

BBRC

Bioscience Biotechnology
Research Communications

SPECIAL ISSUE NUMBER-1 (2018)

Print ISSN: 0974-6455

Online ISSN: 2321-4007

CODEN: BBRCBA

www.bbrc.in

University Grants Commission (UGC)
New Delhi, India Approved Journal

An International Peer Reviewed Open Access
Journal For Rapid Publication

Published By:

Society for Science & Nature (SSN)

Bhopal, India

Indexed by Thomson Reuters ISI ESCI

Online Content Available: Every 3 Months at www.bbrc.in



SOCIETY
FOR SCIENCE AND NATURE

SCIENCE FOR LIFE

Registered with the Registrar of Newspapers for India under Reg. No. 498/2007
Bioscience Biotechnology Research Communications
SPECIAL ISSUE NUMBER-1 (2018)

Generic tool for better learning educational paradigms using mind mapping strategies: Cloud framework Raju Govindaraj Babukarthik, Satheesh Kumar and J. Amudhavel	1-4
Energy Efficient Intelligent Decision-Making System for Coal Mining Raju Govindaraj Babukarthik, Satheesh Kumar and J. Amudhavel	5-9
Secured Rapid Data Retrieval for Cloud Framework (RDRCF) using Multi Objective Particle Swarm Optimization Raju Govindaraj Babukarthik, Satheesh Kumar and J. Amudhavel	10-14
Intelligent Decision Making System encompassing Security Framework for VM scheduling Raju Govindaraj Babukarthik, John Monica, Gnanasekaran Sambasivam and J. Amudhavel	15-18
Technical Analysis and Appraisal of Statistical Tools in Fluid Flow Turbulence with Applications J. Amudhavel, L. S. S. Reddy, M. Rajeswari and R. G. Babu Karthik	19-23
Deep insight in to the methods and applications on Aerodynamics of Flight J. Amudhavel, V. Srikanth, T. Kalai priyan and A. Bhuvaneshwari	24-28
A Detailed Analysis of Computational Fluid Dynamics Methods and Its Engineering Applications J. Amudhavel, L. S. S. Reddy, R. S. Raghav and G. Sambasivam	29-33
Fluid Dynamics in advanced medical Diagnostic Technologies: An in-depth Analysis J. Amudhavel, V. Srikanth, R. G. Babukarthik and G. Sambasivam	34-38
Investigation of various numerical methods for Computational Fluid mechanics J. Amudhavel, L.S.S. Reddy, T. Asvany and D. Saravanan	39-43
Shuffled frog-leaping algorithm SFLA for solving Load Balancing Problem through Fog Computing in Cloud Servers T. Asvany, J. Amudhavel and Sujatha Pothula	44-47
Solving Virtual Machine placement in Cloud data centre based on Novel Firefly algorithm T. Kalai priyan, J. Amudhavel and Sujatha Pothula	48-53
A novel NM-ABC algorithm to solve Economic Power Dispatch Problem with Transmission Loss and Valve Points M. Rajeswari, J. Amudhavel and P. Dhavachelvan	54-61
Enhanced Hybrid QoS Framework for Intelligent Decision Making (IDM) in Fog Computing Gnanasekaran Sambasivam, J. Amudhavel and Raju Govindaraj Babukarthik	62-64
Securing SQL Injection flaw R. Rajadurai, V. Bhadrinath, B. Anandhavel, N. Danapaquiame and J. Amudhavel	65-74
Secure Data Process in Distributed Cloud Computing P. Sharmila, Dr. N. Danapaquiame, R. Subhapriya, A. Janakiram and J. Amudhavel	75-84
A Novel Clustering Algorithm for Big Data: K-Means-Fuzzy C Means A. Manikandan, N. Danapaquiame, R. Gayathri, E. Kodhai and J. Amudhavel	85-93
Blocking Abusive and Analysis of Tweets in Twitter Social Network Using NLP in Real-Time R. Rahin Batcha, K. Prem Kumar, N. Danapaquiame, J. Arumugam and D. Saravanan	94-103
Frequent Item set Using Abundant Data on Hadoop Clusters in Big Data N. Danapaquiame, V. Balaji, R. Gayathri, E. Kodhai and G. Sambasivam	104-112

Continued Inside Cover

Published By:
Society For Science and Nature
Bhopal, Post Box 78, GPO,
462001 India

ISSN 0974-6455



9 770 974 1645 002

Registered with the Registrar of Newspapers for India under Reg. No. 498/2007
Bioscience Biotechnology Research Communications
SPECIAL ISSUE NUMBER-1 (2018)

Topmost-K Frequent Itemsets mining using compressed POCTree and Nodesets Boina Nataraj, Pothula Sujatha and J. Amudhavel	113-122
An enhancement of data locality in Hadoop distributed file system A. Siva Krishna Reddy, Pothula Sujatha, Prasad Koti, P. Dhavachelvan and J. Amudhavel	123-133
Job Scheduling Optimization using BAT Algorithm in Hadoop Environment Raghav R. S, J. Amudhavel and P. Dhavachelvan	134-139

CONTENTS



SPECIAL ISSUE • NUMBER 1 • 2018

Generic tool for better learning educational paradigms using mind mapping strategies: Cloud framework <i>Raju Govindaraj Babukarthik, Satheesh Kumar and J. Amudhavel.....</i>	1-4
Energy Efficient Intelligent Decision-Making System for Coal Miningk <i>Raju Govindaraj Babukarthik, Satheesh Kumar and J. Amudhavel.....</i>	5-9
Secured Rapid Data Retrieval for Cloud Framework (RDRCF) using Multi Objective Particle Swarm Optimization <i>Raju Govindaraj Babukarthik, Satheesh Kumar and J. Amudhavel.....</i>	10-14
Intelligent Decision Making System encompassing Security Framework for VM scheduling <i>Raju Govindaraj Babukarthik, John Monica, Gnanasekaran Sambasivam and J. Amudhavel.....</i>	15-18
Technical Analysis and Appraisal of Statistical Tools in Fluid Flow Turbulence with Applications <i>J. Amudhavel, L. S. S. Reddy, M. Rajeswari and R. G. Babu Karthik.....</i>	19-23
Deep insight in to the methods and applications on Aerodynamics of Flight <i>J. Amudhavel, V. Srikanth, T. Kalaipriyan and A. Bhuvanewari.....</i>	24-28
A Detailed Analysis of Computational Fluid Dynamics Methods and Its Engineering Applications <i>J. Amudhavel, L. S. S. Reddy, R. S. Raghav and G. Sambasivam.....</i>	29-33
Fluid Dynamics in advanced medical Diagnositic Technologies: An in-depth Analysis <i>J. Amudhavel, V. Srikanth, R. G. Babukarthik and G. Sambasivam.....</i>	34-38
Investigation of various numerical methods for Computational Fluid mechanics <i>J. Amudhavel, L.S.S. Reddy, T. Asvany and D. Saravanan.....</i>	39-43
Shuffled frog-leaping algorithm SFLA for solving Load Balancing Problem through Fog Computing in Cloud Servers <i>T. Asvany, J. Amudhavel and Sujatha Pothula.....</i>	44-47
Solving Virtual Machine placement in Cloud data centre based on Novel Firefly algorithm <i>T. Kalaipriyan, J. Amudhavel and Sujatha Pothula.....</i>	48-53
A novel NM-ABC algorithm to solve Economic Power Dispatch Problem with Transmission Loss and Valve Points <i>M. Rajeswari, J. Amudhavel and P. Dhavachelvan.....</i>	54-61
Enhanced Hybrid QoS Framework for Intelligent Decision Making (IDM) in Fog Computing <i>Gnanasekaran Sambasivam, J. Amudhavel and Raju Govindaraj Babukarthik.....</i>	62-64
Securing SQL Injection flaw <i>R. Rajadurai, V. Bhadrinath, B. Anandhavel, N. Danapaquame and J. Amudhavel.....</i>	65-74
Secure Data Process in Distributed Cloud Computing <i>P. Sharmila, Dr. N. Danapaquame, R. Subhapriya, A. Janakiram and J. Amudhavel.....</i>	75-84
A Novel Clustering Algorithm for Big Data: K-Means-Fuzzy C Meansg <i>A. Manikandan, N. Danapaquame, R. Gayathri, E. Kodhai and J. Amudhavel.....</i>	85-93
Blocking Abusive and Analysis of Tweets in Twitter Social Network Using NLP in Real-Timeg <i>R. Rahin Batcha, K. Prem Kumar, N. Danapaquame, J. Arumugam and D. Saravanan.....</i>	94-103
Frequent Item set Using Abundant Data on Hadoop Clusters in Big Data <i>N. Danapaquame, V. Balaji, R. Gayathri, E. Kodhai and G. Sambasivam.....</i>	104-112
Topmost-K Frequent Itemsets mining using compressed POC Tree and Nodesets <i>Boina Nataraj, Pothula Sujatha and J. Amudhavel.....</i>	113-122
An enhancement of data locality in Hadoop distributed file system <i>A. Siva Krishna Reddy, Pothula Sujatha, Prasad Koti, P. Dhavachelvan and J. Amudhavel.....</i>	123-133
Job Scheduling Optimization using BAT Algorithm in Hadoop Environment <i>Raghav R. S, J. Amudhavel and P. Dhavachelvan.....</i>	134-139

EDITORIAL

SPECIAL ISSUE ON EMERGING TRENDS IN ENGINEERING SCIENCES

Bioscience Biotechnology Research Communications is soliciting high quality manuscripts presenting original contributions for its special issue on *Emerging Trends in Engineering Sciences*. This special issue aims to provide a forum that brings together researchers from academia, practicing engineers from industry, standardization bodies, and government to meet and exchange ideas on a very interesting area of research: Mind map autonomous text based learning strategies and implementation of mind mapping as meta-learning.

This special issue aims to gather recent quality work in the topic of current research trends in the Engineering Science which provides the work with a discussion and evaluation forum. This special issue wants to focus on received original scientific contributions in the form of theoretical and experimental research and case studies. Some of the important research issues discussed in this special issue are Mind map autonomous text based learning strategies and implementation of mind mapping as meta-learning strategy which encourage students cognitive and Meta cognitive text-learning strategy use. Another important research deals with novel NM-ABC algorithm to solve Economic Power Dispatch Problem with Transmission Loss and Valve Points which presents an investigation on meta-heuristic approach which integrates Artificial Bee Colony (ABC) algorithm with Nelder-Mead method (NM-ABC) to solve Economic Power Dispatch (EPD) problem. Some of the recent Techniques such as The Frequent Itemsets Mining (FIM), Association Rule Mining (ARM) algorithm, Pre-Order Coding tree (POC), Topmost-K Frequent Itemsets mining (TKFIN) we briefly analysed and discussed.

This special issue has called for papers from researchers working on making inroads in this fascinating area so that we can gather a state-of-the-art account. All submissions are well supported by proof of the concept studies with a direct and simulated comparison to the technical solutions, mathematical models, concepts, standards, designs, implementations, running experiments or applications.

Dr. Jafar A. Alzubi

Associate Professor,
School of Engineering,
Al-Balqa Applied University,
Jordan.

Dr. Ramani Kannan

Senior Lecturer,
Department of Electrical and Electronics Engineering,
Universiti Teknologi PETRONAS,
Malaysia.

Generic tool for better learning educational paradigms using mind mapping strategies: Cloud framework

Raju Govindaraj Babukarthik¹, Satheesh Kumar² and J. Amudhavel³

¹Part time Research Scholar Category-B, Bharathiar University, Coimbatore, Tamil Nadu, India

²Department of Computer Application, Bharathiar University, Coimbatore, Tamil Nadu, India

³Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

ABSTRACT

Neuro-education is a growing field that seeks to mix the cooperative fields of neuroscience, cognitive science psychology and education for improved understanding of how we learn and how information is used to create enhanced teaching methods and educational policy. Recent advanced in Artificial Intelligence (AI) and functional neuro imaging procedures (fMRI) software models is used greatly for understanding of mind. Even though various type of graphic organizer is widely studied, various publication of mind map research is executed in secondary or higher education and is concentrated on mathematics. In this proposal our aim to investigate the applicability and efficiency of the cloud framework for better learning tool is using the mind mapping and Meta learning strategy. Incorporating the optimization algorithm. mind map strategy greatly support autonomous text based learning strategies and implementing mind mapping as meta-learning strategy in class thereby encourage students cognitive and meta cognitive text-learning strategy use.

KEY WORDS: ARTIFICIAL INTELLIGENCE (AI), FUNCTIONAL NEURO IMAGING PROCEDURES (FMRI), RESEARCHER-PROVIDED MIND MAPS, TEXT-LEARNING STRATEGIES

INTRODUCTION

The major accomplishment of the human brain lies in processing of information for biological activity towards the environment by utilizing the resources more efficiently. Brain can be stated as a multifaceted computational system working on hierarchical architecture with

basic unit as neurons associated as neural networks. The encompassing view of brain is based on the knowledge of the nature, functions, and properties of the components, their interactions, and their relationship to the whole. Resembling same heterarchy and hierarchy between functional modules and structural modules. Mind is stated as a process of brain structural neuronal circuitry

ARTICLE INFORMATION:

*Corresponding Author: r.g.babukarthik@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF : 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

and controlled by synaptic accuitances of hormones and neuro transmitters. Mind works at some situation on the basis of perceives, senses, thinks, understands, plans and perform to impulse of the world person living.

Still exact scientific theory of mind understanding and it is processing is not yet defined, recent advanced in Artificial Intelligence (AI) and functional neuro imaging procedures (fMRI) software models is used greatly for understanding of mind, thus there is an need for architecture model to sustain engineering methodology for explaining process of mind activities.

In this paper, the following process is as follows:

- Development of a cloud based architecture (framework) for education paradigms.
- Implementation of mind mapping and Meta learning strategies.
- Implementation of optimization algorithm for optimal retrieval of data.
- Experimental verification of proposed framework for various clouds.
- Analysis of the proposed framework based on performance factors.

It envisaged the government funded and non-funded universities, engineering college, arts and science colleges, and higher secondary school for promotion of better learning education (Anusha et al. 2015).

From 1950s, theories of operational memory have posited an inhabitant of discrete, lithe representations helping fragmentary maintenance of information believed in mind. This representation is stated by various names such as stimulus samples (Estates 1950), slots (Zhang 2008), chunks (Miller et al 1956) and resources (Bays et al 2008). The service is instantiated as cycles of time (Dharshini et al. 2016) –sequence stimulating process (Vergauwe et al 2009). In evolutionary analysis, selection is stated as a cognitive control process thereby redirecting maintenance (Govindharaj et al. 2016), preference of information based on needs of stored information (Muther et al 1965, Suchow et al 2016). Neuro-education is a growing field that seeks to mix the cooperative fields of neuroscience (Ahilandeswari et al. 2016), cognitive science psychology and education for improved understanding of how(Shanmugam et al. 2017) we learn and how information is used to create enhanced teaching methods, (Rajadurai et al. 2017), educational policy and curricula (Thomas et al 2010).

Text based learning survey states the advantage of organized and well- instructed utilization of graphic organizer or maps (Dexter et al 2011). Based on this dimensional spatial arrangement of word cluster (Stull et al 2007) lead to an organizational learning strategy (Dansereau et al 2009), envasing to dynamic knowledge transformation close to deeper level learning and text

processing (Slotte et al 1999). Some researchers used graphic organizer as meta-learning strategy leading to learn independently and meaningfully, using vast strategy set (Chmielewski et al 1998). Even though various type of graphic organizer is widely studied (Novak et al 2002), various publication of mind map research is executed in secondary or higher education and is concentrated on mathematics (Brinkmann et al 2003, Eddie et al 2017).

MATERIAL AND METHODS

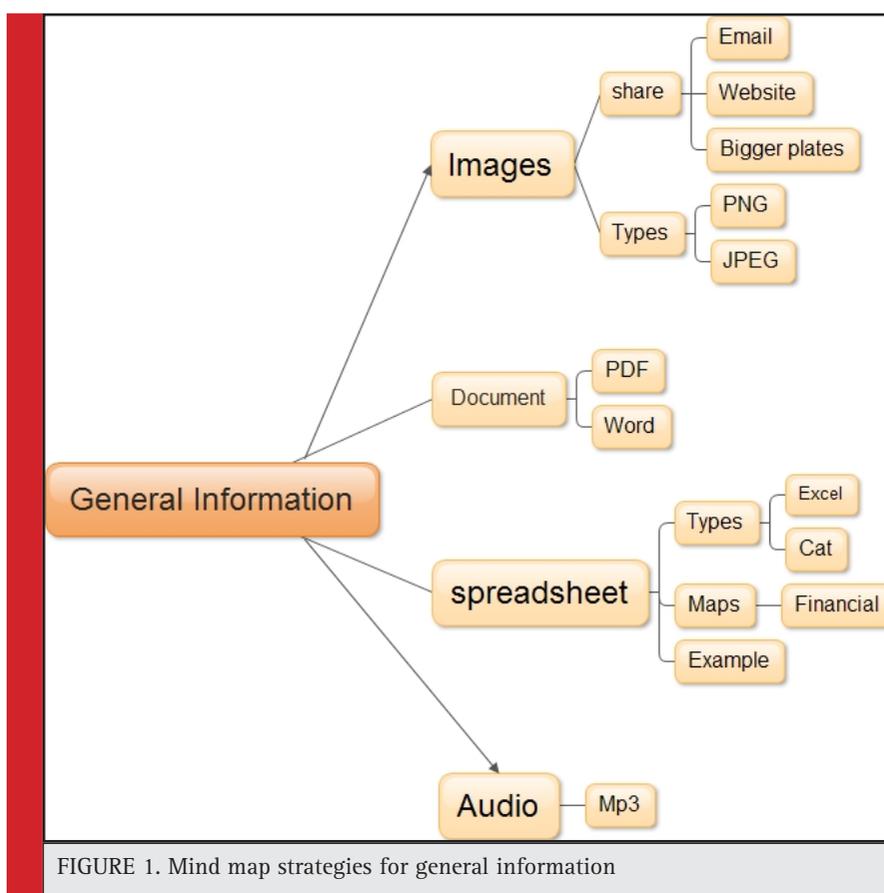
The primary problem is to investigate the applicability and efficiency of the cloud framework for better learning tool using the mind mapping and Meta learning strategy. Incorporating the optimization algorithm for efficient processing and retrieval of information, the optimization algorithm is applied to NP-hard problems (Travelling Salesman Problems (TSP) and Knapsack Problems (KP)) leading to a better performance analysis. MPSO is a variation of PSO where parent swarm is divided into multiple child swarms. The particles very close to each other then they create a child swarm. So, it keeps away from convergence. After convergence, it try to again grouping the child swarm into a new child swarm. The new swarm is started with the particles from the existing swarm.

Three other main factors in favor of the MPSO approach are the following:

- Mind is considered a highly dynamic, uncertain, complex environment; this requires entities capable of flexible autonomous actions.
- Brain/neocortex is naturally modelled as regions/groups of neurons cooperating and competing in order to solve complex problems.
- Data and control are distributed within brain/mind.

Working with mind map strategy greatly support autonomous text based learning strategies and also implementing mind mapping as meta-learning strategy in class thereby encourage students cognitive and meta cognitive text-learning strategy use, leading to greater recall performance during independent text-learning. It is assumed that those student working with student generated mind maps greatly enhanced deeper-level-text-learning strategy use and increase in recall performance compare to student working with student generated maps.

Mind is stated as a process of brain structural neuronal circuitry and controlled by synaptic accuitances of hormones and neurotransmitters. To investigate the applicability and efficiency of the cloud framework for better learning tool using the mind mapping and Meta



learning strategy. Incorporating the optimization algorithm. mind map strategy greatly support autonomous text based learning strategies and also implementing mind mapping as meta-learning strategy in class thereby encourage students cognitive and meta cognitive text-learning strategy use.

