, to utilize sambiloto leaves extract as engineered from eco accumulator.

The purposes of this study were: (1) to examine the contents in mahogany leaves as a source of alternative electrical energy, (2) to understand the process of utilization of electrical energy from mahogany leaves as eco accumulator, and (3) to recognize the role of eco accumulators in supporting the supply of electricity in Indonesia.

II. METHODS

In carrying out this study, several methods were implemented, those are: literature study by reading books or scientific journals and articles that are related to forest rehabilitation, sambiloto plant, and eco accumulator. Primary data retrieval is done by doing an experiment to make eco accumulator model from sambiloto leaves extract. This experiment is done by getting sambiloto leaves in Bukit Suligi Educational Forest and making eco accumulator model from sambiloto leaves extract in Forestry Vocational School of Pekanbaru on December 20th – 26th 2016. After that, primary data will be used as a comparison with literatures (textbooks, journals or other literature sources).

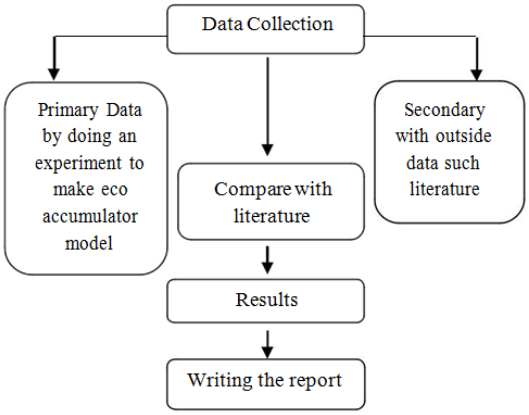


Figure 1. Methodology diagram

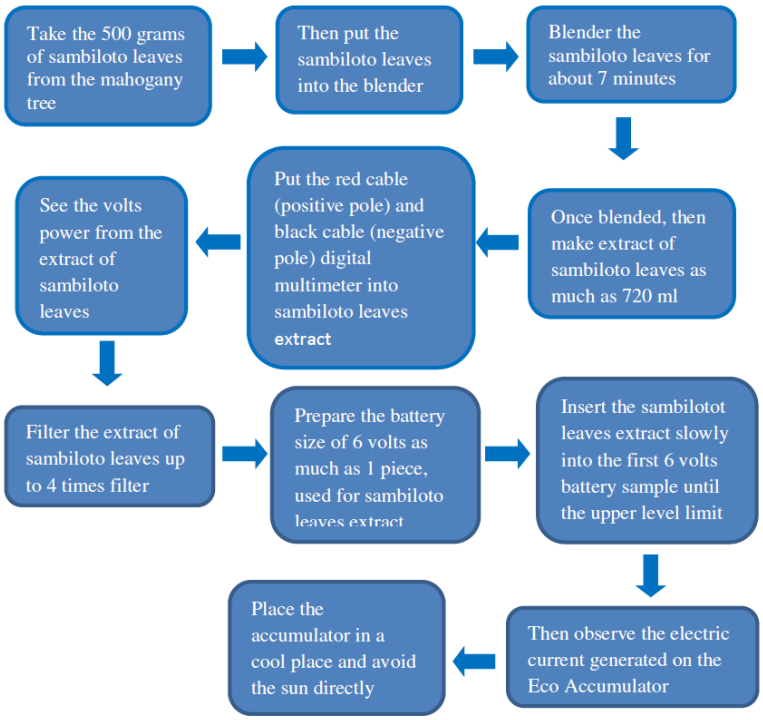


Figure 2. Research flowchart

 ISBN: 978-81-932966-9-1

III. RESULTS AND DISCUSSION

From the research, we tested the electric power generated from sambiloto leaves extract of 720 ml by using a multimeter tester tool. Here is the table of research results

Table 1. Research results

Condition of sambiloto leaves extract	Electrical power
Before entering into the accumulator	1.30 volts
Observation after 5 minutes	3.77 volts
Observation after 10 minutes	3.80 volts
Observation after 15 minutes	3.80 volts
Observation after 20 minutes	3.87 volts

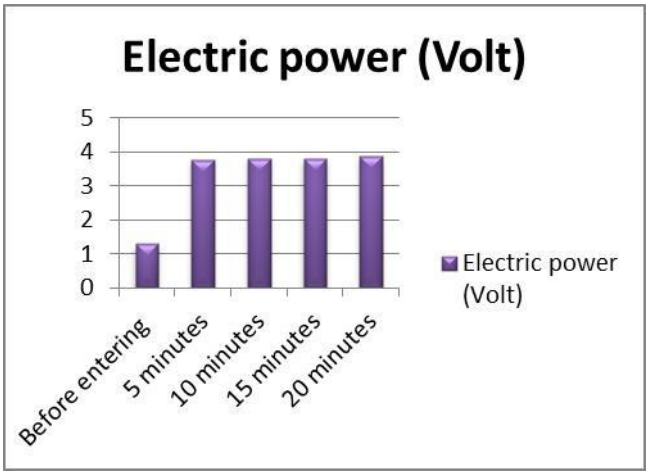
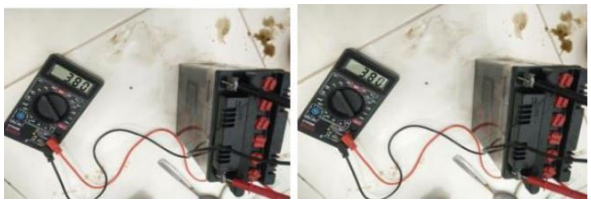


Figure 3. Research result diagram

(a) (b)



(c) (d)

(e)

Figure 4. Electricity experiment : (a) before entering into the accumulator, (b) after 5 minutes, (c) after 10 minutes, (d) after 15 minutes, (e) after 20 minutes.

Sambiloto leaves extract contained electrical conductivity of 1.3 volts before it was inserted into the accumulator. After it had been inserted into the accumulator of 6 volts for 20 min, the electrical conductivity increased to 3.87 volts. This was because when the leaves extract has not been inserted into the accumulator or still in the jar, electric power generated only 1.3 volts as it used only 3 cells. When the leaves extract was inserted into a 6-volt accumulator, electric power reached 3.87 volts which was equivalent to two 1.5 volts dry batteries. Since the accumulator had 6 pieces of cells the electricity generated was two times when it was not in the input to the accumulator.

The electric current in sambiloto leaves extract was caused by the fact that sambiloto leaves contained lead (Pb) heavy metal which was absorbed by the leaves from vehicle fumes and air pollution resulting from residual combustion. Lead (Pb) has toxic and carcinogenic properties. The lead content (Pb) inside the mahogany leaves is 40.28 ppm and Lead (Pb) is commonly used as the poles on the battery (Sedi 2014).

In a positive polar, accumulator using lead peroxide and negative pole using lead plate. When accumulator is used, chemical reactions caused sediment in negative (reduction) and positive electrode (oxidation). This could be the reason why sambiloto leaves extract can deliver electric current, caused by lead content (Pb) in the content. This suggests that the extract of sambiloto leaves can be used as an accumulator filler and a substitute of H2SO4 solution. Accumulator containing sambiloto leaves extract can conduct electricity well to accumulator with 6 volt capacity.

Therefore, the extract of sambiloto leaves can be utilized as a new alternative energy source to be used as accumulator filler solution. From this research, it is hoped that local community of Bukit Suligi Educational Forest is aware of the benefits of sambiloto leaves so that they will not carry

out illegal logging which can cause severe damage to the forest. There is a potential benefit to forest communities both in the economic and forest conservation aspect in order to support better environmental management and sustainable energy development.

IV. CONCLUSION

Sambiloto plant in Bukit Suligi Educational Forest is a potential source of green energy but this species has started to decrease due to illegal logging. Therefore, it is hoped that utilization of sambiloto leaves as an alternative source of electrical energy will make people aware of the benefits of this tree. Further research is however needed on the ability of electric conductivity of sambiloto leaves extract and the factors that influence the conductivity of sambiloto leaves extract so that it can produce an optimal energy.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Bukit Suligi Educational Forest Office Station, Forestry Vocational School of Pekanbaru, and also for Department of Forest Management, Faculty of Forestry, Bogor Agricultural University, Indonesia for supporting facilities and funding for this research.

REFERENCES

[1] Bukit Suligi Educational Forest. 2015. The Information of Bukit Suligi Forest Area. Pekanbaru (ID) : Center of Forestry and Environment Training Pekanbaru.

[2] Juwito AF, Haryono T. 2013. Optimization of Renewable Energy in Generating

[3] Electrical Energy Toward Energy Independent Village in Margajaya Village. JNTETI. 2 (3) : 40- 45.

[4] Latif M, Nazir R, Reza H. 2013. Analyzing the accumulator charging process on a horizontal axis wind turbine prototype in Purus Beach Padang. Jurnal Nasional Teknik Elektro. 2 (1) : 1 - 8.

[5] Widodo PS. 2012. Power Plant with Local Energy Potential as a Manifestation of Electricity Distribution in Disadvantaged and Remote Villages. Vokasi. 8 (3) : 151 – 164.

Development of Mahogany Leaf (Swietenia macrophylla) as A Source of Green Accumulator to Help Rehabilitation in Bukit Suligi Educational Forest

Mahtuf Ikhsan
Faculty of Forestry, Bogor Agricultural University, Indonesia
mahtuf123@gmail.com

Abstract:-- **The problems now faced by Bukit Suligi Educational Forest are approximately 70 % FTC (Forest Training Facility) has been burned and palm oil is illegally planted by the surrounding community. From burned area, rehabilitation is the real choice to preserve Bukit Suligi Educational Forest. According to the problem, a study needs to be done to prove that there are trees in the forest can produce an alternative energy for the community. This is what will change the perception of people who think the value of forest is lower than palm oil plantation. Utilization of mahogany leaf (Swietenia macrophylla) as one of alternative sources that can generate electricity based on the characteristics of mahogany leaf produces a bitter taste in the tongue entirely. This research is done by getting mahogany leaf in Bukit Suligi Educational Forest and making green battery model from mahogany leaf extract in Forestry Vocational School of Pekanbaru. Based on the results, 720 ml green battery model can generate electricity of 3,87 volts. Therefore, mahogany leaf extract is potential for a new alternative electrical energy. From this research, it can be expected that community around the forest will take an initiative to restore degraded forests with mahogany plant.**

Index Terms: mahogany leaf, green battery, rehabilitation

I. INTRODUCTION

Indonesia's forest wealth is one of the greatest biodiversity content in the world. Its diversity spreads across Indonesia, especially in Kalimantan and Sumatera. Various types of plants have benefits that can be used as foods, traditional medicine and renewable energy alternative that have not been optimized as a whole.

Mahogany trees are often found in forests and seaside where all parts of this plant has its own benefit and content ranging from roots, stems, fruit and leaves. One of the lesser known pieces of benefit and usefulness is the mahogany leaf, which is often wasted in a useless manner without meaningful utilization that has more value and extraordinary benefits. One of the benefits that should be developed besides the utilization of mahogany leaves as an air pollutant absorber is mahogany leaves can serve as a source of electrical energy in Indonesia. Utilization of extract from mahogany leaves can be used as renewable energy to overcome the increasing electricity demand in Indonesia.

The supply of community electricity based on PLTMH resources can be set in 200 W per house connection (220 V, 1A). The availability of electrical energy from PLN installed capacity, amounting to 72.85 % of energy generated from fossil fuels. The main allocation of rural household using generally is for night time lighting, with consideration during the day which most people work. New local potential energy sources can be realized into power plants, among others : solar, wind, and microhydro energy. EBT (New and Renewable Energy), especially from local potential, needs to be studied and utilized as primary energy source for electricity generation (Widodo 2012).

Based on the study of chemistry, it can be seen that mahogany leaves which is alkaline with a bitter leaf state along with other supporting solutions such as metal lead (Pb) can be utilized as electrolyte solution on the battery cell. Accumulator with strong electrolyte solution is not environmentally friendly and also dangerous. Therefore, the author is interested in reviewing and applying the utilization of wasted mahogany leaves as an environmental friendly electrolyte solution to the batteries as well as learning for inland communities which is not yet reached by electricity in order to utilize mahogany leaf extract as engineered from green accumulators. The purposes of this study are to know the content in mahogany leaves as a source of alternative electrical energy, to know the process of utilization of electrical energy from mahogany leaves to be used as green accumulator, and to know the role of

green accumulators in supporting the supply of electricity in Indonesia.

II. METHODS

In carrying out this study, several methods were implemented, those are: literature study by reading books or scientific journals and articles that are related to forest rehabilitation, mahogany plant, and green battery. Primary data retrieval is done by doing an experiment to make green battery model from mahogany leaf extract. This experiment is done by getting mahogany leaf in Bukit Suligi Educational Forest and making green battery model from mahogany leaf extract in Forestry Vocational School of Pekanbaru on December 20th – 26th 2016. After that, primary data will be used as a comparison with literatures (textbooks, journals or other literature sources). The last stage is writing the paper.

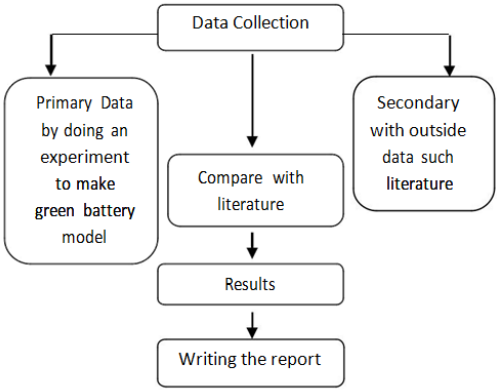


Figure 1. Methodology diagram

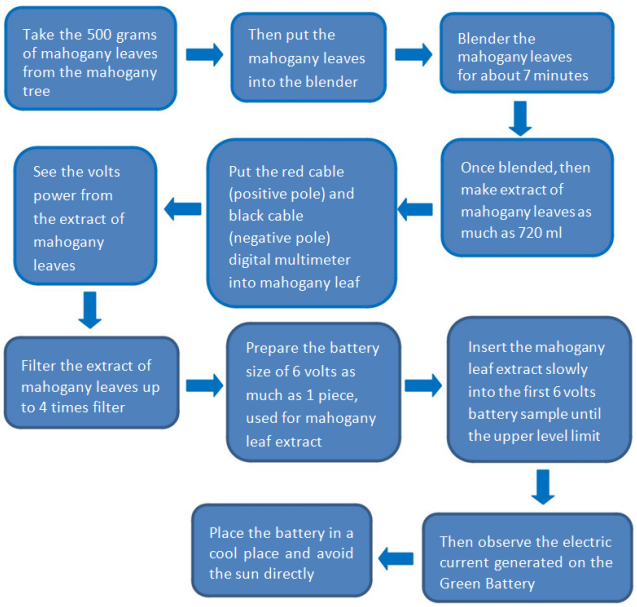


Figure 2. Research flowchart

III. RESULTS AND DISCUSSION

From the research, we tested the electric power generated from mahogany leaf extract of 720 ml by using a multimeter tester tool. Here is the table of research results.

Table 1. Research results

Condition of mahogany leaf extract	Electrical power
Before entering into the accumulator	1.30 volts
Observation after 5 minutes	3.77 volts
Observation after 10 minutes	3.80 volts
Observation after 15 minutes	3.80 volts
Observation after 20 minutes	3.87 volts

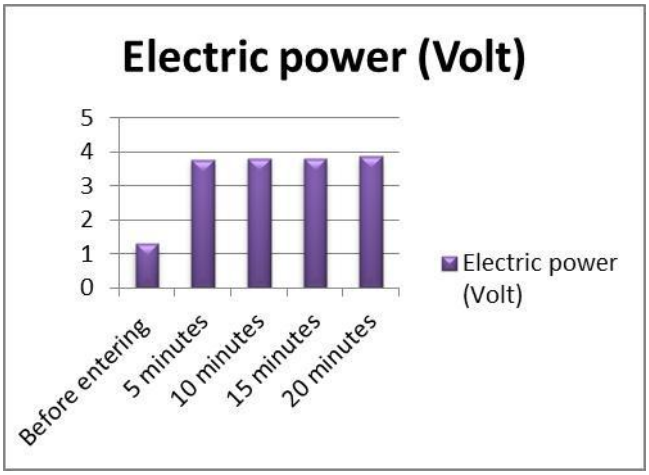
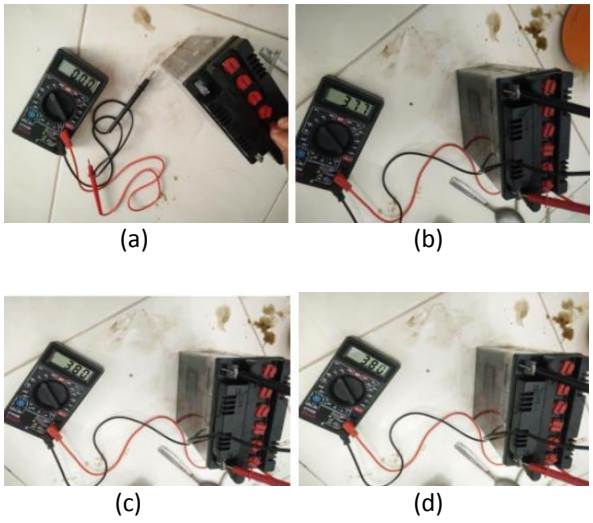


Figure 3. Research result diagram



 ISBN: 978-81-932966-9-1

(e)

Figure 4. Electricity experiment : (a) before entering into the accumulator, (b) after 5 minutes, (c) after 10 minutes,(d) after 15 minutes, (e) after 20 minutes.

Samples of mahogany leaves taken in Bukit Suligi Educational Forest as much as 500 g, then blend it for 7 minutes. This is intended to extract the mahogany leaf soluble in water. This process produces 720 ml of mahogany leaf extract, then test its electrical conductivity using multimeter tester by inserting 720 ml of mahogany leaf extract into the jar. The next step is to insert the copper plate as positive pole and zinc as negative pole, connecting positive pole to red cable on multimeter tester and negative pole with black wire. Read the readable scale and it turns out the readable scale is 1.30 volts or almost equivalent to one large dry battery size of 1.5 volts. This process is done to determine the ability of mahogany leaf extract to deliver electric current without input into the accumulator. The next process is to enter the extract of mahogany leaves into the accumulator, before the first input extract of mahogany leaves filtered as much as 4 times for the leaves of mahogany leaves do not go into the accumulator. After input, electrical conductivity test using multimeter tester showed 3.87 volts in 20 minutes observation.

Table 1. showed that the accumulator contains mahogany leaf extract before it is inserted into the accumulator contained electrical conductivity of 1.3 volts. Once it inserted into the accumulator of 6 volts in 20 minutes observation, then the accumulator electrical conductivity increased to 3.87 volts. This is because when the extract of mahogany leaves has not been inserted into the accumulator or still in the jar, electric power generated only 1.3 volts because it uses only 3 cells while when inserted into a 6 volt accumulator, electric power can reach 3.87 volts or equivalent to two of 1.5 volts dry batteries. This because accumulator has 6 pieces of cells so that the electricity generated can be 2 times when it is not in the input to the accumulator.

The electric current in mahogany leaf extract is caused by mahogany leaves containing lead (Pb) heavy metals which are absorbed by mahogany leaves from vehicle fumes and air pollution resulting from residual combustion. Lead (Pb) has toxic and carcinogenic properties. The lead content (Pb) inside the mahogany leaves is 40.28 ppm and Lead (Pb) is commonly used as the poles on the battery (Sedi 2014).

In a positive polar, accumulator using lead peroxide and negative pole using lead plate, when accumulator is used chemical reactions that cause sediment in negative electrode (reduction) and positive electrode (oxidation). This is the reason why mahogany leaf extract can deliver electric current, caused by lead content (Pb) in the content. This shows that the extract of mahogany leaves is very well used as an accumulator filler and substitute of H2SO4 solution. Accumulator containing mahogany leaf extract can conduct electricity well to accumulator with 6 volt capacity.

Therefore, the extract of mahogany leaves can be utilized as a new alternative energy source to be used as accumulator filler solution. From this research, we hope that community is aware of the benefits of mahogany leaves so that they will not do illegal logging which can cause severe forest damage in Bukit Suligi Educational Forest. This chance is also providing a great benefits to forest communities both in the economic aspect and in the aspects of forest conservation in order to support better environmental management and sustainable energy development.

IV. CONCLUSION

From the results of research, we can conclude that mahogany leaf is one of the potential sources from Bukit Suligi Educational Forest and now the number of this species has started to decrease due to illegal logging. Therefore, the utilization of mahogany leaves as an alternative source of electrical energy will make people aware of the benefits of mahogany leaves. The results of study showed that mahogany leaf extract can generate electricity power up to 3.87 volts. Mahogany leaf extract can be a substitute of H2SO4 solution as a filler solution on wet batteries because it can conduct electricity well. For the recomendation, further research is needed on the ability of electrical conductivity on mahogany leaf extract and the factors that influence the conductivity of mahogany leaf extract so that it can produce an optimal energy.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Bukit Suligi Educational Forest Office Station, Forestry Vocational School of Pekanbaru, Tanoto Foundation and also for Faculty of Forestry, Bogor Agricultural University, Indonesia for supporting facilities and funding for this research.

REFERENCES

[1] Adinugraha HA, The effect of NPK seeding and fertilization on the growth of broad leaf mahogany seedlings, Jurnal Pemuliaan Tanaman Hutan, vol. 6, no. 10, pp. 1-9, 2012.

[2] Bukit Suligi Educational Forest, The Information of Bukit Suligi Forest Area,, Pekanbaru (ID) : Center for Forestry and Environment Training Pekanbaru, 2015.

[3] Juwito AF, Haryono T, Optimization of Renewable Energy in Generating Electrical Energy Toward Energy Independent Village in Margajaya Village, JNTETI, vol. 2, no. 3, pp. 40-45, 2013.

[4] Latif M, Nazir R, Reza H, Analyzing the accumulator charging process on a horizontal axis wind turbine prototype in Purus Beach Padang, Jurnal Nasional Teknik Elektro., vol. 2, no. 1, pp. 1 – 8, 2013.

[5] Widodo PS, Power Plant with Local Energy Potential as a Manifestation of Electricity Distribution in Disadvantaged and Remote Villages, Vokasi, vol. 8, no. 3, pp. 151 – 164, 2012.

Development Approach of a Machine Learning Algorithm for Mooring Line Integrity Monitoring

[1] Seong Min Oh, [2] Do Hyoung Shin (Corresponding Author)
[1] Inha University, Republic of Korea, [2] Inha University, Republic of Korea
[1] 22161463@inha.edu, [2] dshin@inha.ac.kr

Abstract- The purpose of this study is to propose a development approach for a machine learning algorithm to monitor the integrity of fixed mooring lines supporting a floating structure. The study proposed a plan to utilize simulations for the training data of a machine learning algorithm, analyzed the system, condition, and response parameters of floating structures and mooring lines, and discussed machine learning training with the collected data.

Index Terms—Mooring line, Monitoring, Maintenance, Machine Learning, Algorithm

I. INTRODUCTION

Floating structures are in increasingly greater use for the development of marine resources. They are supported by fixed mooring lines. Since damage to mooring lines can lead to damage to floating structures, stable mooring is essential to the maintenance of floating structures. [1] In fact, 23 mooring systems have been destroyed and caused human and property damage since 2000[2] with approximately 1500 mooring lines repaired or replaced.[3] Given the expectation that such damage cases will increase according to the rising demand for floating structures, the maintenance of mooring lines will become even more important.

The integrity of mooring lines is, however, not monitored effectively due to limitations with measurement. This study thus set out to propose an approach to the development of a machine learning algorithm to promote the effective monitoring of fixed mooring lines in integrity..

II. CONVENTIONAL MONITORING METHODS FOR MOORING LINES

Various methods have been attempted to monitor the integrity of mooring lines including sound waves, sensors and remotely operated vehicles (ROVs).

A.A Method with a Simple Sonar Probe

A simple sonar probe assesses the integrity of mooring lines by checking their displacement with sound waves. It is simple and wonderful, but it can fail to detect damage to mooring lines in mud due to tension. Another disadvantage of the method is that if it cannot check mooring lines for about two weeks, average period of marine climate changes and in this period, there can be heavy typhoon, the structure can be damaged.

B.A Method with an ROV

ROVs are unmanned undersea vehicles that are operated remotely. They are put in the water to check the clinometer and the integrity of mooring lines. It is, however, available only when the weather is fine and can be difficult to operate due to current disturbance.

C. A Method with an Instrumented Mooring Line

This method attaches sensors to mooring lines to check their integrity. It is a good approach, but it is impossible to figure out whether abnormal signals from mooring lines are due to a problem with the sensors or with the mooring lines.

D. FPS Offset Monitoring and Line Failure Detection

It detects damage to mooring lines with the offset yield strength of floating production systems (FPSs). It, however, depends on the observer's experiences relatively.

Reviewed conventional monitoring methods and found that there were still considerable limitations with measuring mooring lines and examining their integrity. This study thus proposed an approach to the development of a machine learning algorithm to overcome the old problems.

III. CONCEPT OF MACHINE LEARNING APPLICATION TO MONITOR THE INTEGRITY OF MOORING LINES

Fixed mooring lines and floating production systems show different behavioral responses according to system parameters such as the length of mooring lines and the weight of floating production systems and condition parameters such as wave height and current strength.

The major elements affecting the integrity of mooring lines are the section area and weight of mooring lines. Over time, damage to the section of the mooring line occurs, and biological pollution materials such as seaweeds

are attached to the mooring line. The section area and weight of mooring lines are known as initial values at the time of installation, but

IV. GENERATION OF LEARNING DATA THROUGH SIMULATIONS

Programs to interpret the behavioral responses of mooring lines such as CHARM3D are used to generate learning data for machine learning through simulations. Such interpretation programs require input information including system and condition parameters and use them to compute response parameter values (Fig. 1).

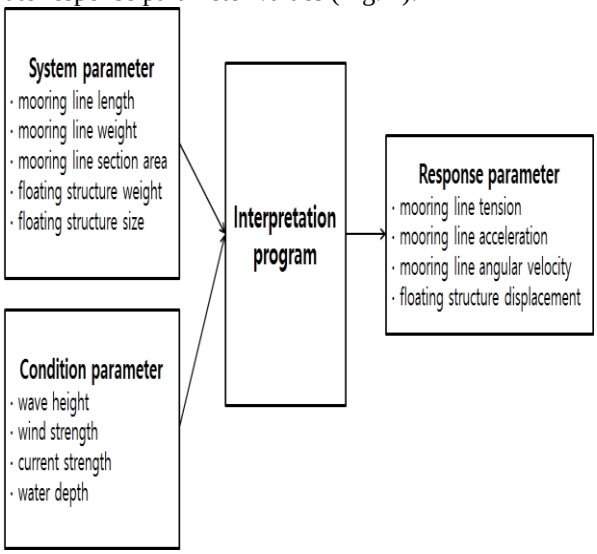


Fig. 1. Data generation through simulations

It is thus necessary to set the type and scope of system and condition parameters to be used in a simulation in advance. And consider whether response parameters can be measured in real system.

The initial values of system parameters are known at the stage of design and making. The concerned parameter values can be adjusted and entered in the program to reflect damage to the section area of mooring lines and changes to their weight. The representative system parameters of floating structures include the weight and size of floating structures and the length, section area and weight of mooring lines. Since the weight and size of structures and the length of

mooring lines remain the same over time, they can be used as fixed values in simulations. The section area and weight of mooring lines, however, change over time as they have been explained above, which raises a need to consider the scope of damage to the section area and changes to the weight and reflect diverse values randomly altered from their initial values in simulations.

The representative condition parameters of floating structures include the wave height, wind strength, current strength, and water depth, of which the wave height, wind strength, and current strength continue to change and can be measured with all kinds of measuring sensors. In simulations, those values should be changed randomly and diversely to reflect the situations that continue to change. The deep sea, where floating structures are commonly installed, undergoes few changes to the water depth and can thus be used as a fixed value in simulations.

It is easy to measure the displacement of floating structures, one of the response parameters, with GPS and a clinometer, but it is very difficult to measure the displacement of mooring lines underwater. There are limited response parameters of mooring lines that can be measured relatively easily, and they include tension, acceleration, and angular velocity. It will be thus realistic to use such response parameters as the displacement of floating structures and the tension, acceleration, and angular velocity of mooring lines in the training of machine learning.

Table 1 presents the parameters that should be used in simulations to generate training data for machine learning.

Table 1 Necessary parameters for simulations

Classification	Variable
System variables	Mooring line length
	Mooring line weight
	Mooring line section area
	Floating structure weight
	Floating structure size
Condition variables	Wave height
	Wind strength
	Current strength
	Water depth
Response Variables	Mooring line tension
	Mooring line acceleration
	Mooring line angular velocity
	Floating structure displacement

V. MACHINE LEARNING TO MONITOR THE INTEGRITY OF MOORING LINES

The section area and weight of mooring lines can be estimated by developing a machine learning algorithm based on abundant pattern information collected in simulations. A

machine learning algorithm uses the condition and response parameter values and the known system parameter values as inputs. And estimates the section area and weight of mooring lines, the unknown system parameter values (Fig. 2).

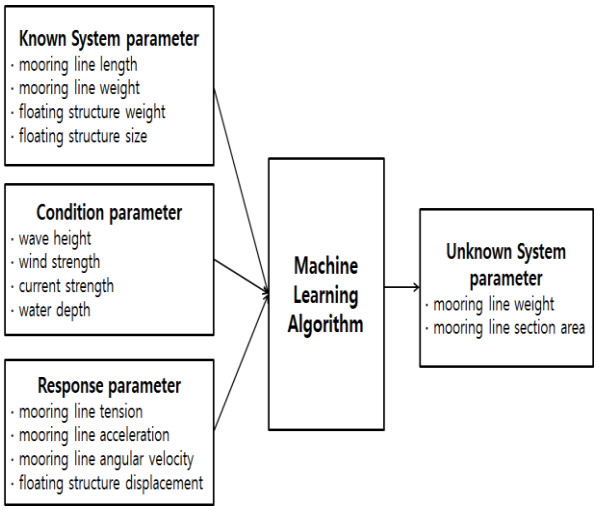
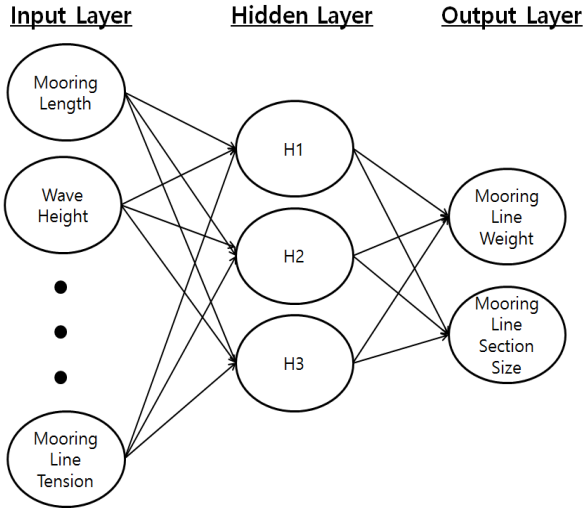


Fig. 2. Estimation of section area and weight of mooring lines through machine learning

There are various types of machine learning, of which deep learning will be proper to fulfill the goal of monitoring the integrity of mooring lines. Based on the artificial neural network theory, deep learning uses input/output layers like human neurons and multiple hidden layers and is effective for solving complex non-linear problems.

It is thus required to designate the input/output and hidden layers of an artificial neural network in order to implement an effective algorithm. Since input values include the condition and response parameter values and the known system parameter values, the input layer will contain the length of mooring lines, the weight of floating structures, wave height, and cable tension. The output layer will contain the weight and section area of mooring lines, the result values sought after by this study (Fig. 3).



nodes. It is thus needed to designate them randomly many times and adjust them according to the data characteristics. In addition, several hyper-parameters should be adjusted including the weights to connect different layers and the learning rate of back-propagation to increase the accuracy of an algorithm.

In an actual measurement of a structure's behavioral responses, measuring noises happen often. Such noises should be reflected in a machine learning algorithm test, in which Gaussian random noises can be used. It will be a useful approach to a test in a situation closer to the reality to insert Gaussian random noises into the data generated in simulations artificially, simulate them like actual measurement data, and test an algorithm with such data.

IV. CONCLUSION

This study proposed a development approach for a machine learning algorithm to monitor the integrity of fixed mooring lines supporting a floating structure. Huge amounts of actual data are needed according to the various situations of fixed mooring lines to develop a machine learning algorithm, but it is practically impossible to obtain them. Thus, in this study, training with data generated in simulations. In addition, the study analyzed the system parameters of floating structures and mooring lines (length, section area, and weight of mooring lines and the weight and size of floating structures), the condition parameters (wave height, wind velocity, current strength and water depth), and response parameters (tension, acceleration, and angular velocity of mooring lines and the displacement of floating structures), as well as proposed parameters that should be used in a simulation.

Deep learning, a type of machine learning, was deemed to be appropriate for the goal of monitoring the integrity of mooring lines. A deep learning-based machine learning algorithm used the condition and response parameter values and the known system parameter values as inputs and estimated the section area and weight of mooring lines, the unknown system parameter values. The use of Gaussian random noises was also proposed to test the machine learning algorithm in a situation close to the reality.

Based on this study, a machine learning algorithm is being developed for the monitoring of mooring line soundness, and it is expected that the maintenance of floating structures will be more effectively performed.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2017R1A2B4011753) they vary with time. Today there are no methods to measure such changes directly. If there are massive data of response parameter values according to system and condition parameters, machine learning can be possible with such

data. In such a case, machine learning can help to estimate the section area and weight of mooring lines according to the given condition and response parameter values and the known system parameter values including the length of mooring lines. The purpose of this study was to propose a development approach for a machine learning algorithm to estimate the section area and weight of mooring lines, unknown system parameters, with condition and response parameters and known system parameters to monitor the integrity of mooring lines. The problem with the approach was the impossibility of collecting enormous amounts of actual data according to the various situations of fixed mooring lines Thus, in this study, using simulation data performed computer program in various conditions as learning data."

REFERENCES

[1] R Venkatesan, N Vedachalam, P Murugesh, P Kaliyaperumal, CK Kalaivanan, T Gnanadhas and MA Atmanand, "Reliability analysis and integrity management of instrumented buoy moorings for monitoring the indian seas", underwater technology vol33 No2 pp115~126, 2015

[2] Mousavi ME and Gardoni P, "A simplified method of reliability and integrity based design of engineering systems and its application to offshore mooring systems", Marine Structures 36 PP88-104

[3] Maslin E, "Mooring line failures attract industry attention", OE Digital, 2013

Feasibility Study of Plastic in Concrete

[1] Dr. Kanish Kapoor, [2] Bhupesh Kumar Gupta, [3] Puneet Dhawan
[1] Assistant Professor, [2] Post graduate Scholar, [3] Assistant Professor
[1][2][3] School of Civil Engineering, Lovely Professional University, India

Abstract- The use of the plastics increasing day by day owing to their advantages such as cheapness, endurance, lightness, hygiene and design adaptability. The negative side of the picture is that, it is biodegradable product and anesthetic views plastic cannot be disposed up to open environment; they can affect harmfully to environment, soil and sub soil. They release toxic gases when exposed to open environment, can block sewer lines, drainages and other waterways. Plastics are also harmful to human health; they may contain or leads to harmful acids. Thus recycling of plastic is essential in today's condition. This paper focuses on the analysis of results interpreted by various researchers by experimentation on polyethylene terephthalate (PET) as concrete component in various forms. The various mechanical and durability properties of concrete made with incorporation of plastics are presented.

I. INTRODUCTION

Plastics are synthetic or semi synthetic organic solid materials, typically polymers of high molecular weight [1]. Owing to their advantages such as cheapness, endurance, lightness, hygiene and design adaptability plastic have been widely used. Some of most common examples of plastic materials are plastic bottles, PVC pipes, plastic bags, electrical plastic circuit, etc. Central Pollution Control Board CPCB has stated that, the demand of plastic product is increasing rapidly at a rate of 22% annually [2, 3]. Depending upon the molecular structures and bonding plastic otherwise known as plastic polymers which are most commonly used, categorized into polyethylene terephthalate (PET or PETE), polyethylene (PE), high-density polyethylene (HDPE), polyvinyl chloride (PVC), low-density polyethylene (LDPE), polypropylene (PP), polystyrene (PS).These makes 90% of the total plastic use worldwide [4]. From the total municipal solid waste, waste plastic represents of about 8 wt% which includes different kinds of plastics: 40.5 wt% HDPE and L/LDPE, 19.6 wt% PP, 11.9 wt% PS/EPS, 10.7 wt% PVC, 8.1 wt% PET, and about 5 wt% ABS and 4.2 wt% other polymers. Due to the complex nature and composition, structural deterioration of the polymeric components and the contamination with various organic, inorganic or biological residues, treatment of these wastes is very difficult [5].

Today sustainability has got top priority in construction industry. From the very beginning of 21st century lot of researches has been done to use this plastic waste as concrete component, thereby providing a sustainable option to deal with plastic waste. Recently, the study of possibility of disposal of plastic waste, one of the components of municipal solid waste in concrete specifically in self compacting concrete (SSC), light weight concrete (LWC) and in pavements is becoming a major research issue. It can also be used as a component of a composite construction material, as an inorganic filling material, and aggregate of concrete. For the highly convenient nature, resistance towards degradation and serious worldwide environmental and health concerns make the study of disposal of plastic highly essential. A number of researches were made on this, particularly on reuse and recycling of plastic waste. Management requires a framework that includes collection of waste outside or inside municipal waste stream, separation, land fillings, energy recovery and recycling [6]. Recycling of plastics encompasses four stages: collection, separation, processing/manufacturing and marketing [7]. The industry of construction engineering seems to be appropriate for disposal of plastic wastes due to its high consumption capacity. The construction engineering area can consume a large amount of PET [9]. The performance of PET as concrete component in various forms had been done in concrete technology. For example, E. Rahmani et al., Alexandra M. da Silva et al. and Nursyamsi have investigated PET as aggregate for concrete.

This paper focuses on the analysis of results interpreted by various researchers by experimentation on polyethylene terephthalate (PET) as concrete component in various form. Most researchers shows that the addition of PET affects various concrete component as workability, compressive strength, modulus of elasticity, split tensile strength, thermal conductivity and slightly enhance abrasion and flexural strength, resistance to ultrasonic wave and sulfuric acid attack , shock resistance to impact with a given energy. Additionally researches recommended that surface of the plastic does not react with the matrix. Thus the surface of the matrix must be treated with a reactive material, such as silica fume (SF), metakaolin (MK), slag, so that the pozzolanic reaction enhances the

strength of the concrete by reacting with the surface coated material.

2. RECYCLING PROCESS

Using plastic waste as a concrete component has made a huge revolution in the field of recycling the plastic waste. The advanced engineering of safe and beneficiary disposal of plastic waste had also proved that the plastic waste can be used in concrete in various forms. The different forms of plastic waste which can be used are plastic aggregates (coarse & fine), plastic fibres, polyester resins and reinforcing materials. There are two ways to produce recycled fibre from mechanical recycling; directly extrude flakes into fibre & first convert flakes into pellets or chips and then melt-extrude pellets or chips into fibre [10]. The approaches that have been proposed recycling of waste polymers include: primary recycling referring to the "in-plant" recycling of the scrap material of controlled history. Mechanical recycling; polymer is separated from its associated contaminants and it is reprocessed by melt extrusion. Chemical recycling: leading in total depolymerization to the monomers, or other secondary valuable materials [11].

3. INFLUENCE OF PLASTICS ON THE PROPERTIES OF CONCRETE

3.1 Slump test

It is found from the previous study that plastic aggregates have an outer smoother surface and they cannot absorb water, which leads to an excess of water which improves the workability. More over the shape, size and roughness of the components particles also have a marked influence on the consistency of mix. Table-1 shows a comparison of maximum slump value from some previous studies.

Table-1: Summary of the slump test results by previous authors

Author	Slump range (mm)
Brahim Safi et al. [12]	240-378
R.V. Silva et al. [13]	135-141
Ali Sadrmomtazi et al. [14]	603-700
Md Jahidul Islam et al. [15]	50-160
Wasan Ismail Khalil et al. [16]	90-97

3.2 Dry bulk density

The lower unit weight of concrete components makes the concrete lighter than standard concretes. Plastic aggregates have apparently a lower specific gravity than natural aggregates which makes concrete lighter. Table-2 shows maximum reduction percentage of dry density from some researches.

Table-2: Summary of the value of bulk dry density reported by previous authors

Author	Range of dry bulk density (kg/m^3)
Semiha Akcaozoglu et al. [9]	1679-1937
Alexandra M. da Silva et al. [17]	1580-1610
Md. Jahidul Islam et al. [15]	1925-1980
Nursyamsi et al. [18]	1741-1802

3.3 Compressive strength

Compressive strength of concrete incorporating plastic aggregate has comparatively lower compressive strength than standard concrete mix. Due to the smoother surface, bonding between cement paste and plastic components is weaker and also presence of extra water in the transition zone makes the bonding weaker. Tabe-3 highlights the outcomes of compressive test of some authors.

Table-3: Summary of the value of compressive strength reported by previous authors

Author	28 days Compressive strength(MPa)
Mariaenrica Frigione [19]	43.2-67.5
Luis Ferreira et al. [20]	20-37
E. Rahmani et al. [21]	28.91-59.51
MB Hossain [22]	14-18
Alexandra M. da silva et al.[17]	1.9-3.6

3.4 Split tensile strength

Like compressive property of concrete the splitting strength of LWAC also reduce. The negative effect of a smooth surface texture on the bond strength is highlighted due to the increase surface area of PET particles compare

to sand. Table-4 highlights the outcomes of split tensile test of some author.

Table-4: Summary of the value of split tensile strength reported by previous authors

Author	28 days Split tensile strength(MPa)
Mariaenrica Frigione [19]	4.1-6
Luis Ferreira et al. [20]	1.5-2.8
MB Hossain [22]	1-2

3.5 Flexural strength

By incorporating plastic aggregate, the flexural strength of concrete is comparatively reduced. The increased porosity and lesser strength of plastic aggregates are attributed to this loss. Additionally considerable improvement in flexural energy values for all concrete containing plastic in coarser form and also an improvement in the deflection value. This property can be explained by the ability of plastic aggregate to prolong crack propagation interval due to their non-brittle characteristics. Table-5 highlights the outcomes of split tensile test of some author.

Table-5: Summary of the value of flexural strength reported by previous authors

Author	Maximum reduction in flexural strength (%)
J.M.L Reis et al. [23]	38
Alexandra M. da silva et al.[17]	54
Ali Sadrmomtazi [14]	55

3.6 Modulus of elasticity

Previously exists researches shows an increasing trend in modulus of elasticity of concrete mix incorporating plastic particles. For the specimens containing 10% plastic has about eight times higher than that of the specimen containing 0% plastic aggregate. The deformation produced in the concrete is partially related the elastic deformation of the aggregate [22]. Table-6 shows outcomes of some authors for modulus of elasticity.

Table-6: Summary of the value of Modulus of elasticity reported by previous authors

Author	Range of modulus of elasticity of concrete (GPa)
Luis Ferreira et al. [20]	18-33
E. Rahmani et al. [21]	15-36
M B Hossain [22]	0.34-40

3.7 Thermal conductivity

Temperature inside test specimens is increased gradually, at a slower rate when compared with the increase in the oven temperature for LWAC. A considerable reduction in thermal conductivity with the increase in Plastic content is observed as there is formation of huge amount of cavities in the structure of concrete containing PET aggregate. Porosity is also another factor affecting the thermal conductivity of concrete and enclosed pores reduces the conductivity due the low thermal conductivity of air. The magnitude of difference increases with the incorporating ratio as; firstly the higher porosity increases due to the replacement of NA by PWA, which facilitated the propagation of heat inside the concrete specimens and secondly the exothermic thermal decomposition of plastic aggregate that generates additional heat[24,16].

3.8 Carbonation depth

Carbonation depths of mortars increases depending on increasing CO2 amount penetrated into samples which increase in time. Plastic and sand aggregates used together do not combine with each other sufficiently; thus mixes have more porous and therefore more CO2 penetrated in the M3 and M4 mortars from their pores. It is also depends upon the curing condition, carbon depth increases when concrete is cured in progressively drier environments. Table-7 shows outcomes of some authors for carbonation depth.

Table-7: Summary of the value of carbonation depth reported by previous authors

Author	Maximum Carbonation depth in mm(28 days)
R.V. Silva et al. [13]	36.9
Semiha Akcaozoglu et al. [9]	2.5

CONCLUSION

Judicious use of natural resources is the first and foremost requirement for sustainable development. In the search for alternatives to natural materials, the use of plastics as a substitute of aggregate serves two requirements, one it eats up the waste secondly it saves the natural aggregate. Furthermore, it is important to take care of the bond between waste plastics and concrete matrix. This is due to the reason that the surface of the plastic does not react with the matrix. Thus the surface of the matrix must be treated with a reactive material, such as silica fume (SF), metakaolin (MK), slag, so that the pozzolanic reaction enhances the strength of the concrete by reacting with the surface coated material

REFERENCES

1. P. Singh, And V.P Sharma, "Integrated Plastic Waste Management: Environmental and Improved Health Approaches". Procedia Environmental Sciences, Vol. 35, 2016, pp. 692-700.

2. Times of India: PTI, "60 Cities generate over 15,000 tonnes of plastic waste per day", April 30, 2015.

3. Press Trust of India: "15,343 tn plastic waste generated in India everyday: Dave, New Delhi". August 2016.

4. W.C. Li, H.F. TSE, And L.FOK, "Plastic waste in marine environment: A review of sources, occurrence and effects". Science of the Total Environment, Vol. 566-567, 2016, pp. 333-349.

5. Moinuddin Sarker, And Mohammad Mamunor Rashid, "Thermal Degradation of Poly (ethylene terephthalate) Waste Soft Drinks Bottles and Low Density Polyethylene Grocery Bags". International Journal of Sustainable Energy and Environment, Vol. 1, No. 3, April 2013, pp. 78-86.

6. K.S. Rebeiz, And A.P. Craft, "Plastic waste management in construction: technological and institutional issues". Resources, Conservation and Recycling, Vol. 15, 1995, pp. 245-257.

7. Chong-Qing Wang, Hui Wang, And You-Nian Liu, "Separation of polyethylene terephthalate from municipal waste plastics by froth flotation for recycling industry". Waste Management , Vol. 35, January 2015, pp. 42-47.

8. Eva Sevigne-Itoiz, Carles M. Gasol, Joan Rieradevall, And Xavier Gabarrell, "Contribution of plastic waste recovery to greenhouse gas (GHG) savings in Spain". Waste Management, Vol. 46, December 2015, pp. 557-567.

9. Semiha Akcaozoglu, Cengiz Duran Atis, And Kibilay Akcaozoglu, "An Investigation on the use of shredded waste PET bottles as aggregate in lightweight concrete". Waste Management, Vol. 30, 2010, pp. 285-290.

10. Li Shen, Ernst Worrell, And Mrtin K. Patel, "Open-loop recycling: A LCA case study of PET bottle-to-fibre recycling". Resources, Conservation and Recycling, Vol. 55, 2010, pp. 34-52.

11. D.S. Achilias, C. Roupakias, P. Megalokonomos, A.A. Lappas, And E.V. Antonakou, "Chemical recycling of plastic wastes made from polyethylene (LDPE and HDPE) and polypropylene (PP)". Journal of Hazardous Materials, Vol. 149, 2007, pp. 536-542.

12. Brahim Safi, Mohammed Saidi, Djamila Aboutaleb, And Madani Maallem, "The use of plastic waste as fine aggregate in the self-compacting mortars: Effect on physical and mechanical properties". Construction and Building Materials, Vol. 43, 2013, pp. 436-442.

13. R.V. Silva, J.de Brito, And Nabajyoti Saikia, "Influence of curing conditions on the durability-related performance of concrete made with selected plastic waste aggregates". Cement & Concrete Composites, Vol. 35, 2013, pp. 23-31.

14. Ali Sadrmomtazi, Sahel Dolati-Milehsara, Omid Lotfi-Omran, And Aref Sadeghi-Nik, "The combined effects of waste PET particles and pozzolanic materials on the properties of self-compacting concrete". Journal of Cleaner Production, 2015, doi: 10.1016/j.jclepro.2015.09.107.

15. Md. Jahidul Islam, A.K. Rakinul Islam, And Md. Salamah Meherier, "An Investigation on Fresh and Hardened Properties of Concrete while Using Polyethylene Terephthalate (PET) as Aggregate". International Scholarly and Research & Innovation Vol. 9(5), 2015, pp. 558-561.

16. Wasan Ismail Khalil, And Khalaf Jumaa Khalaf, "ECO-FRIENDLY CONCRETE CONTAINING PET PLASTIC WASTE AGGREGATE". Diyala Journal of Engineering Sciences, March 2017, Vol. 10, No. 01, pp. 92-105.

17. Alexandra M. da Silva, Jorge de Brito, And Rosario Veiga, "Incorporation of fine plastic aggregates in rendering mortars". Construction and Building Materials, Vol. 71, 2014, pp. 226-216.

18. Nursyamsi, And Winner Syukur Berkat Zebua, "The influence of pet plastic waste gredations as coarse aggregate towards compressive strength of light concrete". Procedia Engineering, Vol. 171, 2017, pp. 614-619.

19. Mariaenrica Frigione, “Recycling of PET bottles as fine aggregate in concrete”. Waste Management, Vol. 30, 2010, pp. 1101-1106.

20. Luis Ferreira, Jorge de Brito, And Nabajyoti Saikia, “Influence of curing conditions on the mechanical performance of concrete containing recycled plastic aggregate”. Construction and Building Materials, Vol. 36, 2012, pp. 196-204.

21. E. Rahmani, M. Dehestani, M.H.A. Beygi, H. Allahyari, And I.M. Nikbin, “On the mechanical properties of concrete containing waste PET particles”. Construction And Building Materials, Vol. 47, 2013, pp. 1302-1308.

22. MB Hossain, P Bhowmik, And KM shad, “Use of waste plastic aggregation in concrete as constituent material”. Progressive Agriculture, Vol. 27 (3), 2016, pp. 383-391.

23. J.M.L. Reis, And E.P. Carneiro, “Evaluation of PET waste aggregates in polymer mortars”. Construction and Building Materials, Vol. 27, 2012, pp. 107-111.

24. J.R. Correia, J.S. Lima, And J. de Brito, “Post-fire mechanical performance of concrete made with selected plastic waste aggregates”. Cement & Concrete Composites, Vol. 53, 2014, pp. 187-199.