The figure 1 shows the general mind map strategies stating the content of information such as images, document, spread sheet and audio. Within the images sharing of images and its types is included which in turn states the email share, website share, and bigger plates share, whereas type states the png and jpeg image types.

RESULTS AND DISCUSSION

Working with mind map strategy greatly support autonomous text based learning strategies and also implementing mind mapping as meta-learning strategy in class thereby encourage students cognitive and meta cognitive text-learning strategy use, leading to greater recall performance during independent text-learning. It is assumed that those student working with student generated mind maps greatly enhanced deeper-level-text-learning strategy use and increase in recall performance compare to student working with student gener-

ated maps. To implement the proposed framework in the cloud environment and to verify the above simulation results with the other strategies.

Computer simulation and experimental results is compared for the validation of the framework and its optimization algorithm. Analysis of various cloud frameworks and to design the framework suitable for better learning educational paradigms. To convert the informative text into the Mind map strategies based on the researcher provided MM and to implement MPSO optimization algorithm for optimal retrieval of informative data. Using MPSO algorithm and mind map strategies to stimulate. Carry out experiment and compare with other mind map strategies.

REFERENCES

- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 1(1): 1-7.
- Bays, P.M. and Husain, M, 2008: Dynamic shifts of limited working memory resources in human vision. Science 321, 851-854.
- Brinkmann, A, 2003: Graphical knowledge display - mind mapping and concept mapping as efficient tools in mathematics education. Mathematics Education Review, 16, 35-48.

- Chmielewski, T. L., & Dansereau, D. F., 1998: Enhancing the recall of text: Knowledge mapping training promotes implicit transfer. *Journal of Educational Psychology*, 90(3), 407-413. doi: 10.1037/0022-0663.90.3.407.
- Dansereau, D. F., & Simpson, D. D., 2009: A Picture Is Worth a Thousand Words: The Case for Graphic Representations. *Professional Psychology-Research and Practice*, 40(1), 104-110, doi: 10.1037/A0011827.
- Dexter, D. D., & Hughes, C. A., 2011: Graphic organizers and student with learning disabilities: a meta-analysis. *Learning Disability Quarterly*, 34(1), 51-72.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). Secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Estes, W.K., 1950: Toward a statistical theory of learning. *Psychol. Rev.* 57, 94-107.
- Eddie B. Jackson Jr, 2017: *Concept Mapping: Developing Critical Thinking through Mind Mapping*, Center for Faculty Excellence, United States Pp1-9.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Muther, W.S., 1965: Erasure or partitioning in short-term memory. *Psychon. Sci.* vol:3, Pp429-430.
- Miller, G.A., 1956: The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* vol 63, Pp81-97.
- Novak, J. D., 2002: Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science education*, vol 86(4), Pp548-571 doi:10.1002/sci.10032.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Stull, A. T., & Mayer, R. E., 2007: Learning by doing versus learning by viewing: three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology*, vol 99(4), Pp808-820, doi:10.1037/0022-0663.99.4.808.
- Suchow, J.W and Griffiths, T.L., 2016: Deciding to remember: memory maintenance as a Markov decision process. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.
- Slotte, V., & Lonka, K., 1999: Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology*, vol 24, Pp1-20, doi:10.1006/ceps.1998.0980.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Thomas j. crew and susan H.Magsamen, 2010: *Neuroscience and Education: An Ideal Partnership for Producing Evidence-Based Solutions to Guide 21st Century Learning*, Neuron 67 September Elsevier.
- Vergauwe, E., 2009: Visual and spatial working memory are not that dissociated after all: a time-based resource-sharing account. *J. Exp. Psychol. Learn. Mem. Cognit.* vol 35, Pp1012-1028.
- Zhang, W.andLuck,S.J, 2008: Discrete fixed-resolution representations in visual working memory. *Nature* vol 453 (7192), Pp 233-235.

Energy Efficient Intelligent Decision-Making System for Coal Mining

Raju Govindaraj Babukarthik¹, Satheesh Kumar² and J. Amudhavel³

¹Part Time Research Scholar Category-B, Bharathiar University, Coimbatore, Tamil Nadu, India

²Department of Computer Application, Bharathiar University, Coimbatore, Tamil Nadu, India

³Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

ABSTRACT

Coal is considered as a necessary and rising energy source that continues to create an important contribution to the universal energy mix for various decades; largest amount of energy is wasted for producing huge amount of coal and responsible for large quantity of pollutant such as waste water, waste gas and industrial residue. The main aim of investment is to reduce pollutants and energy consumption, but the essential need of all the coal mining enterprises is to maximize the profit, thus the problem turn out to be multi-objective decision making. Energy efficiency is achieved with the support of coal manufacturing company by proper implementation of design and energy management system (model), a model must offer a framework for measurement of energy use, analysis, identifying and implementing energy efficient possibility. In this project we proposed a intelligent decision making system for energy savings and emissions reduction using Huddle PSO techniques, thereby providing a best decision making support on environmental investment for achieving Maximum profit and energy saving, optimal solution is provided by considering various parameter in terms of technology, power (diesel and electricity) and pollutants for reducing energy consumption and minimization of pollutants.

KEY WORDS: ANT COLONY OPTIMIZATION (ACO), BRANCH BOUND (BB), MULTI-OBJECTIVE MIXED INTEGER NON-LINEAR PROGRAMMING (MMINLP), PARTICLE SWARM OPTIMIZATION (PSO)

INTRODUCTION

Generally coal mining sector produce coal as fuel for all industries and housing utilization, thereby large amount of energy is wasted for the process of largely produc-

tion of coal and is mainly responsible huge quantity of pollutant such as waste water, waste gas and industrial residue. It is estimated that in 2012 for each 1000 kg of coal produced, 33 kWh of electricity and 24 kg of coal is utilized (Yu et al 2014). The largest coal producing

ARTICLE INFORMATION:

*Corresponding Author: r.g.babukarthik@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal
NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

country china has come across so much of issues in terms of energy and environment from various coal mining industries (Jinanjun tu et al 2014). In addition to energy consumption, coal industries also face other issues in terms of pollutant, discharge of industrial waste water of 1.42 billion tones, waste gas of 32.49 billion m³ and 385.37 million tones of solid wastes in 2012 (Zhenyi Guo et al 2013b).

The latest study of coal mining areas focuses on energy savings and emission reductions; energy reduction strategies concentrated on following aspect they are: namely, system, organization, technology, management, talent and investment (Li et al 2010). Used simulation system for recognizing low-carbon in coal mining areas (Yu et al 2012)], the major theoretical factor to accomplish energy savings and emission reduction in coal mining focus on circular economy or ecological industrial chain (Fang 2007). From enterprise perceptive focuses on reduction of energy usage and emissions (Sheng et al 2010), using clean production and pollution reduction technologies (Zhang et al 2010), and better usage of production supplementary system such as drainage, ventilation and lighting systems (Cheng et al 2011).

Investment in latest technologies and equipment also need to consider for reduction of energy, such as energy saving equipment, energy saving technologies and waste water treatment method. The main aim of investment is to reduce pollutants and energy consumption, but the essential need of all the coal mining enterprises is to maximize the profit, thus the problem turn out to be multi-objective decision making (Chad et al 2017).

The objective function is non-linear and hence the decision making problem turn out to be mixed integer non-linear programming (MINLP), therefore investment in emission reduction and energy savings in coal mining is termed out to be multi-objective mixed integer non-linear programming (MMINLP).

The primary process for energy saving and emission reductions in coal mining areas is

- Using emission as a supplementary product in manufacturing of raw materials.
- Modification of equipment and techniques.
- Radical construction of environmental amenities to reduce wastes.
- Improved monitoring, management of energy consumption and pollutant waste.

The manufacture procedures involved in coal mining include auxiliary production (Gayathri R et al. 2016), basic production and service production (Hwang 1979). The basic production in coal mining areas includes mine field construction, extraction, excavation, elevation and underground transportation. Whereas auxiliary and service production include drainage, underground ventilation, power

supplies, materials and industrial water. In addition maintenance of laneways, electro-mechanical equipment and industrial buildings (Dharshini G et al. 2016).

Coal manufactures starts with selection of coal and washing, most of the country have there own coal manufacturing plants, mine coal fired boilers external sales and pit coal-fired power plants final end point of raw coal after manufacturing. The main energy consuming sources in coal mining areas are electricity, gasoline, raw coal and diesel (Govindharaj I et al. 2016). Electricity is used for equipment in coal manufacturing process, boilers consume raw coal for supplying heat, gasoline is used in office vehicles and diesel is used for transportation vehicle (Rajadurai R et al. 2017).

The main pollutant of coal mining areas in manufacturing process is waste water, waste gas and solid waste. The direct burning of coal gangue produces harmful gases (carbon monoxide, hydrogen sulphide, carbon dioxide, nitric oxide and sulphur dioxide) (Sasidevi V et al. 2017).

For single objective problem optimal solution is achieved by gradient and simplex based predictable optimization techniques or evolutionary algorithm (reduced gradient method) (Burdett et al 2016), Branch Bound (BB) method (Pinto et al 2015) or BB with success scalarizing function approach (Park Lee et al 2013), constraint generation algorithm (Deep et al 2009) and many mixed coded evolutionary algorithm, example genetic algorithm (Lin et al 2015), Particle Swarm Optimization (PSO) (Lin et al 2015) and Ant Colony Optimization (Sahoo et al 2014) with various hybrid algorithm (Rao et al 2009). Thus for multi-objective problems pareto optimal solution is achieved that is one solution is better than other solution (Anusha B et al.2015), the solution providing best solution is termed as pareto solution (Yu et al 2014).

MATERIAL AND METHODS

Profit Maximization: Maximizing profit is the main aim of coal mining companies and is stated as a product of coal out (Xt) and cost per unit of coal output (pt) in year t; total profit in the year t used for energy saving and emission reduction. Hence spending in energy saving and emission reductions, cost reduction of energy consumption and cost reduction of emission treatment.

Minimization of energy consumption: In regards to energy saving rules and policies coal mining companies must have to avoid wasteful production equipment thereby reducing energy consumption per unit coal manufactured and replacing the more energy consuming equipment, which in turn leads to reduction of energy consumption.

Minimization of pollutants: During the coal mining process some of pollutant are waste water, So₂, solid

waste, dust and smoke and dust need to reduce by spending in pollution treatment and various emissions technology.

Constraints: Total coal resource constraints, Annual production capacity constraints, Coal output constraints in the production process, Annual investment constraints, Constraints on investment frequency, Nature of the decision variables.

Proposed Algorithm: *Huddle PSO for multi-objective Problems* is a variant of particle swarm optimization, where multiple sub swarms is used as a replacement for a single swarm and each sub swarm focus on a specified region, it is helpful for the problems which has multiple optimal solutions. *Charged particles*: For maintaining the diversity charged swarms is used and charged particles repel the same charged particles, so that the sub swarms interact with the others within the search space, because of the repulsion of the charged particle, the swarms explore the different region, creating a high possibility for the swarms not to group in the same place. But a problem might arise in this method and neutral (charge less) particles is not repelled or attracted by other charged particles. *Sub-group formation*: cooperative swarm approach, in which population is divided into a large number of sub-groups (sub swarms). The information is exchanged among the particles in the sub swarms and it is regrouped to improve the global searching ability, use of diverse sub-swarms increases performance when optimizing multi-modal functions. To enhance the diversity of Evolutionary Algorithms (EV) for addressing the Dynamic Optimization problems (DOP), the cooperative PSO is used. SPSO is introduced by Parrott and Li, algorithm changes the number and size of the swarms dynamically by creating an ordered list of particles and ranked them by their fitness with spatially nearer particles is grouped to a particular species. In dynamic environment, the charged PSO approach is used to find the multiple optima at the same time with multiple swarms.

Main swarm and sub swarm formation: In the cooperative PSO algorithm, the main swarm M contains c particles with (xi, vi) where xi is the position, vi is the velocity of the particles. The particles in the main swarm is moved to any place by the momentum and the cognitive components of the velocity update function. The sub swarms (Q) is generated from the main swarm (M).

$$Co_p \text{ cost} = \sum_{i=1}^n M_z(v_{ij}, P_{ij}) + \sum_{i=1}^n S_z(v_{ij}, P_{ij}) \quad (1)$$

$$M_z = \sum_{i=0}^n InitV_{ij} + InitP_{ij} \int \nabla i, j \in n \quad (2)$$

$$S_z = \sum_{i=0}^n S_{1(v_{ij}, p_{ij})} + S_{2(v_{ij}, p_{ij})} + \dots + S_m(v_{ij}, p_{ij}) \int \nabla i, j \in n \quad (3)$$

In the above equations, *co_pso* refers the huddle PSO. M_s Represents the main swarm, S_s refers the sub swarm. In *co_pso* equation, to calculate the velocity and position of the main swarm (M_s) and sub swarm (S_s). In the main swarm equation (1), (2) and (4), velocity and position of the swarm is calculated, using the sub swarm equation velocity and position of all the sub swarms is calculated. To explore the search place, the main swarm is used. The sub swarm is generated from the main swarm and particle in the main swarm does not have change in its fitness value during the execution; a new swarm is generated. *Merge swarm*: When a new sub swarm is formed, then the particle nearer to that is get merged into the sub swarm and it is deleted from the main swarm. Sub swarm has a center (radius). The center of the sub swarm is defined by the particle which has the best fitness value. The radius of the sub swarm is the farthest distance of any particle in that swarm. *Absorb swarm*: If some particle is moved to the sub swarm, it gets absorbed into that. Each sub swarm acted individually like the main swarm. These several sub swarms is formed in the region to create more diversity than the normal PSO. This helps to focus on different areas in the space and it is good for finding the optimal solution.

Intelligent decision making system can be used in all coal mining enterprises for achieving better decision in terms of maximum profit, minimization of energy and reduction of pollutants. The optimal solution provides the enterprise to compare with the existing systems and technology used. Importance of the proposed work,

- Using emission as a supplementary product in manufacturing of raw materials.
- Modification of equipment and techniques.
- Radical construction of environmental amenities to reduce wastes.
- Improved monitoring, management of energy consumption and pollutant waste.

100 independent trail need to be performed in the computer environment on the basis of multi-objective and cooperative optimization model, the best case, average case and worst case, standard deviation have to be calculated. The obtained optimized values is matched with coal mine reduction of energy consumption that is value is obtained in terms of *profit priority, energy saving priority, environmental priority and equal priority* based on the objectives. A setup is created to implement the proposed framework in the cloud environment and to verify the simulation results. Computer simulation and experimental results will be compared and case study is performed in the real time scenario for environment decision making. The profit priority in terms of every month, every 3 month, 6 month and year in terms of investment, technology used and usage of secondary

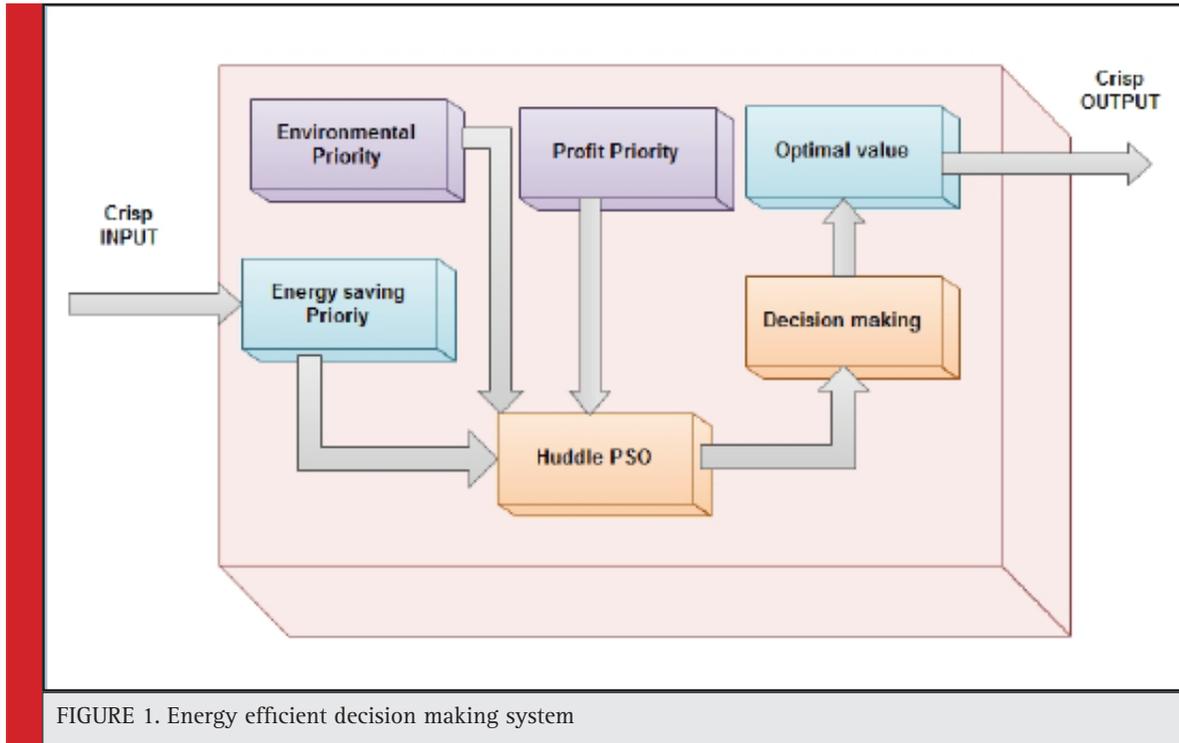


FIGURE 1. Energy efficient decision making system

products during coal mining. On this basis maximum energy saving, minimization of emission and reduction of pollutant is state for all the coal mining enterprises. All the coal mining manufactures, government and small scale coal mining enterprise can utilize for better decision making and reduction of energy consumption. The figure 1 shows the energy efficient decision making system using the huddle PSO and thus the optimal value is obtained based on the decision making system which is to be used in the coal mining enterprises. It consists of crisp input that is all the necessary input is fed into the system based on the energy saving priority, profit priority and environmental priority. The huddle PSO solve the problems the problem based on this constraint and decision making table is generated for analysis, from the table optimal value is chosen and crisp output is produced for coal mining enterprises.

RESULTS AND DISCUSSION

An Intelligent Decision Making System for energy savings and emissions reduction (Cloud based architecture) is proposed. Implemented Huddle PSO algorithm to solve the Multi-objective problems (mixed integer model). Optimal solution is selected by using proposed techniques considering various parameters using Huddle PSO. Analysis of the proposed decision making systems and evaluate the performance. Performance analysis on coal mining company using (Intelligent Decision making system) energy efficient model need to be carried out in the future.

REFERENCES

- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Ahиландeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 1-5.
- Burdett, R. L., Kozan, E, 2016: A Multi-Criteria Approach for Hospital Capacity Analysis. *European Journal of Operational Research*, vol 255(2), Pp505-521.
- Chad O, Hargave, Craig A, James, 2017: Infrastructure-based localisation of automated coal mining equipment, *springer* Pp 1-10.
- Cheng, Y.-P., Wang, L., & Zhang, X.-L, 2011: Environmental impact of coal mine methane emissions and responding strategies in China. *International Journal of Greenhouse Gas Control*, vol 5(1), Pp157-166.
- Deep, K., Singh, K. P., Kansal, M., & Mohan, C, 2009: A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Applied Mathematics and Computation*, vol 212(2), Pp 505-518.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). Secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.