In-silico approach to study the mitochondrial localisation of a Fanconi Anemia protein

Shubhrima Ghosh
Enzyme and Microbial Biochemistry Laboratory, Department of Chemistry,
Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India
Email- shubhrima.ghosh@gmail.com

***Abstract*- Fanconi's Anemia (FA) is an inherited genomic instability disorder caused by mutations in genes regulating replication-dependent removal of interstrand DNA crosslinks. The symptoms include bone marrow failure, congenital defects, and cancer predisposition. Fifteen FA gene products have been identified, which cooperate in a common DNA damage-activated signalling pathway regulating DNA repair (the FA pathway). Eight FANC proteins (A/B/C/F/G/I/J/L) assemble into a nuclear complex and two other proteins, FANCD1/BRCA2 and FANCD2 act downstream of the core complex. This FA protein complex is required for the monoubiquitination of the FANCD2 and FANCI protein in response to DNA damage. The inactivation of the FA pathway has also been observed in a wide variety of human cancers and is implicated in the sensitivity of cancer cells to DNA crosslinking agents.**

The central FA protein FANCD2 is known to relocate to chromatin upon DNA damage and is thus a key player in the Fanconi Anemia pathway. It consists of 44 exons, encodes a novel 1451 amino acid nuclear protein, and has two protein isoforms. Till date FANCD2 has been implicated to have a definite role in nuclear DNA damage repair. But the role of FANCD2 in mitochondria is yet to be elucidated. In this study, we aim to find out the mitochondrial localization mechanism of FANCD2 and its role in mitochondria, specifically in mitochondrial DNA damage repair. Our in silico findings have predicted the localisation of FANCD2 in the mitochondria and we have been able to identify a 30 amino-acid mitochondrial localisation signal (MLS). Further studies have helped in the characterisation of specific amino acids in the MLS for their role in mitochondrial localization and shed light on their structure-function relationships. The study may thus help us to find out key regions of the protein which are indispensible for mitochondrial localisation and provide us with better knowledge of mitochondrial DNA damage repair through the FA pathway.

Keywords: Fanconi Anemia, cross-links, localisation, mitochondria, monoubiquitination,

Identifying and Maintaining Overleaded Servers Through Dynamic Placement of Virtual Machines in Cloud

[1] Dr. Md. Mahbubul Alam Joarder, [2] Asif Imran, [3] Lamisha Rawshan
Institute of Information Technology, University of Dhaka, Ramna, Dhaka 1000, Bangladesh

***Abstract*:-- The cloud has become a widespread and commonly used service due to its improved resource utilization. The little maintenance expense and on demand characteristics of cloud has made it one of the most popular commercial infrastructure. However, cloud computing possesses many kinds of technical challenges such as fault tolerance, reliability, availability, integrity etc. But the main problem related to all those is overload incurred by Virtual Machines (VM). This research proposes an approach to identify and manage overloaded servers with dynamic placing of its VMs. It monitors and stores resource usage data. The algorithm checks server resource usage status by calculating the value of skew after a predefined time interval. Next skewness is used to identify the over-loaded and under-loaded vm-instances. Positive skewness indicates overloading and negative skewness indicates under-loading. For experimental purposes, -0.75 is considered as the lower threshold and 1.50 is considered to be the upper threshold. Experimental evaluation show that out of the 40 vm-instances, on average 17 vm's were found to be under-loaded and 10 vm's were found to be over-loaded. The analysis shows that using skewness to aid in migration reduces migration time by about 21% for user processes ranging from 1500 to 3000 users.**

***Keywords*: Cloud Computing, Resource management, Skewness, Virtual machine migration, Overload Detection**

I. INTRODUCTION

Cloud is a group of physical machines pretending to be one computing environment [1]. User see the cloud as an illusion of unlimited computing resources [2]. But the main challenge is to manage the variability and heterogeneity of application requirement [3]. VMs shares the resources of PMs among several users to maximize the use of resources. The mapping of these PMs to VMs is very important to gain better performance of cloud computing [4]. This mapping is completely unknown to the users. They have no knowledge about the location of the PMs from where their VM is running. It is one of the main responsibilities of the cloud provider to meet the resource demand of users. The capability of PM must satisfy the resource demand of VMs running in it. Otherwise, the PMs will degrade its performance due to overloaded VMs [5]. From this the need to a framework that detect load of a PMs and dynamically place VMs in a way to maximize the resource utilization is needed.

This paper will present a framework to detect overloaded PMs and manage VMs re- sponsible for overload. To improve performance of the existing scheme, skewness detection mechanism has been used. Skewness can be of two types. One is positive skewness and the second is negative skewness. Any type of skewness will indicate that the server is either heavily loaded (positive skewness) or under-loaded (negative skewness). As a result, it is an important indicator of the current load distribution of servers. Since load distribution is an important aspect of cloud computing, it can be effectively managed with skewness.

This paper proposes an algorithm to detect the current load on cloud servers. At the same time, it presents a mechanism which shows the Effective X-factor based mechanism has been described for the detection of load on servers in the cloud. Next algorithms for the detection of skewness in the load of the servers are provided. Afterwards an algorithm for Reallocation of resources to solve the issue of overload and under-load has been described. Effective framework to detect the migration of the jobs in the cloud to reduce the overload or under- load condition has been provided. Results have been tabulated which shows the number of migration jobs which are performance affecting can be reduced if skewness is determined and based on skewness the decision are taken. Also, the performance improvement of skewness based reallocation of jobs is seen when results indicate that time taken to re-allocate resources significantly decreases when the output of detected skewness if followed.

The rest of the paper are as follows. Section 2 represents the concise discussion on the existing research in our field of interest. Section 3 showcases the proposed methodology and algorithms and provides justification on the effectiveness of using skewness. Section 4 highlights the experimental test-bed (i.e. environment) which was set up and used for this research. Section 5 shows the obtained

results and presents a discussion on these results. Finally, Section 6 draws a conclusion to the paper and identifies some scope of future research in this field.

II. LITERATURE REVIEW

An inaccurate resource allocation can cause insufficient use of resources, higher cost and low performance [6]. Because of the importance and effectiveness of performance increasing in recent years, researchers have discussed different methods of load detection and dynamic placement of VMs. This section highlights recent works that are performed to identify overloaded servers and managing those VMs of cloud computing. Vijayakumar et al. [7] have presented a model to allocate resources dynamically for data streaming applications. The main objective of their research is to ensure careful allocation of resources to avoid both over provision and the under provision. They detect buffer overflow by comparing time interval between receiving two blocks of streaming data and processing of one block of that data. When any overflow occurs their model increase a fixed amount of CPU based on requirement. They define a fixed amount of CPU percentage for multiplicative increase and additive decrease of CPU. They evaluate their model in both static and dynamic environments. In their research, they used fixed amount of CPU addition which is costly.

Multi-Attribute Utility Theory is used for VM allocation and migration. VM allocation and migration decision is affected by many attribute like resource availability, network band- width, network cost and Service Level Objectives (SLO) violation. In this scheme, nodes are organized using unstructured Peer-to-Peer (P2P) architecture to avoid single point of fail- ure threat. Every physical machine collects information from its neighbored with a constant distance.

III. PROPOSED METHODOLOGY

This section describes the algorithm for detecting overloaded servers and necessary self- healing to recover from failed condition. The algorithm maintains a timer, SpareT ime after which it becomes alive to collect data. It collects I nputRate, OutputRate, M axM emV ector, and ResourceU tilizationV ector over a specified interval.

The LoadDetection algorithm detects overloaded server by calculating the value of skew. The SkewDetection algorithm detects the percentage of unevenness of resource utilization among the PMs. So, it computes resource utilization skew of the cloud by combining re- sources utilization of its PMs. This skew is compared with a threshold. If skewness crosses the threshold, it means uneven resource utilization exists in the cloud. Therefore, perfor- mance of some PMs can be degraded. In this situation VM redistribution is needed to equalize load. Next, it computes X factor values to fix the destination of the overloaded PMs. ResourceM apGenerator uses the X factor values and generate a resource map to reallocate VMs to PMs.

Detection of the skew is a way to understand that all resources occupied load distribution symmetrically. This section contains an algorithm to detect uneven resource utilization (Skew) between PMs.

3.1. Skew Detection Algorithm

Skewness is an important aspect of statistics which will ensure that whether the system is over-loaded, under-loaded or the load is evenly distributed. As a result, if the graph has positive skewness then it is overloaded, on the opposite if it is under-loaded then the graph will have negative skewness. Our load is to use the skewness graph to detect which servers are overloaded and under-loaded and finally evenly distribute the load across the system to ensure evenness. the reason for using skewness is to ensure that the load can be distributed to a greater level of accuracy using less time which will ensure that the system is optimized in performance. The more the number of physical machines, the power of skewness is likely to be even more clear in terms. also skewness will enable lesser number of migrations.

One common property of cloud computing that can cause low performance or failure is uneven distribution of resources. So it is necessary to distribute resources in such a way that all nodes of a system will consume approximately similar amount of memory (uniform load distribution) . In order to prevent unevenness of resource distribution, detection of skew is necessary. Node with enough space will be chosen for migrating an overloaded VMs. Equation 1 contains formulas for detecting resource utilization skew of overall cloud. Skew,

$$S = \frac{m_3}{m_2^{\frac{3}{2}}} \tag{1}$$

Here

$$m_2 = \sum_{i=1}^{i=N} \frac{(U_i - U_{avg})^2}{N} \tag{2}$$

$$m_3 = \sum_{i=1}^{i=N} \frac{(U_i - U_{avg})^3}{N} \tag{3}$$

In equations 2 and 3, Ui indicates memory uses of ith server in the time, t. N is the total number of PMs running in the cloud. At last, skew is generated using the values of m2 and m3. From the value of S, resource redistribution decision will be made. The value of S will help to decide if further resource reallocation is needed or not. When the resource distribution is approximately symmetric, migrations of VMs will be stopped.

Algorithm 1 Find Skew of the Cloud

Input: $N, ResourceUtilizationVector, ResourceRequirement$
Output: $Skew$

```
procedure SKEW DETECTION
    for i ← 1 to N do
        m ← Number of VMs running on i^th PMs
        for j ← 1 to m do
            m2 ← Σ_{j=1}^{j=N} (U_j − U_avg)^2 / N
            m3 ← Σ_{j=1}^{j=N} (U_j − U_avg)^3 / N
        end for
    end for
    S ← m3 / m2^(3/2)
    Skew ← S
    Return Skew
end procedure
```

This algorithm is presented in 1. The Skew Detection algorithm takes 2 inputs- Num-ber of physical machines and Resource utilization metrics. By using equation 2 and 3, it determines the deviation from average usage for every server. It calculates Skew for the cloud.

Algorithm 2 Reallocate Overloaded VMs

Input: $Skew, ResourceUtilizationVector, OverloadedPMsIDVector$
Output: $ResourceReAllocationMap$

```
procedure RESOURCE ALLOCATION
    for i ← 1 to N do
        X_i ← (U_i − U_avg) / D
        m ← Number of VMs running on i^th PMs
        for j ← 1 to m do
            ServerID ← AllocateResource(X_i,
          ResourceUtilization[i][j], OverloadedPMsIDVector)
            ResourceReAllocationMap ← ResourceMap[ServerID][j]
        end for
    end for
    Return ResourceReAllocationMap
end procedure
```

3.2. Placement of Overloaded VMs

In the above section we discussed about the skew detection mechanism. When skewness is not in the range of the threshold value, VM migration will be done using the value of X factor. X factor is the value which de

$$D = \sqrt{\sum_{i=1}^{i=N} \frac{(U_i - U_{avg})^2}{N}} \tag{4}$$

X-factor,

$$X = \frac{(U_i - U_{avg})}{D} \tag{5}$$

ResourceMapGenerator algorithm determines which PM can continue normal work ow of overloaded servers. The input of this algorithm is -Skew, overloaded servers and resource utilization details. It outputs ResourceReAllocationMap from overloaded PMs to VMs. The procedure is depicted in algorithm 2. It sorts the X values list in decreasing order. Add overloaded VMs resource requirement with the low valued PMs. The main purpose is to choose a PM with low valued and enough space for migrating overloaded server. If the PM with low X value have enough space to continue the normal work of overloaded VMs, map the VM with the PM. If the PM with low x value not satisfy the resource requirement, choose the second low valued PM.

IV. EXPERIMENTAL TESTBED (I.E. ENVIRONMENT)

For experimental purpose, BSD Unix systems have been used. More specifically, in BSD Unix system, the researchers installed OpenStack Cloud which had kernel virtual machine (kvm) enabled. as a result of this the virtualization methodology can be used to lauch vm instances and migrate workload into the new vm-instances whenever a machine is overloaded. In the case of under-loading, the vm-instances were reduced in number after migration of the jobs running in under-loaded servers into other vm-instances. As a result, cloud virtualization was used.

Apart from the physical sertup of cloud computing environment, the researchers im- plemented the algorithms in those cloud server to detect the number of U nusedmachines, skewness, mean of the machine load, Standard Deviation, m1, m2 and Z − scores. Based on a series of calculations, the under-loaded and overloaded machines were determined. It was seen that for experimental purposes, 1.5 was considered to be the upper threshold whereas -0.75 was considered to be the lower threshold. Based on this, the overloaded and under-loaded servers were identified as shown in Figure. Skewness was determined for a range of +.5 to -.5 which is normal. Above this range all are asymmetrical. Range is +1.0 to -1.0 is normal also. Hence, above this range all numbers are considered to be asymmetrical as well. Based on the above

determinations, the experiment was repeated for 6 times using the same set of vm-instances in the cloud using various workloads and based on that the under- loaded and overloaded servers were marked. More specifically, 40 vm-instances were tested for this purpose. Hence, the algorithms automatically identified the overloaded and under- loaded vm-instances and the migration algorithm conducted migration more effectively using the skewness. The results of this proposed experimental setup is highlighted in the later sections of this paper.

V.ANALYSIS OF OBTAINED RESULT

This section focuses on the results of proposed algorithms. It shows performance of proposed algorithms by graphical representation. Resource utilization vs. load graph shows the performance of skew detection algorithm. A comparison is given between the traditional approach and our algorithm which shows that the proposed algorithm needs less number of migration to equalize loads between physical machines. As seen in Figure 1, the purple line are showing which machines are under-loaded. On the other hand, the red colored Z-scores represent which marks are over-loaded. As a result, the skew detection algorithm can be effectively used to detect the under-loaded and overloaded vm-instances. For the 40 vm-instances covered in our test case, 21 are shown in Figure 1. The output of all Z-scores for the 40 vm-instances used in our experiment are shown in Figure 2. As we considered 1.5 to be the upper threshold of the Z-scores, it was found that 7 vm-instances are over-loaded. On the other hand -0.75 was considered to be the lower threshold as per out experimentation assumptions. Hence, 9 vm-instances were found to be under-loaded. This shows that number of overloaded machines are lesser than number of under-loaded machines, which is caused mainly due to the type of workload.

Machine Name	Unused Resource, x	Mean, A	STD. DEV	Skewness	Deviation x-A	m2 987.95	m3 34878.3	Z-scores
1	5				-25	625	-15625	-0.79537608
2	6				-24	576	-13824	-0.76356104
3	9				-21	441	-9261	-0.66811591
4	12				-18	324	-5832	1.940717635
5	13				-17	289	-4913	-0.54085573
6	4				-26	676	-17576	-0.82719112
7	11				-19	361	-6859	1.527122073
8	3				-27	729	-19683	-0.85900617
9	1				-29	841	-24389	-0.92263625
10	13	30	31.43167192	1.123189049	-17	289	-4913	-0.54085573
11	4				-26	676	-17576	-0.82719112
12	16				-14	196	-2744	-0.4454106
13	10				-20	400	-8000	-0.63630086
14	17				-13	169	-2197	-0.41359556
15	18				-12	144	-1728	-0.38178052
16	2				-28	784	-21952	-0.89082121
17	5				-25	625	-15625	-0.79537608
18	7				-23	529	-12167	-0.73174599
19	23				-7	49	-343	-0.2227053
20	29				-1	1	-1	-0.03181504
21	45				15	225	3375	0.477225648

Figure 1: Sample test vm-machines with load detection

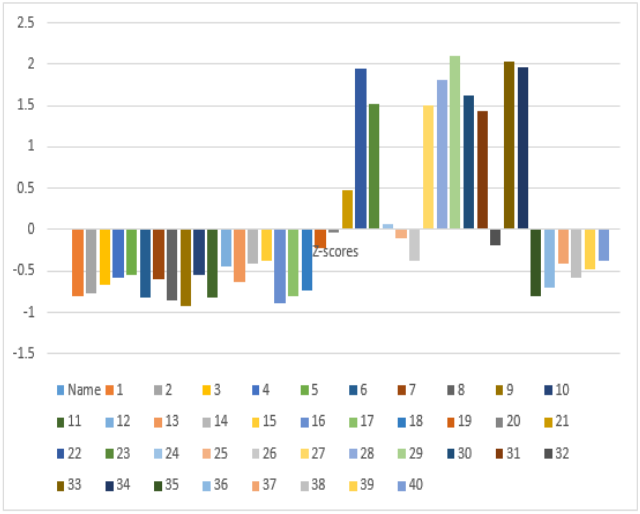
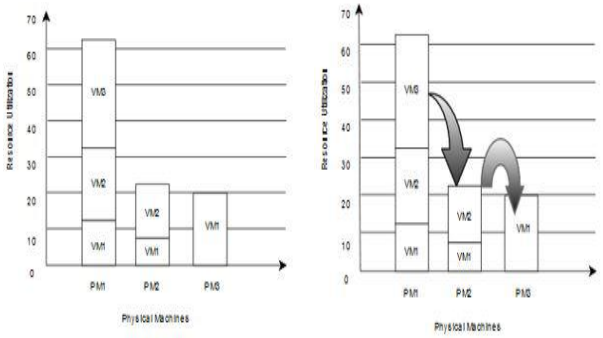
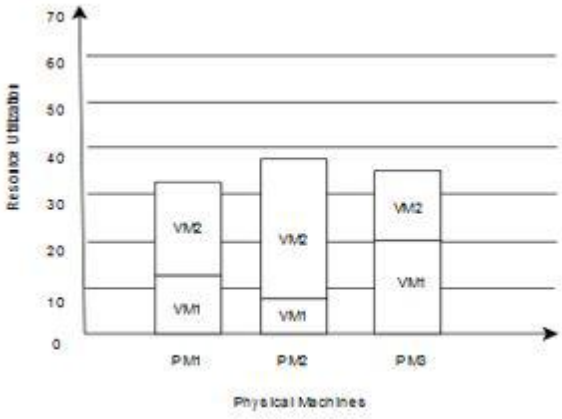


Figure 2: Machine wise identification of vm overloading and vm-under-loading



(a) Figure 1(a) ***(b) Figure 1(b)***



(c) Figure 1(c)

Figure 3: Resource utilization in the physical machines

5.1. Resource Utilization vs Load The main goal of proposed algorithm is to equalize resource utilization among PMs. Because uneven utilization of resources reduce overall performance of cloud as well as service expectation of users. Figure 3 shows a graph containing three physical machines. The Y-axis indicates resource utilization of

respected machines. It indicates resource consumption of three machines. Here, resource utilization of machine 1 is more than average. On the other hand, 2nd and 3rd machines resource utilization is little. Using the skew Ndetection algorithm a resource reallocation map is generated. Figure 3(b) indicates where to allocate the extra load of machine 1. Figure 3(c) presents the utilization of machines after implementing map.

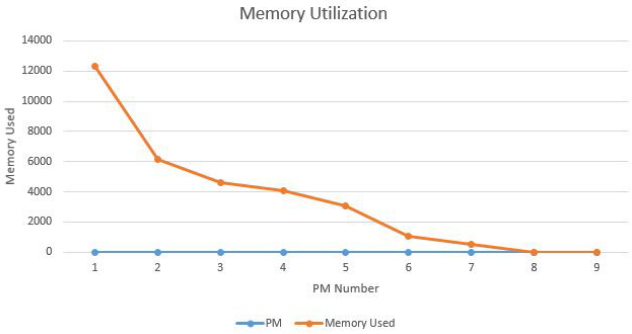


Figure 4: Memory Utilization Before Migration

5.2. Skew detection of cloud

Table 1: Skewness

Memory Uses of PMs	Skew Before Migration	Skew After Migration
12288	9*1.12	9*0.19
6144		
4608		
4096		
3072		
1024		
512		
0		
0		

Table 1 represent values of PMS memory utilization in the cloud. For this simulation two thresholds are selected to compare with the skewness. When skewness is greater than 1 or less than - 1, the resource distribution is considered as highly asymmetry otherwise it is moderately symmetry. Before migration of VMs the skew value of cloud is 1.1278. Which means the resource distribution is highly asymmetry in compared to the threshold. Figure 4 shows a chart with positive skewness. The graph is spotted in respect to memory utilization of all the PMs. So, a long tail is shown in positive side. It indicates that the cloud contains uneven resource distribution. So there exists a possibility of low performance as some of the PMs are highly loaded by its VMs. Figure ?? shows the result of resource distribution after implementing our algorithm. Now the skewness is 0.19. The resource distribution is now symmetric.

Table 2: Skew Detection execution Time

User	Average Execution Time (with skew millisec)	Average Execution Time(millisec)
500	2.9	2.9
1000	5.8	6.1
1500	11.5	13.6
2000	15.5	19.4
2500	24.89	28.7
3000	31.5	39.7

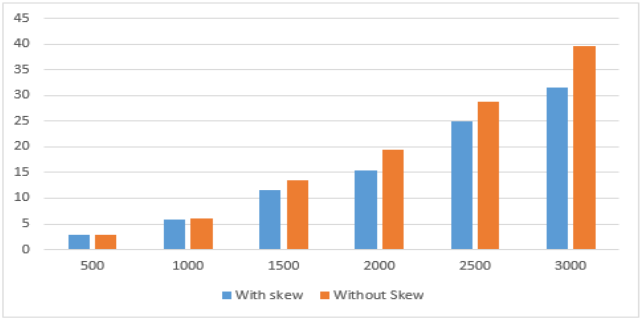


Figure 5: Comparison of execution times when migration is done with skewness and without using skewness

Table 3: Migration Number

User Number	Migration Number(with skew)	Migration Number
500	0	0
1000	2	3
1500	6	10
2000	8	15
2500	11	16
3000	18	25

5.3. Comparison with Traditional Approaches

5.3.1. Migration Number

As shown in Figure 5, the comparison of time taken for migration using the skewness and without skewness technique is highlighted. It is seen that for a variety of number of user processes ranging from 500 to 3000 processes, in all cases of higher number of user processes, the time taken for migrating vm-instances is significantly lower when skewness is used compared to that when skewness is not used. The performance of using skewness is justified when it is seen that for a higher number of users, the time difference for migration of all overloaded and under-loaded instances is almost 21% lower when skewness is deployed compared to when skewness is not deployed.

The performance of using skewness is personified when larger number of user processes are involved similar to real life cloud scenarios. The justification for this is that if

skewness is not used, then brute-force technique is needed to detect overloaded and under-loaded vm's and migrate those. On the other hand, when skewness is used, the vm-instances who are under-loaded and overloaded are first pinpointed which makes the job of the migration algorithm easier since it has a specific cluster of vm-instances to migrate. As it has a lesser number of vm-instances under consideration compared to brute-force, it can complete migration procedures much faster compared to when skew is not used.

Figure 5 shows that our algorithm significantly decrease the average number of over- loaded servers in the system. It presents that our algorithm is highly effective to proactively preventing overload of servers. Without load detection the algorithm tries to migrate VMs of PM as soon as the load of PM is below the threshold. Figure shows the migration num- bers of traditional approach and proposed approach for a certain amount of given load. It is clear that our approach significantly reduces the migration number of VMs. Table 3 shows the migration number in traditional approach and proposed approach. First column of the table contains VM numbers against a certain load. Second and third column contains migration numbers with and without load detection.

5.3.2. Average Execution Time

Table 2 shows average execution time against user load of the skew detection algorithm. First column shows user number. Second column shows average execution time of the pro- posed approach and the third column shows average execution time in traditional approach. From this table it is clear that the proposed approach needs less migrations to equalize the load in all PMS.

VI. CONCLUSION

This paper have proposed a methodology to detect overloaded servers which are running in multiple PMs. When any overloaded server is found, it dynamically maps VMs to PMs. The main goal of this work is to avoid failures due to overload by users, which implies to increase performance of a hybrid cloud computing infrastructure. To achieve this goal, three algorithms are developed- Failure Detection, Skew Detection and Resource Allocation. These algorithms give overloaded servers enough resources to continue their running task without any interruption. Resource are distributed in a way that resource utilization skew of every physical machine will be minimized. Which improves overall working efficiency of the cloud.

REFERENCES

[1] D. Barrett, Security architecture and forensic awareness in virtualized environments, Cybercrime and Cloud Forensics: Applications for Investigation Processes: Applications for Investigation Processes (2012) 129.

[2] A. M. Sampaio, J. G. Barbosa, Dynamic power-and failure-aware cloud resources allocation for sets of independent tasks, in: Cloud Engineering (IC2E), 2013 IEEE International Conference on, IEEE, 2013, pp. 1–10.

[3] Z. Gong, P. Ramaswamy, X. Gu, X. Ma, Siglm: Signature-driven load management for cloud computing infrastructures, in: Quality of Service, 2009. IWQoS. 17th International Workshop on, IEEE, 2009, pp.

1–9.

[4] H. Kim, M. Kang, S. Kang, S. Oh, A novel adaptive virtual machine deployment algorithm for cloud com- puting, in: proceedings of International Conference on Information Science and Industrial Applications (ISI 2012), Philippines, 2012.

[5] S. Prathima, S. S. Ali, Dynamic resource allocation using virtual machines for cloud computing envi- ronment.

[6] M. Arias, E. Rojas, J. Munoz-Gama, M. Sepu´lveda, A framework for recommending resource allocation based on process mining, in: 3th International Workshop on Decision Mining and Modeling for Business Processes (DeMiMoPˆa AZ15), 2015.

[7] S. Vijayakumar, Q. Zhu, G. Agrawal, Dynamic resource provisioning for data streaming applications in a cloud environment, in: Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on, IEEE, 2010, pp. 441–448.

An Approach to Measure Similarity of Software Projects at the Design Phase

[1] Asif imran
Institute of Information Technology, University of Dhaka

Abstract:-- **Estimation of software comparability is one of the ideal approach to utilize experiences of effectively developed software. Experiences obtained from previous projects can help software industries to deliver software project in a short pe- riod of time. Although various approaches have been proposed throughout years on software similarity that helps to utilize the previous projects knowledge, none of these are based on design diagrams. However, research on measuring software similarity based on design paradigm is expected to use the historic projects knowledge in early phase of software development. This paper proposes an approach of measuring similarity by developing a tool named Software Design Similarity Measurement Tool (DeSiMeT). DeSiMeT ascertains a similarity score between two software projects utilizing class diagram, sequence diagram and state transition diagram. An experimental analysis has been conducted by running the tool using seven software projects to verify the approach. The analysis of the tool appears in a precision of 0.83, recall of 1 and F-measure of 0.91 which concludes that the tool performs well as a novel work.**

Index Terms—**Software Similarity, Historic Project, Software Design, Class Diagram, Sequence Diagram, State Transition Diagram**

I. INTRODUCTION

In today's fast changing business environment, software industries attempt to rapidly develop their products so that they can speed up the delivery of their latest innovations to customers. This makes software development more chal- lenging as software developers need to design, implement, and test complex software systems as early as possible. As a result, software companies are in search of some solutions that can help to deliver good quality and error free software in right time. Experiences obtained from historic projects can help software industries to adopt shorter release cycles. In this regard, measurement of software similarity is one of the best way to use experiences of already developed software [1]. Proper identification of similar software helps to select development methods, design patterns and reliability testing models as well as is applicable in the areas of data mining, software testing, plagiarism detection and software security [2]. This helps software industries to decrease their efforts, costs and time of development cycle [3].

In this research, a method to measure similarity between software projects were proposed based on software design paradigm. The proposed method computes similarity between intended software project and historic software projects to use previous knowledge. More precisely, the problem can be addressed by the following questions:

- How to utilize the experiences of historic software

Here, we use design paradigms of different projects to measure similarity among those. The proposed method solely focuses on the quantitative similarity measure. For this pur- pose, we use different design diagrams i.e. class diagram, sequence diagram and state diagram. To measure similarity score, a tool named Software Design Similarity Measurement Tool (DeSiMeT) is developed.

II. RELATED WORK

In the software engineering literature, a less work has been done that addressed the issue of measuring similarity between two software projects. Some works have been done based on fuzzy logic, Euclidean Distance, Graph Matching and Source code corresponded but these are not enough to solve the existing problems. Two approaches for measuring similarity between software projects based on fuzzy C-means clustering and fuzzy logic were presented in [7]. The proposed approaches overcame the problems of nearest neighborhood techniques. First approach was developed based on identification features of fuzzy sets and second approach was based on partition matrix that is obtained by fuzzy C-means. They stated that first approach outperforms second approach based on some experimental results. This approach is not applicable for linguistic values and only suitable for numerical and categorical data. Some specific research has been done for computing difference between class diagrams. A generic difference algorithm proposed for computing similarity of two UML models which were encoded in XML files from design diagram [14]. The implemented algorithm performed well on runtime for small documents but not good for a large documents. A

comparative result were presented using basic graph by denoting node and edge. In this approach, at first the elements of each document were detected and then calculated similarity by a defined function that worked with some predefined criteria. Weight was defined for each criteria in a way that may mislead to a missed correspondences. To optimize cost-resource for cloud environment an empirical analysis investigation was proposed in [13]. Authors presented a comparison analysis between open source cloud and organizational cloud to increase the performance of open source cloud by optimizing cost and resource. However, performance,of open source cloud can be enhanced by applying software similarity approach to find out the best result. Although various approaches have been proposed throughout years, none of these are based on design diagrams. However, research is expected to use the historic projects knowledge in software development. In this research, we mainly focused on the design phase of SDLC.

III. PROPOSED APPROACH

The novelty of this research is to measure similarity of software projects based on some design diagrams in the design phase of SDLC. Firstly, in order to measure design similarity, we consider several design diagrams: (a) Class Diagram (CD) (b) Sequence Diagram (SD) and (c) State Transition Diagram (STD). These diagrams cover the overall design of any object oriented software project to the maximum extent. Later on, the diagrams are parsed by an open source automated tool i.e. StarUML and converted to XML format. Then the XML files are parsed by XML parser and compared using the structure and some comparison criteria. To demonstrate the approach, we have developed a tool and named it as Software Design Similarity Measurement Tool (DeSiMeT).

A. Overview of DeSiMeT

Top level view of the proposed tool's architecture is shown in the Fig. 1. The tool developed in a way that measured similarity between current project and historic project based on design diagrams those are declared in the previous section. Selected design diagrams are converted to XML format using an open source converter as the tool takes XML file as input.

B. Similarity Measurement Approach

Main purpose of this research is to find out the best similarity match of a current project with some historic projects. The proposed approach to measure similarity between software projects based on design depends on different UML diagrams which describe the external behavior of a project.

a) Structural Matching: In the view of a class diagram, the whole system can be compared by relationships of classes. At first, class diagrams of a system are converted to XMLs and inputted to the tool. Then elements are parsed by an XML parser to proceed the next step. Class diagrams are considered as a graph where classes are denoted as node and relationships are denoted as edge.

($A \xrightarrow{Ag} D$, $A \xrightarrow{As} C$, $B \xrightarrow{Ag} C$, $D \xrightarrow{As} C$, $D \xrightarrow{G} E$ and $C \xrightarrow{As} E$)

Historic class also contains five classes and relation among the classes are:

($B \xrightarrow{Ag} A$, $B \xrightarrow{As} C$, $A \xrightarrow{Ag} C$, $A \xrightarrow{G} G$, $A \xrightarrow{G} F$ and $C \xrightarrow{As} F$)

b) Criteria Matching:

In the view of class each of the classes a similarity can be measured by comparing some criteria that is presented by Ketle et al. [14]. Due to the generic approach a function is defined by setting some criteria i.e. number of operations and number of attributes. In every case, graph matching can not provide accurate similarity score. Thus, quantitative value need to be considered as an important fact of measuring similarity of class diagram. A criteria matching algorithm is developed that is shown in Algorithm 2. For measuring criteria similarity, a class of first diagram is compared with all of the classes of the second diagram and stored.The same class can not be matched with any other class of the second diagram. The criteria similarity score is Ncalculated (Algorithm 2 Line 19). Finally, the total similarity between two class diagrams are measured by integrating the score of structural matching and criteria matching.

Algorithm 1 Structural Matching

```
1: input: XMLs (CD, SD, STD)
2: output : finalScore_structural
3: initialize mat[u][v] ← 0
4: initialize setofrelations ← R
5: for edges(u, v) in diagrams do
6:     mat[u][v] ← getValue()
7: end for
8: procedure GETVALUE
9:     if relations ∈ R then
10:        setValue = value
11:    end if
12: end procedure
13: finalScore_structural ← matchBFS()
```

1) Similarity of Sequence Diagram: SD is used to show the interaction among the objects in a given scenario based on a time sequence. It is a logical view of a system under development that is typically associated with the use cases. A sequence of messages are exchanged among the

participating instances through the interactions of the system and some actors or different subsystems or classes [17]. Sequence diagram is considered as a part of project to measure similarity as it represents the dynamic interactions of classes in execution [11]. Similarity measurement of sequence diagram is similar as class diagram using two phases such as :structural matching and criteria matching. The phases are described in the following sections.

a) Structural Matching: For structural matching, sequence diagram is also considered as a graph to measure similarity like class diagram. At first, the UML sequence diagrams of the system are converted to XMLs and inputted to the tool. The XMLs are parsed by the XML parser and a n n matrix is generated like class diagram. The lif elines are considered as node and sendMessage and replyMessage are considered as edge of the graph for generating matrix. Fig. 3 presents a sample example of sequence diagram and fig. 4 presents the generated matrix of fig. 3. For generating matrix, the values of edges are set as prime number to keep track the multiple call between two lif elines. Different values are set for sendMessage and replyMessage where sendMessage is defined by 2 and replyMessage is defined by 3 because these two edges are not same and address a different meaning for sequence diagram. Generated matrix is compared by the same algorithm that is used in class diagram is presented in Algorithm 1. The algorithm describes that all matrix is generated (Algorithm 1 Line 6). Value of edges are set based on message type between lif elines (Algorithm 1 Line 9-10). Similarly, for measuring similarity between two matrices of

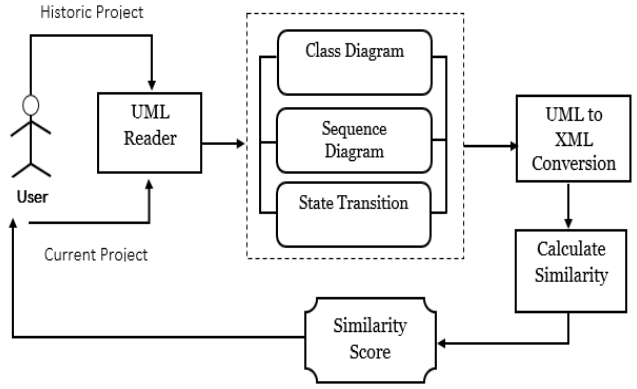


Fig. 1: Overview of Framework DeSiMeT

	A	B	C	D	E
A	0	0	2	5	0
B	0	5	0	0	0
C	0	0	0	0	2
D	0	0	2	0	3
E	0	0	0	0	0

Current Project

	B	C	A	G	F
B	0	2	5	0	0
C	0	0	0	0	2
A	0	5	0	3	3
G	0	0	0	0	0
F	0	0	0	0	0

Historic Project

Fig. 2: Generated Matrices of Fig. 2

sequence diagram, a customize Breath First Search algorithm is used (Algorithm 1 Line 13) that is proposed by smith et al. [20]. b) Criteria Matching: Criteria matching of sequence diagram is a statistical similarity measured based on some criteria i.e. number of lif elines, number of sendMessage and number of replyMessage. Algorithm 2 is used to measure criteria matching like class diagram. However, a function is defined for this purpose and a sequence diagram from first project is compared with all of the sequence diagram of the second project and store the max value. If a match is found in the next step then it is ignored as it is already compared. Then, the similarity score is calculated (Algorithm 2 Line 19). Finally, similarity of sequence diagrams are measured by integrating the score of structural matching and criteria matching.

2) Similarity of State Diagram: STD describes the behavior of a system using states of the system and transitions between states [18]. It shows different states of an entity as well as how an entity respond to events by changing the states. Similarity of state diagram is measured into two phases like class diagram.

a) Structural Matching: State diagram is considered as a graph to measure similarity where the states are considered

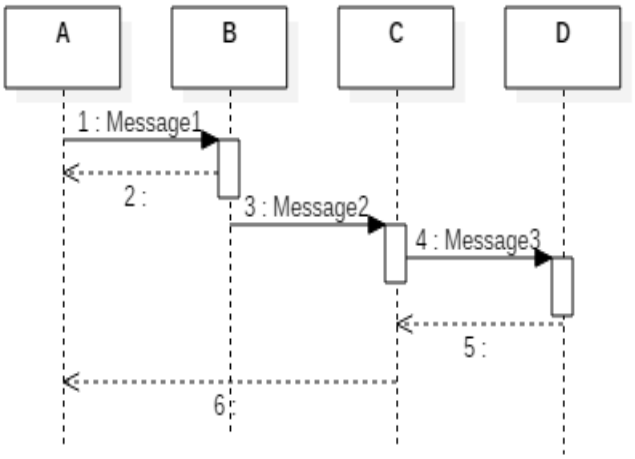


Fig. 3: Example of Sequence Diagram

	A	B	C	D
A	0	2	0	0
B	3	0	2	0
C	3	0	0	2
D	0	0	3	0

Fig. 4: Generated Matrix of sequence diagram

as node and transitions are considered as edge of the graph. At first, the UML state diagrams of the system are converted to XMLs and inputted to the tool. Then, the XMLs are parsed by the XML parser and a n n matrix is generated like class diagram. The value of edges are set as: start state to

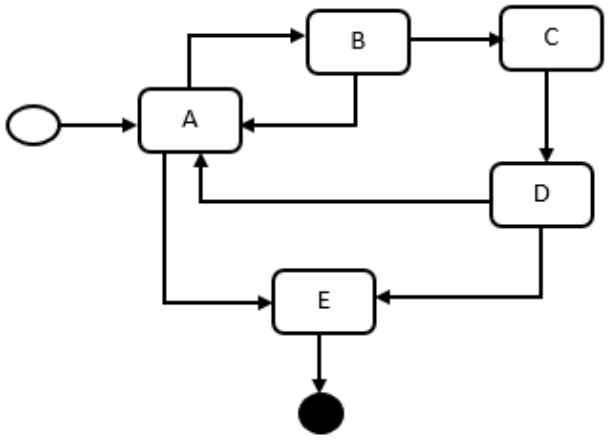


Fig. 5: Example of State Diagram

general state is defined by 1, general state to general state is defined by 2 and general state to final state is defined by 3. Algorithm 1 is also used for structural matching of state diagram. Similarly, for generating matrix the function is used (Algorithm 1 Line 6). The value of edges are set based on transition type between state that is shown (Algorithm 1 Line 9-10). The matrices are compared like class diagram that is used (Algorithm 1 Line 13).

b) Criteria Matching: Some criteria i.e. number of states, and number of transitions are defined for measuring the statistical similarity of two state diagrams. A function is defined based on these criteria and the similarity is computed using the algorithm 2 like other two diagrams. However, a state diagram from first project is compared with all of the state diagram of the second project and store the max value. If a match is found in the next step then it is ignored as it is computed. The similarity score is calculated (Algorithm 2 Line 19). Then, the similarity of state diagrams are measured by integrating the score of structural matching and criteria matching. Finally, total similarity score of two projects is calculated by integrating the similarity score of class diagram, sequence diagram and state transition diagram.

IV. CASE STUDY

The goal of this case study is to evaluate approximation of the proposed approach. The experiment have been conducted on some design diagram of software projects. A tool DeSiMeT has been implemented in java for this purpose.

A. Dataset

The analysis was performed on 7 different software project requiring different diagrams those are used in this research. These projects have been collected from the student of Institute of Information Technology, University of Dhaka. These projects are Inventory Management System (IMS), Student Management System (SMS), AmaderChakri.com (AC.com), Program Office Management System (POMS), Library Circulation System (LCS), Cricket Circle (CC) and Cloud Portal(CP). The project set is converted to XMLS using StarUml before running DeSiMeT as it takes XMLs as input. Table I shows the dataset in details. Project name, number of classes in class diagram, number of sequence diagrams and number of state diagrams of each project are presented in this table.

B. Study Result

For experimental result, dataset projects were run using DeSiMeT. The similarity score was measured between current project and historic project. In the dataset, IMS is the selected as current project and other projects as historic projects. For each historic projects, similarity values of class diagram, sequence diagram and state diagram were measured by comparing with current project that is presented in Table II. Here, first column presents the current project and second column presents historic project. Similarity values of Class Diagram (CD), Sequence Diagram (SD) and State Transition Diagram (STD) are also presented that were obtained from DeSiMeT. In the last column, similarity score of two projects are presented that is calculate d from the average value of CD, SD and STD.

C. Analysis

For the justification of DeSiMeT, an empirical analysis was performed. Table III presents the expected result and actual result of this tool. The expected similarity result was identified from a manual analysis that was performed by some software design experts. The actual result was generated from DeSiMeT and the similar project were chosen based on a threshold value that is greater than or equal 0.6 (threshold 0:6).

TABLE I: Result Analysis

CP	HP	Actual Result (threshold $\geq$ 0.60)	Expected Result
IMS	SMS	Yes	Yes
	AC.com	Yes	No
	POMS	Yes	Yes
	LCS	Yes	Yes
	CC	No	No
	CP	No	No

 ISBN: 978-81-932966-9-1

Now, from the actual and expected result that is shown in the table III, the precision and recall of DeSiMeT can be measured. Let, tp =true positive, fp =false positive, fn =false negative. From Table I, tp = 5, fp = 1, fn = 0. Thus,

$$Precision = \frac{tp}{tp + fp} = \frac{5}{5 + 1} = 0.83$$

As, DeSiMeT provides 1 false negative result, it possesses the maximum recall. Using the precision and recall, the F-measure or the balanced F-score (F1 score) can be calculated.

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = 2 \cdot \frac{0.83 \cdot 1}{0.83 + 1} = 0.91$$

V. CONCLUSION

In this paper, a generic approach for measuring similarity between software projects is proposed. The approach is based on a series of design diagrams those are converted to XMLs. One of the challenges to measure similarity between complex types which are represented by XML, is handled by different techniques for getting the best scores as described previously. The task is performed in three steps: similarity of class diagrams in first step, similarity of sequence diagram in the second step and similarity of state diagram in third step are measured. A prototype tool DeSiMeT is developed to prove the feasi-bility of the approach. A case study is presented that evaluates the applicability of the approach. In this case study, seven projects were selected to perform the empirical study. For DeSiMeT, the precision, recall and F-measure were calculated that possesses a precision of 0.83, recall of 1 and F-measure of 0.91. The result shows that the proposed approach performs well as a novel work. In this approach, only three UML diagrams are considered for similarity Measurement, in future more diagrams will consider for performing better result. This work directs to our next work to software reliability model selection based on similarity score between current project and historic project.

REFERENCES

[1] Zeng, Jintao, et al. ”A prototype system of software reliability prediction and estimation.” in proceedings of the Third International Symposium on Intelligent Information Technology and Security Informatics (IITSI), 2010.

[2] Reungsinkonkarn, Arun. ”Hierarchical similarity measurement model of program execution.” in proceedings of the Software Engineering and Service Science (ICSESS), 2013 4th IEEE International Conference on IEEE, 2013.

[3] Salami, Hamza Onoruoiza, and Moataz A. Ahmed. ” UML artifacts reuse:
state of the art.” arXiv preprint arXiv:1402.0157 (2014).

[4] Fenton, Norman, and James Bieman. Software metrics: a rigorous and practical approach. CRC Press, 2014.

[5] Idri, Ali, and Alain Abran. ”A fuzzy logic based set of measures for software project similarity: validation and possible improvements.” Soft- ware Metrics Symposium, 2001. METRICS 2001. Proceedings. Seventh International. IEEE, 2001.

[6] Idri, Ali, and Alain Abran. ”Evaluating software project similarity by using linguistic quantifier guided aggregations.” IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th. Vol. 1. IEEE,
2001.

[7] Azzeh, Mohammad, Daniel Neagu, and Peter Cowling. ”Software project similarity measurement based on fuzzy C-means.” Making Globally Distributed Software Development a Success Story. Springer Berlin Heidelberg, 2008. 123-134.

[8] Shepperd, Martin, and Chris Schofield. ”Estimating software project effort using analogies.” Software Engineering, IEEE Transactions on 23.11 (1997): 736-743.

[9] Tibermacine, Okba, Chouki Tibermacine, and Foudil Cherif. ”A practical approach to the measurement of similarity between WSDL-based web services.” Revue des Nouvelles Technologies de l’Information (2014):
03-18.

[10] Nahar, Nadia, and Kazi Sakib. ”SSTF: A novel automated test gener- ation framework using software semantics and syntax.” Computer and Information Technology (ICCIT), 2014 17th International Conference on. IEEE, 2014.

[11] Nahar, Nadia, and Kazi Sakib. ”Automatic Recommendation of Soft- ware Design Patterns Using Anti-patterns in the Design Phase: A Case Study on Abstract Factory.” 3rd International Workshop on Quantitative Approaches to Software Quality.

[12] Yamamoto, Tetsuo, et al. ”Measuring similarity of large software sys- tems based on source code correspondence.” Product Focused Software Process Improvement. Springer Berlin Heidelberg, 2005. 530-544.

[13] Imran, Asif, Alim Ul Gias, and Kazi Sakib. ”An empirical investigation of cost-resource optimization for running real-life applications in open source cloud.” Procedings of High Performance Computing and Simulation (HPCS), 2012 International Conference on. IEEE, 2012.

[14] Kelter, Udo, Jrgen Wehren, and Jrg Niere. ”A Generic Difference Algorithm for UML Models.” Software Engineering 64.105-116 (2005): 4-9.

[15] Szlenk, Marcin. ”Formal semantics and reasoning about uml class diagram.” Dependability of Computer Systems, 2006. DepCos- RELCOMEX’06. International Conference on. IEEE, 2006

[16] Chen, Peter. ”Entity-relationship modeling: historical events, future trends, and lessons learned.” Software pioneers. Springer Berlin Heidel- berg, 2002. 296-310.

[17] Firley, Thomas, et al. ”Timed sequence diagrams and tool-based anal- ysisa case study.” UML99 The Unified Modeling Language. Springer Berlin Heidelberg, 1999. 645-660.

[18] Klein, Cornel, Christian Prehofer, and Bernhard Rumpe. ”Feature Spec- ification and Refinement with State Transition Diagrams.” FIW. 1997.

[19] Dong, Jing, Dushyant S. Lad, and Yajing Zhao. ”DP-Miner: Design pattern discovery using matrix.” Engineering of Computer-Based Sys- tems, 2007. ECBS’07. 14th Annual IEEE International Conference and Workshops on the. IEEE, 2007.

[20] Smith, S., and D. R. Plante. ”Dynamically recommending design pat- terns.” SEKE. 2012.

Sentiment Analysis of Reviews Using NIPA and Lexicon Approach

[1] Tirath Prasad Sahu , [2] Sarang Khandekar
[1][2] Department of Information Technology National Institute of Technology Raipur, India
[1] tirsahu.it@nitrr.ac.in, [2] sarang.sarang.khandekar@gmail.com

Abstract- Sentiment analysis can be a very useful aspect for the extraction of useful information from text documents. The purpose of sentiment analysis is how people think for a particular online review i.e. product review, movie review etc. Sentiment analysis is the process of classifying the reviews as positive or negative. The web is enriched with huge amount of reviews that can be analyzed to make it meaningful. Sentiment analysis can prove to be useful specifically for marketing and business as it greatly influences the people and market. This paper presents the use of lexicon resources for sentiment analysis of different publically available reviews. The presented work is divided into divided into three steps. In the first step, the polarity shift is handled by negations (N), intensifiers (I), punctuations (P), and acronyms (A) are taken into consideration. In the second step, the opinion words are extracted to compute the sentiment score. In the third step, different machine learning algorithms are applied and the proposed algorithm is evaluated using the various parameters like accuracy, precision, and recall. The experimental result shows that the proposed approach is remarkable in identifying the sentiments of reviews.

Keywords: Sentiment Analysis, Machine learning, Performance Measures; Online Reviews, Lexicon Resource.

1. INTRODUCTION

Sentiment analysis is becoming a trending topic for many researchers. Sentiment analysis is a special case of text classification which aims to categorize an opinion based on polarities of these opinions for e.g. positive or negative. People can express their opinion on the web in the form of reviews. These reviews can be from different fields such as product review, movie review, and also tweets from twitter can be a review. Basically sentiment analysis is done to express what a person thinks for a particular product, movie etc.

In recent years lots of people are expressing their opinion on the web. Everyday millions of people express what they think in the form of review using online platform such as blogs, forums, social networking sites like twitter etc. These platforms help people to connect to other people whom they don't even know and get the opinion from them. Basically these platforms serve as intermediately between the end user and the service provider. From the end users perspective these platforms helps to get an idea about the product and from the service provider point of view it helps raise the standard of their product and services.

The data on the web is increasing everyday as many people are using it a platform to express their reviews and opinion. Because of this it becomes a very hectic task for people to take instant decision about the reviews for a particular. People become confuse on which review they can trust and on which review they cannot. Due to this problem an automatic system designed to make the process of analysis, summarization and classification easy. The general approach for sentiment analysis follows the Bag of Word (BOW) Approach. In this approach a document is divided into bag of words, these are then mapped into a feature vector and finally classified by the machine learning algorithms such as support vector machine (SVM), Naïve Bayes etc. But this BOW approach failed to generate the desired results. BOW does not capture word order and semantic relation between these words. These measures can play an important role in polarity computation. Many researchers have been carried in the field of sentiment analysis which tries to improve BOW in combination with linguistic knowledge [1-8]. However, due to a few constraints in BOW, all of these efforts showed negligible effect in improving the classification accuracy. Another technique which can be used for sentiment classification one of which is supervised technique. In supervised approach, lexicon based approach is used in which the lexical resource such as SentiWordNet is considered which assigns a polarity score to each word in the review, these scores can be further to determine the overall polarity of the review.

One of the most prominent issue in sentiment analysis is polarity shift problem. Many approaches have been suggested in the literature to overcome the polarity shift problem [9-14] . However, most of them required either complex linguistic knowledge or extra human annotations. Such high-level dependency on external resources makes the system difficult to be widely used in practice.

Different levels in which Sentiment Analysis can be achieved are aspect level, sentence level and document level. In document level Sentiment Analysis, the approach tries to find out if the complete document conveys a positive or a negative characteristic. In sentiment analysis at sentence level opinions are expressed for a single sentence or review. The proposed method use sentiment analysis at sentence level using the lexical resource SentiWordNet 3.0 to decide the overall polarity of the review. Lexicon based Sentiment Analysis concerns with the presence of different words in the dictionary. Different features present in these lexicons contain can be part of speech words, respective sense number, sentiment value, and subjective words. Using these features the polarity of the review is determined. The assumption is that the accuracies generated by proposed approach increases with increase in training data.

2. RELATED WORKS

In this paper, sentiment analysis is done on different textual reviews. The proposed approach makes the use of lexicon resource to calculate the polarity of the review along with the machine learning approach. A lot of researches have been done on sentiment analysis in recent years. The center research areas are sentiment classification and feature-based sentiment analysis. Sentiment classification explores ways to characterize each review as positive or negative. Research effort on classification at the document level involves [1, 6, 10, 15-19].