- Fang, Y., Cote, R. P., & Qin, R, 2007: Industrial sustainability in China: practice and prospects for eco-industrial development. *Journal of environmental management*, vol 83(3), Pp 315-328.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Hwang, C.-L., Tillman, F. A., & Kuo, W, 1979: Reliability optimization by generalized Lagrangian-function and reduced-gradient methods. *Reliability, IEEE Transactions on*, vol 28(4), Pp 316-319.
- Jinanjun tu, 2014: CCIA Chinese Coal Industry Association. *China Coal Industry Report*.
- Li, F, 2010: The target and potential study on energy saving and emission reduction of coal enterprises. *Energy Technology and Management*, vol 1, Pp 1-3.
- Lin, Q., Li, J., Du, Z., Chen, J., & Ming, Z, 2015: A novel multi-objective particle swarm optimization with multiple search strategies. *European Journal of Operational Research*, vol 247(3), Pp 732-744.
- Park Lee C., Y., Kim, Y., & Park, G, 2014: An access network design problem with end-to-end QoS constraints. *Omega*, vol 48(0), Pp 36-48.
- Rao, S. S, & Rao S, 2009: *Engineering optimization: theory and practice*: John Wiley & Sons.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 3(1): 7-21.
- Sahoo, L., Banerjee, A., Bhunia, A. K., & Chattopadhyay, S, 2014: An efficient GA-PSO approach for solving mixed-integer nonlinear programming problem in reliability optimization. *Swarm and Evolutionary Computation*, vol 19, Pp 43-51.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Sheng B Yang, L, 2010: Low-carbon development paths discussion of coal mining area. *China energy Journal*, vol 2, Pp 5-7.
- Yu, S., Wei, Y.-M., Guo, H., & Ding, L, 2014: Carbon emission coefficient measurement of the coal-to-power energy chain in China. *Applied Energy*, vol 114, Pp 290-300.
- Yu S, & Wei Y.m2012: Prediction of China's coal production-environmental pollution based on a hybrid genetic algorithm-system dynamics model. *Energy Policy*, vol 42, Pp 521-529.
- Yu, S., Wei, Y.-M., Guo, H., & Ding, L, 2014: Carbon emission coefficient measurement of the coal-to-power energy chain in China. *Applied Energy*, vol 114, PP 290-300.
- Zhenyi Guo, 2013b: NBSC (National Bureau of Statistics of China) (2013b). *National Bureau of Statistics of China. China Energy Statistical Yearbook: China Statistical Press*.
- Zhang J, 2010: Technology and approach of developing low-carbon economy for coal enterprise. *Coal Mining Technology*, vol 15(2), Pp 102-103.

Secured Rapid Data Retrieval for Cloud Framework (RDRCF) using Multi Objective Particle Swarm Optimization

Raju Govindaraj Babukarthik¹, Satheesh Kumar² and J. Amudhavel³

¹Part time Research scholar category-B, Bharathiar University, Coimbatore, Tamil Nadu, India

²Department of Computer Application, Bharathiar University, Coimbatore, Tamil Nadu, India

³Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

ABSTRACT

Cloud computing services is often related to the process of handling enormous amount of data, often Cloud storage leads to many security problems such as data integrity checking, confidentiality and privacy issues, organizations as to assure that cloud service provider as current privacy and security policies for sharing of data with various organization. In this paper, we proposed architecture for cloud security and privacy issues enabling the services for faster access of data providing confidentiality, integrity, authentication and authorization. The experiment is performed using the proposed framework in cloud environment and is computed in software as a service layer. Optimal web service is retrieved using multi objective particle swarm optimization based on QoS factor, the performance is analyzed and the proposed techniques out performed well.

KEY WORDS: BOUNDARY VALUE ANALYSIS, INFRASTRUCTURE AS A SERVICE, PLATFORM AS A SERVICE, QUALITY OF SERVICE, SOFTWARE AS A SERVICE

INTRODUCTION

Cloud computing (Mell et al 2011) delivers on-demand, rapid and scalable operation to a collective pool of shared resources, without considering the organization of those resources (Buyya et al 2010, Zhang et al 2010). Services provided in the cloud are: Software as a Service (SaaS)

it provide all the running software as service thereby user no need to install the required software, Platform as a service (PaaS) a accessible platform for developing conventional software, Infrastructure as a service (IaaS) infrastructure is provided in form of virtual machine, networks and storage. Three types of Clouds is classified so for, public cloud providing service to all user and

ARTICLE INFORMATION:

*Corresponding Author: r.g.babukarthik@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

it is controlled by single entity or organization, private clouds it is owned by a single organization for own purpose (Shanmugam M et al. 2017), Hybrid cloud is stated as a collection of public clouds and private clouds (Dean et al 2004). Cloud computing services is often related to the process of handling enormous amount of data (Geodendorf et al 2013) thereby leading to minimal operational of huge data (big data) (Rajadurai et al. 2017). However this farm out practices is generally seen as path to transfer operative and implementation risks for the service provider, the actual certainty is that proper security mechanism doesn't exist for the data still (Chang et al 2015). Cloud storage leads to many security problems such as data integrity checking, confidentiality and privacy issues when data is hosted in cloud storage server which is accessible by all (Philip Chen et al 2014). For justification tactic guaranteeing privacy and security in usage of cloud services, (Phillips-wren 2014) proposed that an organizations as to assure that cloud service provider as current privacy and security policies for sharing of data with various organization, and (Goodendorf et al 2013) also supported this process by asserting that user and cloud service provider need to clearly mention the data privacy controls. Instead of prescribed articles leading to more cost effective than the common security needs.

Security and privacy: Security states the amount to which user's trust the services provided by the cloud and is secure for storing and sharing of confidential data. Hence separate care need to be taken for introducing security in the cloud computing this is mainly due to the usage of confidential data, once data is stored in the cloud more risk is involved in sending confidential data thereby affecting the usage of cloud services (Sasidevi et al. 2017), its application and secure wireless data transfer. Implementation of low level security in cloud computing greatly affects the usage of user using the cloud services (Anusha B et al. 2015). Privacy refers to the Computers in Human Behavior 93–98 (Schepman et al 2012) and the 94 I. (Zhu et al 2006) extend to which cloud services is safe and secure the confidential information, whenever user use the services thereby leads to high risk of monitoring and recording the user personal information (Dharshini et al. 2016). From the above analysis it is clear that two important aspects has to be consider in cloud computing namely security and privacy issues for data storage and service provider.

Data confidentiality: It depicts the safe guarding of confidential data from the unauthorized user accessing the data since all the service is provided in the cloud computing informs of services. Medical information stored need to be transmitted upon the delicate nodes and the personal information (patient details) about the individual patients need to be kept secret (Govindharaj et al.

2016), thereby protecting from the unauthorized access else leading dangerous to the patient life (El-Gazzar et al 2017). Proper encryption techniques need to be used for ensuring confidentiality of the data (Fujisaki et al 2004). **Data integrity:** It states the procedure taken to protect the original content of a message and its consistency and accuracy. It relates to single messages and streams of messages (Giniat et al 2011, Wu et al 2012). Moreover data confidentiality does not provide any protection mechanism from safeguarding against external modification of the data, there is a frequent chance of modifying the data when transmitted in the networks and generally modification is like appending a packet of information with the original data, hence proper authentication protocol need to be implemented (Khalil et al 2014).

Data Freshness: This technique can efficiently make sure that the confidentiality and integrity of data is protected from copying and replaying of existing data, thereby ensuring original data is not altered or recycled and its frame remains correct. Two types of data freshness techniques is currently used weak freshness restrict to the limited to data frame ordering and strong freshness guarantees frame ordering and delay (Fujisaki et al 2004, Zhang et al 2014). **Availability of the networks:** It ensures that 24 x 7 network is available for the entire cloud user, since it transmit highly confidential, sensitive and potentially risk data. It states the principal that network is available at any time for all cloud users, even in case of emergency (Rocha et al 2011).

Data Authentication: Application residing in the cloud requires data authentication and a mechanism is needed to verify that all the information is sent only from the known data center and not from anywhere else. Network and active nodes involved in transmission of data need to calculate Message Authentication Code (MAC) by sharing secret key. Exact calculation of MAC code guarantee that message is sent only from trustworthy node (Zhang et al 2014). **Secure Management:** Proper encryption and decryption techniques is needed for delivering the secret key over the network and it is taken care by the coordinator node, its role is to add and remove the node in a secure manner during grouping and removing of various nodes (Zhang et al 2014). **Dependability:** All the system in the cloud need to be reliable and dependable, any failure in retrieving the original data stated as a major concern (Bertino et al 2008) and hence proper error-correcting code techniques need to be used (Tran et al 2011).

Privacy rules: It is essential to protect private data from unauthorized usage which is a global concern; the significant of privacy measure is to set up the rules/policies, regarding who is having access to the sensitive data thereby protecting user privacy (Khan et al 2012). Right now various sets of polices/ regulation exist all over the world.

MATERIAL AND METHODS

Proposed an architecture for cloud security and privacy issues thereby enabling the services for faster access of data, providing confidentiality, integrity, authentication and authorization. *Data sharing management*: It states the process of creating the data into the shareable resource thereby achieving availability of data to all the cloud users, with generic policies for sharing of data leads to transparency and openness of data. Data management helps to identify the sensitive data and eliminating duplicate data based on rules/policies. *Activity monitoring*: It monitor the user activity and copying of user action, it captures each and every complete details of the user activity usage data, applications, commands, visited URL, check box clicked and text typed.

Access management: It states only authorized user is accessing the right resource at correct time for right purpose, that is act of controlling the user and it is achieved by authenticating the user, thereby authorizing the user to perform/access certain specific task. *Security risk management*: security generally states the degree of resistance against vulnerable and valuable asset, in cloud it refers to protection of confidential data and information infrastructure. Risk is analyzed on the basis of potential threat to the data, impact and vulnerabilities. *Privacy*: privacy measure is to a set of rules/policies, regarding who is having access to the sensitive data thereby protecting user privacy. Various sets of polices/ regulation is framed for the cloud users. *User management*: It ensures that the entire user is provided with required facilities by user self- provisioning, advanced provisioning and dynamic provisioning. Using the above framework in cloud software as a service, web service is computing using three QoS factor and for optimal selection of web service particle swarm optimization algorithm is used.

PSO Algorithm: PSO is initialized by random particles and searches the optimal solution by updating their position and velocity. Two best values are updated at each iterations of (fitness) pbest that is tracked by the PSO with the best value, obtained so far by any particle in their population (gbest). When a particle takes part of the population as its topological neighbourhood, the best value is local best and also called as lbest. After finding the pbest and gbest, the particle updates its velocity and position using corresponding methods. Therefore over the search environment, the particles move through the N-dimensional problem space in the direction of better search environment. Velocity and position is calculated and the performance of each bird is measured in terms of fitness function, each and every bird moves in a multi-dimensional space with adaptable velocity, the particles remember the best position in the past and it is consider as the best position attained ever by the parti-

cles, thereby supporting the birds (particles) to search in the multidimensional space faster. The Boundary Value Analysis techniques uses three methods Region_bc(Rg), Replicate_bc (Rp) and Evade_bc(Ev)

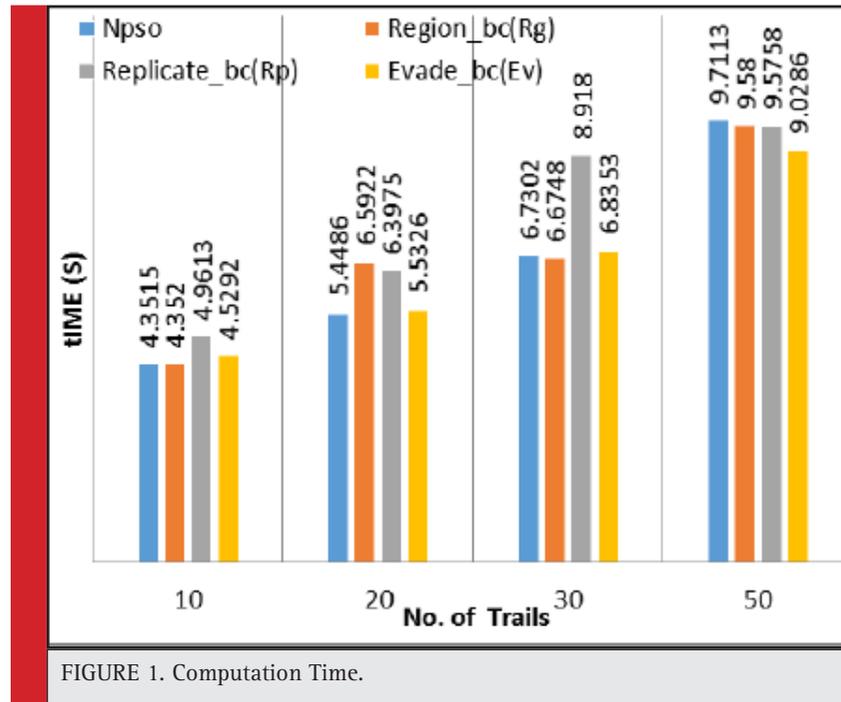
Basic PSO structure is as follows:

Table 1. Pso Parameters		
Sl. No	Parameters	Description
1.	C	Number of particles
2.	Xi	Initial displacement
3.	Vi	Initial velocity
4.	Xk	Current searching point
5.	xk+1	Updated searching point
6.	Vk	Current velocity
7.	vk+1	Updated velocity
8.	C1,c2	Acceleration coefficients
9.	W	Inertia weight
10.	K	Iteration value
11.	r1,r2	Random values
12.	Pi	Personal best
13.	Li	Local best

- Step 1: create the position and the velocity of the particles in initial swarm.
- Step 2: Update the position and velocity of individual using the relations.
- Step 3: The position and the velocity of each individual. Calculate the corresponding fitness values using fitness function of the optimization problem and call BVA techniques.
- Step 4: Update the values of pi (pbest) and li(lbest).
- Step 5: If the discontinue criteria (for example a given maximum number of iterations) is not met, go to step2; otherwise terminate the algorithm.

```

Algorithm: PSO
Input: number of particles, k(iteration), c1, c2
Output: velocity, position
Step1: Initialization of parameters
        Number of particles, k , acceleration co-efficient(c1,c2)
Step2: calculation
        For each particle
        {
                Calculate the fitness value
                Call BVA techniques
                Pbest=pbest+current value
        }
        End for
        Select the best fitness
        Update velocity
        Update position
Step3: Output
        Velocity and position
    
```



The experiment is performed using the cloud lab setup with Xen cloud platform tool the necessary VM images and software is installed, the proposed framework is implemented and using software as a service layer the experimented is performed. Using the QoS constraints (Availability, Response Time and Throughput) for the web service data set (The dataset is a set of normalized data containing about 1000 web services and its various QoS values such as Availability, Response Time and Throughput) using the proposed algorithm the value is retrieved, the various web service retrieval time is evaluated using the BVA techniques and is compared with the NPSO (Normal PSO). It is clear that proposed BVA techniques out performed well. The figure 1 shows the computation time for various using the proposed framework.

RESULTS AND DISCUSSION

Data stored in cloud has more risk when sending confidential data this in turn affect the usage of cloud services and its cloud application. Cloud storage leads to many security problems such as data integrity checking, confidentiality and privacy issues. The proposed architecture for cloud security and privacy issues enable the services for faster access of data providing confidentiality, integrity, authentication and authorization. The framework includes data sharing management, activity monitoring, access management, security risk management, privacy and User management. Web service is computed based on QoS factor, the optimal selection of web service is

performed using multi objective particle swarm optimization. For various web services its retrieval time is evaluated and is compared with the NPSO (Normal PSO). It is clear that proposed techniques out performed well.

REFERENCES

- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 1(1): 1-7.
- Bertino E, Shang R, Wagstaff Jr.S.S, 2008: An efficient time-bound hierarchical key management scheme for secure broadcasting, IEEE Trans. Dependable Secur. Comput. Vol 5, Pp 65-70.
- Buyya R, Broberg J, Goscinski A.M, 2010: Cloud computing: Principles and paradigms, Vol. 87, John Wiley & Sons.
- Chang V, 2015: Towards a Big Data system disaster recovery in a Private Cloud. Ad Hoc Networks 000, Pp1-18.
- Dean J, Ghemawat S, 2004: MapReduce: Simplified data processing on large clusters, 6th Symposium on Operating System Design and Implementation (OSDI 2004), San Francisco, California, USA, December vol 6-8, pp. 137-150.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). Secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(1): 1-6.
- El-Gazzar R, Hustad E, Olsen D.H., 2017: Understanding cloud computing adoption issues: a Delphi study approach, Journal System Software vol 118, Pp64-84.

- Fujisaki E, Okamoto T, Pointcheval D, Stern J, 2004: RSA-OAEP is secure under the RSA assumption, *J.Cryptol.* vol 17 Pp 81–104.
- Giniat E.J, 2011: Cloud computing: innovating the business of healthcare, *Healthcare Finance. Manag: J.Healthc. Financ. Manag. Assoc.* vol 65, Pp130–131.
- Goodendorf, L, 2013: *Managing Big Data. Information Security, Fourth Qua*, Pp29–33.
- Khalil I, A.Khreishah, M.Azeem, 2014: Consolidated identity management system for secure mobile cloud computing, *Computation Network.vol65* Pp 99–110.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Khan A.R., 2012: Access control in cloud computing environment, *ARNJ.Eng.Appl. Sci.* 7 Pp 613–615.
- Mell P, T. Grance, 2011: The NIST definition of cloud computing.
- Philip Chen, C. L. & Zhang, C.-Y. 2014: Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Inf. Sci.(Ny)*. Vol 275, Pp 314–347.
- Phillips-Wren, G, Iyer, L. S., Kulkarni, U. & Ariyachandra, 2015: T.Business analytics in the context of big data: A roadmap for research". *Commun. Assoc. Inf. Syst.* Vol 37, Pp 448–472.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Rocha F, S.Abreu, M.Correia, 2011: The final frontier: confidentiality and privacy in the cloud, *Computer* vol 44(9) Pp44–50.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 7-21.
- Schepman, A., Rodway, P., Beattie, C., & Lambert, 2012: An observational study of under graduate students' adoption of (mobile) note-taking software. *Computers in Human Behaviour*, vol 28(2), Pp 308–317.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Tran D.H., Nguyen H.-L, Zha, 2011: Towards security in sharing data on cloud-based social networks, in: *Proceedings of the 8th International Conference on Information, Communications and Signal Processing (ICICS)*, pp 1–5.
- Wu R, 2012: *Secure Sharing of Electronic Medical Records in Cloud Computing*, Arizona State University.
- Zhang Q, L. Cheng, R. Boutaba, 2010: Cloud computing: state-of-the-art and research challenges, *J. Internet Serv. Appl.* Vol 1 (1) Pp 7–18.
- Zhu K, Kraemer K. L, & Xu, S, 2006: The process of innovation assimilation by firms in different countries: A technology diffusion perspective on e-business. *Management Science*, vol 52(10), Pp 1557–1576.
- Zhang R, L.Liu, R.Xue, 2014: Role-based and time-bound access and management of EHR data, *Secur. Commun. Netw.vol 7*, Pp994–1015.

Intelligent Decision Making System encompassing Security Framework for VM scheduling

Raju Govindaraj Babukarthik¹, John Monica², Gnanasekaran Sambasivam³ and J. Amudhavel⁴

¹Part time Research Scholar category-B, Bharathiar University, Coimbatore, Tamil Nadu, India

²Department of Computer Science and Engineering, Pondicherry University, Puducherry, India

^{3,4}Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

ABSTRACT

Cloud security states a set of policies, security, access controls for protecting the data and application related to cloud infrastructure but still there exists a major issues in the security and policy aspect of cloud. In this paper we proposed a framework for cloud security and policy issues, comprising various layer in the cloud. Using the security aspect VM scheduling is performed with optimization algorithm. The proposed work perform VM scheduling and VM management using the frame work VM scheduling is performed and in turn frame work consists of Permission Manager, Access control list, Credential validation and Permission Store. Proposed scheduling algorithm is compared with the random scheduling algorithm based on computational time and the analysis shows that it outperforms better.