Most of the methods are based on sentence and document level classification which involves reorganization of opinion words. Corpus based method and dictionary based method are the two type of method that are basically used. In corpus based method co-occurrence patterns of words are found out to determine the sentiments of words. The research effort in [19-21] shows this work. In dictionary based method synonyms and antonyms have been used from WordNet to decide word sentiments of opinion words. Researches carried out for this approach includes [22-25]. The work in [13] presents the idea of summarization and sentiment analysis. Lexicon-based approach is being used to decide whether the review expressed is positive or negative. The authors of [25] used a similar approach. In [26], Authors have improved these approaches by a more sophisticated approach called relaxation labeling.

The authors of [20] Studies sentence level subjectivity classification which decides whether a sentence is a subjective sentence or a sentence based on facts. Researches in Sentence level sentiment is done in [20, 24-28]. These approaches are different from our approach; at first the sentence is preprocessed to remove unnecessary data from it. Secondly the polarity detection is performed with the use of lexicon dictionary and lastly the result of this is applied to the learning algorithm.

3. PROPOSED APPROACH

Lexicon based approach is one of the approach which can be used for sentiment analysis. In the proposed approach each word in the sentence is taken into account and then the proposed approach is applied to compute the weight of the opinion words and finally the overall polarity of the sentence is computed. For the computation of the polarity the lexicon resource SentiWordNet is used. The main work on this paper focuses on determining the polarities of the sentences using the following steps: 1) Selecting Features 2) Applying NIPA's 3) Scoring Reviews 4) Machine Learning Classification. The review datasets from different fields are used which are available publically. Finally the reviews are classified into two classes i.e. positive (1) and negative (0). The proposed approach is depicted in figure 1.

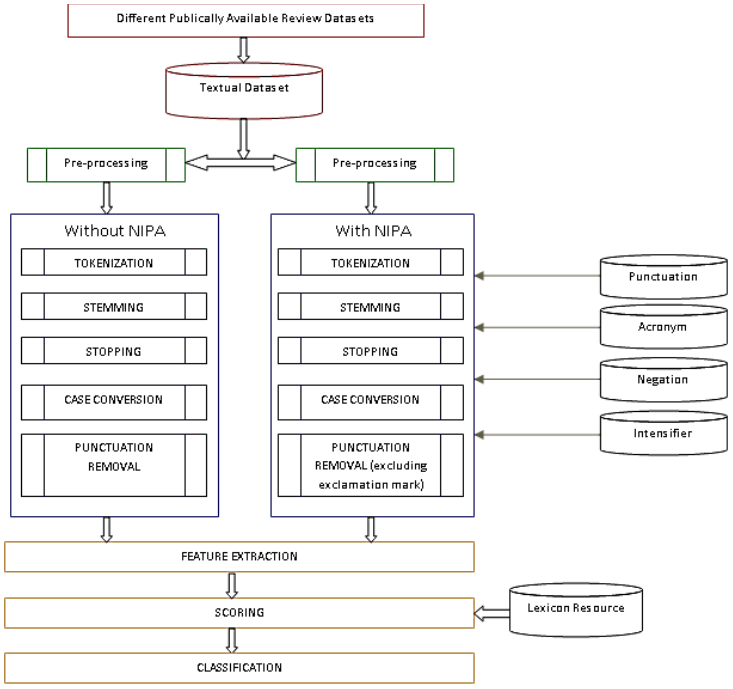


Figure 1: Proposed Approach

3.1 Pre-Processing

The dataset used are generally unwanted and noisy. The main reason for pre-processing is to remove these unwanted and noisy data. Pre-processing is basically done to speed up the process of categorization of the polarities of the sentences or reviews. In this paper, the pre-processing consists of the following steps.

(i) Tokenization: Tokenization is the process which breaks

stream of text into a list of words.

(ii) Stemming: Stemming makes the process of sentiment analysis easy. This process reduces the words to their root words. For e.g. argue, argued, argues, arguing all are reduced to their root word argu.
iii) Stopping: Stop word basically includes verbs, articles, preposition and so on. These stop words are not involved in analyzing sentiments and therefore these words are removed from the reviews. These words are created as the requirement occurs. For e.g. the review "Twilight is the best series that I have ever seen" will be processed as "Twilight best series ever seen".
(iv) Parts of Speech Tagging (POS): In this process all the words in the text are marked in accordance to their particular POS like verb, adverb, pronoun, noun, determinant and adjective. For e.g. "Berry Alan likes to be the fastest man alive and saves the city from meta humans" is tagged as "Berry (Adverb) Alan (Adverb) likes (Verb) to be (Verb) the (Determiner) fastest (Adjective) man (Noun) alive (Adjective) and (Conjunction) saves (Verb) the (Determiner) city (Noun) from (Preposition) meta (Adjective) humans (Noun) ".
(v) Case Conversion: This process converts the reviews into lower case.
(vi) Punctuation Cleaning: The presence of punctuation in the text do not gives any useful information. Therefore in this step all the punctuation characters are removed from the review.

3.2 Feature Extraction

After the completion of pre-processing step, the further steps are to analyze the reviews to find out the entire pattern that affects the process of computation of polarities of review. Here the words with some opinion which are present in SentiWordNet are considered as features in the process of polarity computation. Every review is thoroughly analyzed, from which only those words which helps in the computation of the polarity are chosen.

3.3 Lexicon Resource

A lexical resource includes a list of terms along with the respective polarity of each term. Many Lexical resources have been developed and are made publically available. In this paper, we have used SentiWordNet 3.0 [29]. Especially for the purpose of sentiment analysis SentiWordNet3.0 is designed. An objective score, a positive and a negative score is allocated to ever synset in this lexicon resource. Various semi-supervised ternary classifiers are used in order to calculate these scores. These ternary classifiers have a special capability to decide whether a term is objective, positive or negative.

3.4 NIPA Approach

(i) Coping with Punctuation

In polarity computation a crucial role can be of Punctuations symbols but these punctuation are often being ignores. The symbol 'Exclamation (!)' should also be taken into consideration. People use to express a surprise, astonishment and other such strong emotions by using exclamation mark. So while calculating polarity this measure can play a very vital role. Positive weight of τ is assigned to it as it shows a positive emotion.

(ii) Coping with Acronym

While giving a review many people use the acronyms. In order to save the time and effort people use the acronyms. Acronyms are short forms of a complete term. Before calculating the polarity of a sentence, the acronyms should be considered. Top 100 acronyms are used from netlingo [32] which are generally used by people when a review is given. The acronyms from netlingo are taken into account and then these acronyms are converted into their actual meaning. While computing the polarities, generally these acronyms are ignored this is the reason why this process in important. For e.g. GR8. This acronym is largely used by the person which is the short form for Great. This acronym GR8 is generally ignored while calculating the polarity of a review, which may affects the calculation phase. Every word in the acronym is considered while calculating the polarity when these acronyms are converted into their actual meaning.

(iii) Coping with Negation

While deciding the polarity of a review, the major problem that can arise is polarity shift problem. While performing sentiment analysis coping with negation is a major problem. The negation terms such as not, cannot, would not etc are considered as stop words which are often ignored by many researchers. A review can completely change its meaning when negation terms are ignored. If negation terms are ignored a completely a review can be interpreted with a completely different meaning. For e.g. let us consider the following review "This movie is not liked by anyone" The polarity of the word 'liked' is positive and the polarity of the word 'not', is negative. It can easily be interpreted that the overall polarity of this review is negative. But many researchers ignore the negative polarity term. They consider the word 'not', as a stop word and do not include this word while determining the polarity. But they take the review as "This movie is liked by everyone". When the word like 'not' is ignored, the review changes its perception from negative to positive. If the researchers consider the word 'not', the review remains intact with the actual perception i.e positive. This problem is overcome by the proposed approach. The negation term is replaced by a '^' symbol

whenever a review encounters a negation term. A stop words list is made which contains negation words. Review changes its polarity when these negation terms are not considered. For this modifications are done in the lexicon resource. These negation words are assumed as negative words and hence assigned a negative weight of -1 in the lexicon resource.

(iv) Coping with Intensifier

The sentiment of a review is amplified or diminished by the intensifiers. While deciding the polarity of a review a critical role can be played by intensifiers. The proposed approach considers some intensifier which has been used by people while giving a review. The intensifiers that are taken are 'So', 'Absolutely', and 'Totally'. Through literature [33], it is inferred that 'Totally' is used along with the negative terms so a negative weight –λ is assigned to it. The term 'So' is used along with the positive term so positive weight +λ is allocated to it. Absolutely gives a neutral perception therefore a neutral weight is allocated.

3.6 Allocating Weights to PIN's

The proposed method has done some assumptions and allocated polarity to Punctuation 'Exclamation', Negation and the Intensifier, which are then updated in the lexicon resource, these measures are then considered while computing the polarity. The polarity allocation is as follows:

$$Weights = \begin{cases} \delta \; If \; Words = Negation, \\ \lambda \; If \; Words = intensifier, \\ \tau \; If \; Word = \; Exclamation, \end{cases}$$

Where δ= -1, λ= 1 or -1 and τ= 1.

β = -1 since the negation words are assumed as negative words and therefore allocated a negative weight of -1 in the lexicon.

γ = -1 or 1 since assumptions are done and inferred that 'Totally' is used along with the negative terms so a negative weight of -1 is allocated to it. The term 'So' is used along with the positive terms so a positive weight +1 is allocated to it. Absolutely gives a neutral perception therefore a neutral weight is allocated.

α= 1 since assumption are made that exclamation symbol is generally used by the people to express a surprise, astonishment and other such strong emotions.

3.6 Algorithm Applied

In the proposed approach the following ALGORITHM 1 is used to compute the sentiment polarity.

Algorithm 1: Algorithm for sentiment polarity detection

INPUT: $Review = \sum_{i=1}^{j} R\text{eview}_i$

1. Initialize every variable TermScore(α), TermSentimentScore(β), EventualSentimentScore(γ) to 0.

2. Compute EventualSentimentScore(γ) based on TermScore(α) and TermSentimentScore(β).

3. **While** end of sentence **do**

4. Compute TermSense

$$\text{TermSense} = (\text{Positive Synset}) - (\text{Negative Synet}) \quad (1)$$

5. Compute TermScore(α)

$$TermScore(\alpha) = \sum_{k=1}^{n} \frac{TermSense(k)}{TermFrequency(k)} \quad (2)$$

6. Compute TermSum

$$TermSum = \sum_{l=1}^{n} \frac{1}{TotalTermSense(l)} \quad (3)$$

7 Calculate TermSentimentScore(β)

$$TermSentimentScore(\beta) = \sum \frac{TermScore}{TermSum} \quad (4)$$

8. Calculate FinalSentimentScore(μ)

$$FinalSentimentScore(\gamma) = \sum_{\beta=1}^{m} TermSentimentScore \quad (5)$$

9. **END**

10. According to γ allocate the polarity

The table 1 describes the variables used in the algorithm along with their definition

Table 1: Variables used and their definition

Variable	Definition
TermSense	TermSense is the difference between the positive and negative weighted terms in accordance to the number of different perception allocated to that term in the lexicon resource.
TermScore(α)	Termscore is the aggregation of TermSense to the total number of different perception allocated to a term.
TermSum	TermSum is the weighted average of total number of different perception allocated to a term.
TermSentimentScore(β)	TermSentimentScore is the aggregation of the TermScore to the TermSum of a term.

FinalSentimentScore(γ)	FinalSentimentScore is the aggregation of all TermSentimentScore of each term that are in a review.

3.7 Classifiers

In text categorization machine learning techniques such as naive Bayes (NB) [30] and support vector machines (SVM) [31] are proven to be the best. Pang et al. [6] shows how feasible are these classifiers in sentiment classification. The work done by pang & Lee on the movie review dataset, using Naïve Bayes, and Support Vector Machine gives a better result than those given by human generated baselines. Therefore in this work, these three classifiers are used and using these, the evaluation parameters are calculated such as accuracy, precision and recall.

4. OUTCOME AND ANALYSIS

4.1 Dataset Description

Various products, movies and social networking review from twitter are taken as datasets which are available publically on the web. Products review such as Books, Movies, Hotels, Car are taken from SFU review dataset. Positive review and negative review are evenly divided in this review corpus. Movie review datasets has been taken from Pang and Lee [34]. This dataset is evenly divided into positive and negative reviews. 700 positive reviews and 700 negative reviews are present in this dataset. Dataset from epinion [35] has been considered which contains reviews for Automobiles. The dataset is also evenly divided into positive and negative reviews. This dataset is distributed as 692 positive and 692 negative reviews. I phone dataset is also extracted from Twitter [36]. This dataset is distributed into 240 positive and 239 negative reviews. Two classes are classification is used for these reviews i.e a positive and a negative class. A positive class will be having a score of $\boldsymbol{Score > 0}$ and a negative class will be having a $\boldsymbol{Score < 0}$. This dataset is divided into 66% of training data and remaining is the test data. 10 fold cross validation is also performed with this data.

4.2 Evaluation Parameters

Table 2 represents the contingency table to find out the evaluation measures like accuracy, recall and precision.

1. Precision

Precision is defined as the ratio between true positive (TP) and both true positive (TP) and false positive (FP). The equation is:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (6)$$

2. Recall

Recall is defined as the ratio between true positive (TP) and both true positive (TP) and false negative (FP). The equation is:

$$Recall = \frac{True\ Positive}{True\ Positive + Fasle\ Negative} \quad (7)$$

Table 2: Contingency Table

		Correct Labels	
		Positive	Negative
Classified Labels	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

3. Accuracy

Accuracy is defined as the ratio of number of correctly classified results and the total number of cases examined. The equation is:

$$Accuracy = \frac{True\ Negative + True\ Positive}{TrueNegative + FalsePositive + FalseNegative + TruPositve} \quad (8)$$

4.3 Case Analysis

The illustration of the proposed work is shown along with an example and also shown how this example is labeled and scored. The accuracy of the proposed algorithm is also shown when this algorithm is applied to various datasets and how well the algorithm works on these review datasets is shown

Let us consider the following review from the review dataset:

"i remember really enjoying this movie when i saw it years ago. I guess my memory really sucks. There is very very little that is funny in caddyshack. The laughs are few and far between and what there are really aren't that great. Caddyshack as the name implies more or less centers on one young caddy working at an exclusive country club.michael o keefe plays said caddy. Why they cast this unknown fairly untalented actor in the lead role is completely beyond me . The movie doesn t seem to have a real plot, just a series of scenes that are little more than opportunities for the rest of the cast to mug at the camera.

The only real story if you can call it that was a subplot involving the mentally disturbed greens keeper bill murray who is having his own private little war against a gopher who is ruining the course. Most of the marginal laughs come from rodney dangerfield and ted knight mugging and overacting for the camera with painfully limited success. Bill murray is slightly amusing in places but fairly wasted. The biggest waste of all is chevy chase who didn t even crack a smile on my face with his character s lame zen like approach to golfing. There are a few decent scenes involving the interaction between dangerfield and knight but they are far too infrequent to carry the movie. I guess that s what you get for basing a story around an unknown kid. Everything about this movie was just completely implausible as far as the comedy was concerned. maybe if you were drunk out of you mind or high off some sort of illegal narcotic this thing might be funny but for the rest of us stay the hell away from caddyshack."

By just having a glance at this review is can be easily inferred that the above review shows us a negative perception. A negative emotion is shows by the reviewer for the movie caddyshack. The original review of this is also negative one. Further analysis has been done on how the proposed algorithm performs when this algorithm is applied on the above review. When the algorithm is applied following results are generated in Table 3.

Table 3: Illustration of result for a movie review

ID	POS	Positive POS	Negative POS	Score	Review
1984	120	56	64	-6.2525	0

4.4 Experimental Result

The extensive experiment has been carried out for both NIPA method and without NIPA method using SVM and Naïve Bayes classifiers. The experimental results from SVM and Naïve Bayes are presented in table 4 and table 5 respectively. Various performance measures like accuracy, precision and recall of both NIPA and without NIPA are shown when the methods are applied on various datasets.

Table 4: Comparing Results using SVM for two approaches

	Accuracy With NIPA	Precision With NIPA	Recall With NIPA	Accuracy Without NIPA	Precision without NIPA	Recall without NIPA
Book	90.33	87.50	87.50	61.29	60.00	48.00
Car	73.52	66.60	33.33	55.55	50.00	20.00
Hotel	88.23	80.00	80.00	58.82	25.25	16.16
Phone	81.81	75.00	90.00	48.48	60.00	66.66
Epinion	97.91	97.22	94.54	82.14	82.52	74.73
Twitter	94.00	96.55	93.33	81.38	80.00	78.68
Movie	91.82	96.00	88.88	80.56	87.66	83.07

Table 5: Comparing Results using Naïve Bayes for two approaches

	Accuracy With NIPA	Precision With NIPA	Recall With NIPA	Accuracy Without NIPA	Precision without NIPA	Recall without NIPA
Book	91.30	84.00	95.2	54.34	29.54	54.34
Car	76.00	55.00	78.57	60.00	36.08	60.00
Hotel	82.00	84.21	96.40	62.00	38.86	62.00
Phone	91.83	96.40	90.00	57.14	32.74	57.15
Epinion	91.82	90.01	89.85	59.40	35.35	59.45
Twitter	91.83	94.31	91.55	62.42	39.00	62.47
Movie	94.20	91.93	94.87	81.35	84.53	78.20

The result shows that the proposed NIPA approach gives better performance than the approach that does not use NIPA.

5. CONCLUSION AND FUTURE SCOPE

In this work, Sentiment Analysis is done using Lexicon approach. The proposed methods are applied on different datasets i.e. product, movie and also tweets. The approach with NIPA gives better accuracy, precision and recall than the approach that does not uses NIPA approach. In future, the work will be done for stock market reviews through which people will know on which area they can invest on. Other available lexicon resources will also be considered for sentiment analysis such as AFFIN lexicon [38] and MPQA lexicon [39].

REFERENCES

1. Dave, K., S. Lawrence, and D.M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. in Proceedings of the 12th international conference on World Wide Web. 2003. ACM.

2. Gamon, M. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. in Proceedings of the 20th international conference on Computational Linguistics. 2004. Association for Computational Linguistics.

3. Kennedy, A. and D. Inkpen, Sentiment classification of movie reviews using contextual valence shifters. Computational intelligence, 2006. 22(2): p. 110-125.

4. Na, J.-C., et al., Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. 2004.

5. Ng, V., S. Dasgupta, and S. Arifin. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. in Proceedings of the COLING/ACL on Main conference poster sessions. 2006. Association for Computational Linguistics.

6. Pang, B., L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. 2002. Association for Computational Linguistics.

7. Whitelaw, C., N. Garg, and S. Argamon. Using appraisal groups for sentiment analysis. in Proceedings of the 14th ACM international conference on Information and knowledge management. 2005. ACM.

8. Xia, R., C. Zong, and S. Li, Ensemble of feature sets and classification algorithms for sentiment classification. Information Sciences, 2011. 181(6): p. 1138-1152.

9. Councill, I.G., R. McDonald, and L. Velikovich. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. in Proceedings of the workshop on negation and speculation in natural language processing. 2010. Association for Computational Linguistics.

10. Das, S. and M. Chen. Yahoo! for Amazon: Extracting market sentiment from stock message boards. in Proceedings of the Asia Pacific finance association annual conference (APFA). 2001. Bangkok, Thailand.

11. Ikeda, D., et al. Learning to Shift the Polarity of Words for Sentiment Classification. in IJCNLP. 2008.

12. Li, S. and C.-R. Huang. Sentiment Classification Considering Negation and Contrast Transition. in PACLIC. 2009.

13. Li, S., et al. Sentiment classification and polarity shifting. in Proceedings of the 23rd International Conference on Computational Linguistics. 2010. Association for Computational Linguistics.

14. Wilson, T., J. Wiebe, and P. Hoffmann, Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational linguistics, 2009. 35(3): p. 399-433.

15. Gamon, M., et al. Pulse: Mining customer opinions from free text. in international symposium on intelligent data analysis. 2005. Springer.

16. Hearst, M.A., Direction-based text interpretation as an information access refinement. Text-based intelligent systems: current research and practice in information extraction and retrieval, 1992: p. 257-274.

17. Pang, B. and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. in Proceedings of the 43rd annual meeting on association for computational linguistics. 2005. Association for Computational Linguistics.

18. Riloff, E. and J. Wiebe. Learning extraction patterns for subjective expressions. in Proceedings of the 2003 conference on Empirical methods in natural language processing. 2003. Association for Computational Linguistics.

19. Turney, P.D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. in Proceedings of the 40th annual meeting on association for computational linguistics. 2002. Association for Computational Linguistics.

20. Hatzivassiloglou, V. and J.M. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. in Proceedings of the 18th conference on Computational linguistics-Volume 1. 2000. Association for Computational Linguistics.

21. Wiebe, J. and R. Mihalcea. Word sense and subjectivity. in Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. 2006. Association for Computational Linguistics.

22. Andreevskaia, A. and S. Bergler. Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses. in EACL. 2006.

23. Fellbaum, C., Wordnet: An electronic lexical database (isbn: 0-262-06197-x). 1998, MIT Press.

24. Hu, M. and B. Liu. Mining and summarizing customer reviews. in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004. ACM.

25. Kim, S.-M. and E. Hovy. Determining the sentiment of opinions. in Proceedings of the 20th international conference on Computational Linguistics. 2004. Association for Computational Linguistics.

26. Popescu, A.-M. and O. Etzioni, Extracting product features and opinions from reviews, in Natural language processing and text mining. 2007, Springer. p. 9-28.

27. Nasukawa, T. and J. Yi. Sentiment analysis: Capturing favorability using natural language processing. in Proceedings of the 2nd international conference on Knowledge capture. 2003. ACM.

28. Wilson, T., J. Wiebe, and R. Hwa. Just how mad are you? Finding strong and weak opinion clauses. in aaai. 2004.

29. Baccianella, S., A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. in LREC. 2010.

30. Lewis, D.D. Naive (Bayes) at forty: The independence assumption in information retrieval. in European conference on machine learning. 1998. Springer.

31. Joachims, T., Text categorization with support vector machines: Learning with many relevant features. Machine learning: ECML-98, 1998: p. 137-142.

32. Netlingo, Popular text terms, Top 100 Acronyms<http://www.netlingo.com/top100/popular-text-terms.php>

33. Dushku S., Common Intensifiers and Evaluation Phrases in Spoken English <http://www.paulslals.org.uk/Dushku_handout.pdf>

34. Pang, B., Lee, L., (Introduced June, 2004). 1000 Positive and 1000 Negative Reviews. Retrieved February, 2005 from Cornell University, Cornell Natural Language Processing <https://www.cs.cornell.edu/people/pabo/movie-review-data/>

35. Carnegie Mellon University, 1382 Ford Automobile Review, Text Analytics <http://boston.lti.cs.cmu.edu/classes/95-865-K/HW/HW3/FordAutomobile>

36. Carnegie Mellon University, 479 Twitter Tweets, Text Analytics <http://boston.lti.cs.cmu.edu/classes/95-865-K/HW/HW3/Twitter/ >

37. Multi Perspective Question Answering (MPQA) Online Lexicon <http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/>

38. Nilson Finn A,AFFIN , Informatics and Mathematical Modeling, Technical University, Denmark, <http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010>

Cloud Robotics: A Brief Review

[1] Shweta Meshram, [2] Anita Murmu, [3]Rajesh Doriya
[1][2][3] Department of Information Technology
[1][2][3] National Institute of Technology, Raipur-492010, India

Abstract:-- **The cloud computing has been emerged has one of the driving technology in various fields. Cloud robotics is one of the such field where the robots making the use of cloud computing technology in order to provide various robotics services. These services may be computation offloading, data storage or other computationally algorithm such as map-building, path planning, navigation, etc. Since it's a very new area of research for the robotics community, very few significant research to work has been reported. In this paper we are reporting some of the significant advancement made in the area of the cloud robotics.**

Keywords—Cloud computing, ROS, DAvinCi, Robot-Cloud, Rapyuta, C2TAM, RoboEarth

I. INTRODUCTION

The cloud computing is one of major area now a day which provide the delivery computing services like servers, storage, database, networking, software, analytics over the internet that lead to store, manage and process data. because of their highly computational services and applications the scope of the cloud computing day by day increases. cloud computing services benefit in cost, speed, global scale, productivity, performance, reliability. this features leads our mind to concentrate on this side. as we are concerning about the cloud robotics it highly recommends computational and operation time should powerful with the common mechanism and functionality. the cloud computing will provide accurate storage and space for the robotic application.

As we are taking about the robot to manage and store their services ROS (Robotic Operating System) is comes in discussion. ROS is an open source meta operating system which is used by the robots. ROS provide the service to operating system with hardware abstraction to the user by hiding low level device control .it uses the peer to peer network topology. ROS distributed framework of process that are designed for specific purpose .it also provide the set of tools and libraries used to build robotic application .and provide the interface in package format for message passing between the hardware and robotic software.

For the extreme and efficient work, we propose some framework first is DAvinCi Framework that provide the scalability and parellelism advancement of cloud computing for service robots in large enviroment. It uses Hadoop Distributed file system for data storage and Hadoop Map–Reduce for parallel processing of the sensor data. It implements the such system in hadoop clusters and ROS for messaging network and also implement a FastSLAM algorithm for the significant performance in excution time for building a large area in eight node Hadoop clusters. this will introduce in software as a service (SaaS). This will reduce the burden of robotic application and extra weigthage.