KEY WORDS: IAAS INFRASTRUCTURE AS A SERVICE, PAAS PLATFORM AS A SERVICE, PSO PARTICLE SWARM OPTIMIZATION, SAAS SOFTWARE AS A SERVICE, VM VIRTUAL MACHINE

INTRODUCTION

Cloud computing offer services to user to store and analysis of data in the data center provided by the third party, the security concern arise as major issues for cloud provider. Many techniques are provided for cloud computing, mobile cloud computing and fog computing. A new cipher text based encryption techniques is used in the mobile cloud computing due to the light weight nature and hence it is rapidly deployed in any environment. It offers same size of cipher texts and key using AND gate

structure, this model is proved using the selective security model, it clear that same security model is extended to other real world environment (Odelu et al 2017).

The identity based data possession technique has various flaws and it is overcome by identity based protocol using the identity based signature and provable data possession techniques, hence is clear security greatly lies in the provable data possession protocols (Shanmugam et al. 2017) , it provide a valuable support to the cloud computing environment (Liu et al 2017). Describing the semantics of password techniques is used for managing

ARTICLE INFORMATION:

*Corresponding Author: r.g.babukarthik@gmail.com

Received 27th April, 2017

Accepted after revision 29th July, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF : 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

the password to achieve better security, analysis is performed in huge data in real password stored in the various websites and it is clear that 76% password is contain numeric and adding pinyin and digits (Rajadurai *et al.* 2017). Further suggested useful information for cloud user regarding the understanding of semantic pattern in the password and the policies used in storing the passwords (Han *et al.* 2017).

Tree based model for cluster adapted access control for mobile cloud control for providing better access control (Ren *et al.* 2017). Quantum computation protocol for providing better trusted center for active participants, getting permission with the quantum channel that is quantum state and is verified based on the outcomes, analysis and proof (Tan *et al.* 2016). VM Cloning technique is used for the high performance interconnected and storage:Virtual machine fork constantly replica (duplicate) vm into n number of vm running in the different host machine, all replica as identical state matching with spontaneous meaning of work creation. Virtual machine fork enable creation and deployment of several tasks which requires fast instantiation of stateful worker in cloud environment (eg: load sharing, deterministic job placement, or parallel computing system. The stateful clone service the user of cloud computing in adhoc manner for handling state of the application and cycle provisioning (Andrés Lagar-Cavilla *et al.*).

Hybrid virtualization Technology which could be able to migrate between heterogeneous various virtual machines: Virtualization technology grown faster with the aid of hardware supporting virtualization technologies, it is the most important underlying technology of the various IT applications in terms of cloud computing and green IT (Sasidevi *et al.* 2017). Many processor developing enterprises initiated several hardware supports in favor of virtualization technologies in order to prevail over software virtualization technologies, as of now, full virtualization techniques is supported by hardware. Resourceful sever virtual machine is merged with the para virtualization with hardware support virtualization (Anusha B *et al.* 2015), In proposed virtualization technology complexity us minimized to a greater extended than in the conventional server machine, proper security mechanism need to be deployed for accessing the virtual machines (Inhyuk *et al.* 2010).

Some of the enterprise-class server usage such as network routers and network attached storage system is have more advantage from the virtualization techniques, moreover present virtual machine communication technology as major performance outlay whenever deployed in highly collaborative appliances components, presented fido a secure virtual machine communication initiated by relaxed trusted model between various software components for achieving high performance (Burtsev *et al.* 2012). Efficient utilization of virtualization

and dynamic scaling of resource in order to meet the green cloud computing: Concentrated on privacy and security aspect of scaling of resources with main objective of minimization of energy consumption in addition to full equipment life cycle as a major promoter for green cloud application design. With major objective is to minimize the resource in a well organized, cost effective and effective, several advance green It is virtualization, material recycling, power management and telecommuting (Dharshini *et al.* 2016).

Secure virtual platform is possible to run several storage device and interconnected physical computers. Cloud computing uses the concept of sharing resources equally with the nodes, cloud virtualization is stated as a process of running more logical computer systems on hardware based on equal resources sharing, in turn leading to reduction on efficiency of server virtualization techniques. Proposed a novel technique for securely connecting various nodes with minimum number of servers (Yamini *et al.* 2010).

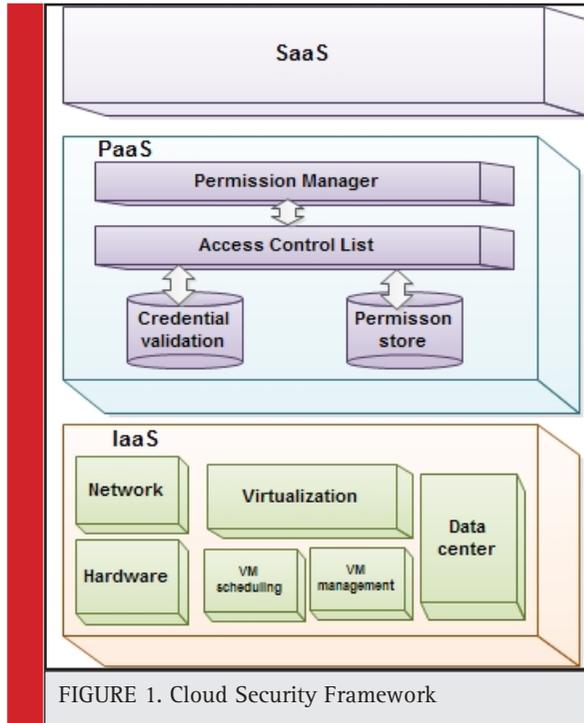
A new framework work for security based energy management in a single and multi-layered operating system structures is proposed, it provide a combined model to separate and distribute energy based on mechanism energy efficient resource allocation and accounting. The framework work based on recursive energy consumption into consideration, used in virtualization layer or following driver components. Secure protocol implementation aim hypervisor based virtual machine consisting of two major components that is guest level and host level (Govindharaj *et al.* 2016), guest level energy management trust on useful virtualization of actual physical energy offered by virtual machine monitor, host level subsystem enforces energy constraint on machine -wide thereby rendering guest operating systems and service components, energy-aware guest operating system is stated as a fine-grained application of particular energy management (Jan Stoess *et al.* 2007).

MATERIAL AND METHODS

Proposed a framework for solving the security issues in the cloud computing concentrating on the various layer of the cloud, in cloud everything is serviced in form of service thereby security and policies plays a major role in the data storage, VM scheduling and VM management.

Permission Manager: provide permission for the various service and user for authenticating, authorization to access the cloud data, it provide all the authenticating service to the user with the help of access control list.

Access control list: It maintain complete details of the access for all the cloud user, list is maintain for the all the cloud user listing the various permission and provide support to the permission manager.



Credential validation: It states the security checking of policy setting, a complete list of auditing event is generated based on credentials which is submitted by the user for login request. It occur with respect to the local computer authoritative and domain controller authoritative

Permission Store: It is used to provide permission for cloud users using permission data available in the permission manager database (example automatic startup and sending of SMS, secure from malicious app, remove the permission). The figure 1 state the cloud security frame work for VM scheduling.

Particle Swarm Optimization: Position of a particle is represented as candidate solution and the swarm (population) of the particles are known as searching agents. In PSO, each bird is known as individual (particle). All particle moves on its own way with the position and velocity searching for best value in the solution space. Here, each particle has its own solution. Each agent (particle) searches foods in its own way and finds the best solution in a search space. All the particles communicate with each other to find the best solution. Each particle has its own solution. For each generation, the position of the particle is updated; the best solution for a particle which is achieved so far is known as individual best (pbest) and the most excellent value of any agent within the swarm is known as global best (gbest). The new location of a particle is calculated based on its preceding position and velocity

Modified_PSO_VMscheduling_Algorithm

Input: No. of servers, user.

Output: Energy, memory.

Step 1: Initialization

Server, topology, user, data center components.

Step 2: Parameters Setting

Non-linear power model, core switch, aggregation switch, task.

Step 3: Execution

Scheduler

PSO, Round Robin,

Energy model

eDVFS, DNS.

Step 4: Report

Individual server load

Individual VM load

Step 5: Output

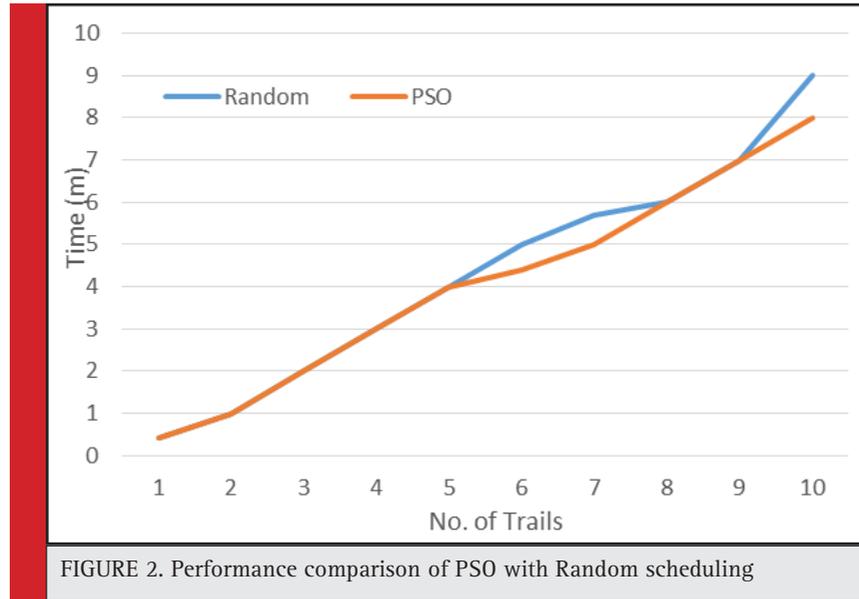
Total energy consumed

Energy of servers

The experiment is evaluated using the Green Cloud it is a packet level simulation tool using the NS2 libraries files for energy-aware scheduling of data centers in the cloud computing. It compromising of various parameter of cloud such as switches, servers, links used for communication and consumption of energy. It used for resource allocation and monitoring, workload scheduling. The computation time for modified PSO and random scheduling algorithm is evaluated. 10 trail is performed and the computation time is evaluated by increasing the number of server for a fixed number of user, from figure 2 it is evident that modified PSO outperform well compare to the random scheduling algorithm. The modified PSO is compared with random technique and it is clear the modified PSO computation time is very minimum compared to random method.

RESULTS AND DISCUSSION

Cloud computing offer various services for users, still the security and policies in the cloud need to be considered. Many approaches are proposed for secure data transfer and VM scheduling based on various policies, we proposed secure VM scheduling for optimized scheduling using modified Particle Swarm Optimization (PSO). The various parameter of cloud such as switches, servers, links used for communication and consumption of energy. It used for resource allocation and monitoring, workload scheduling. The computation time for modified PSO and random



scheduling algorithm is evaluated. It is evident that computation time of PSO is considerably low compared to the random scheduling algorithm, in further we plan to apply for the other combinatorial optimization problems.

REFERENCES

- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Andrés Lagar-Cavilla H, Joseph A. Whitney, Adin Scannell, Philip Patchin, Stephen M., 2009: SnowFlock: Rapid Virtual Machine Cloning for Cloud Computing, *ACM APRIL*.
- Anton Burtsev†, Kiran Srinivasan. Prashanth Radha krishnan, Lakshmi N, 2012: Fido: Fast Inter-Virtual-Machine Communication for Enterprise Appliances.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). Secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Han G, Y. Yu, X. Li, K. Chen and H. Li: Characterizing the semantics of pass-words: The role of Pinyin for Chinese Netizens, Elsevier, *Computer Standard & Interfaces*, vol 54, Pp 20-28.
- Inhyuk Kim, Taehyoung Kim, Young Ik Eom 2010: NHVM: Design and Implementation of Linux Server Virtual Machine using Hybrid Virtualization Technology, *IEEE International Conference of Computational Science and Its Applications*.
- Jan Stoess. Christian Lang. Frank 2007: Energy Management for Hypervisor-Based Virtual Machines, *USENIX Annual USENIX Association Technical Conference*.
- Liu. H, Y. Mu, J. Zhao, C. Xu, H. Wang, L. Chen and Y. Yu, 2017: Identity-Based Provable Data Possession Revisited: Security Analysis and Generic Construction, Elsevier, *Computer Standard & Interfaces*, vol 54, Pp 10-19.
- Odelu, V, A. Kumar Das, Y.S. Rao, S. Kumari, M. Khurram Khan and K. Choo, 2017: Pairing-based CP-ABE with constant-size ciphertexts and secret keys for cloud environment, Elsevier, *Computer Standard & Interfaces*, vol 54, Pp 3-9.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Ren W, Liu R, Lei M, Choo K, 2017: SeGoAC: A tree-based model for self-defined, proxy-enabled and group-oriented access control in mobile cloud computing, Elsevier, *Computer Standard & Interfaces*, vol 54, Pp 29-35.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Tan X, X. Zhang and T. Song, 2016: Verifiable delegated quantum computation with χ -type entangled states, Elsevier, *Computer Standard & Interfaces*, vol 54, Pp 40-49.
- Yamini B, D.Vetri Selvi, (2010) "Cloud Virtualization: A Potential Way to Reduce Global Warming", *IEEE International Conference of Computational Science and Its Applications*.

Technical Analysis and Appraisal of Statistical Tools in Fluid Flow Turbulence with Applications

J. Amudhavel^{1*}, L. S. S. Reddy², M. Rajeswari³ and R. G. Babu Karthik¹

^{1,2}Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

³Department of Computer Science, Pondicherry University, Puducherry, India

ABSTRACT

To carry out research, a scientist has to employ different techniques of research so as to stimulate different physical systems using the basic laws of physics. The most important part of research is the use of statistical tools based on different laws to examine data. Fluid flow can be approximated by linear systems of equations. However, some flows react to excitation in ways that can be analyzed by various statistical tools. Fluid flow turbulence is a field under fluid mechanics which is of great importance and has led to the extensive research and the use of various statistical tools and their applications in turbulent flow.

KEY WORDS: ANALYSIS METRICS, ANALYSIS TOOLS, FLUID APPLICATIONS, FLUID MECHANICS, TURBULENCE FLOW

INTRODUCTION

In fluid mechanics, fluid flows are approximated by linear system equations. However, some systems may react to excitations, and the excitation can be analyzed especially if it is weak for example, emission of acoustic waves from a turbulent jet. Spectra and correlations depend on the second moments of joint probability density whereby to relate the probability distributions and the correlations, elements of probability theory are used. These elements include central limit theorem and the normal distribution. The Fourier transform of the auto covariance describes all the frequency contents of

a process (Davis 1977). Fourier transform for the cross-covariance for two random functions at different times describes the cross-spectral density. However, when the measures are done independent of time origin, then the process taken into consideration are stationary. Cross-covariance and cross spectra play a very important role in deriving propagation speed, the decay times and many other properties of the random process (Faragher, J., 2006; Anusha B et al. 2015). The relationship between an input and output statistical measures in a linear system is a useful application of cross spectra and cross-covariance. For instance, if you know the statistical properties of the input, then the properties of the output

ARTICLE INFORMATION:

*Corresponding Author: info.amudhavel@gmail.com

Received 27th April, 2017

Accepted after revision 29th July, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

will be found so easily. There are various statistical tools that are used to analyze processes and behavior of fluids in fluid mechanics (Jovanović, et al., 1993).

ENSEMBLE AVERAGE

The concept of ensemble average is very important in fluid mechanics and is based on the existence of independent events. It allows one to form an average for different processes which are time dependent. It uses the auto covariance function to perform statistical measures whereby the function provides information about the average dependence of a process (Dharshini et al. 2016). Ensemble average helps to draw dynamics of a system through observation of excitation and response.

This tool of averaging involves using a random function that cannot be predicted from its past. A turbulent jet is a perfect example of a random function involving functions of space and the velocity field. There is a constant supply of air to the jet, but the velocity varies with time in an unpredictable manner. Therefore, the macroscopic boundary conditions may be independent of time. Time and velocity are the useful parameters that enable one to describe flows. Time averages are not used where the macroscopic boundary conditions change with time; this is where the ensemble averages come in. Noise and oscillation of an ensembled macroscopic process differ depending on the physics of the underlying process.

REYNOLDS AVERAGING

The method of Reynolds averaging is practically applied in all areas of fluid dynamics. The geophysical fluid dynamics is a field where this method is very diverse and very imperative. There is a variety of ways in which you can define the averaging operator and each way has some advantages and disadvantages. We do not average the turbulence data over infinity of realizations when dealing with the practical applications (Ueda and Hinze, 1975). For example, in laboratory environment, turbulence in a wind tunnel is studied, it is easy to get information and data which are important when it comes to computing various calculations like ensemble averages over a finite number of realizations that are almost similar with except few less important information and details (Govindharaj et al. 2016). It is difficult to obtain multiple realizations of the similar situation in studies of the uncontrolled atmosphere. However, the best that can be done is simulating multiple realizations of the similar situation when using the high-resolution models that have slightly different initial conditions.

When the Reynold's averaged equations are applied, the solution will depend on the grid size used in the numerical model of the atmosphere. The model used

plays a critical role in the forecasting and simulating the distributions of the present prognostic and diagnostic variability which appear on the grid (Van Thanh, 1982). The solution from this method largely depends on the grid size. For instance, when Reynolds-averaged equations are applied in a numerical model of the atmosphere, then the results found are used in forecasting or simulating the distributions of different diagnostic and prognostic variables that occur over the grid. For example, in the free atmosphere, we do have the sub grid-scale fluxes that usually represent the vertical exchanges that are linked with cumulus convection (Govindharaj et al. 2016). A grid model, which is known for resolving cumulus convection, then the cumulus parameterization, can be used to obtain the fluxes.

In a model having the rough and coarse grid, the averaging operator should be defined regarding spatial averages over the grid cells. Always, this will involve the use of the method of finite-volume so as to accommodate all sizes of grids. For instance, if the horizontal grid size that the model has is on the order of 100 km or larger, then the grid cell will contain have a wide number of the cumulus clouds. The grid-cell averages can be regarded as representing the collective effects of the many clouds that are co-existing in the grid cell at a particular time. If the grid size of the model is minimized to a quite small value, for example, 1 km, the spatial averages over individual grid cells will always be a reflect of the presence of the larger individual cumuli, which would be marginally represented on such a grid (Gayathri et al. 2016).

The averaging distance is not a must that it has to be the same as the grid size. We could choose to associate the averaging distance with a physical length scale, which should be independent of the somewhat arbitrarily chosen grid size (Lyons et al., 1991). When averaging distance is larger than the grid size then we can reduce the grid size without necessary changing the averaging length. You need to ensure that the Solutions are smooth when the grid scale becomes much finer than the averaging length. Having a smooth solution will result to a waste in the fine resolution (Lyons et al., 1991). Finally, suppose we have the averaging operator that is defined as an ensemble average. The ensemble average will remain spatially and temporally smooth when the grid size is being reduced. This is attributed to the individual cumuli occur in different places and at different times in the different realizations that make up the ensemble. Individual cumulus clouds are usually produced by the individual members of the ensemble in different places at any given time, but the large-scale spatial average of a variable. For example, the precipitation rate is presumably be approximated to be of the same to all members in the ensemble.

CENTRAL LIMIT THEOREM

This theorem states under certain the normal distribution will always to be approached by the probability distributions. There are different versions of this theorem that depend on the restrictions and conditions imposed. In this theorem, one assumes to have two stochastic variables x and y and is significant to draw conclusions about statistical properties of the variables. One variable, for instance, x must be specified to have probability densities with zero means and have equal deviations so that the other variable, in this case, y will not be infinite. This postulates that the x variables should have identical distributions with means of zero. The additive increments usually determine the most of the values of the stochastic variables (Shanmugam *et al.* 2017). For example, the momentum of a molecule in a dilute gas. The momentum component denoted by x describes the vector sum based on all momentum increments which are brought about by collisions with other molecules in the past. Spherical coordinates can be introduced commonly known as the Maxwell-Boltzmann distribution of molecular velocities that occurs in a dilute gas. Because the gas is at rest, the distributions will have a zero mean (Karlsson and Johansson 1986). On the other hand, there is no normal distribution of the velocities of the turbulent fluid and this is caused by the momentum increments in a given fluid particle at different time are not independent. The eddies and convergence regions tend to interact and be coherent eliminating the independence of interactions with the neighboring fluid particles that is experienced by a fluid particle (Rajeswari *et al.* 2017).

The local motions at some scales may be nearly independent resulting in the turbulent fields not being Gaussian despite having a Gaussian distribution. This makes a crucial difference that gives a significant property and characteristic of the dynamics of turbulence (Kendall *et al.*, 1948). Central limit theorem plays a crucial role in analyzing the dynamics of turbulence (Rajadurai *et al.* 2017). Although the non-Gaussian properties have to be included, the effects of turbulence are a major concern and with the aid of experimental verification, this Gaussian distribution may be regarded as adequate approximation.

LOGNORMAL DISTRIBUTION

To start with, the normal distribution is the sum or the difference of different parameters. On the other hand, the lognormal distribution is computed by finding the product or the quotient the different parameters in place. If a bunch of variables is being multiplied together, then the resultant distribution approaches lognormal as the number of variables increases. The central limit theorem is a tool used to describe a sum of random vari-

ables which tends to a Gaussian distribution as their sum number increases towards infinity. For a group of positive random variables, it is evident that their sum tends faster to a log-normal distribution although the sum will eventually tend to a Gaussian distribution. However, the log-normal distribution is close to the distribution rather than to any Gaussian distribution more if the variables are numerous enough. Although the log-normal distribution regarded to be due to the product given by the random variables or non-linearity of the system, it is quite observable for an additive process which is typical of all linear systems (Jovanović, *et al.*, 1993).

The method of lognormal distribution is a common technique when it comes to model the lives of units in which their failure modes are of fatigue and stress nature. A high number of the mechanical systems use the lognormal distribution and therefore, have wide use of the application in the field of fluid mechanics. The Weibull distribution is a good partner for the lognormal, especially when you are trying to model these types of units. The lognormal distribution and normal distribution have certain similarities. A random variable that has log normally distributed is that whose logarithm of the random variable is of normally distributed. The mathematical reasoning used for the construction of the probability by plotting of scales is an ultimate example of the mathematical similarities between the two distributions. The bias of parameter estimators used is very similar for these two distributions (Sasidevi *et al.* 2017).

SPECTRAL ANALYSIS

In the early 1990s Multitaper Spectral Analysis was introduced. It was performed by applying a single taper to the data, and then Fourier transformation is done yielding periodogram estimates which can then be smoothed using a window. The tapering was done to reduce leakage which could eventually lead to a bias of the estimated spectrum. Using a window to smoothen helped to reduce the variance of the estimated spectrum, however, the cost of reduced frequency resolution will be affected. Multitaper spectral analysis can perform tapering using more than one taper, and then the results from the separate periodogram estimates of each taper are averaged instead of windowing.

On doing this, the results obtained for a given bandwidth indicates that the multitaper method has a greater resolution and furthermore, produces a better combination of low bias and low variance. Utilizing the multitaper approach is necessary especially when the spectral and cross-spectral analysis is appropriate. This tool is necessary when it comes to giving a description of the global behavior and characteristics of complex nonlinear flows in fluid mechanics by decomposing the flow

into various modes which are determined by spectral analysis of a technique called Koopman operator. This operator is an infinite-dimensional linear operator associated with any full nonlinear system. It has different modes which are referred to as Koopman modes that are linked with particular characteristics that can be observed. These modes may be determined directly when you have the numerical or experimental data by using a variant of a standard Arnoldi method.

Each of the modes has a temporal frequency and rate of growth. This normally seen as a nonlinear generalization of global eigenmodes. These modes are always an alternative to proper orthogonal decomposition whereby when you are having a look at periodic data, the discrete temporal Fourier transform is reduced. A dynamic mode is a form of an algorithm used to find the Koopman modes. The jet in crossflow is an ultimate example that illustrates the Arnoldi method. It shows that this method catches the dominant frequencies and at the same time it elucidates the associated spatial structures.

EIGENVECTOR ANALYSIS

Eigenvector analysis method is a useful technique that is used in fluid mechanics. The gradient of a velocity vector field is regarded as an asymmetric tensor field which is being used to give a critical insight for the vector field that would otherwise be difficult to infer especially when using the technique of the traditional trajectory-based vector field visualization. The structures in the eigenvalue and eigenvector fields are described as of the gradient tensor. The function of these structures is well described which include describing behaviors portrayed by the velocity field that can represent either a 2D or 3D compressible flow in a two-dimensional manifold.

Eigenvalue manifold illustrates the structure in the eigenvalue field and furthermore enables clear novel visualizations, which then depicts the relative strengths in the physical components in the vector field. These components include isotropic scaling, anisotropic stretching, and rotation. An eigenvector analysis was developed basing on the concept of the eigenvector manifold and gives additional insight into 2D asymmetric tensors fields.

The results of this tool give rise to the realization of the simple and intuitive geometric of the eigenvectors. This is important when it comes to measuring of the signed distance of a tensor from being symmetric. Eigenvectors are a continuous extension of the complex domains called the pseudo-eigenvectors. The evenly spaced tensor lines are utilized following the use of pseudo-eigenvectors so as to illustrate the local linearization of tensors everywhere inside complex domains simultaneously. Both eigenvalue and eigenvector mani-

fold are supported by a tensor re-parameterization that has physical significance. Re-parameterization allows tensor analysis to be done on physical quantities such as vortices, expansion, contraction and deformation which describes the physical interpretation of the tensor-driven vector field analysis in fluid mechanics.

CONCLUSION

In order to carry out research, there are various tools that are necessary for it to be effective and efficient. A scientist must call upon various techniques to stimulate physical systems while using the basic laws fluid mechanics. The use of these statistical tools based on the different laws enables an engineer to examine data, make observations and draw conclusions before implementing it. The tools play an important role in designing of equipment that work based on the principle of fluid mechanics. Thus when carrying out research on the physical tools then having the statistical tools is one thing that you can't do away with. We do have a variety of the statistical tools that you can consider when looking for a perfect statistical tool to use. All you need to ensure is that you are using the right statistical tool.

REFERENCES

- Ahilandeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 1-5.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Davis, R.E., 1977. Techniques for statistical analysis and prediction of geophysical fluid systems. *Geophysical & Astrophysical Fluid Dynamics*, 8(1), pp.245-277.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Faragher, J., 2006. The implementation of probabilistic methods for uncertainty analysis in computational fluid dynamics simulations of fluid flow and heat transfer in a gas turbine engine (No. DSTO-TR-1830). DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION VICTORIA (AUSTRALIA) AIR VEHICLES DIV.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 22-29.

- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Jovanović, J., Durst, F. and Johansson, T.G., 1993. Statistical analysis of the dynamic equations for higher-order moments in turbulent wall bounded flows. *Physics of Fluids A: Fluid Dynamics*, 5(11), pp.2886-2900.
- Jovanović, J., Durst, F. and Johansson, T.G., 1993. Statistical analysis of the dynamic equations for higher-order moments in turbulent wall bounded flows. *Physics of Fluids A: Fluid Dynamics*, 5(11), pp.2886-2900.
- Karlsson, R.I. and Johansson, T.G., 1986. LDV measurements of higher order moments of velocity fluctuations in a turbulent boundary layer. In *3rd International Symposium on Applications of Laser Anemometry to Fluid Mechanics* (pp. 12-1).
- Kendall, M.G., Stuart, A. and Ord, J.K., 1948. *The advanced theory of statistics* (Vol. 1, pp. 42-46). London: C. Griffin.
- Lyons, S.L., Hanratty, T.J. and McLaughlin, J.B., 1991. Large-scale computer simulation of fully developed turbulent channel flow with heat transfer. *International journal for numerical methods in fluids*, 13(8), pp.999-1028.
- Lyons, S.L., Hanratty, T.J. and McLaughlin, J.B., 1991. Large-scale computer simulation of fully developed turbulent channel flow with heat transfer. *International journal for numerical methods in fluids*, 13(8), pp.999-1028.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Rajeswari, M., Amudhavel, J., Pothula, S. and Dhavachelvan, P., 2017. Directed Bee Colony Optimization Algorithm to Solve the Nurse Rostering Problem. *Computational intelligence and neuroscience*, 2017.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Ueda, H. and Hinze, J.O., 1975. Fine-structure turbulence in the wall region of a turbulent boundary layer. *Journal of Fluid Mechanics*, 67(1), pp.125-143.
- Van Thinh, N., 1982. A study of the intermittent phases in the wall region of a turbulent flow. *Structure of turbulence in heat and mass transfer*, pp.103-116.

Deep insight in to the methods and applications on Aerodynamics of Flight

J. Amudhavel^{1*}, V. Srikanth¹, T. Kalai priyan² and A. Bhuvaneshwari²

¹Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

²Department of Computer Science, Pondicherry University, Puducherry, India

ABSTRACT

Understanding the behavior of airflow around objects plays an important role in the calculation of forces and the moments that act on it. To design or construct bridges, buildings, aircraft, vehicles as well as wind turbines, aerodynamics is necessary so as to minimize fuel consumption and emissions. This topic provides vital information concerning atmospheric and weather phenomena. Applications of Aerodynamics are wide-ranging environmental effects such as pedestrian level wind to the transport sector in airplanes and vehicles are widely discussed in this paper.

KEY WORDS: AERODYNAMICS, AERODYNAMICS OF FLIGHT, AIRFOIL, AERODYNAMICS OF FLYING BIRDS AND INSECTS AND THE GROUND AERODYNAMICS

INTRODUCTION

Aerodynamics is a sub-discipline of fluid mechanics is a study that deals with forces and the motion of objects that goes through the air. It explains the way objects can move through air ranging from a rocket blasting off, an airplane, a flying kite, and even cars because they surrounded by air. To predict the complex trajectories of these objects that move across air, the knowledge about the instantaneous fluid forces is necessary. Getting the forces that act on any moving body might be considered as classical problem, but analytical based results already exist though with special limits which are the inviscid rotational flow

(Anderson Jr 2010) and Stoke's flow around the rotating airfoil. The fluid forces on a moving object through air rely on the kinematic variables like the angle of trajectory and the translational velocity (Sasidevi et al. 2017). However, it is evident that most objects that move in a fluid encounter unsteady flows that might be difficult to know which quasi-steady model on inviscid theory will apply. Bernoulli's principle plays an important role in the construction and operation of objects that move across air with the sum of static and dynamic pressures being constant (Rajadurai et al. 2017).

For instance, for a fluid that exerts static pressure, when it starts to move, this pressure will decrease as the

ARTICLE INFORMATION:

*Corresponding Author: info.amudhavel@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

dynamic pressure increases (Sedov, 1980). Therefore, a greater speed will lead to a greater dynamic pressure which in turn results to a lesser static pressure. The engineers while constructing an airplane majorly used the concept of slower and faster currents to keep the plane a loft. The presence of viscosity near the surface of the object creates havoc with these proportions of Bernoulli principle. Viscosity causes the air to stick on the surface of the object thereby when the air flow is slowed near the surface it will form a boundary layer of slow motion. For instance, at the flow begins, a laminar flow will be experienced at the leading edge of the airplane wings, and it increases with the width of the layer up to a point where the flow becomes turbulent (Shanmugam *et al.* 2017).

These scenarios affect the coefficients of drag and lift which play a fundamental role in the operation of an airplane.

There are four types of forces that act on a flying object. These are lift, thrust, weight and drag and they make an object to move faster or slower, up or down depending on the amount of each force. To start with the weight which is the gravitational force acting on the moving object depending on its mass? A flying object must overcome this downward force before it flies. The lift force is opposite of force in that it pushes the object up and for an object to fly upwards, it must have a higher lift as compared to its weight. For instance, a hot air balloon gets its lift when the hot air in the balloon will rise and carry the balloon up with it (Gayathri *et al.* 2016).

For a helicopter, the motion of air through the rotor blades moves the helicopter upwards. On the other hand, drag is a force that pulls back the object from moving providing resistance to motion (Ho *et al.*, 2003). The shape of any object determines the drag it is going to experience, for instance, narrow surfaces will have less drag than wide ones. Lastly, thrust force is the opposite of drag and pushes the moving object forward. An object moving forward must have a greater thrust than drag, and it will fly until the drag causes it to slow down while landing.

AERODYNAMICS OF FLIGHT

Flight operations and performance of planes depend on the theory and applications of aerodynamics. Bernoulli's principle which describes the relationship between the fluid velocity and its internal pressure helps to explain why an airfoil develops aerodynamic force (Ahilandeswari *et al.* 2016). Fluid flow pressure has both dynamic and static pressures whereby the static pressure decreases as velocity increases which cause the dynamic pressure to increase.

In an airplane, an additional lift is created by the lower surface of the rotor blades such that the air striking on the undersides is deflected downwards (Nitsche and Weiser 1987). Under flying conditions, deflection of air and impact of pressure from the lower surface of the rotor blade contributes a small portion of the lift. Most of the lift is as result of decreased pressure on the upper side of the plane. This pressure difference between the sides causes the plane to move upwards.

Although the weight of the airplane is fixed, normally thought to consist the helicopters weight, fuel, and its occupants, the weight can be influenced by aerodynamic loads (Govindharaj *et al.* 2016). Therefore it's a must that the rotor system generates enough lift required to offset the total weight of the plane together with its occupants. Airflow affects the reaction of the rotor system. There is the relative wind which is the air that flow about the plane. The movement of this airflow will create a relative wind which flows in parallel but opposite to the plane. The wind passing through the blade of the rotor can either be horizontal or vertical. The airflow across the rotor system when the plane is moving forward is equal to the velocity of the plane (Dharshini *et al.* 2016). Autorotation is another major part in aerodynamics where the plane utilizes the action of air moving up through the rotor instead of using the engine power. It is normally allowed for freewheeling that is enabling the rotor to continue turning even if the engine is not running. Induced drag is another property that affects the operation of the plane whereby the airflow circulation generates around the rotor blade which in turn creates lift. There is high pressure beneath the plane compared to the top surface, and the tips and this will cause a spiral or vortex whenever lift is being produced.

AIRFOIL

An airfoil is that surface which can produce more lift than drag when passing through the fluid (air) at a certain angle. They are the ones that more often produce a lift and are used majorly to create thrust (propulsion), for stability and also for control. Rotor blades are an example of an airfoil which has been carefully structured so as to accommodate a specific set of flight characteristics (Anusha *et al.* 2015).

There are various terminologies that are more often used while dealing with an airfoil. The blade span which is the length that the rotor blade has from the tip of the blade to the center of rotation is also determined and is an imperative factor for the creating a relation between Aerodynamics and Fluid Dynamics. The chord line is the line that joins the trailing edge of the airfoil to the leading edge. For an airfoil to work perfectly, there are two angles that enable it to produce the lift that is required.

The “angle of incidence” which is the angle that is in between the chord line and the rotor hub affects the coefficient of lift by changing the lift produced by the airfoil. The angle of attack increases the lift unless it goes beyond the critical angle.

AERODYNAMICS OF FLYING BIRDS AND INSECTS

Birds and insects are designed to flight not only because they have wings, but also, they possess a range of characteristics that enable them to fly and balance their movement through the air. These characteristics include having hollow bones, less body weight allowing them to overcome gravity so as to remain aloft (Gormont 1973). The bird’s sternum is so huge in proportion as compared to its whole-body in terms of body size. Therefore, it enables it to achieve the required thrust which is imperative to keep it flying. The light feathers also allow it to have an optimal lift with a minimum drag.

The wings of a bird are curved at the top to maximize on Bernoulli Effect. As the air moves over the edges of the wing, it divides, and the air on the top will have to go a distance which is greater because of the curve and therefore has to be of higher velocity than that on the lower side (Furukawa *et al.*, 1998, June). As shown by Bernoulli, the air above the wings which moves at a relatively higher velocity exerts less pressure than the air below it moving relatively slower. When the external factors are not present, the air above tries to move with the same time limit as the time above the wings, so the only way to achieve this is by increasing velocity which in turn lowers the pressure according to the Bernoulli’s principle. The pressure difference between the two regions keeps the wing aloft and creates a lift force. Everyone has noticed a bird beat its wings, this is not just for the sake of it, but it does so to help it in movement (Andersen *et al.*, 2005). The down strokes propel the bird and help it rise above the ground as it gains aerodynamic lift which pushes its wings upwards in the preparation of the next down stroke. The bird will experience aerodynamic drag; however, to counter this, the birds fold its wings so as to decrease the wingspan thereby experiencing less resistance. Another way that birds gain balance in the air is by moving their wings backward and forward (Rajeswari *et al.* 2017).

The birds flap their wing in a parallel direction to the ground so as to fly very slowly. The movement of birds through the air was the motivation that humans sought to invent flight objects to help themselves move from one place to another (Shyy *et al.*, 1999), (Shyy *et al.*, 2010). Witnessing the astonishing aerodynamics effects, Icarus and Daedalus constructed a set of bird-like wings to enable them to escape from prison by flying away (Ansari. *et al.*, 2006).

AIRPLANES: HOW THEY GET AND STAY ALOFT WHILE REMAINING STABLE

There are several factors that counter a successful flight. Once engineers and pilots take to air, for the plane to get aloft and stay aloft, it is subjected to drag, weight, lift and thrust, and these forces have to be well-balanced for it to operate successfully. The design of the wings wholly takes an advantage of the Bernoulli principle. The wing has a shape of a long bird with a large forward end and a narrow tip pointing towards the rear. Its wings are asymmetrical, and the bottom part of it is flat making up the airfoil with the camber (the greater curvature witnessed on the upper side in comparison to the lower side). A straight line which is imaginary is the one that connects the spot. This is a spot point where the air hits the front which is called the stagnation point. According to Bernoulli’s principle, the shape of the airflow is the major factor that results to the spread of laminar flow around the spot. The currents are moving at a lower velocity hence a greater pressure is exerted beneath the airfoil as compared to the faster currents above it. This will definitely give the aircraft a lift (Shyy *et al.*, 2007). The size of the wing is also a vital parameter that influences the degree of lift that the aircraft usually experiences. A longer wing will exert a greater total force exerted beneath it and the greater the ratio of this pressure to that of the air above. Due to the presence of flaps at the trailing edge, there is usually a variation in the size of a modern aircraft’s wing.

Flaps are usually designed with different properties that are displayed at different stages of the flight. For instance, they provide lift during takeoff, but when the plane is already at its stable flight the flaps increase drag thus the pilot has to retract them. While preparing to land on the ground, as the aircraft slows and it has to descend, the plane extends these flaps to provide stability (Pesavento 2005). It also has to assist in the decreasing the speed since the resistance will increase.

Speed also affects the extend of the lift force. The faster the speed of air as it moves over the wing, the faster the craft will move. Therefore the pilot will have to manipulate this by lowering or making the power of the engine very high. The engine power plays a critical role when it comes to controlling the speed of the plane’s propellers during a turn. Airfoil’s angle of attack is an imperative component during the lift stage of a plane. This is the orientation of the airfoil about the direction of air flow. Increasing the angle of attack will enable the aircraft to have an extra lift (Spalart and Allmaras 1994). This is because it is responsible for moving the stagnation point away from the leading edge down that occurs along the lower surface. This is important when it comes to increasing the low-pressure area of the upper

surface. This will in turn increase the pressure difference that will definitely provide an air lift of the plane. However, a further increase in the angle of attack will have an adverse effect to the boundary layer of slow-moving air. This will make the aircraft to stall.

The engine gives power to the propellers, which in turn will give the aircraft a propulsive force. The propeller blades have miniature wings pivoted at the center and are also powered by the engine so to result to a rotational motion. The blades of the aircraft wings are designed with a convex forward surface, and a trailing edge that is narrow. They have angle of attacks that are adjusted at different points so as to meet different needs during each stage of the flight. During a stable flight, the angle of attack is increased so as to enable the propeller blades to sharply face against the airflow (Menter 1992). During takeoff and landing, the pilot dramatically reduces the angle of attack by reversing the direction of the propeller blades and produces a lurching.