The another framework that we used is Rapyuta helps robot to offload heavy computation by providing secured customizable computing environment in the cloud. Rapyuta will allow to outsource some or all robot onboard computational process to a commercial data centers like Google App Engine, Amazon Machine Image(AMI). It uses the PaaS (Platform as a service) cloud computing platform includes an operating system, an execution environment a database and communicational servers. the main component of Rapyuta is computing environment, communication protocol like internal communication protocol and external communication protocol, core and task like master task set, robot task set, environment task set, handling large binary message, communication with RoboEarth and virtual network.

The RoboEarth is one of the most frequent framework can be use the internet to share knowledge and to help each other accomplish complex task. the vision is to implement

the roboearth create an internet for robots, a worldwide common knowledge base where robots can share knowledge about objects, environment and action with other robots. RoboEarth uses the internet to create a giant open source network database that can be accessed and continually updated by robots around the world.

II. PRELIMINRIES

In this section, we have provide the overview of cloud computing, and its services. And also provided the overview of ROS (Robotic Operating System).

A. CLOUD COMPUTING

Cloud computing uses the service driven business model [1], where the infrastructure providers lend their pool of resources in data centers (e.g., hardware, networks, storage, servers, applications, and services) to the service providers as a utility. Virtualization technology plays a vital role in providing cloud resources. The service provider can pay for the infrastructure based on their usage with less investment. The resources are

readily available from the infrastructure provider to meet the increase in service demand hence provides high scalability. Most of the services that are hosted in the cloud environment are web based and are easily accessible through the internet. Some of the popular cloud computing products are Amazon EC2, Windows Azure, and Google App Engine[no]. NIST [2] defined cloud computing as "Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." The distributed File System like HDFS, GFS, and Map Reduce framework are the major technologies that provides distributed storage and parallel computing in cloud computing environment. The well–known service models of cloud computing are

1) Infrastructure as a Service (IaaS): in which users are provisioned with infrastructures like network, storage, hardware and computing nodes as a service, also the resources may be 3 virtualized one. The maintenance of the resources like hardware, software updates are serviced by the infrastructure providers.

2) Platform as a Service (PaaS): Service providers can deploy their own applications, servers to be accessed ubiquitously by other users

3) Software as a Service (SaaS): This allows users to access the already deployed applications in the cloud through web browsers.

B. ROS: Robotic Operating System

Programming for robots is difficult as it requires excellent knowledge from wide areas like hardware, kinematics, probability, etc. Considering the complexity in robotics like sensor selection, an interfacing mechanism with sensors, actuator modeling mechanism, the reliable control strategy for actuators, and the creation of flexible control policies it requires the breadth of expertise from diverse fields which is beyond the capabilities of any single researcher. Hence simulation tools play a major role in research, education and design process of complex systems like robotics. Various robotic simulators are available like Carmen [3] for single robot, Gazebo [4] for multi–robot, Morse [5] for heterogeneous humanoid robots, Webots [6] for mobile robots. The Robotic middleware's and frameworks also play a significant role in making the software work on other platforms, dealing with inter–process communication and thus provides interoperability and promote software reuse. YARP [7], MOOS [8], MRDS [9] and Player Stage [10] are one among few. Nowadays ROS [no] is mostly used by most of the Robotic platforms. According to ROS wiki [no] more than 90 robot platforms officially support ROS. ROS is an open source meta–operating system, which runs alongside the traditional operating system which is used by robots. ROS provides service similar to an operating system with hardware abstraction to the user by hiding low–level device control. In a Robot, there are multiple processes (nodes) running, and each node is responsible for a unique task like collecting sensor data, controlling motors and running localization algorithms. ROS uses a peer to peer network topology. ROS is a distributed framework of processes that are individually designed for specific purpose and can be loosely coupled at\runtime. ROS provides a set of tools and libraries used to build robotic applications. The widely used robotic algorithms for motion planning, navigation and mapping are implemented, debugged are made into a standard package which can be used by other systems. In addition to that, the interface for message passing between the hardware and robotic software also available in package formats. It also provides tools and libraries for writing and running code across multiple computers. ROS is designed to be thin so it can be integrated with any other robotic frameworks like the Player. The programming for ROS can be written in any language like C++, python, and lisp. The messaging mechanism used by the nodes for their communication are as shown in Figure 4. A node can use Publish/Subscribe model for asynchronous communication, in which a node can publish a topic which is listened by multiple nodes. It uses service based model for one to one synchronous communication between nodes. Publish/Subscribe model is

shown above in Figure 4. Any node can publish their message on some topic. Nodes whichever interested on a particular topic can subscribe to that topic. The master node acts as a moderator, and provides name service through which the publishers and their topics are exposed to the subscribers. The actual data does not pass through the master node. Rosbridge [11] is a JSON based API used to bridge non–ROS program to ROS. Any language or protocols that can send JSON are able to interact with ROS using rosbridge. Thus rosbridge enables programming robotic application in any language of user's choice. It allows message handling over web sockets, HTTP5 and POSIX IP sockets. rosbridge comes along with rosjs the application layer library which provides support for remote robotic laboratory and human–robot interaction. Rosjava [12] is a pure implementation of ROS in java. It provides a client library so that java programmers can easily interface with ROS messages, topics, and services. The primary goal of Rosjava is to make ROS android friendly. Rosjava nodes can run in thread pool instead of a process in other client libraries.

III. PROPOSED FRAMEWORKS IN CLOUD ROBOTICS

C. DAvinCi Framework

The DAvinCi [13] (Distributed Agents with Collective Intelligence) framework address the challenges of data sharing, cooperative perception and collective intelligence. The robotic tasks such as vision processing and mapping are computationally intensive tasks, which can be done by increasing the processing power of on-board computers which in turn requires large amount of robot's power. Also, this process is repetitive where networked robots are used to build the map of the heterogeneous environment. DAvinCi uses distributed ROS for sensor data collection and interaction among robots. It uses Hadoop Distributed file system for data storage and Hadoop Map–Reduce for parallel processing of the sensor data. It exposed SLAM as a cloud service which can be accessed by multiple robots. DAvinCi contain multiple robots connected with Wi-Fi and each robot have different type of sensors for image acquisition and localization. All this sensor information from robots are communicated through the cloud using a gateway. By uploading the information from multiple robots to a single controller, the computationally intensive process map building is done in the cloud which is reused by all robots on-demand basis.

The DAvinCi Server allows multiple robots to share their sensor information using clouds. The Subscriber node registered for the topics from multiple robots and save the information on HDFS. The Map Reduce Framework allows parallel processing of the SLAM algorithm on the Hadoop cluster.

The main components of DAvinCi architecture is explained as follows

- DAvinCi Server: It acts as a proxy that connects robots to the backend computation and storage clusters. It acts as service provider to the robots by maintaining the ROS name service which can be queried by ROS on the robots. The Robots communicate to the server using ROS 9 messages which are wrapped inside HTTP for communicating to server through Wifi.The server collects data from robots by subscribing to the topics and then pushed the data to HDFS and also it triggers some map/reduce tasks to process the data. It exposes service to other users using http over the internet

- Hadoop Clusters: It provides computation and storage resources for the DAvinCi platform. The HDFS file system running in the clusters provides the massive storage. The Map–Reduce framework running in the clusters allows to execute the robotic algorithms in parallel. Thus it saves time for the computationally intensive tasks.

- ROS: ROS is a loosely coupled distributed framework for robotic environment

- FastSLAM implementation using Hadoop: The Grid based FastSLAM algorithm is implemented using Map Reduce Framework thus reducing the computation time by splitting the work across clusters.

Limitation: The communication between robots and DAvinCi server involves huge volume of data includes images or maps requires the channel should be reliable. The network latency for communication to the cloud is not considered. In addition to that the implementation of this framework is not publicly available and only a single master node manages all the inter process communication which can be prone to single point of failu

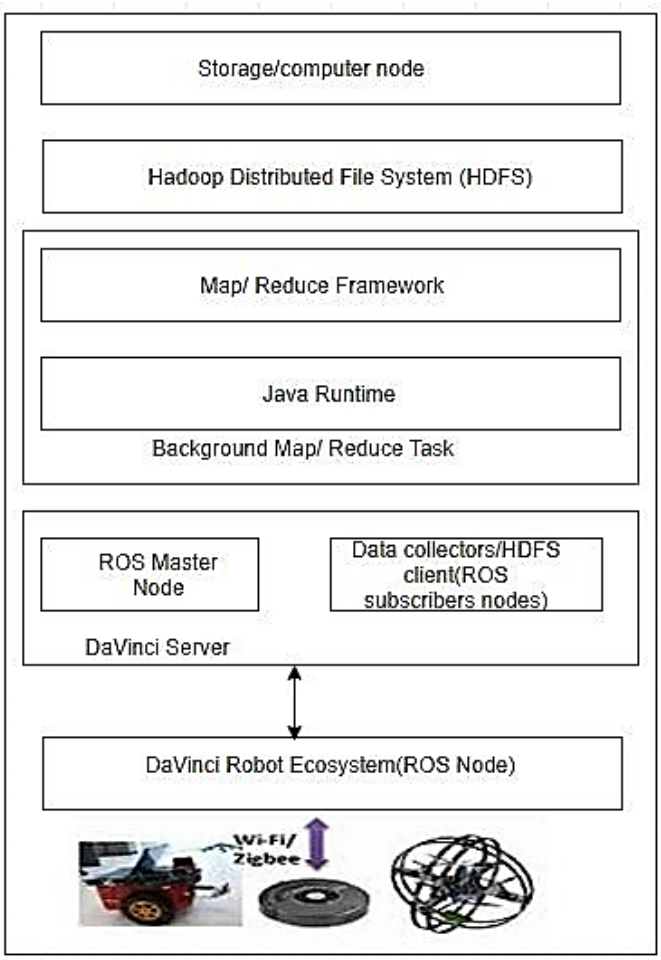


Fig. 1.Architecture of the DAvinCi Cloud computing Platform

D.Robot-Cloud Framework

Robot and Automation system [14] that really on data or code from a network to operation, i.e., where not all sensing, computing, and is integrated into a system. The cloud can also improve robots and Automation system by providing access to: Dataset, publication, models, and simulation tools Dataset, publication, models, and simulation tools, Open computations for design and system, and Open source. Cloud Robot and Automation system often include some capacity provides where network access is unavailable and unreliable.

The cloud robots and Automation system is use the wireless network, big data, cloud computing, open source, and other shared resources to improve performance in an application including assembly, driving, package delivery, and surgery. Cloud based formation control of ground Robot sharing data through collective robot learning can also improve the capacities of Robots with resource.

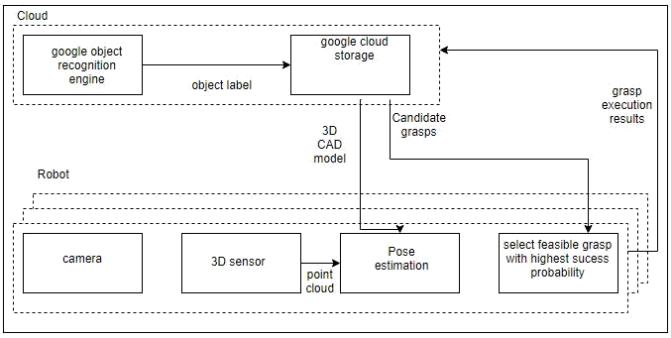


Fig. 2. System Architecture for cloud-based object recognition for grasping

E. Rapyuta Framework

The Rapyuta [15] is an open source, Platform as a Service (PaaS) framework for robotics application. In Rapyuta, the robots will get dynamically allocated computing environment for offloading their computation tasks. The computational environments are tightly interconnected which makes a suitable framework for multiple robot coordination. This interconnection allows robots to share their services and information with other robots. The computing environment has high bandwidth access to the knowledge repository so that the robots can bene- fit from other robot's experience. Rapyuta Architecture consists of computing environments, communication protocols, set of core tasks and command data structure.

1) Computing Environments: The computing environments are light weight Linux container used for offloading the computation tasks. Any ROS node can run in a computing environment and uses ROS inter–process communication. A Linux container offers an environment which is similar to a virtual machine, without hardware virtualization. So all the process runs inside the same kernel but it provides isolation of process and resources within a single host. Linux containers allow scalability by configuration of disk quotas, memory limits, I/O rate and CPU quotas.

2) Communication protocols: Endpoints which is a process, forms the basic building blocks of Rapyuta communication protocol. Each endpoint contains Ports and Interfaces. Ports are used for internal communication between other Rapyuta processes (endpoints). Interfaces are used for external communication between non Rapyuta process that may run in robots or computing environment.

The interface can serve as service provider or service client, and topic publisher or topic subscriber. Another RPC communication takes place between endpoints and main controller process. 10

3) Command Data structure: Rapyuta's command data

structure contains 4 components.

User: Generally, a human having one or more robots want to communicate with cloud. Each user has a unique API key which is used by robots while connecting to the cloud for authentication.

Load balancer: Load balancer is used to manage machines which are running in computing environment. It contains the representation of containers running in a machine, it also assigns new container for a machine.

Distributor: The distributor is used to distribute the connections from a robot to other robot endpoints.

Network: It is referenced by the other 3 components and thus provides the abstraction of the platform. The internal and external communications are organized using this command data structure.

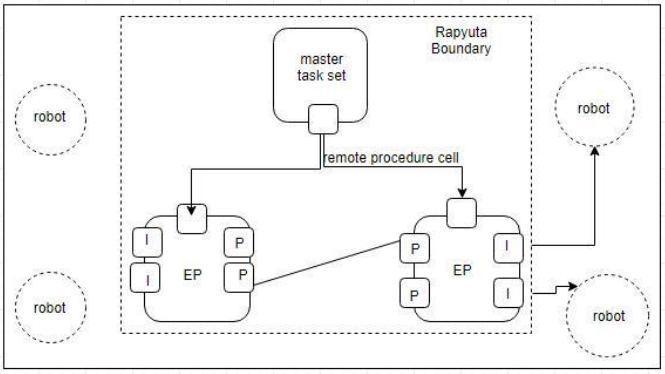


Fig.3. The basic communication channels of Rapyuta: The Endpoints (EP) are connected to the Master task set using

a. two-way remote procedure call (RPC) protocol. Additionally, the Endpoints have Interfaces (I) for connections to robots or (ROS) nodes, as well as Ports (P) for communicating between Endpoints. The dotted lines represent the external communication, dashed lines represent the ROS-based communication between ROS nodes and Rapyuta, and finally all solid lines represent the internal communication between Rapyuta's processes.

4) Core Task sets: In Rapyuta, the task sets which is a set of functionalities used for system administration. In a process, more than one task sets can be combined based on the required

functionality. The four task set are
Master Task set: It is the main controller for monitoring command data structure. Also only a single copy of master task set should run inside Rapyuta. It's responsible for organizing connection between Robots and Rapyuta, it also routes the configuration request from robots and monitors the network of other task sets.

Robot Task set: It defines the essential requirements for communicating with robots. It does the conversion of data messages between robots and other end points. It should be able to forward configuration request to the master

Environment task set: Its run as process inside every computing environment. It defines the how to

communicate with the computing environment which includes how to launch and stop ROS nodes, adding and removing parameters and communication with ROS nodes and other endpoints

Container task set: Its run as a process inside every machine used to define how to start and stop computing environments. Limitation: The failure of the master task may result in the complete failure of entire Rapyuta. The network connectivity issues when connecting to cloud and security are not considered.

F. C2TAM Framework

C2TAM [16] tries to explore offloading the computationally intensive task to the cloud. As robots have limited resources in their onboard computers, C2TAM tries to move some part of computationally intensive SLAM computation to the cloud. SLAM is one of the computationally intensive tasks and also has strong real–time constraint, so moving all computation of the SLAM to the cloud would not be effective. C2TAM propose a method to partition the real–time SLAM algorithm that allows part of the computation should be moved on to the cloud. So to have a better understanding of SLAM, the main components of visual SLAM is explained in the following section and in the next section which SLAM components are executed onboard or offloaded to the cloud. The Main Components of SLAM are

1) Mapping: It builds the map, a model of the scene from the sensor data. It involves updating the camera pose and estimating the map sequentially whenever sensor data

arrives. C2TAM uses the optimization approach proposed by recent works, that involves two threads, one for camera pose updation and another for map estimation which reduces the computational time without loss of accuracy.

2) Tracking: The tracking component tracks the camera pose for each time step from the given map. It has strong real–time constraints and should be done at high frequency as it tracks visual features from the image sequence.

3) Relocation: The tracking failure can happen due to high acceleration motion, blur or lack of visual features. Once tracking failure occurred the relocation component tries to relocate the camera and restart tracking

4) Place Recognition: It happens once there is a tracking failure. When a robot is put in an unknown environment it tries to recognize the place from the database of maps by checking with each location. It differs from relocation in the sense that relocation having a prior knowledge of location, and it tries to relocate by small camera motion and tries to check nearby places as it happens immediately after tracking failure.

5) Map Fusion: It merges two independent maps when overlapping area is detected by place recognition. The point correspondence between two maps is identified and the 3D points in common key frames are projected. The relative transformation from the composition is calculated and finally, the duplicate points are removed. The rest of the points are

transformed according to the relative motion between two maps and are merged into a single map

C2TAM provides SLAM as cloud service: C2TAM recommends moving non real–time computationally expensive tasks 8 to be moved to the cloud which reduces the payload and power consumption in robots due to its onboard computation. SLAM is distributed, so different visual sensors from cameras of different robots can connect to the mapping server in cloud. The centralized map building provides massive storage of robotic sequences which can be used as the knowledge base for training purpose. It allows cooperative SLAM, where multiple robots can build a joint map of the environment. The server can fuse the maps. This section explains working of each SLAM components with respect to the cloud service.

Mapping as a cloud service: Mapping the most computationally intensive task, that can tolerate network delays and it does not need any real– time constraint, so mapping is the perfect candidate for moving to cloud.

Tracking as a client: Tracking involves tracking the camera pose which has real–time constraint that it has to operate at a frame rate. Network delays are not acceptable during tracking and it can't be moved to cloud. It inputs the new key frames to mapping component in the cloud. It returns back the map once the current map is optimized.

Relocation as a client: There are two possible reasons for relocation. i) Due to sudden motion or occlusion, the tracking node which is operating successfully over a map is lost. In this case, relocation needs to be done only on current map as the camera tracking is in the previous map, so it can be done on tracking client. ii) The second case, occurs when the tracking thread is just started or lost for a long time. In this case, relocation should look for correspondence in a large number of maps, so it has to be done on mapping node in raw cloud.

Place recognition as two-step process: Place recognition is computationally demanding as it needs mapping with a large database. Moreover, this process is sensitive to network delays. So in the first step, the mapping server which is in the cloud coarsely relocates the camera from the large number of maps and sends the filtered map to the client. The client will do the second step of finely relocating from the filtered map.

Limitations: C2TAM only tries to reduce the computational task on the robot. It demonstrates building maps from different sensors deployed in the same environment. Moreover, the cloud server tries to look for similarities and fuse the maps, which are built by different robot in different environment. But there is no

parallelizing of mapping tasks, which can even reduce the map building time further.

G. RoboEarth Framework

In the complex and unstructured environment, the information generated by one robot should be reused by other robots in order to make the robots adjusts to the dynamic environment quickly. However, the data generated by robots with different hardware configuration are not same which prohibits the reuse of data. RoboEarth [17] focuses on how robots should store and share the knowledge among themselves which can speed up their learning

process. Thus RoboEarth provides a platform for robots to collect, stores, reuse and sharing of data independent of hardware over the network. The information stored also linked with their relative data. Example the CAD Model of an object is linked with its name, its properties, and instructions to manipulate the object. Robots using the RoboEarth can adapt to the complex environment quickly and also able to extract the action recipe of other robots and execute a task which is not planned at design time. RoboEarth mainly aims at how a robot can download the action recipe from the web and do the task on its own. To better illustrate the scenario let us take an example where Robot A which serves in a hospital can share the knowledge about location of bed, bottle, patient, and also how to open the cupboard, locate and pick the bottle and serve a patient. When Robot B of another type is instructed to serve the patient it can download the action recipe stored by Robot A and used this as prior information to do the task. RoboEarth is composed of three layered architectures. The top layer is the Server layer which contains the database of objects, its associated properties, various environments and action recipes with their semantic information. It organizes the data in the hierarchical tables in distributed manner using Apache Hadoop. It also has a centralized graph database which contains the relationship among the tables. It uses REST–based interface for storing the information to the server. RoboEarth also provides the web interface for accessing the database services by the user. The next layer is generic component layer which forms the part of robots control software that is independent of hardware. The action recipes from the database layer are interpreted in this layer and provide reasoning, modelling, and learning capabilities to the robot. For executing an action recipe, this layer checks whether all the required constraints are met by the robot and it receives input from the perception of robot and associate the data from the database and controls the action.

The third layer is abstraction layer provides generic interface to hardware related functionalities of the robots. Limitation: The data is linked with semantics like a bottle is used to store liquid, which is usually placed in the cupboard. Robots are learning from the semantic information rather than associating an object with an environment which may need further research. Sharing knowledge among robots over the web may create risks in terms of legal, moral, secrecy, privacy, safety aspects which needs further research

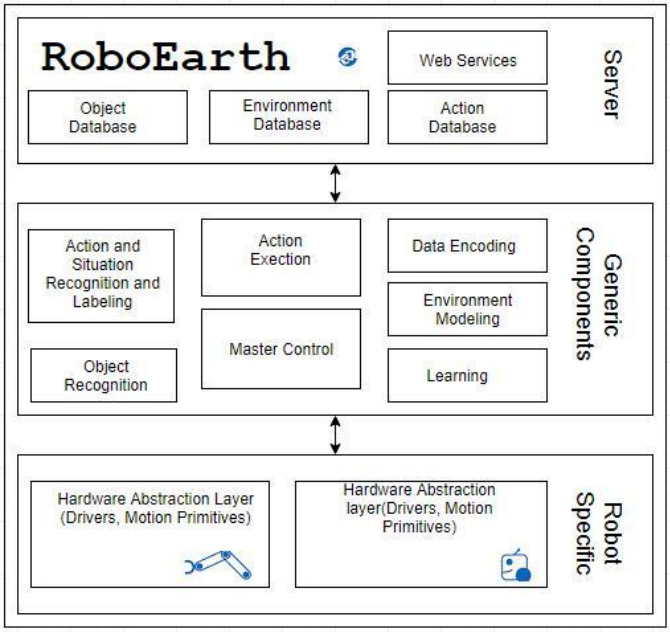


Fig.4. The RoboEarth systems architecture designed to allow robots to share data and learn from each other

IV. DISCUSSION AND RECOMMENDATION

In the discussion there are two reason that we are concluded with is the first that the cost of onboard computation and storage is high. This will lead the robot costly and it will cause the addition weight and extra space which will affect the time to perform an operation and speed will less. Generally, because of this reason robot affect their mobility and due to more functionality it constrains their potential.

The second reason which affect their intelligence is lack of mechanism and communication medium which affect the connection between the robot with hardware and software components.

The social robotic architecture should allow the team of heterogeneous robots able to communicate and coordinate on a specific task. As robots are controlled over the internet then Service Oriented Robotics come into existence. Service oriented architecture (SOA) in robotics provides various advantages. Each robotic component like sensors, actuators is implemented as a service with standard interfaces. Thus any device can invoke the same service irrespective of the robotic platforms. Service oriented robotics permits different robotics applications can be developed using the set of available services within a short span of time. It

allows programmers without robotic knowledge to build robotic applications by using the services.

This paper presents the proposed framework that provides robotic services with security. It explains why security is needed for robotic communication and how we can provide

security. It also lists some of the tasks that can be provided as cloud robotic services and which level of security each service is needed. It also requires some more in depth analysis on how can we provide security for robotic services and to which level

V. CONCLUSION

This paper provides the brief introduction to cloud computing, cloud robotics and how cloud robotics has evolved. It also explores why robots need the cloud for their tasks. This paper presents facts about existing cloud robotics framework and their advantages over the other. It explores the details of SOA, and how it can be used in cloud robotics application. Moreover, this paper gives a brief about cloud robotics architecture, challenges in cloud robotics, and various communication mechanisms for a robot to communicate to the cloud. ROS provides easy integration with other frameworks, and gives support for reuse with the help of existing packages thus promotes robotic research. The next section gives details about existing cloud robotics framework. The C2TAM framework help in deciding which tasks of SLAM should be offloaded to the cloud. The DAvinCi framework analyze how we can support parallel processing for the robotic application using Map Reduce Framework. The RoboEarth framework presents the details about how the information can be stored and can be reused among multiple robots. The Rapyuta framework presents how resources can be dynamically allocated and robots to share their services in multi–robot environment. The final section of this paper presents the proposed framework that provides robotic services with security. It also lists some of the tasks that can be provided as cloud robotic services and which level of security each service is needed. It also requires some more in depth analysis on how can we provide security for robotic services and to which level. Our future work will be based on this review and we will be providing security enabled cloud robotic services using ROS framework.

REFERENCES

[1] Q. Zhang, L. Cheng, and R. Boutaba, “Cloud computing: state-of-the-art and research challenges,” Journal of internet services and applications, vol. 1, no. 1, pp. 7-18, 2010.

[2] P. Mell, and T. Grance, “The NIST definition of cloud computing,” 2011.

[3] M. K. Srinivasan, K. Sarukesi, P. Rodrigues, M. S. Manoj, and P. Revathy, "State-of-the-art cloud computing security taxonomies: a classification of security challenges in the present cloud computing environment." pp. 470-476.

[4] M. Chraibi, H. Harroud, and A. Maach, "Classification of security issues and solutions in cloud environments." p. 560.

[5] G. Echeverria, N. Lassabe, A. Degroote, and S. Lemaignan, "Modular open robots simulation engine: Morse." pp. 46-51.

[6] O. Michel, "Webots: a powerful realistic mobile robots simulator."

[7] G. Metta, P. Fitzpatrick, and L. Natale, “YARP: yet another robot platform,” International Journal of Advanced

Robotic Systems, vol. 3, no. 1, pp. 8, 2006.

[8] P. M. Newman, “MOOS-mission orientated operating suite,” 2008.

[9] K. Johns, and T. Taylor, Professional microsoft robotics developer studio: John Wiley & Sons, 2009.

[10] B. Gerkey, R. T. Vaughan, and A. Howard, "The player/stage project: Tools for multi-robot and distributed sensor systems." pp. 317-323.

[11] C. Crick, G. Jay, S. Osentoski, B. Pitzer, and O. C. Jenkins, "Rosbridge: Ros for non-ros users," Robotics Research, pp. 493-504: Springer, 2017.

[12] L. Turnbull, and B. Samanta, "Cloud robotics: Formation control of a multi robot system utilizing cloud infrastructure." pp. 1-4.

[13] M. T. Gettman, M. L. Blute, G. K. Chow, R. Neururer, G.

Bartsch, and R. Peschel, “Robotic-assisted laparoscopic partial nephrectomy: technique and initial clinical experience with

DaVinci robotic system," Urology, vol. 64, no. 5, pp. 914-918, 2004.

[14] B. Kehoe, S. Patil, P. Abbeel, and K. Goldberg, "A survey of research on cloud robotics and automation," IEEE

Transactions on automation science and engineering, vol. 12, no. 2, pp. 398-409, 2015.

[15] G. Mohanarajah, D. Hunziker, R. D'Andrea, and M. Waibel, "Rapyuta: A cloud robotics platform," IEEE

Transactions on Automation Science and Engineering, vol. 12, no. 2, pp. 481-493, 2015.

[16] L. Riazuelo, J. Civera, and J. Montiel, "C 2 tam: A cloud framework for cooperative tracking and mapping," Robotics and Autonomous Systems, vol. 62, no. 4, pp. 401-413, 2014.

[17] M. Waibel, M. Beetz, J. Civera, R. d'Andrea, J. Elfring, D. Galvez-Lopez, K. Häussermann, R. Janssen, J. Montiel, and A. Perzylo, "Roboearth," IEEE Robotics & Automation

Magazine, vol. 18, no. 2, pp. 69-82, 2011.

The Interplay of Text Genre and Comprehension-checking Behavior: Moroccan EFL University Students as a Case Study

Mohammed Msaddek
Delegation of Salé, Morocco
msadek60@gmail.com

Abstract:-- **It is true that textual processing entails the methodical usage of monitoring strategies intended to facilitate the attainment of an effective, adequate comprehension. Indeed, English as a foreign language (EFL) learners, in their endeavor to generate efficient sense of any written discourse, are expected to employ a large corpus of strategic processes. Granted this, the current study depicts the typical influence of text genre (i.e., narrative, expository) on the use of the monitoring techniques among Moroccan EFL learners (n=113) at the pre-test and post-test levels. Two English department groups were targeted: Control group (n=50) and treatment group (n=63). In order to reach comprehensive findings, three research instruments such as reading tests, strategy instruction and reading comprehension texts (i.e., narrative, expository), were tapped. The outcomes showed that text genre does, to some extent, impact the frequent use of monitoring during the reading process. Lastly, the study concludes with some pertinent implications and suggested perspectives.**

Keywords: executive control, expository text, narrative text, monitoring, text analysis

I. INTRODUCTION

It is commonly admitted that being strategic, independent readers is an important precondition to an academic success (Hong-Nam & Leavell, 2011; Bharuthram, 2012; Kolić-Vehovec, et al., 2011). Indeed, the methodical use of effective strategies in the course of reading written texts can both facilitate the comprehension process for the learners in various ways and improve their reading capabilities to a greater extent. Thus, the lack of an appropriate understanding of the content of written materials can be attributable to the inefficient use of monitoring strategies. Most EFL learners do not seem to apply them efficiently for ensuring a thorough understanding. They only employ some strategic processes, often unconsciously, so as to make sense of the textual content in a general way. This engenders a certain type of inaccuracy and ineffectiveness in terms of comprehending the target message conveyed by the author/writer via his/her text. Accordingly, it can be assumed that the use of monitoring is a potentially promising means for attaining an accurate, comprehensive understanding of the written text (i.e., narrative, expository).

Actually, a large corpus of previous extended reading research conducted on the perceived interconnectedness between text genre (i.e., narrative, expository) and strategies, has been restricted to the exploration of the influence of text type on prediction (e.g., Afflerbach, 1990), inferencing (e.g., Baretta, et al., 2009), selective attention (e.g., Goelman, 1982), and recalling (e.g., Zabrucky & Ratner,1992; Wolfe & Mienko, 2007). Further, while a grouping of reading research specialists (e.g., Best, at al., 2008) investigate the interplay of the variable of text type and decoding skills and background knowledge in the reading process, other researchers (e.g., Khansir & Mohammadi Fard, 2014) tackle the text genre impact on vocabulary acquisition techniques among learners. However, the effect of text type on comprehension monitoring has been addressed by very few researchers (e.g., Zabrucky & Ratner, 1992; Zabrucky & Moore, 1999). Hence, the current study intends to address this issue in a Moroccan EFL context in an attempt to depict the link relating between text typology (i.e., narrative, expository) and learners‟ strategic monitoring behavior.

II. THEORETICAL FRAMEWORK

A. Executive Control

Executive control refers to the process of performing a particular cognitive task in an organized, regulated and efficient way. Wang, et al. (1997) consider it to be “a student‟s capacity to plan, monitor, and, if necessary replan comprehension strategies in the service of understanding”. Most important, executive control enables EFL learners to methodically apply a wide variety of strategies in the learning as well as the reading tasks. It is deemed a sophisticated kind of metacognitive control

which allows learners to undertake and approach different tasks with a certain amount of efficiency and accuracy.

For Rubin (1992), executive control consists of three major processes which are reflected in the setting of goals, the monitoring of performance or comprehension of any problems that arise, and the making of decisions as to appropriate subsequent action. In very simpler terms, executive control, as an effective way of organizing and controlling one‟s learning process, incorporates three primary strategies which can be referred to as planning, monitoring and evaluating. For instance, in the process of reading, learners are highly expected to formulate goals for their reading task, monitor the progress of comprehending the content and assess the effectiveness of their reading performance. These strategies form an essentially significant portion of the executive control since they assist learners to attain an overall interpretation of the text content.

B. Monitoring in Text Reading

Monitoring, as a „regulatory‟ and executive strategy, refers to "one‟s awareness of comprehension and task performance while in the process of performing a specific task" (Nietfeld, et al., 2005: 9). It is, indeed, an essential metacognitive strategy which enables learners to be fully aware of the process of understanding. In other terms, monitoring, when undertaking a given reading task, gives learners the opportunity to ensure that the process of comprehension is effectively undertaken. For Nietfeld, et al. (2005), who basically underscore the critical significance of the monitoring strategy in dealing with different academic tasks, effective monitoring "aids students in keeping track of ongoing cognitive processes and using regulatory strategies to solve problems" (p.9).

According to Ley and Young (2001), monitoring has been associated with achievement (p.98). This reveals that the utilization of this „self-regulatory‟ strategy in attempting to approach academic reading is a foundational prerequisite to achieving a successful comprehension. To put it succinctly, conducting a given reading task without monitoring the extent of the progress that is taking place will not assist learners to attain an efficient understanding. Within this framework, Puntambekar (1995) reflects the fact that younger and less skilled learners do not use many monitoring strategies. This amply evinces that the development of monitoring strategies, as noted by Puntambekar (1995), can be closely related to two major variables which are age and aptitude. That means that mature and skilled learners are able to effectively monitor their comprehension of diverse texts (i.e., narrative, expository), whereas younger and less skilled learners cannot use the monitoring strategies as successfully as their counterparts (skilled learners).

In view of this, monitoring, as "a crucial component of metacognition" (Schmitt & Newby 1986: 30), significantly assists learners to tackle the reading texts of narrative and expository type with greater effectiveness and efficiency. Nonetheless, monitoring strategies, as confirmed by Brezin (1980), are frequently implemented by experienced learners. This explicitly stated fact reflects that this type of strategies cannot be used by „unskilled‟ and „inexperienced‟ learners more efficiently unless they are provided with adequate strategy training. Hence, an extensive use of the monitoring process by learners is associated with the attainment of an effective understanding of the content included in diverse textual input (i.e., narrative, expository).

Research Objectives & Research Questions

This exploratory study seeks to reveal the typical influence of text genre (i.e., narrative, expository) on the monitoring techniques among Moroccan EFL learners (n=113). It also aims at disclosing whether the variable of strategy training can impact English department university learners‟ monitoring behavior as pertains to text genre (i.e., narrative, expository). .For bringing these objectives into a succinct, plain perspective, some research-based instruments such as reading comprehension texts (i.e., narrative, expository), strategy instruction and narrative and expository reading tests (i.e., pre-test, post-test) were put into action by the researcher for the assurance of quantitative and qualitative data. Hence, two prime research questions are formulated.

a- To what extent does text genre influence Moroccan EFL university learners‟ strategic monitoring?

b- To what extent can Moroccan EFL university learners‟ monitoring behaviour be impacted by reading strategy training whilst coping with text type (i.e., narrative, expository)?

Research Hypotheses

In light of the above-stated research questions, two prime research hypotheses have been formulated. They are brought forward as follows:

1.Moroccan EFL university students‟ strategic monitoring is not genre-specific.

2.Strategy training does not have a significant effect on EFL university learners‟ monitoring behaviour with regard to text genre (i.e., narrative, expository).

III. METHOD

A. Participants

One hundred and thirteen Moroccan EFL university students, studying at the first-semester level at the Faculty of Letters and Human Sciences, Mohammed V- Agdal in Rabat, participated in the current study. Two groups were addressed. The first group (n=63) was exposed to the instructional treatment and the second group (n=50) did not receive any strategy intervention. The targeted learners in the two EFL groups (i.e., control, experimental) are not repeaters and they have a comparable educational background. The prime impetus for selecting these two groups is to draw a parallel between the experimental and control groups with respect to comprehension-checking at the pre- and post-test levels.

B. Procedure

The current exploratory study adopts a pre-post-test design by assigning a pre-test and a post-test to the control and treatment groups. The two groups (control & experimental) were pre-tested on narrative and expository written discourse and given a „self-report questionnaire‟ at pre-testing. Afterwards, the control group (50) was instructed in reading comprehension without receiving any strategy intervention. As to the intervention group (63), it was instructed in strategic monitoring for a semester-long period (Fall Term/ 2012). The procedure was conducted through the assignment a wide series of narrative and expository written texts with a view to consolidating the treatment group‟s reading strategic behavior and analysis/ synthesis skills. At the conclusion of the undertaken intervention, the controls as well as the treatment group were presented with narrative and expository post-tests. This was accompanied with a „self-report questionnaire‟ for gauging the learners‟ strategic monitoring throughout the course of text processing (i.e., narrative, expository).

The data obtained through the „self-report questionnaire‟ were computed through the Excel software Program (version 2007) with a view to measuring the strategic monitoring behaviour among both groups (i.e., control, experimental) in coping with text genre (i.e., narrative, expository). This was effected at the pre-test and post-test levels for investigating whether EFL learners‟ monitoring behaviour is „genre-sensitive‟. The strategic monitoring processes of the target learners were numerically counted. Further, illustrative figures reflecting the attained findings were used. In fact, the frequency of monitoring strategy use for both the narrative and the expository written texts included in the pre-test and post-test was tabulated with percentages.

IV. FINDINGS & DISCUSSION

4.1. The Interplay of Self-monitoring & Text Genre among EFL learners

Based on the findings reached, it is of paramount significance to state that the utilization of self-monitoring during the reading act among Moroccan EFL learners cannot be impacted by the text type variable. This is displayed in the two following figures.

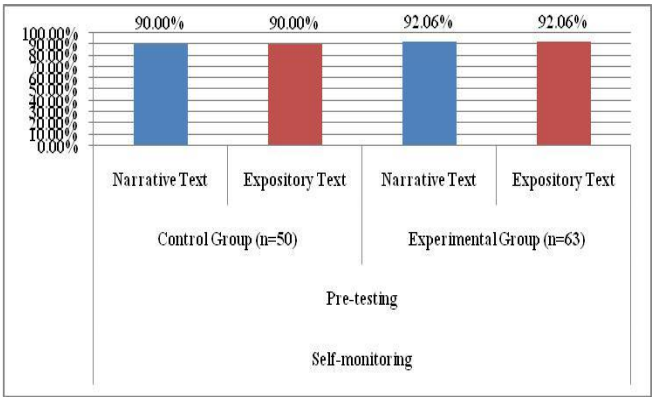


Figure 1. Self-monitoring use among EFL groups at pre-testing level

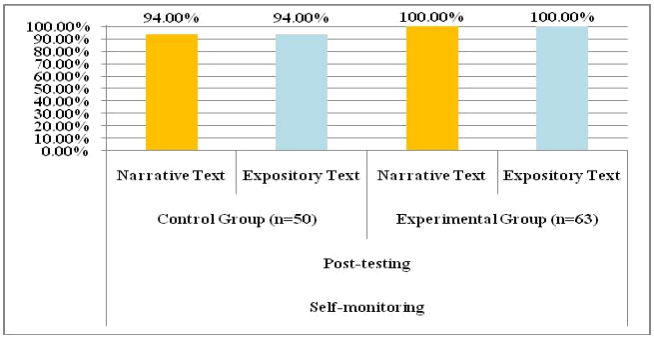


Figure 2. Self-monitoring use among EFL groups at post-testing level

As plausibly indicated in the two figures, self-monitoring was recruited with similar frequencies in reading narrative and expository texts among the control and experimental groups. For instance, the control subjects resorted to the strategy of monitoring during reading the narrative as well as the expository texts with a frequent occurrence of 90% and 94% at the pre-test and post-test levels respectively. As

to the treatment group, it depended on self-monitoring with percentage of 92.06% in tackling the narrative and expository texts at the pre-test. At post-testing, a heavy reliance of this comprehension-checking strategy was reflected among the treatment subjects. This shows that text genre does not have any marked impact on the sampled EFL students‟ self-monitoring in text reading.

4.2. The Interplay of Self-questioning & Text Genre

As was the case with self-monitoring, the strategy of self-questioning, as a basic monitoring move, does not appear to be genre-oriented. This is plainly illustrated in the ensuing figures.

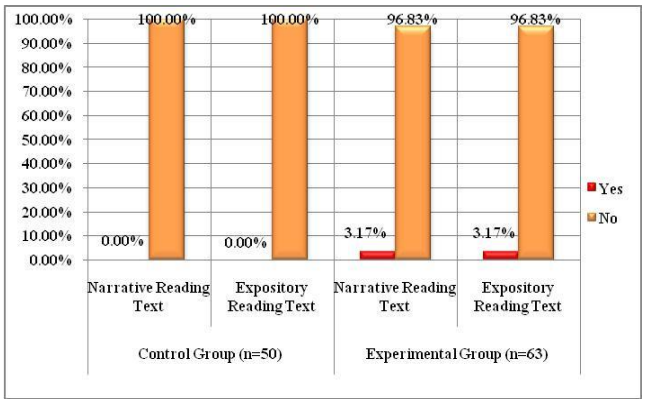


Figure 3. Usage of self-questioning among EFL groups at pre-testinglevel

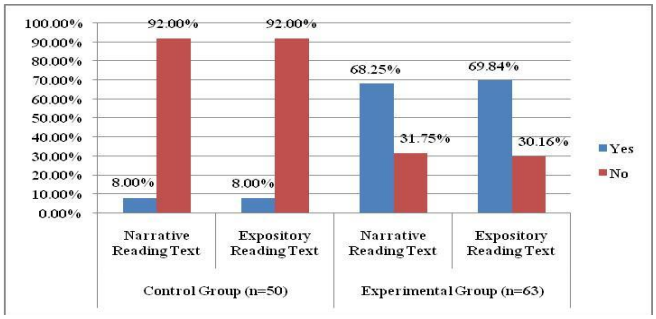


Figure 4.Usage of self-questioning among EFL groups at post-testing level

It is manifest that in engaging in the analysis of the narrative and expository texts, the EFL learners belonging to the control group did not use self-questioning for the sake of facilitating understanding. They only employed this strategy with a rate of 8% at post-testing in dealing with both texts (i.e., narrative, expository). On the other hand, at the pre-test, the experimental group had recourse to self-questioning in synthesizing the narrative and expository textual input with a frequency of 3.17%. Yet, at the post-test, the experimental group‟s use of self-questioning increased with occurrences of 68.25% and 69.84% for the narrative and expository texts respectively. This, indeed, does not prove that self-questioning is genre-sensitive given the slight difference between the frequent uses of this strategy in coping with both text types (i.e., narrative, expository).

4.3. The Interplay of rereading & Text Genre among EFL learners

Text rereading is deemed a substantial element of the comprehension checking act for the sake of effective textual analysis. The frequent use of text rereading among EFL learners is showcased in Figures 5 & 6.

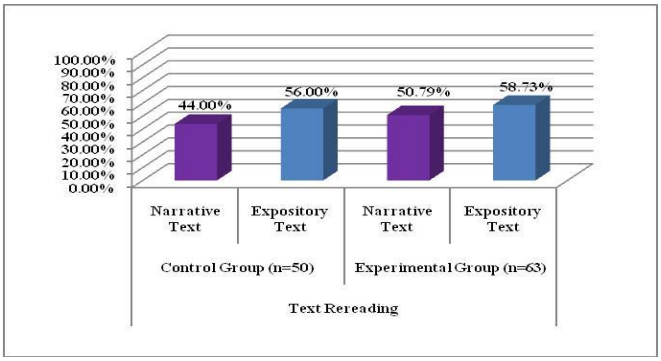


Figure5. Reliance on of text-rereading among EFL groups at pre-testing

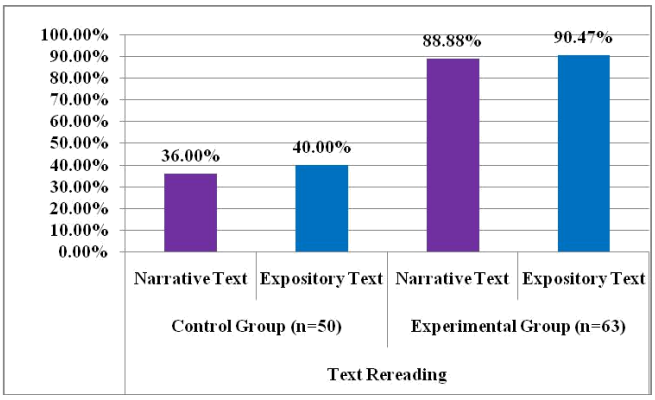


Figure 6. Reliance on of text-rereading among EFL groups at post-testing

As manifested in Figure 5, most EFL control group tended to reprocess the presented texts at the pre-test level with percentages of 44% and 56% for the narrative and expository written discourse respectively. Further, whereas 36% of the controls declared that they re-inspect the narrative text, 40% of the same group reread the expository text at post-testing. Also, a great number of the

experimental subjects did engage in reprocessing the target written texts. Respectively, 50.79% and 58.73% of this group managed to reread the assigned narrative and expository written discourse at pre-testing. Rereading the two texts of narrative and expository type at post-testing was performed by 88.88% and 90.47% of the participants belonging to the experiment group. Clearly, the dominant use of rereading is embodied in processing expository texts among both groups along the pre-post-test continuum.

It is of note that the use of some monitoring reading strategies (i.e., self-monitoring, self-questioning) were not subject to the impact of text type across the pre-post-test levels. However, the strategy of rereading is genre-dependent since it was used with high frequencies in coping with the expository texts across the pre- post-test stages. This indicates that the variable of genre impacts the usage of rereading. Further, the treatment group‟s exposure to the strategy training improved their rereading strategy use. Yet, the frequency of using this strategy remained heavily dependent on the agent of text genre (e.g., narrative, expository). In fact, considering that strategy intervention is a contributing constituent to the improvement of the learners‟ strategy use across diverse texts (e.g., narrative, expository), the text genre does impact only rereading. As to self-monitoring and self-questioning, they are not influenced by text type (i.e., narrative, expository). This is in accord with the postulate stated by some researchers (e.g., Zabrucky & Ratner,1992; Zabrucky & Moore, 1999) that text type does not influence learners‟ monitoring act.

The fact of the matter is that whilst the execution of some monitoring tactics (i.e., self-monitoring, self-questioning) is not governed by the type of the text (i.e., narrative, expository) being processed among EFL learners, rereading remains subject to the text genre impact given the difficulty that typically characterizes expository texts. The latter entails the frequent use of rereading. This backs up the view of a grouping of educational researchers (e.g., Geva and Reyan, 1985; Best et al., 2008; Yoshida, 2012) that expository texts are more difficult to process and analyze than narrative texts.

V. CONCLUSION

The present study sought to gauge the text genre impact (i.e., narrative, expository) on English department learners‟ comprehension monitoring within the Moroccan context. It also evinces the extent to which strategy intervention can influence the targeted learners‟ comprehension monitoring as concerns text typology (narrative & expository). These two postulates are deemed as a guiding signpost for conducting a comprehensive investigation.

Apparently, the cognitive engagement in text processing varies according to genre since the accessibility to textual input of expository type imposes more frequent use of rereading on EFL learners for the goal of attaining a certain kind of sufficiency in terms of comprehension This is supported by many researchers (e.g., Best et al., 2008; Zhou & Siriyothin, 2011). Thus, the claim put in Hypothesis 1 that Moroccan EFL university students‟ strategic monitoring is not genre-specific is partly confirmed. Indeed, rereading, which is a constituent element in comprehension monitoring, is dependent on genre as opposed to self-monitoring and self-questioning which are remote from „genre-specificity‟.

Though the use of self-monitoring and self-questioning was substantively enhanced among the experimental subjects at post-testing, they were made use of in interpreting the content of both types of texts by both EFL groups with roughly similar frequencies. This underlying principle places a tremendous stress on the absence of the genre impact on the EFL learners‟ monitoring behavior (self-monitoring & self-questioning). Yet, EFL learners‟recourse to rereading is governed by genre (i.e., narrative, expository) since the characteristic features and the inherent content of the expository written text entail heavier dependency on rereading that enables an effective understanding. This is in accordance with the view stated in the second research hypothesis that strategy instruction does not have an impact on English department learners‟ monitoring behaviour with regard to text genre (i.e., narrative, expository).

The implications drawn from the study are as follows. It is clear that exposure to reading strategy instruction helps EFL learners enhance their awareness of comprehension monitoring moves used in each given text type (i.e., narrative, expository). Thus, learners can foster the potent capability of effectively tracking their grasp of the content of differing texts. Also, improving EFL learners‟ perception of „genre awareness‟ can direct their monitoring processes towards the optimal way of tackling the given text (narrative & expository) in an efficient fashion. Overall, it is suggested that other Faculties of Letters & Humanities as well as faculties of Sciences be prospective case studies for confirmatory purposes. It is further recommended that a comparative investigation between EFL male learners and EFL female learners at the level of comprehension monitoring as to narrative and expository texts be conducted for the sake of research findings expansion.

REFERENCES

[1] Afflerbach, P., "The influence of prior knowledge and text genre on readers‟ prediction strategies". Journal of Reading Behaviour, vol. XXII , no. 2, pp.131-148, 1990

[2] Baretta, L., Tomitch, L. M. B., MacNair, N., Lim, V. K., and Waldie, K. E. "Inference making while reading narrative and expository texts: An ERP study". Psychology and Neuroscience, vol. 2, no. 2, pp.137-145, 2009

[3] Best, R. M., Floyd, R. G., and McNamara, D. S. "Differential competencies contributing to children‟s comprehension of narrative and expository texts". Reading Psychology, vol.29,no.2, pp.137-164, 2008

[4] Bharuthram, S. "Making a case for the teaching of reading across the curriculum in higher education". South African Journal of Education, vol.32, no. 2, pp.205-214,2012.

[5] Brezin, M., J. "ECTJ/ERIC-IR young scholar paper: Cognitive monitoring: From learning theory to instructional applications". Educational Communication and Technology, vol.28, no.4, pp.227-242,1980.

[6] Geva, E., and Ryan, E. B. "Use of conjunctions in expository texts by skilled and less skilled readers". Journal of Reading Behavior, vol.17, no.4, pp. 331-346, 1985.

[7] Goelman, H. "Selective attention in language comprehension: Children‟s processing of expository and narrative discourse". Discourse Processes, vol. 5, no.1, pp.53-72, 1982.

[8] Hong-Nam, K., and Leavell, A., G. "Reading strategyinstruction, metacognitive awareness, and self-perception of striving college developmental readers". Journal of College Literacy & Learning, 37, 3-17,2011

[9] Khansir,A.A. and Mohammadi Fard, E. "The impact of narrative and. expository text genres on incidental vocabulary acquisition of Iranian EFL learners". International Journal of Language Learning and Applied Linguistics World, vol.6, no.4, pp.491-506,2014

[10] Kolić-Vehovec, S., Bajšanski, I., and Zubković,B., R. "The role of reading strategies in scientific text comprehension and academic achievement of university students". Review of Psychology, vol.18, no.2, pp.81-90,2011.

[11] Ley, K., and Young, D., B. "Instructional principles for self-regulation". Educational Technology Research and Development, vol.49, no.2, pp.53-103,2001.

[12] Nietfeld, J., L., Cao, L., and Osborne, J., W."Metacognitive monitoring accuracy and student performance in the post-secondary classroom". The Journal of Experimental Education, vol.74, no.1, pp.7-28,2005

[13] Puntambekar, S. "Helping students learn how to learnn from texts: Towards an 1TS for developing metacognition".Instructional Science, vol.23, no.1, pp.163-182,1995.