THE GROUND AERODYNAMICS

Basing on the above discussions of aerodynamics it's clear that the behavior of objects off the ground is due to Aerodynamics. Aerodynamics contributes to the effective functioning of wheeled transport on Earth's surface such as bicycle and automobiles. For instance, in bicycles, the cyclist accounts for more than half of the drag, and therefore, the position of the rider on airflow is highly important (Amitay et al., 2001). It is a significant factor that designers have put into consideration so that all possible barriers to airflow are eliminated. As the wheel of a bicycle rotates, the airflow that is behind the spoke will separate. This will create a turbulence which in turn leads to a resistive force to motion. To counter this problem, many bicycles today are designed in a way such that the length of the spoke is reduced (Aungier 2006). On the other hand, the use of vehicles which are machine-powered has helped a lot in achieving extraordinary speeds of up to 260km/h. This was possible with the aid of aerodynamic concept in design.

CONCLUSION

The physical objects on earth experience forces including the force of gravity that tends to retain them pressed down. The air surrounding these objects prevents lift force from raising the objects high above the ground just because this force is way lower than the gravitational force. The downward force of air is less unlike that of water because air pressure is greater in water.

REFERENCES

- Ahиландесwari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 2(2): 1-5.
- Amitay, M., Smith, D.R., Kibens, V., Parekh, D.E. and Glezer, A., 2001. Aerodynamic flow control over an unconventional airfoil using synthetic jet actuators. *AIAA journal*, 39(3), pp.361-370.
- Andersen, A., Pesavento, U. and Wang, Z.J., 2005. Unsteady aerodynamics of fluttering and tumbling plates. *Journal of Fluid Mechanics*, 541, pp.65-90.
- Anderson Jr, J.D., 2010. *Fundamentals of aerodynamics*. Tata McGraw-Hill Education.
- Ansari, S.A., Żbikowski, R. and Knowles, K., 2006. Aerodynamic modelling of insect-like flapping flight for micro air vehicles. *Progress in Aerospace Sciences*, 42(2), pp.129-172.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Aungier, R.H., 2006. *Preliminary Aerodynamic Design of Axial-Flow Turbine Stages*. ASME press.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Furukawa, M., Inoue, M., Saiki, K. and Yamada, K., 1998, June. The role of tip leakage vortex breakdown in compressor rotor aerodynamics. In *ASME 1998 International Gas Turbine and Aeroengine Congress and Exhibition* (pp. V001T01A054-V001T01A054). American Society of Mechanical Engineers.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 22-29.
- Gormont, R.E., 1973. A mathematical model of unsteady aerodynamics and radial flow for application to helicopter rotors (No. D210-10492-1). *BOEING VERTOL CO PHILADELPHIA PA*.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Ho, S., Nassef, H., Pornsinsirak, N., Tai, Y.C. and Ho, C.M., 2003. Unsteady aerodynamics and flow control for flapping wing flyers. *Progress in Aerospace Sciences*, 39(8), pp.635-681.
- Menter, F.R., 1992. Improved two-equation k-omega turbulence models for aerodynamic flows.
- Nitsche, W. and Weiser, N., 1987. Identification of aerodynamics flow transitions using pvdf foils. *Ferroelectrics*, 75(1), pp.339-343.

Pesavento, U., 2005. Unsteady aerodynamics of falling plates.

Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 9-16.

Rajeswari, M., Amudhavel, J., Pothula, S. and Dhavachelvan, P., 2017. Directed Bee Colony Optimization Algorithm to Solve the Nurse Rostering Problem. Computational intelligence and neuroscience, 2017.

Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 7-21.

Sedov, L.I., 1980. Two-dimensional problems of hydrodynamics and aerodynamics. Moscow Izdatel Nauka.

Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 1-8.

Shyy, W., Aono, H., Chimakurthi, S.K., Trizila, P., Kang, C.K., Cesnik, C.E. and Liu, H., 2010. Recent progress in flapping wing aerodynamics and aeroelasticity. Progress in Aerospace Sciences, 46(7), pp.284-327.

Shyy, W., Berg, M. and Ljungqvist, D., 1999. Flapping and flexible wings for biological and micro air vehicles. Progress in aerospace sciences, 35(5), pp.455-505.

Shyy, W., Lian, Y., Tang, J., Viieru, D. and Liu, H., 2007. Aerodynamics of low Reynolds number flyers (Vol. 22). Cambridge University Press.

Spalart, P.R. and Allmaras, S.R., 1994. A one equation turbulence model for aerodynamic flows. RECHERCHE Aerospatale-French Edition-, pp.5-5.

A Detailed Analysis of Computational Fluid Dynamics Methods and Its Engineering Applications

J. Amudhavel¹, L. S. S. Reddy², R. S. Raghav³ and G. Sambasivam¹

^{1,2}Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

³Department of Computer Science, Pondicherry University, Puducherry, India

ABSTRACT

This paper focuses on the various areas in which fluid mechanics plays a major role. This subject is highly relevant to the operation of engineering systems such as the brakes in a car, railcar doors, press and hydraulic actuators. Fluid mechanics has a wide range of applications in nearly all areas including mechanical, electrical, chemical and biotech engineering. In the field of medicine for instance, circulation of blood and other body fluids solely depend on the understanding of principles of fluid mechanics. The applications that are considered outstanding in the field include water distribution system, dam irrigation, siphon airfoils, pumps, drafts and hydraulic design of water treatment work. Flow in tubes, pascal's principle, Bernoulli's principle and many other applications of fluid mechanics are encountered in the normal daily life activities.

KEY WORDS: AIR FLIGHT, BIO-MEDICAL SERVICE, FLUID MECHANICS, HYDRAULIC BRAKE SYSTEM AND SIPHON

INTRODUCTION

Fluid mechanics is a branch of science that focuses on the study of fluids and analysis of various types of forces that act on them (Chanillo, S., Van Schaftingen, J. and Yung, P.L., 2016). It has a wide area of specification which includes the study of fluid behavior, fluid kinematics, fluid dynamics, fluid in motion and the study of effects that forces have on the flow of fluids. It is an

ancient field of science that is still used in the modern technology in understanding behavior of the real fluid.

The knowledge of fluid flow was first exhibited in the design of boats, arrows, spears and any hydraulic projects for water supply. The basic and important principle of hydrostatics was given by the Archimedes' principle which focuses on floating bodies which led to the development of the law of buoyancy (Daneshvar, F., Nejadhashemi, A.P., Woznicki, S.A. and Herman, M.R.,

ARTICLE INFORMATION:

*Corresponding Author: info.amudhavel@gmail.com

Received 27th April, 2017

Accepted after revision 29th July, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

2017). This principle states that a body will experience a force that is equal to the weight of the fluid when it is immersed in a fluid (Anusha B *et al.* 2015).

The continuity equation that is widely used in fluid mechanics is very important in analyzing the velocity of fluids and the area that it occupies (Davis, R.E., 1977). The response of fluids to pressure is another major aspect that led to many applications of this subject (Dharshini G *et al.* 2016). The continuity equation plays a major role in fluid mechanics especially in the study of the behavior of fluids in both statics and dynamics as a sub discipline of fluid mechanics.

APPLICATIONS

A. In Refrigerators and Air Conditioners

Working of refrigerators and air conditioners depend on the behavior of the fluids circulating in them. These fluids are known as refrigerants. Refrigerator makes use of the changes in temperature when it is either compressed or expanded to cool the contents in the chamber (Faragher, J., 2006). This fluid then takes in heat from the evaporator and releases it to the surrounding. On the other hand, in an air conditioner, the fluid absorbs heat from the room and releases it to the surrounding making use of the temperature difference between the two locations to keep the room cool (Govindharaj I *et al.* 2016). These two processes depend solely on the circulation of the refrigerant, and that is where fluid mechanics come in.

B. Air flight

Bernoulli's principle has a wide range of applications in the field of engineering and transport. One of the most ordinary applications is in air flight. This principle utilizes the relationship of pressure and velocity of a moving body to allow the airplane to move fast (Jovanovic, J., 1984). The wings of an airplane is made in a way such that it is slightly curved at the upper part while the bottom is flat whereby the air moves slowly at the bottom and faster at the top. This builds more pressure at the bottom than at the top creating a lift force that allows the plane to move across the sky (Ahilandeswari T *et al.* 2016). There are many forces also that act on an airplane. These forces include gravitational forces due to its weight, lift which is the opposite of gravitational force, thrust and drag forces. Drag force offers resistance to the thrust force and therefore at any given time, thrust must be greater than drag for the airplane to keep moving.

C. Sailing

Just like the airplane wing, sailing uses Bernoulli's principle to enhance its movement. Boats that are used in

sailing mostly have two parts, the keel which points in the south direction and a sail which points to the north (Ueda, H. and Hinze, J.O., 1975). When the wind is moving in the direction that faces the side of the sail, the air in the keel moves faster due to the increased distance due to curvature (Rajeswari, M *et al.* 2017). Therefore, a force which is perpendicular to the direction of flow of wind acts on the sail. The keel resists any movement in the horizontal direction and only allows the boat to move forward (Gayathri R *et al.* 2016). In order to sail the boat backwards, then the forces that pushes it in a direction that is perpendicular to the wind should be greater than the push force that keeps in forward direction.

D. Pump & motor

A pump is another crucial application of fluid dynamics that's used in very many areas. The pump is used to create a pressure difference to enhance the movement of the fluid from one position to another. The source of the fluid should be a region of higher pressure than the end in which the fluid is being delivered to (Faragher, J., 2006). The pump develops this pressure difference with the based on both Pascal and Bernoulli and pascal principle. There are special types of motors which are used for pumping fluids and are called hydraulic motors (Shanmugam M *et al.* 2017). The working of hydraulic motors is just similar to that of a hydraulic cylinder. Despite the fact that many hydraulic pumps cannot be back driven, a modification was made on the hydraulic motors so that it can work on both sides of the motor, i.e. forward and backward direction.

E. Hydraulic brake system

A major application of fluid mechanics in the transport industry is the hydraulic brake system that involves applying energy to stop an object (Rajadurai R *et al.* 2017). When the brake pedal is pushed, the small piston is pushed. This process is initiated by the hydraulic fluid that transmits pressure called the brake fluid. The small piston will therefore transmit pressure to the brake fluid which in turn presses larger piston of the system through the brake pads. The brake drum comes into contact with the brake pads slowing the car, and finally, it stops.

F. Jack

Fluid mechanics is widely applied in the making of machines for lifting of loads from one position to another. A Jack is a machine that utilizes pascal principle in order to lift heavy loads. The working principle of the Jack depends on balancing of hydraulic forces ((Faragher, J., 2006).). It is made of screw threads and a hydraulic cylinder and employs both mechanical and hydraulic forces to function. The jacks used in cars and houses to lift heavy equipment are mostly mechani-

cal jacks and are rated by their capacity to lift. Bottle jacks and floor jacks use hydraulic force to lift objects (Sasidevi V *et al.* 2017). Hydraulic jacks are preferred to mechanical jacks because they can lift heavier loads higher.

G. Waterwheel

Developing a waterwheel was the greatest achievement of engineering innovations in the ancient times (Ueda, H. and Hinze, J.O., 1975). It consists of a number of buckets which are placed along a rim that allows one to raise water from a well that is deep below the surface or any other water source and be able to scatter to other points. This was also a very significant invention in the field of agriculture. Water flowing through a steep terrain consists of kinetic energy which is then utilized to turn wheels and does other work like grind grain (Lyons, S.L., Hanratty, T.J. and McLaughlin, J.B., 1991).

H. House fan & Turbine

Fresh air and heat regulation in a house is initiated by fans. Any moving fluid possess kinetic energy, in this case the turbine regulates heat by converting energy possessed by the fluids in motion into mechanical energy that is used to moving parts. This is done by passing the fluid in a continuous contact with the moving blades and fans (Jovanović, J., Durst, F. and Johansson, T.G., 1993). A perfect example of this turbine is a house fan although it works in opposite direction. In this case, the fan energizes the nearby air while the turbine absorbs mechanical energy from the passing air that is surrounding it. Turbines have a wide range of application with the most significant one being used in the extraction of hydroelectric power.

I. Hydroelectric dams

Hydroelectric dams are the most significant application of fluid mechanics. They generate energy that is enough to supply an entire country using an entirely renewable source. They are impressive in power generation and are equally large. The dams in most cases, are made of concrete and steel making them strong enough to hold a large amount of water available for production of energy. At the top of the reservoir, the water possesses an enormous potential energy, and by allowing streams of water to flow downwards, it will acquire a great deal of kinetic energy which in turn powers the turbines (Lyons, S.L., Hanratty, T.J. and McLaughlin, J.B., 1991). The rotating turbines solely powered by running water can generate electric power.

J. Air compressor

Electrical energy from a motor powered electrically, gasoline engine or a diesel engine can be converted into a

form of kinetic energy by using a machine known as an Air compressor. This machine performs its operation by pressurizing and compressing air and then it is released in rapid bursts. Its main purpose is to deliver air into a nail gun. Air compression procedures come in two major ways; that which utilizes positive displacement and the other uses negative displacement. This technique is useful in high-pressure gas cylinders, filling tires and also useful in producing medium-pressure air suitable for industrial processes such as oxidation.

K. Gears

Gears form the most significant part in vehicles and other industrial machines. For instance, modern cars are able to transfer the rotation produced by the engine to the wheels to enable movement with aid of these gears. Moreover, gears not also transfer rotation of engine but also equate the torque produced by the engine so as to create a mixture of speed and acceleration for the entire part of the car. Gears are responsible for the entire movement of the car unless of course, they used rubber bands.

L. Water Slides or roller coasters

The models of cars which use coasters are driven by gravity. They have wheels that have less friction between the wheels and the car and enhance movement as they roll along the track. The coaster car rises higher in the air at the beginning of the ride possessing the greatest potential energy enough to release it from top of the hill. Gravitational force converts potential to kinetic energy pulling the car down. Friction between the track and the wheels decreases gradually keeping the car in motion (Faragher, J., 2006).

On the other hand, the same principle applies to water slides. While climbing the stair, potential energy is gained which is then converted into kinetic energy as you slide down. The shape of the slide determines the potential energy generated in a particular slide. Therefore, potential energy is the one that brings the difference between different slides. In this case, water acts like a lubricating medium between the slides and the body. The slides consist of a continuous flow of water from top to bottom

HYDRAULIC PRESSES

Pascal's principle is applied in operation of many machines that apply the concept of a hydraulic press. A hydraulic press has a wide range of applications including hydraulic jack that is used to raise various loads off the floor like cars in a mechanic workshop. This involves a chamber beneath the floor of the shop with two hydraulic cylinders which are placed side by side. Each of the cylinders is connected to a piston and the

floor between the two cylinders is controlled by valves. When a force is exerted by pressing piston of one cylinder, a uniform pressure pushes the second cylinder, and the output raises the car.

A. Siphon

A siphon is a type of a pump which utilizes the pressure difference to transfer fluids from one region of high pressure to another of low pressure. A siphon works based on both Pascal's and Bernoulli's principles. It has a variety of applications like drawing water from a tank on the roof to the ground, drawing gas from the fuel tank of a car using a siphon hose. When air from one end of the siphon is sucked, the hose creates a low-pressure region, and the content of the pipe will eventually come out from the end with a relatively lower pressure.

A WIND TUNNEL

Airflow between objects can be well described by a wind tunnel by studying behavior of wind that is flowing through a room. A rudimentary wind tunnel aids in studying the characteristics of air that flows in contact with objects such as automobiles and aircraft. This wind tunnel makes use of properties of fluid mechanics such as the Bernoulli's principle. The aim of this study tunnel is to be able to test and study the interaction of airflow and solids in relative motion. In this experiment, the aircraft can be moving against a stationary aircraft moving against the flow of air or the other choice is having a stationary aircraft so that the airflow moves across it. However, it poses little danger to expose stationary craft to a high-speed simulating that of that of an aircraft in motion.

A. Air-jet weaving machine

This is a modern type of weaving machine that is used to produce the highest insertion performance and is used to manufacture light and medium weight fabrics.

B. Creating a draft

The use of aerodynamics is among the most known applications of Bernoulli's principle. It explains why a curtain in the shower tends to billow inward when the water is turned on. It is also able to explain why an open door and window at the same time create a draft. With the door, open, air flows from the high-pressure region of the room to the low-pressure hallway, and soon wind will begin to blow through the room (Lyons, S.L., Harratty, T.J. and McLaughlin, J.B., 1991).

C. Fluid dynamics in medical research and bio-medical service

Fluid mechanics plays a fundamental role in the development of the biomedical device, development of micro-

electronic and microfluidic device. The new oxygen-air mixing device with a different design which has been successfully innovated with the current technology. The new design for instant graphs which is used for the treatment of an aneurysm also employs fluid mechanic principles (Sefiane, K., 2010). It has also led to the electronic fluid flow detector which is used to measure fluid flow in a hospital setting.

D. Natural hazard analysis and environmental flow

The field of fluid mechanics is important when dealing with simulation of a land slide that generates Tsunamis by free surface flow evaluation using the volume of fluid technique. It is also applied in the leak and spill-related fire analysis and smoke assessment (Sefiane, K., 2010).

E. Elevators

Most elevator systems are driven either by hydraulic or electric motion. The hydraulic ones are usually in buildings with a maximum of five floors like hospitals, small apartment homes, and clubs (Van Thinh, N., 1982). The speed of the elevator is around 0.75m/s, and it doesn't need any overhead gear. It allows people to move from one floor to the other without any effort.

F. Cranes

The lifting machinery normally seen on many construction sites. Telescopic cranes rely on hydraulics to function effectively. They have a high rating of efficiency since the moving parts are reduced by the oil-based fluids. The tower cranes which are self-erecting devices reach unsupported heights of 80m. The tower cranes use hydraulic rams to lift materials upwards or reverse the direction when the procedure is completed. It has allowed engineers to deal with the problem of increased heights of modern buildings.

CONCLUSION

Fluid mechanics is fundamental unit of science which has a wide range of applications in nearly all areas including mechanical, electrical, chemical and biotech engineering. Some of these applications are experienced in our day to day lives. The applications that are considered manipulative in the field include water distribution system, dam irrigation, siphon airfoils, pumps, drafts and hydraulic design of water treatment work. Both Pascal's and Bernoulli's Principle play a fundamental role in our day to day activities.

REFERENCES

Ahilandeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Envi-

- ronment. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(2): 1-5.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 1(1): 1-7.
- Chanillo, S., Van Schaftingen, J. and Yung, P.L., 2016. Applications of Bourgain–Brézis inequalities to fluid mechanics and magnetism. *Comptes Rendus Mathématique*, 354(1), pp.51-55.
- Daneshvar, F., Nejadhashemi, A.P., Woznicki, S.A. and Herman, M.R., 2017. Applications of computational fluid dynamics in fish and habitat studies. *Ecohydrology & Hydrobiology*, 17(1), pp.53-62.
- Davis, R.E., 1977. Techniques for statistical analysis and prediction of geophysical fluid systems. *Geophysical & Astrophysical Fluid Dynamics*, 8(1), pp.245-277.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(1): 1-6.
- Fragher, J., 2006. *The implementation of probabilistic methods for uncertainty analysis in computational fluid dynamics simulations of fluid flow and heat transfer in a gas turbine engine* (No. DSTO-TR-1830). DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION VICTORIA (AUSTRALIA) AIR VEHICLES DIV.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(1): 7-13.
- Jovanovic, J., 1984. *Statistical analysis and structure of wall turbulence* (Doctoral dissertation, PhD thesis, Mechanical Engineering, University of Belgrade, Yugoslavia).
- Jovanović, J., Durst, F. and Johansson, T.G., 1993. Statistical analysis of the dynamic equations for higher-order moments in turbulent wall bounded flows. *Physics of Fluids A: Fluid Dynamics*, 5(11), pp.2886-2900.
- Lyons, S.L., Hanratty, T.J. and McLaughlin, J.B., 1991. Large-scale computer simulation of fully developed turbulent channel flow with heat transfer. *International journal for numerical methods in fluids*, 13(8), pp.999-1028.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 9-16.
- Rajeswari, M., Amudhavel, J., Pothula, S. and Dhavachelvan, P., 2017. Directed Bee Colony Optimization Algorithm to Solve the Nurse Rostering Problem. *Computational intelligence and neuroscience*, 2017.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455-8907, 3(1): 7-21.
- Sefiane, K., 2010. On the formation of regular patterns from drying droplets and their potential use for bio-medical applications. *Journal of Bionic Engineering*, 7, pp.S82-S93.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 1-8.
- Ueda, H. and Hinze, J.O., 1975. Fine-structure turbulence in the wall region of a turbulent boundary layer. *Journal of Fluid Mechanics*, 67(1), pp.125-143.
- Van Thinh, N., 1982. A study of the intermittent phases in the wall region of a turbulent flow. Structure of turbulence in heat and mass transfer, pp.103-116.

Fluid Dynamics in advanced medical Diagnostic Technologies: An in-depth Analysis

J. Amudhavel^{1*}, V. Srikanth², R. G. Babukarthik³ and G. Sambasivam⁴

^{1,2,3,4}Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

ABSTRACT

Fluid dynamics is a branch of science that describes the motion of liquids and gasses and their interaction with solid bodies surrounding them. Grotberg discussed various methods in a broad interdisciplinary field which impacts every aspect in our daily lives including medicine, biology, environment, security, manufacturing and transportation. To effectively predict the flow of blood in a human body, one has to carefully study and understand fluid dynamics and its contribution to the sub-discipline of biology. The quest for deeper understanding has inspired many advances in the field of medicine and realizing of new diagnostics methods to address flow in the human body, animal swimming and flight, just to name a few.

KEY WORDS: FLUID DYNAMICS, INTERDISCIPLINARY, SUB-DISCIPLINE, DIAGNOSTICS METHOD AND HUMAN BODY

INTRODUCTION

Fluid dynamics have a big role in the growth of any living organism that is in chemical and physical processes in biological systems since many of the fluid forces move to guide growth and improvement. All biology includes fluids, circulation, and cleansing of blood, transporting nutrients, flying of birds, pollination all the way to sea animals. The process of fertilization and reproduction depend on fluid flow. The body plan itself depends on fluid dynamics so as to develop. Generally, biological fluid flows are present in all living organ-

isms because they undergo the process of reproduction, metabolism, feeding, growth and locomotion (Anusha B et al. 2015). For proper functioning of the major part of human beings such as lungs, kidneys, liver, eyes and brain critically depend on the fluid transport processes to enable rapid exchange of molecules between blood and the tissues are stated in Laine et al.

The development of any organism is highly affected by its external environment which includes external flow. Cartwright et al. discussed about Embryonic mollusks, fish, and amphibians depend on the fluid flow in the external environment for their eggs to be hatched.

ARTICLE INFORMATION:

*Corresponding Author: info.amudhavel@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

Water plants, corals, and sponges depend on the behavior, velocity of water and wind current for their development. The major biological processes such as diffusion, osmosis, and dialysis can't function in the absence of fluids. Therefore, Cartwright et al. says fluid dynamics is the heart of all living organisms. In this research paper, I have discussed the various situations in which biology depends on fluid dynamics. For example, how the body regulates blood flow, removal of waste products and the various applications in the field of medicine that solely depend on fluid flow (Dharshini G et al.2016).

PHYSICAL PROCESSES AND DRIVING MECHANISMS

For proper function of a living organism, there are various processes that take place to initiate development. Turgor pressure is responsible for maintaining the shape of the plant and allows them to be stiff and unrolled, the hydrostatic skeleton in insects solely depend on the fluids in the body of the organism. Osmotic pressure is a physical mechanism generally for setting and balancing of fluids in motion. Cortical flow in amphibians is caused by the buoyancy forces which further causes the appearance of instability in fluid flow such as the Taylor instability which enables the heavier fluids to underlie the lighter one common in the lungs and structures of fossil Ammonoids (Govindharaj I et al. 2016). Generation of fluid within a cell is done by a mechanism called Actomyosin. However, it is varied from one organism to the other. Another mechanism for moving fluid is the cilia which are present in the kidney, ear, and brain. Liquid crystallization is also involved in the supra-molecular assembly of tissues like molluscan nacre whose function is to combine and mix materials to take them to their respective parts. Reynolds number is used to gauge the turbulence in fluids whereby the flow of fluid around the body ceases when driving stops and the value within which development stops.

CYTOPLASMIC STREAMING

Fluid mechanics plays an important role in shaping the structure of an organism right from the smallest building block i.e. the cell to the whole body. Cytoplasmic streaming which is commonly known as cyclosis is the continuous circulation of fluids within the cell and occurs in various organisms such as fungi, protozoa, and ameba. It is a crucial process in cell locomotion of these organisms. The materials that originate from an anucleate region are actively transported by a mechanism called actomyosin which is dependent on the cytoplasmic streaming (Ahilandeswari T et al. 2016).

Also, cytoplasmic streaming is experienced by green plants during their cellular growth. For instance, in a

Chara coralline species of alga, the fluid velocity gives a Reynolds number of around 0.1 greatly outpaces molecular diffusion for single cells. Streaming transports and mixes the nutrients in a way that enhances fitness and cell growth.

CIRCULATION OF BLOOD IN THE HEART

The circulatory system consists of many tubes with fluid flowing through them. The heart contracts and relaxes to generate regional pressures to ensure that blood moves in one direction. Hassan et al. discussed about the heart, lungs and the blood vessels collectively involve in maintaining the cells and organs healthy by influencing the fluid dynamics of blood.

The author Castro et al. discussed about Hemodynamic is an area which has been widely studied and many applications of fluid dynamics major in this part. Study of the blood flow has been a source of many research and fluid-dynamical problems that has contributed to the understanding of the contribution of fluid dynamics in the field of medicine. Comprehension of elasticity and fluid motion in networks such as arteries and small capillaries has helped in creation of the circulatory system. Display of self-excited oscillations induced by coupling, is vital in the circulatory system and forms one of the diagnostic tools to be analyzed. On the other hand, there are many fluid dynamical challenges as a result of blood rheology and the effects it experiences from the transport of the blood cells (Gayathri R et al. 2016).

Hemodynamic is now being studied extensively, and we have to appreciate its importance in connection with the cardiovascular system discussed in the year 2005 by Cebra et al. Meng et al stated in research shows that in the early stages of development, the forces in the developing heart are more firm compared to the cell's mechanical sensitivity with Reynold's number of 0.1. This indicates that nature is using the force surge to regulate cellular differentiation and genetic definition at an early stage of human heart development. The primary role of a small embryonic heart is solely to regulate heart morphogenesis and not to pump blood for transport (Rajeswari, M et al. 2017). The heart only will beat for transport when the diffusive transport of materials ceases therefore before the fetus is born, food, water, and oxygen are circulated by diffusion. Flow of fluids is now much acknowledged for its contribution in the development of the vascular system.

THE MOVEMENT OF BLOOD AND URINE

Kidneys clean and filter blood, while fluid forces help in their growth and functioning discussed by Smith et al in the year 2011. Walls of the kidney tubes have mech-

ano-sensory cilia which are sensitive to deflection which vary with the flow. At a low Reynold's number, the flow is regular as in Hagen-Poiseuille principle on flow in a pipe. The polycystic kidney disease (PKD) is a disorder that results when the tube sizing fails to sense the developmental system. When the mechano-sensing cilia grow beyond their normal size, it will end up disrupting the topology of the structure hence causing renal malfunction. The flow of amniotic fluid is maintained by urine I the embryo (Shanmugam M *et al.* 2017).

SHEDDING AND METAMORPHOSIS

Most body structures' development depends on the movement of gases and liquids and the fluid dynamics involved in it is a great deal. Molting/shedding is associated with regeneration of exoskeleton in anthropoids and roundworms and also metamorphosis in insects. Fluid flow, in this case, is necessary for building and also to breaking down the existing exoskeleton before it is removed. The molting fluid dissolves the old cuticle so as to form a new one. For instance, in a silk moth, the molting fluid is emitted from the exuvial slot when the epidermis withdraws itself from the mature pupal cuticle. The fluid goes on for around three weeks of its adult life discussed 2004 by Hoi *et al.* On around the end of the second week of the pharate development, the molting gel liquefies to hydrolyze proteins and chitin of the endocuticle (Rajadurai R *et al.* 2017). By the end of the third week; the endocuticle disappears leaving a thin exocuticle and epicuticle. The molting fluid, however, is absorbed and secreted through osmosis.

REGULATION OF OXYGEN AND HEAT LEVELS BY EXTERNAL FLOW

Mollusks and amphibians have one evident characteristic of laying gelatinous eggs. Oxygen flows to the embryo via the extraembryonic liquid around it. If no fluid is flowing, oxygen diffuses across although insufficient oxygen will pass through. The whole embryo rotates, and the ciliary pulse directions align itself on the skin of the embryo so as to get enough oxygen. These embryo movements are observed in cephalopod mollusk, snails, squids, and fish. Other types of fish are found to be fluttering their fins to increase their oxygen levels (Sasidevi V *et al.* 2017). These structures are perfect examples of developmental fluid dynamics in the real world.

LARVAL SWIMMING AND PLANKTON DYNAMICS

Mobile and sessile organisms utilize the flow in their surroundings to their advantage. Many larvae adapt to

their environment by the swimming response they experience in their surroundings. The larva swims less when there are rapid currents or aligns themselves in a manner that orients them on the shear. Some of these organisms choose to be on the surface and attach themselves to other water plants depending on the velocity of the flowing water.

As they develop through their larval stages Reynold's number becomes greater than one ($Re > 1$) whereby inertia will be dominant and viscous forces will no longer be useful. Therefore, as they grow, their propulsion mechanism changes from swimming based on the drag forces but a real intermittent swimming. This becomes more advantageous to these organisms because the viscous forces reduce as their Re increases. Another example is the mollusks which often use the technique of jet propulsion whereby water is squirted off their veil cavity as they clap their shells collectively in a simultaneous manner. The smaller ones don't use the inertia form effectively because of the greater viscous forces making them sedentary.

BIOMATERIALS AND FIBROUS COMPOSITES

Everyone has come across some objects that have fiber structures set in a matrix of a similar fashion like plywood and fiberglass. Examples of these natural composites include plant cell wall, arthropod cuticle, cornea, and bone. The structure of these composites suggests liquid that went through crystallization process and solidified to form a particular pattern. For instance, in a mother of pearl, the fluid which occurs in a liquid filled extra pallial space goes through self-organization to form a shell structure. Biomaterials such as teeth and bones are formed through this same process. Therefore, it can be seen that fluid flow plays a major role in the development of living organisms.

TRANSPORTATION OF FLUIDS BY THE PHYSIOLOGICAL PROCESSES

Nutrients are transported through the human body by a random thermal molecular motion called diffusion. This process plays an important role in the exchange of nutrients and waste products between the blood and the tissues. The single-celled organisms solely depend on diffusion for uptake of nutrients, unlike the multicellular organisms which require faster means of transportation like the circulatory systems in mammals. Osmosis is another important process driven by the imbalance in water concentration where water molecules move from a region of highly concentrated region to a lower concentrated region through a semi permeable membrane. Dialysis is another important process used by the kid-

ney to cleanse the blood. In many hospitals, the dialysis machinery is important while treatment of individuals with kidney problems.

THE LYMPHATIC SYSTEM

Fluid flow guides the growth and functioning of the lymphatic system. The lymph system is fed by fluids flowing from the capillary vessels which flow from the body tissues to the lymph system. The fundamental role of the lymph system is to maintain a balance of the interstitial fluid discussed by Swartz et al. Guidance of growth and organization of the lymphatic capillary network may affect the interstitial flow. The lymph flow makes use of osmotic pressure for the interstitial fluid to flow from one region to another in the lymphatic system. It is necessary to note that the transcapillary exchange of fluid is important to control the volume of the fluid space in situations such as hemorrhage, hemodialysis, and fluid replacement therapy.

ANALYSIS

Fluid dynamics research has a detailed history ranging from a single organism, aquatic bacteria, and fish to aerial birds and insects. A lot of work has been performed on physiology-where biology comes into contact with fluid dynamics discussed Denny by 2014. The appreciable numbers of examples of biological processes that apply fluid dynamics discussed above are in human and biological life as a whole with each process seen in isolation. It can be seen that fluid dynamics shapes the development of the cell itself and many organs carrying fluids in the body of a grown up. Moreover, it is noted that the fluid environment of an organism has an impact on its normal functioning and the organism has to adjust to adapt to any given environment.

It is presented that some organisms use fluid flow in their external surrounding through protective mechanisms such as the nest for termites and bees which provide them with the required temperature for brood development. This is an extended adaptation involved in the development of fluid dynamics.

It is worth recalling the importance of osmotic pressure, buoyancy joins with liquid crystallization and other physical systems which are effective driving mechanisms for biological morphogenesis and plasminogen. When the process of fluid flow goes wrong, fluid dynamics is intimidated which will ultimately result in the development of a disease. So as to understand and correct these problems, a deep understanding of fluid dynamics is necessary. Diseases such as PKD and hydrocephaly are fluid-dynamical problems resulting from genetic errors in the makeup of cilia which produces flow.

CONCLUSION

It is important to understand the fluid mechanical concepts that are applied in various areas ranging from the biological oceanography, locomotion, cardiovascular physiology to diffusion and osmosis within the cell and across cell membranes. All organisms are filled with fluid compartments and live in an environment surrounded by fluids. Fluid Dynamics is responsible for the movement of an organism, feeding, exchanging heat and sense information. The functional design of these organisms and how they behave in their environment is heavily explained by fluid mechanics. Therefore, biology and fluid dynamics work hand in hand so as to understand the physical phenomena. The interdependence between biology and fluid dynamics need to be solid so as to bring comprehension of various physical steps taken and morphogenesis in biology. The cases discussed above explore definitely how fluid forces perform to guide development in living organisms. Fluids make the greatest portion in living organisms. Therefore, fluid dynamics is crucial in chemical and physiological processes shaping the development.

REFERENCES

- Ahिलandeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 2(2): 1-5.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Cartwright, J.H., Piro, N., Piro, O. and Tuval, I., 2007. Embryonic nodal flow and the dynamics of nodal vesicular parcels. *Journal of The Royal Society Interface*, 4(12), pp.49-56.
- Cartwright, J.H., Piro, O. and Tuval, I., 2009. Fluid dynamics in developmental biology: moving fluids that shape ontogeny. *HFSP journal*, 3(2), pp.77-93.
- Cartwright, J.H., Piro, O. and Tuval, I., 2009. Fluid dynamics in developmental biology: moving fluids that shape ontogeny. *HFSP journal*, 3(2), pp.77-93.
- Castro, M.A., Putman, C.M. and Cebal, J.R., 2006. Computational fluid dynamics modeling of intracranial aneurysms: effects of parent artery segmentation on intra-aneurysmal hemodynamics. *American Journal of Neuroradiology*, 27(8), pp.1703-1709.
- Cebal, J.R., Castro, M.A., Burgess, J.E., Pergolizzi, R.S., Sheridan, M.J. and Putman, C.M., 2005. Characterization of cerebral aneurysms for assessing risk of rupture by using patient-specific computational hemodynamics models. *American Journal of Neuroradiology*, 26(10), pp.2550-2559.
- Denny, M., 2014. *Biology and the mechanics of the wave-swept environment*. Princeton University Press.

- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Grotberg, J.B., 2001. Respiratory fluid mechanics and transport processes. *Annual review of biomedical engineering*, 3(1), pp.421-457.
- Hassan, T., Timofeev, E.V., Saito, T., Shimizu, H., Ezura, M., Matsumoto, Y., Takayama, K., Tominaga, T. and Takahashi, A., 2005. A proposed parent vessel geometry-based categorization of saccular intracranial aneurysms: computational flow dynamics analysis of the risk factors for lesion rupture. *Journal of neurosurgery*, 103(4), pp.662-680.
- Hoi, Y., Meng, H., Woodward, S.H., Bendok, B.R., Hanel, R.A., Guterman, L.R. and Hopkins, L.N., 2004. Effects of arterial geometry on aneurysm growth: three-dimensional computational fluid dynamics study. *Journal of neurosurgery*, 101(4), pp.676-681.
- Laine, G.A., Hall, J.T., Laine, S.H. and Granger, J., 1979. Trans-sinusoidal fluid dynamics in canine liver during venous hypertension. *Circulation research*, 45(3), pp.317-323.
- Meng, H., Wang, Z., Hoi, Y., Gao, L., Metaxa, E., Swartz, D.D. and Kolega, J., 2007. Complex hemodynamics at the apex of an arterial bifurcation induces vascular remodeling resembling cerebral aneurysm initiation. *Stroke*, 38(6), pp.1924-1931.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Rajeswari, M., Amudhavel, J., Pothula, S. and Dhavachelvan, P., 2017. Directed Bee Colony Optimization Algorithm to Solve the Nurse Rostering Problem. *Computational intelligence and neuroscience*, 2017.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Smith, D.J., Smith, A.A. and Blake, J.R., 2011. Mathematical embryology: the fluid mechanics of nodal cilia. *Journal of Engineering Mathematics*, 70(1-3), pp.255-279.
- Supatto, W., Fraser, S.E. and Vermot, J., 2008. An all-optical approach for probing microscopic flows in living embryos. *Biophysical journal*, 95(4), pp.L29-L31.
- Swartz, M.A., Kaipainen, A., Netti, P.A., Brekken, C., Boucher, Y., Grodzinsky, A.J. and Jain, R.K., 1999. Mechanics of interstitial-lymphatic fluid transport: theoretical foundation and experimental validation. *Journal of biomechanics*, 32(12), pp.1297-1307.
- Waite, L. and Fine, J.M., 2007. *Applied biofluid mechanics*.

Investigation of various numerical methods for Computational Fluid mechanics

J. Amudhavel^{1*}, L.S.S. Reddy¹, T. Asvany² and D. Saravanan¹

¹Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

²Department of Computer Science, Pondicherry University, Puducherry, India

ABSTRACT

Computer science has got a critical role when it comes to computational fluid dynamics field although it has been looked down upon for a long period of time. Computer science involves using sophisticated techniques to extract information, analyze it and render to the users from both the academia as well as the industry. It allows simulation tools to be integrated into process chains to come up with flexible modular software concepts that are efficient to manage and to work with. The growing complexity in the application areas led to the need to use numerical algorithms and hardware architectures. This sophisticated software is needed in fluid mechanics so that large sets of data resulting from numerical simulations and parallelization can be handled in a more effective way. For instance, the complex software codes are combined with the upcoming multi-core and complex software when it comes implementation and use of the numerical algorithms in the efficient way possible. Furthermore, these algorithms are data-intensive thereby, posing a big challenge while storing and accessing data. The space-tree grids together with the field of partitioned fluid structure interactions are examples that illustrate how and where computer science contributes in simulation environment.

KEY WORDS: MULTI-CORE, PARALLELIZATION, SIMULATION, FLUID MECHANICS, COMPLEXITY

INTRODUCTION

Modeling together with the classical arithmetic have really dominated this field of fluid mechanics for several years, but computer science came in when the algorithmic and numerical scheme was finally found. Although computer science is associated with more programming, it applies mostly to the less technical part of fluid

mechanics. For a successful simulation of pipelines and loops require computer science for effective running of the cycle which involves implementation, exploration and visualization, validation and finally embedding. Computer science come in where cache awareness is required as well as parallelization and performing of multicore capabilities that reach beyond the colorful fluid dynamics. Computer science involves using

ARTICLE INFORMATION:

*Corresponding Author: info.amudhavel@gmail.com

Received 27th April, 2017

Accepted after revision 29th July, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

sophisticated techniques to extract information, analyze it and render to the users from both the academia as well as the industry. It allows simulation tools to be integrated into process chains to come up with flexible modular software concepts that are efficient to manage and to work with. The standard stages for pipeline simulation and numerical algorithms are outlined below (Anusha *et al.* 2015).