[14] Rubin, D.L.. "Nonlanguage factors affecting undergraduates‟ judgements of nonnative English-speaking teaching assistants". Research in Higher Education, vol.33, no.4,pp.511-531,1992

[15] Schmitt, M. C., & Newby, T. J. "Metacognition: Relevance to instructional design".Journal of Instructional

Development, vol.9, no.4, pp.29-33,1986.

[16] Wang, M. C., Haertel, G. D., & Walberg, H. J. Learning influences. InH. Walberg & G. Haertel (Eds.), Psychology and educational practice(pp. 199–211). Berkeley, CA: McCuchan Pub,1997

[17] Wolfe, M. B. W., & Mienko, J. A. "Learning and memory of factual content from narrative and expository Text". British Journal of Educational Psychology,vol.77,no.3,pp.541-564,2007

[18] Yoshida, M. The interplay of processing task, text type, and proficiency in L2 reading. Reading in a Foreign Language, vol.24, no.1, pp.1-29, 2012.

[19] Zabrucky, K. M., & Moore, D. "Influence of text genre on adults' monitoring of understanding and recall". Educational Gerontology, vol.25, no.8, pp.691-710,1999.

[20] Zabrucky, K., and Ratner, H. H. Effects of passage type on comprehension monitoring and recall in good and poor readers. Journal of Reading Literacy Research, vol.XXIV, no.3, pp.373-391, 1992.

[21] Zhou, L., and Siriyothin, P. "Effects of text types on advanced EFL learners‟ reading comprehension".Journal of Language and Culture, vol.30, no.2, pp.45-66,2011.

Unconstrained Optimization using Improved Differential Evolution Algorithm with a Varying Non-Linear Decreasing Probability Rule

[1] Tapan Prakash, [2] Sugandh P. Singh, [3] V. P. Singh
[1][2][3] Dept. of Electrical Engg., National Institute of Technology, Raipur, India

Abstract: **In this paper, an improved differential evolution (IDE) algorithm with a varying non-linear decreasing probability rule is used for the optimization of unconstrained problems over continuous space. In IDE, a new mutation strategy is embedded into DE algorithm and combined with basic mutation strategy, DE/rand/1/bin, through a non-linear decreasing probability rule. The aim of this paper is to vary this probability rule from linear to higher order non-linear ones and observe its effect on the performance of IDE. For linear decreasing probability rule, the power is increased from 1 to 10 converting the linear rule into higher order non-linear ones. The algorithm is tested on standard benchmark functions to show the effects of the change in probability rule.**

Keywords: **Crossover; Differential evolution; Improved differential evolution; Mutation; Probability rule.**

I. INTRODUCTION

Large scale global optimization problems and the methods to obtain their solution has become one of the prime concerns of our scientific community. There has been a significant change, since few years, in obtaining the methods for the solution of complex non-linear problems of the practical world with the rise of evolutionary computational methods [1]. For any optimization problem, generally, the motive is to optimize certain criteria regarding properties of a system by varying some of the parameters within the system. The main concerns while obtaining optimal solution using any optimization technique are: it should reach true global optima, fast convergence and it should have minimum number of control parameters. These ideas formed the basis for advent of evolutionary algorithms (EAs) [1]. Through the years, problem specific several EAs have been developed to counter the global optimization problem in more efficient and effective way.

Out of different EAs developed, differential evolution (DE) is considered as one of the most reliable and efficient algorithm because of its simplicity and robustness [2-4]. DE includes initialization, mutation, crossover and selection. The basic mutation strategy employed in DE is the difference of two randomly selected vectors which acts as the random variation for the third vector. Recently, there has been a wide spread applications of DE in different areas like electric power system [5-9], control system and robotics [10, 11], electromagnetism [12, 13], engineering design [13], pattern recognition and image processing [14], etc.

With proven efficacy and reliability, DE has some limitations. DE lacks adequate search capability [15]. DE also encounters premature convergence and stagnation [16]. This problem becomes more severe with badly tuned control parameters [17]. To eliminate all of these limitations, various variants of DE with different mutation schemes and wise selection of control parameters have been reported in literatures [14, 18-23]. An improved DE (IDE) has been introduced with new triangular mutation scheme and a restart mechanism [24].

In this paper, IDE algorithm [24] with variable linear to high non-linear decreasing probability rule is tested on high dimensional standard benchmark functions. The variation from linear to high non-linear decreasing probability rule is carried out by varying the power of the rule in the range [1, 10] while keeping all other parameters (of the algorithm) constant. This work is aimed to find out the optimal type of non-linear decreasing probability rule. The outcomes (figure-of-merit (FOM) and standard deviation (SD) for ten consecutive runs) are obtained on benchmark functions for two different dimensions i.e. 10 and 20.

In section 1, introduction is presented. Section 2 describes differential evolution algorithm. The IDE is discussed in section 3. Section 4 provides results and discussion. Finally, the whole work is concluded in section 5.

2. DIFFERENTIAL EVOLUTION ALGORITHM

DE is a stochastic, population based optimisation algorithm, first introduced by Storn and Price in 1996 [25]. It was developed to optimise real parameter, real valued functions. It is in the class of evolutionary algorithms whose basic steps include initialization, mutation, recombination or crossover and selection. Detailed description of the basic steps of DE is outlined below:

A. Initialization

If there is a function $f(x)$ to be optimised having C number of real parameters, then, the population having R solutions is initialized as:

$$x_i^{G_n} = \left[x_{i,1}^{G_n}, x_{i,2}^{G_n}, \cdots, x_{i,C}^{G_n} \right], \quad i = 1, 2, ..., R \tag{1}$$

where, G_n is the generation number. Upper and lower bounds for each parameter are defined as:

$$x_j^L \le x_{i,j} \le x_j^U, \quad i = 1, 2, ..., R; \quad j = 1, 2, ..., C$$

The initial values are randomly generated within the defined bounds $\left[x_j^L, x_j^U \right]$. The fitness function values, $f\left(x_{i,j}\right)$ for $i = 1, 2, ..., R$ and $j = 1, 2, ..., C$, are evaluated with randomly generated population.

B. Mutation

Mutation is a process which explores the search space. For a given parameter vector, $x_i^{G_n}$, three vectors, $x_{r1}^{G_n}$, $x_{r2}^{G_n}$ and $x_{r3}^{G_n}$, are randomly selected such that the indices i, $r1$, $r2$ and $r3$ are distinct. The weighted difference of the two vectors is added to the third vector as follows:

$$v_i^{G_n+1} = x_{r1}^{G_n} + M_F \left(x_{r2}^{G_n} - x_{r3}^{G_n} \right) \tag{2}$$

where, M_F is the mutation factor which is a constant value in the range [0, 2]. $v_i^{G_n+1}$ is called mutant vector.

C. Recombination

Recombination incorporates successful solutions from the previous generations. The trial vector, $u_i^{G_n+1}$, is developed from the elements of the target vector, $x_i^{G_n}$, and the elements of the mutant vector, $v_i^{G_n+1}$. The elements of target vector enter into the trial vector with a probability according to the following rule:

$$u_i^{G_n+1} = \begin{cases} v_{i,j}^{G_n+1} & \text{if } rand_{i,j} \le CR \text{ or } j = I_{rand} \\ x_{i,j}^{G_n} & \text{Otherwise} \end{cases} \tag{3}$$

where, CR is crossover rate, $CR \in (0, 2)$, and $rand_{i,j} \in [0,1]$. I_{rand} is a random integer in the range $[1, C]$ which ensures that $v_i^{G_n+1} \ne x_i^{G_n}$.

D. Selection

The target vector, $x_i^{G_n}$, is compared with the trial vector, $v_i^{G_n+1}$, in terms of figure-of-merit and the one with the lower figure-of-merit is admitted to the next generation according to following rule:

$$x_i^{G_n+1} = \begin{cases} u_{i,j}^{G_n+1} & \text{if } f\left(u_{i,j}^{G_n+1}\right) \le f\left(x_{i,j}^{G_n}\right) \\ x_{i,j}^{G_n} & \text{Otherwise} \end{cases}$$

Mutation, recombination and selection are continued until a stopping criterion is reached.

1. Improved Differential Evolution (IDE) Algorithm

There is no assurance of reaching global optima due to local optima trappings in DE. In IDE, new triangular mutation scheme and restart mechanism in addition to adjustment of control parameters are introduced to improve its performance. The detailed description of IDE is given below:

A. Initialization

The population is generated randomly according to (1) and(2). The fitness function values, $f\left(x_{i,j}^{G_n}\right)$ for $i=1,2,...,R$ and $j=1,2,...,C$, are obtained with randomly generated population.

B. Crossover and Mutation

A dynamic crossover probability scheme is used in order to maintain a balance between global exploration capability and local search exploitation capability. It is expressed as follows:

$$C_R = C_R^{\max} + \left(C_R^{\min} - C_R^{\max}\right)\left(1 - \frac{G_n}{Ng}\right)^k \quad (6)$$

where, C_R is the crossover rate with $C_R^{\min}$ and $C_R^{\max}$ as its minimum and maximum values, respectively, Ng is the total number of generations, and k is a positive number whose value is constant (and taken as 4 in IDE).

A new triangular mutation scheme is adopted in IDE for the enhancement of local search capability and it is expressed as:

$$v_i^{G_n+1} = \overline{x}_c^{G_n} + 2M_F\left(x_{best}^{G_n} - x_{worst}^{G_n}\right) \quad (7)$$

where, $\overline{x}_c^{G_n}$ is a convex combination vector, and $x_{best}^{G_n}$ and $x_{worst}^{G_n}$ are the best and the worst individuals among the randomly selected three vectors, respectively. $v_i^{G_n+1}$ is the mutant vector. The convex combination vector, $\overline{x}_c^{G_n}$, is described as the linear combination of three vectors by the following rule:

$$\overline{x}_c^{G_n} = p_1 x_{best}^{G_n} + p_2 x_{better}^{G_n} + p_3 x_{worst}^{G_n} \quad (8)$$

where, $x_{best}^{G_n}$, $x_{better}^{G_n}$ and $x_{worst}^{G_n}$ are the tournament best, better and worst among randomly selected vectors, respectively. The real weights p_i satisfy $p_i \geq 0$ and $\sum_{i=1}^{3} p_i = 1$. The real weights p_i are given by $p_i = m_i / \sum_{i=1}^{3} m_i$, where $m_1 = 1$, $m_2 = rand(0.75,1)$ and $m_3 = rand(0.5, m_2)$. The mutation scheme mentioned in (7), is embedded with basic mutation scheme of DE algorithm through a non-linear decreasing probability rule. If $\left(U(0,1) \geq \left(1 - \frac{G_n}{Ng}\right)^a\right)$, where $U(0,1)$ is a random number in the range [0, 1], Ng is total number of generations, and a is degree of nonlinearity (that is taken as 2 IDE), then, the mutation scheme given by (7) is applied otherwise the basic mutation scheme is applied. The basic mutation scheme of DE is expressed in (3).

C. Selection

A greedy selection scheme described by (5) is employed to select the new mutant vector, x_i, which will take part in next generation.

D. Restart Mechanism

A restart mechanism is applied for all solution vectors that should satisfy the following condition:

If $\left|f(x^{current}) - f\left(x^{previous}\right)\right| \leq \delta$ for $K = 20$ generations, then restart mechanism starts. $f\left(x^{current}\right)$ and $f\left(x^{previous}\right)$ are current and previous fitness function values, respectively, and δ is a predetermined level of tolerance (10e-6) for predetermined allowable number of generations, K. If $\left(U(0,1)\right) \geq 0.5$, then, random mutation scheme is applied otherwise modified breeder genetic algorithm (BGA) mutation scheme is applied. The random mutation is expressed as:

 ISBN: 978-81-932966-9-1

$$x_{i,j} = \begin{cases} x_j^{\min} + rand_{i,j}\,(x_j^{\max} - x_j^{\min}) & j = I_{rand} \\ x_{i,j} & \text{otherwise} \end{cases}$$

The modified BGA mutation is expressed as:

$$x_{i,j} = \begin{cases} x_j^{\min} + rand_{i,j}\,(x_j^{\max} - x_j^{\min})\omega & j = I_{rand} \\ x_{i,j} & \text{otherwise} \end{cases}$$

where, ω can be calculated from a distribution as follows:

$$\omega = \sum_{k=0}^{15} \omega_k \times 2^{-k}, \quad \omega_k \in (0,1)$$

Before mutation, the value of ω_k are set to 0.

2. RESULTS AND DISCUSSIONS

IDE algorithm is used to minimize five standard benchmark functions. The considered functions are expressed as follows:

A. *Ackley Function*

$$f(x) = -20\exp\left(-0.2\sqrt{\frac{1}{D}\sum_{i=1}^{D} x_i^2}\right) - \exp\left(\frac{1}{D}\sum_{i=1}^{D}\cos(2\pi x_i)\right) + 20 + e$$

B. *Greiwank Function*

$$f(x) = \sum_{i=1}^{D}\frac{x_i^2}{4000} - \prod_{i=1}^{D}\cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$$

C. *Rastrigin Function*

$$f(x) = \sum_{i=1}^{D}\left(x_i^2 - 10\cos(2\pi x_i) + 10\right)$$

D. *Schefe2.22 Function*

$$f(x) = \sum_{i=1}^{D}|x_i| + \prod_{i=1}^{D}|x_i|$$

E. *Sphere Function*

$$f(x) = \sum_{i=1}^{D} x_i^2$$

All functions defined above from (12)-(16) have optimum value as $f(x^*) = 0$ where x^* denotes value of decision vector at optimum point.

Table I and Table II shows the values of figure-of-merit (FOM) and standard deviation (SD) for different benchmark functions with the value of a in the range [1, 10] for a dimension $C = 10$. From the tables I and II, it is observed that the minimum values of the functions occur a maximum of two times for $a = 3$ and $a = 8$. Table III and Table IV presents the value of FOM and SD for same benchmark functions taking the value of a in the same range of [1, 10] for dimension $C = 20$. From the tables III and IV, it is seen that the minimum values of the functions occur four times when $a = 2$ and two times each when $a = 1$, $a = 3$ and $a = 10$. If the dimension of the problem is small, then, it is more justified to keep a non-linear cubic decreasing probability rule as it is producing better results than a non-linear quadratic decreasing probability rule. Moreover, if the dimension of the problem is large, then, in most cases, a non-linear quadratic decreasing probability rule is producing better results but in some of the cases, a non-linear cubic decreasing probability rule is producing better results. So, keeping a non-linear cubic decreasing probability rule can be better choice for obtaining optimum results.

Table I: FOM and SD

$C = 10$		
$a = 1$		
Functions	FOM	SD
Ackley	2.8826e-07	1.6312e-07
Griewank	6.6946e-15	9.7119e-15
Rastrigin	18.5057	**6.7869**
Schefel2.22	2.2499e-14	1.5584e-14
Sphere	5.7718e-25	1.0257e-24
$a = 2$		
Ackley	1.3658e-07	6.6412e-08

 ISBN: 978-81-932966-9-1

Griewank	6.1062e-16	1.2725e-15
Rastrigin	18.8479	8.1271
Schefel2.22	1.1437e-14	1.2305e-14
Sphere	5.2725e-26	1.2569e-25
$a = 3$		
Ackley	1.4718e-07	8.9614e-08
Griewank	4.3298e-16	6.8728e-16
Rastrigin	17.30231	8.5221
Schefel2.22	7.8860e-15	9.7570e-15
Sphere	**1.0113e-26**	**1.5578e-26**
$a = 4$		
Ackley	1.2793e-07	1.1570e-07
Griewank	4.1078e-16	9.3042e-16
Rastrigin	14.67632	9.9639
Schefel2.22	6.0032e-15	8.7704e-15
Sphere	1.0897e-26	1.9499e-26
$a = 5$		
Ackley	1.1536e-07	7.0967e-08
Griewank	1.2212e-16	1.1040e-16
Rastrigin	**12.98910**	8.5741
Schefel2.22	3.6154e-15	5.2630e-15
Sphere	1.3136e-25	4.0082e-25

Table II: FOM and SD

$C = 10$		
$a = 6$		
Functions	FOM	SD
Ackley	9.0409e-08	**4.6091e-08**
Griewank	2.1094e-16	3.2452e-16
Rastrigin	16.8611	13.945
Schefel2.22	1.4459e-13	4.13259e-13
Sphere	3.4589e-15	5.8451e-15
$a = 7$		
Ackley	1.2147e-07	5.6421e-08
Griewank	9.9920e-17	**9.7210e-17**
Rastrigin	19.0973	8.3672
Schefel2.22	5.5564e-15	7.46785e-15
Sphere	3.5055e-15	4.3662e-15
$a = 8$		
Ackley	1.1614e-07	8.7394e-08
Griewank	3.8857e-16	7.6784e-16
Rastrigin	20.80790	10.4787
Schefel2.22	**2.9772e-15**	**2.7300e-15**
Sphere	2.0685e-15	2.8359e-15
$a = 9$		
Ackley	**7.9390e-08**	5.7516e-08
Griewank	1.3322e-16	1.3647e-16
Rastrigin	19.72020	12.5272
Schefel2.22	3.88840e-15	2.9423e-15
Sphere	5.3270e-15	4.7036e-15
$a = 10$		
Ackley	9.5557e-08	7.8186e-08
Griewank	**8.8817e-17**	1.0202e-16
Rastrigin	18.9178	8.6115
Schefel2.22	4.38591e-15	8.97118e-15
Sphere	3.6358se-15	6.7057e-15

Table III: FOM and SD

$C = 20$		
$a = 1$		
Functions	FOM	SD
Ackley	**0.00240**	**0.0012**
Griewank	4.2594e-11	6.2037e-11
Rastrigin	86.8771	22.1304
Schefel2.22	3.3878e-05	4.7351e-05
Sphere	1.6048e-10	1.9678e-10
$a = 2$		
Ackley	0.0887	0.2697
Griewank	4.4121e-11	4.7068e-11
Rastrigin	61.8010	36.9642
Schefel2.22	**1.1784e-05**	**1.0917e-05**
Sphere	**3.1629e-11**	**2.9448e-11**
$a = 3$		
Ackley	0.1739	0.3596
Griewank	**4.9199e-12**	**3.7453e-12**
Rastrigin	60.6750	42.2462
Schefel2.22	2.2856e-05	3.6476e-05
Sphere	1.7287e-10	2.6856e-10
$a = 4$		
Ackley	0.5516	0.4836
Griewank	2.0077e-11	3.4836e-11
Rastrigin	53.2954	34.5025
Schefel2.22	2.4033e-05	3.1772e-05
Sphere	1.7892e-09	3.4348e-09
$a = 5$		
Ackley	0.4652	0.4976

Griewank	1.4614e-11	2.4408e-11
Rastrigin	40.2245	36.3431
Schefel2.22	1.5357e-04	4.2347e-04
Sphere	5.3192e-09	1.1168e-08

Table IV: FOM and SD

$C = 20$		
$a = 6$		
Functions	FOM	SD
Ackley	0.5275	0.5853
Griewank	2.0302e-11	2.0847e-11
Rastrigin	33.3329	22.2339
Schefel2.22	3.3878e-05	2.7880e-05
Sphere	7.9934e-06	9.4408e-06
$a = 7$		
Ackley	0.6117	0.5602
Griewank	2.9231e-10	8.8748e-10
Rastrigin	45.9989	35.2808
Schefel2.22	0.0097	0.0306
Sphere	2.6715e-05	2.8202e-05
$a = 8$		
Ackley	0.8175	0.4804
Griewank	1.4311e-11	1.5698e-11
Rastrigin	52.8582	35.8208
Schefel2.22	2.8045e-05	4.0794e-05
Sphere	3.7689e-05	3.3562e-05
$a = 9$		
Ackley	0.1754	0.3588
Griewank	1.8239e-11	2.6025e-11
Rastrigin	43.0835	29.2105
Schefel2.22	1.9261e-05	3.9341e-05
Sphere	3.9511e-05	1.0206e-04
$a = 10$		
Ackley	0.8164	0.6252
Griewank	5.3996e-11	1.2630e-10
Rastrigin	**28.4443**	**17.7596**
Schefel2.22	5.2898e-04	0.0016
Sphere	3.2697e-05	6.0209e-05

3. CONCLUSION

This work presents IDE algorithm for solving global numerical optimization problems over continuous space with a varying non-linear decreasing probability rule. The non-linear decreasing probability rule allows the new triangular mutation scheme of IDE to be embedded with basic mutation strategy of DE. This probability rule is varied from linear decreasing rule to highly non-linear decreasing rule. This is tested on five standard benchmark functions and the results of FOM and SD are obtained. The results show that if the dimension of problem is small, then, it is more advisable to use a non-linear cubic decreasing probability rule than a non-linear quadratic decreasing probability rule. Moreover, if the dimension of the problem is large, then, in most cases, a non-linear quadratic decreasing probability rule is producing better results but in some of the cases, a non-linear cubic decreasing probability rule can be opted. So, keeping a non-linear cubic decreasing probability rule can be better choice for optimal solution.

REFERENCES:

[1] T. Bäck, *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*: Oxford university press, 1996.

[2] K. V. Price, "Differential evolution vs. the functions of the 2 nd ICEO," in *Evolutionary Computation, 1997., IEEE International Conference on*, 1997, pp. 153-157.

[3] K. Price and R. Storn, "Differential Evolution-a simple and efficient adaptive scheme for global optimization over continuous space," *Berkeley: Intenational Computer Science Institute*, 1995.

[4] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous

spaces," *Journal of global optimization,* vol. 11, pp. 341-359, 1997.

[5] L. D. S. Coelho and V. C. Mariani, "Combining of chaotic differential evolution and quadratic programming for economic dispatch optimization with valve-point effect," *Power Systems, IEEE Transactions on,* vol. 21, pp. 989-996, 2006.

[6] C. Liang, *et al.*, "Parallel optimal reactive power flow based on cooperative co-evolutionary differential evolution and power system decomposition," *Power Systems, IEEE Transactions on,* vol. 22, pp. 249-257, 2007.

[7] C.-T. Su and C.-S. Lee, "Network reconfiguration of distribution systems using improved mixed-integer hybrid differential evolution," *Power Delivery, IEEE Transactions on,* vol. 18, pp. 1022-1027, 2003.

[8] K. Mandal and N. Chakraborty, "Short-term combined economic emission scheduling of hydrothermal power systems with cascaded reservoirs using differential evolution," *Energy Conversion and Management,* vol. 50, pp. 97-104, 2009.

[9] H. Cai, *et al.*, "Application of differential evolution algorithm for transient stability constrained optimal power flow," *Power Systems, IEEE Transactions on,* vol. 23, pp. 719-728, 2008.

[10] C.-H. Chen, *et al.*, "Nonlinear system control using adaptive neural fuzzy networks based on a modified differential evolution," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* vol. 39, pp. 459-473, 2009.

[11] F. Neri and E. Mininno, "Memetic compact differential evolution for cartesian robot control," *Computational Intelligence Magazine, IEEE,* vol. 5, pp. 54-65, 2010.

[12] A. Qing, "Electromagnetic inverse scattering of multiple two-dimensional perfectly conducting objects by the differential evolution strategy," *Antennas and Propagation, IEEE Transactions on,* vol. 51, pp. 1251-1262, 2003.

[13] R. Storn, "System design by constraint adaptation and differential evolution," *IEEE Transactions on Evolutionary Computation,* vol. 3, pp. 22-34, 1999.

[14] S. Das, *et al.*, "Automatic clustering using an improved differential evolution algorithm," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on,* vol. 38, pp. 218-237, 2008.

[15] N. Noman and H. Iba, "Accelerating differential evolution using an adaptive local search," *Evolutionary Computation, IEEE Transactions on,* vol. 12, pp. 107-125, 2008.

[16] S. Das, *et al.*, "Differential evolution using a neighborhood-based mutation operator," *Evolutionary Computation, IEEE Transactions on,* vol. 13, pp. 526-553, 2009.

[17] S. Das and A. Konar, "Automatic image pixel clustering with an improved differential evolution," *Applied Soft Computing,* vol. 9, pp. 226-236, 2009.

[18] A. K. Qin, *et al.*, "Differential evolution algorithm with strategy adaptation for global numerical optimization," *Evolutionary Computation, IEEE Transactions on,* vol. 13, pp. 398-417, 2009.

[19] S. M. Islam, *et al.*, "An adaptive differential evolution algorithm with novel mutation and crossover strategies for global numerical optimization," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 42, pp. 482-500, 2012.

[20] Z. Yang, *et al.*, "Self-adaptive differential evolution with neighborhood search," in *Evolutionary Computation, 2008. CEC 2008.(IEEE World Congress on Computational Intelligence). IEEE Congress on*, 2008, pp. 1110-1116.

[21] J. Brest and M. S. Maučec, "Population size reduction for the differential evolution algorithm," *Applied Intelligence,* vol. 29, pp. 228-247, 2008.

[22] J. Zhang and A. C. Sanderson, "JADE: adaptive differential evolution with optional external archive," *Evolutionary Computation, IEEE Transactions on,* vol. 13, pp. 945-958, 2009.

[23] D. Zaharie, "A comparative analysis of crossover variants in differential evolution," *Proceedings of IMCSIT,* pp. 171-181, 2007.

[24] A. W. Mohamed, "An improved differential evolution algorithm with triangular mutation for global numerical optimization," *Computers & Industrial Engineering,* vol. 85, pp. 359-375, 2015.

[25] R. Storn and K. V. Price, "Minimizing the Real Functions of the ICEC'96 Contest by Differential Evolution," in *International Conference on Evolutionary Computation*, 1996, pp. 842-844.