For the above stages to be implemented, a profound comprehension of the role of the hardware, compilers, together with the implementation is very vital so that it can develop an efficient simulation code from the data. For a code to be applied in a realistic simulation, efficiency is a precondition. Computer science contributes heavily to the study of computational fluid dynamics and aids in efficient and effective numerical simulation of fluid flow (Dharshini *et al.* 2016). Most programs requires an elaborate user interface that is easy to adapt with the knowledge in combinations with expertise of the user although it is simple, it is not understood by everyone. The numerical simulations are commonly applied in physical uses which include the interaction of fluid-structure as an efficient and flexible technique to get important and independent codes.

COMPUTATIONAL FLUID DYNAMICS (CFD)

When it comes to the use of numerical analysis in the analysis and solving problems then the branch of fluid mechanics that can handle that is Computational Fluid Dynamics. It enables quick and efficient simulation of fluid flow and allows one to calculate the forces that act on a fluid. CFD involves replacing the partial differential equations, which are used for the conservation laws. This are combined with the algebraic equations, which are usually solved by digital computers. It is an important field that allows engineers and scientists to perform computer simulations as numerical experiments with a virtual flow representing a real flow. Ferziger & Peric developed a new source in the computational model. These numerical simulations of fluid flow play an enormous role in understanding fluid mechanics. It enables architects to design safe and comfortable structures suitable for living. The designers of vehicles can improve the aerodynamic characteristics. For optimal oil recovery structures, petroleum engineers need to understand the necessary numerical simulations and be able to reduce radiations and other health hazards.

CFD does not rule out measurements completely but enables the process to be cheaper, faster, parallel and multipurpose. Despite the fact that it's difficult to transport the equipment and personnel, the CFD software solves all these problems this is attributed to the way they are portable and very easy when it comes to use.

In addition to that, they are very easy when it comes to modification. Results obtained from the CFD simulation are considered reliable though not wholly due to the inputs involve imprecision, the model at hand at times may be considered inadequate and the accuracy being dependent on the computing power. Cebral and Löhner applied and developed a finite method in the use of CFD simulation. CFD simulations are more reliable for slow and laminar flows rather than the fast and turbulent ones. It is accurate for fluids that are chemically inert and mostly focused on single-phase flows rather than multi-phase flows. The accuracy of the results given by the simulation majorly relies on the assumptions underlying with the mathematical m, the stability of the numerical scheme and the error indicators (Govindharaj *et al.* 2016).

FLUID STRUCTURE INTERACTION

When it comes to coupling of several codes then the MpCCI is the commonly used Software. The MpCCI Software was developed at Fraunhofer Institute purposely. The main purpose of the MpCCI Software was to Compute the algorithms and scientific as a predecessor of the GriSSLi Software (Ahilandeswari *et al.* 2016). This software is known for providing the user with various functions that are critical for the exchange of data, which involves the interpolation procedure. This is imperative when enhancing the mapping of data of the involved servers and computational grids. Therefore, when it comes coupling of codes and enabling a programmer to understand the coupling in a platform-independent manner then this is the software to use. Implementing the coupling control is determined by the coupling strategy applied, which is then implemented in respective solver codes. There are other sophisticated coupling strategies for instance, the multi-grid coupling which not only requires data exchange and but also real control is required. Wesseling (2009) has described various types of principles models which have been provided in this technique. Applications of fluid-structure simulation cover a wide area specifically on the elastic wings and the surrounding air does operate and interact (Rajeswari *et al.*, 2017). Others Include movement of ships, towers, and skyscrapers under various wind conditions and coastal fortifications that exposed to sea waves and many others.

NUMERICAL METHODS USED IN FLUID FLOW

Navier-Stoke's equation governs fluid motion and consists of a set of linear and non-linear differential equations that are derived from the basic conservation laws. The main components of a numerical solution in fluid dynamics include a mathematical model, discretization

method, coordinate and vector system, the numerical grid generation and finally the finite approximations used in discretization process (Gayathri et al. 2016).

Discretization is the central process in CFD and involves using differential equations that has an infinite number of degrees of freedom while trying reducing it so that it can fit in the system which has finite degrees of freedom. From this method, instead of getting the solution all over for a number of times, calculation of these equations using a finite number of locations at specified time intervals will be enough to give the expected results. The partial differential equations will be reduced to the algebraic equations system hence they will be able to be solved by use of computer. This process is very simple. However, it is prone to errors, but the behavior characteristics of these errors is essential that they are manipulated so as to ensure that, correct equations are being solved following a recommended consistency property. One major way of decreasing these errors is by increasing the number of degrees of freedom so as to gain stability and convergence. Laval, B was proposed a numerical based internal and external source flows (Shanmugam et al. 2017).

Once the errors have been minimized, the power of computing machines using suitable CFD software to leverage the result when it comes to solving the problem in an incredible way that is numerically reliable. We used several discretization methods developed to help cope up the issues up that come along while taking the differential equations and converting them into finite degrees of freedom. The most efficient methods include the finite difference methods, finite volume methods, spectral methods and finite element methods (Rajadurai et al. 2017).

FINITE DIFFERENCE METHOD

A second order accuracy of the Finite Difference Method is used apart from the few special situations. This method is still a popular numerical method when finding solutions of PDEs. This is due to the simplicity of this method, its efficiency and has a low computational cost. The geometric inflexibility is the major Challenge of this method since it makes their applications when solving the general complex domains very difficult and complicated. This inflexibility can be reduced by using either mapping techniques or masking the computational mesh to fit onto the computational domain. Tijms, Henk proposed a stochastic model is derived in fluid mechanics domain (Sasidevi et al. 2017).

FINITE ELEMENT METHOD

When it comes to dealing with problems that arise due to Complicated Computational regions then this is the

method to use. The partial differential equations are first to recast into several form whereby it forces the average error to be minute everywhere. Donea & Huerta has introduced this discretization process functions by dividing the computational domain by the elements of rectangular or triangular shape. An interpolation of Polynomial of a lower order within each element and the solution found will be required to obtain the collocation points. This method was adopted by the computational fluid dynamics' community in the 1980s.

SPECTRAL METHOD

The first two methods were found to be efficient only when dealing with low-order equations usually between second and fourth orders, and they usually have a property of local approximation. By saying local indicates that a particular collocation point is hugely affected by a limited number of points, which are neighboring it. In other words, this other method has a property global approximation use whereby its interpolation functions, which can be either polynomials or trigonometric functions which in nature are global. Their primary advantage of this method is on the rate of convergence. The smoothness of the solution obtain relies on the rate of convergence. The smoothness is determined by the number of continuous derivatives it admits. For instance, in an infinitely smooth solution, the exponentially decrease in error is evidently faster than algebraic. Therefore these methods are mostly significant in the calculation and computations which homogeneous turbulence is experienced. One of the models that have adopted the spectral model is the atmospheric model significantly due to their property of convergence discussed by Tijms, in the year 1994. The regular and consistent spherical shape of the computational domain of this method makes it possible to be used in the atmospheric model.

FINITE VOLUME METHOD

When having a look at the applications of aerodynamics where strong discontinuities and shocks in the solution are likely to come from then the finite volume methods is the ultimate method to use. This technique is used to solve an integral form of particular equations so that the property of the local continuity does not have to hold. This is one of the incredible techniques coupled together with the Computer Science technology.

SOLUTION METHOD

Discretization process will yield a large system of linear and non-linear algebraic equations. The Linear equations are computed using the Algebraic equation solvers while the iteration scheme used to solve Non-linear

equations which work by linearizing the equations to give a result in linear systems which are then solved by iterative techniques. Unsteady flows are solved using methods based on marching in time while the steady flows are usually solved by use of the pseudo-time-marching technique or an equivalent iteration scheme method.

COMPUTATIONAL COST

The time required to provide a solution to the system of equations do differ from one method to another. The cheapest technique which uses the per grid point technique is the finite differences method. This is coupled with the finite element method and finally the spectral method. However, a per grid point technique comparison does not take into consideration features of each method and their advantages. For instance, the Spectral methods are preferred when one a high-level accuracy on a per grid point basis than either finite element method or the finite difference method. Therefore, the comparison will be more meaningful if the computational cost to attain a given error tolerance is emphasized. Errors in numerical solutions provided. Numerical solutions involve three kinds of errors which are the Modeling Errors, the discretization errors and the iteration errors. Modeling errors are obtained by the difference of the actual flow & exact solution obtained from the mathematical model. On the other hand, Discretization Errors are found from the difference of the exact solutions of conservations equations and an exact solution found from the algebraic system of equations and is commonly referred to as truncation error. Finally, Iteration Errors are the deviations found by the difference between the exact and iterative solutions of the algebraic equations and are normally referred to as round-off errors.

DISCUSSION

Computer science and numerical techniques provide an insight of the flow patterns that are deemed difficult, impossible and expensive using traditional experimental techniques. The simulations which involve prediction of flow phenomena using CFD software such as Solid Works are efficient and reliable. It is used for all desired quantities because it has a high resolution in both space and time. These simulations are done for the actual flow domain, and virtually any problem in the realistic condition is met. The sources of errors include modeling, iteration, implementation, and discretization. On the other hand, laboratory experiments describe a quantitative description of fluid flow by use of the measurements. The main disadvantage of this is that one can use only one quantity at a single time interval because

it has a limited number of points and time instances obtained. It is solely suitable for laboratory scale model and is limited to a wide range of the problems and certain operating conditions required. The source of errors includes measurement and flow disturbances. The flow patterns given out by the Computer Science and Numerical techniques have proved beyond any doubt to be less expensive as compared to the traditional techniques. This is actually an imperative role of technology in Fluid Mechanics.

CONCLUSION

The approaches used to solve problems in fluid mechanics include experimental methods, theoretical methods, and numerical computational methods. Computational and numerical methods are widely used because it has no restriction to linearity, can explain complicated physics and can handle a large Reynold's number. Simulation as a third pillar of science has its disadvantages including truncation errors, boundary condition problems, the need to incorporate mathematical model and of course the expensive computer and machinery. The role of informatics in computational science and engineering has been restricted to computers and a rather technical programming task. For instance, CFD highlights two conceptual and organizational patterns that are developed and optimized in computer science which aids the study of state-of-the-art flow simulations.

Computational fluid dynamics which at times goes by the acronym CFD is going through a critical expansion regarding disciplines and applications the number of researchers active in the field. The recent developments in the field are some software packages available that can solve fluid flow problems although its market is not that big as compared to the one for structural mechanic's codes. This is because CFD problems are considered to be more difficult to solve. However, it is noticeable that CFD codes and software are slowly being accepted as design tools by most of the industrial users. At the moment, users of CFD require to be more knowledgeable about this software which requires education for both students and those who are working in the field. The CFD Codes and the Software have become imperative design elements for you to consider in an industry, at the moment a high number of industries are considering this techniques when it comes to design purposes.

REFERENCES

Ahиландесwari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455-8907, 2(2): 1-5.

- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Cebal, J.R. and Löhner, R., 2001. From medical images to anatomically accurate finite element grids. *International Journal for Numerical Methods in Engineering*, 51(8), pp.985-1008.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Donea, J. and Huerta, A., 2003. *Finite element methods for flow problems*. John Wiley & Sons.
- Ferziger, J.H. and Peric, M., 2012. *Computational methods for fluid dynamics*. Springer Science & Business Media.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Laval, B., 2008. Numerical computation of internal and external flows: the fundamentals of computational fluid dynamics. *Canadian Journal of Civil Engineering*, 35(7), pp. 756-757.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Rajeswari, M., Amudhavel, J., Pothula, S. and Dhavachelvan, P., 2017. Directed Bee Colony Optimization Algorithm to Solve the Nurse Rostering Problem. *Computational intelligence and neuroscience*, 2017.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Tijms, H.C., 1994. *Stochastic models: an algorithmic approach* (Vol. 303). John Wiley & Sons Inc.
- Wesseling, P., 2009. The basic equation of fluid dynamics. In *Principles of Computational Fluid Dynamics* (pp. 1-52). Springer Berlin Heidelberg.

Shuffled frog-leaping algorithm SFLA for solving Load Balancing Problem through Fog Computing in Cloud Servers

T. Asvany^{1*}, J. Amudhavel² and Sujatha Pothula¹

¹Department of Computer Science, Pondicherry University, Puducherry, India

²Department of Computer Science and Engineering, KL University, AP, India

ABSTRACT

A memetic meta-heuristic based shuffled frog-leaping algorithm (SFLA) has mainly developed to solve combinatorial optimization problems. The main objective of proposed approach parameters are CPU Utilization Rate and Memory Rate. Cloud maintaining large numbers of applications, at a time but some limitation have load balancing of data and hosting of cloud data centers, occurs large network latency. Newly introduced Fog computing is located at network edge in cloud environment. Using SFLA algorithm to solve load balancing problem in a Fog based cloud computing. Based on the computational results benchmark problems are compared with existing algorithms demonstrate the proposed SFLA algorithm effectiveness.

KEY WORDS: BENCHMARK PROBLEMS, CLOUD COMPUTING, FOG COMPUTING, LOAD BALANCING, SHUFFLED FROG LEAPING ALGORITHM (SFLA)

INTRODUCTION

The essential need of the IT organizations a new challenging technical environment is very competitive growth in a world. The success of the internet usage in past years, the resource utilization of a system existed everywhere is known as cloud computing. Cloud computing is a famous technology, which works through the internet and maintain a large applications, where different organization clients shares information in one platform called cloud computing (Gayathri et al. 2016).

Cisco systems introduced a new concept is Fog computing, it is fundamentally extended or a new model of cloud computing environment. This Fog computing is mainly introduce for distributed environment in a wireless data objects, placing the data and cloud edge resources are hosting and running from a centralized cloud. Basically Fog computing systems operate from network ends (Shanmugam et al. 2017).

In use of fog technology the bandwidth gets reduced from cloud over network rather aggregated in data in some access points.

ARTICLE INFORMATION:

*Corresponding Author: asvanytandaani@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

Load balancing technique is mainly used to distribute loads on the servers. To distribute loads on individual nodes to minimize the response time and maximize the throughput. First the clients request is sent to the nearby Fog servers, which they are consistently used frequent data (Rajadurai et al. 2017). Suppose, the requested users didn't get the required resources in nearby Fog servers (who are communicating each other) finally the client request is forwarded to cloud servers. The load balancing technique is implemented using shuffled Frog Leaping Algorithm (SFLA).

LITERATURE SURVEY

In a cloud computing technique different kinds of load balancing algorithm (Scheduling) are used and collected data in a fixed time period. The scheduling algorithms are First Come First Serve (FCFS), Min – Max algorithm and Round Robin algorithm (Anusha et al. 2015). Verma et al. proposed a new scheduling technique for load balancing problem using in a Fog computing environment. They considered mainly utilization of network and maintaining the overall consistency at the same time to meet less complexity of end users (Verma et al. 2017).

Paulin Florence and Shanthy planned to implement the load balance problem in a firefly optimization algorithm. Here they maintain an index table for virtual servers for a client request sequence (Ahilandeswari et al. 2016). Pivot point concept and attribute to calculate from the derived formula (Dharshini et al. 2016). They achieve average time obtained by use of optimization algorithm is 0.934 ms. Bhujbal et al. (2015) proposed a method to solve the load balance problem in a cloud environment using Round Robin Algorithm (Govindharaj et al. 2016). Use of Round Robin algorithm it improves the overall performance and their efficiency. The proposed algorithm is very useful to reduce economic cost and fewer resources required for handling the user requests.

SYSTEM MODEL

The proposed approach is used to design a load balancing problem using Shuffled Frog Leap Algorithm. The main objective of load balancing problem is to maintain the CPU Utilization Rate and Memory Rate. The working procedure of the load balancing problem is consists of 'J' represents the number of client jobs requested to the server 'F'. The client first initiated the request to the Fog servers. Fog server forwarded the client machine request to cloud machine 'C', the cloud administrator checked the user is authenticated or invalid. The user is authenticated, forwarded the authentication back to Fog server to precede the client request.

The load balancing technique applicable to Fog servers $F = \{F_1, F_2, \dots, F_n\}$, because which Fog is free at this time and the users request $J = \{J_1, J_2, \dots, J_n\}$ finished quick in which Fog servers, depending on that SFLA is used. They considered two performance metrics represented as, $\tau_i = \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}$. The α represents CPU Utilization Rate and β represents Memory Rate. The total demand of all jobs can be represented as $\sum_{i=1}^{|J|} J_i \leq \sum_{i=1}^{|F|} F_i + \sum_{i=1}^{|C|} C_i$. The choice between cloud and Fog can be found with $P(F) > 0.08$ and $P(C) > 0.20$ where P represents the probability factor chosen from the proposed work. Finally the Fog sends the response to the particular user.

SHUFFLED FROG – LEAPING ALGORITHM

The shuffled frog-leaping algorithm designed as meta-heuristic concept used to perform a heuristic search using an informed heuristic function to search for a solution of combinatorial optimization algorithm. Based on memes evolution they carried by interactive individuals, and globally exchange information along with themselves.

In SFLA, initial position is formed a generated solutions (randomly) called virtual frogs represented by vector. Whole population is partitioned into m subsets called memplexes. Different types of memplexes, contains n frogs, considered as different cultures of frogs develop independently in the search space. Each memplex is sorted so frogs arranged in decreasing performance ($i = 1, \dots, n$). In each memplex subset called submemplex is constructed. Submemplex contains s frogs randomly selected from memplex to their triangular probability distribution. i.e. $P_i = \frac{2(n+1-i)}{n(n+1)}$, $i = 1, \dots, n$. each submemplex, worst frog P_w will leap to its own experience and experience from best frog P_b in this memplex.

The leaping step size equation as follows

$$S = \begin{cases} \min\{\text{int}[\text{rand}(P_b - P_w)], S_{max}\} \\ \max\{\text{int}[\text{rand}(P_b - P_w)], -S_{max}\} \end{cases} \quad (1)$$

Where rand is a random real number between 0 and 1, and S_{max} is maximum leaping step size. The worst frog new position is calculated by

$$P_n = P_w + S_b \quad (2)$$

The new position is best the old one replace P_w with P_n ; else calculate another new position P_{na} of frog adjust according to better frog P_g in entire population with following leaping step size

$$S_g = \begin{cases} \min\{\text{int}[\text{rand}(P_g - P_w)], S_{max}\} \\ \max\{\text{int}[\text{rand}(P_g - P_w)], -S_{max}\} \end{cases} \quad (3)$$

$$P_{na} = P_w + S_g \quad (4)$$

```

Algorithm : Shuffled Frog-Leaping Algorithm
Begin
Parameter Initialization
Initialize the parameter values F, m, n, s,  $E_p$ ,  $I_s$ ;
( $F = m, n$  and  $s \leq n$ )
Population Initialization
Initialize random (solutions) positions of F frogs;
(For all activities a solution is a vector of start times)
Evaluate Solution
For each frog;
    Calculate its performance;
    (for each performance is taken as the objective function value)
    {
Sort the frogs F in descending order of performance values;
(Re-label the frogs such that frog with best performance has label 1 and the another one with least performance has label F)
Let  $G_b$  be the global best position;
Divide the population into m memplexes each has n frogs;
Frog i is assigned to memplex r where  $r = j \text{ mod } (m)$ 
    {
For each memplex;
    Re-label the frogs 1,2,..., i,..., n
Assign probability to frog
 $i = P(i) = 2(n + 1 - i)/(n(n + 1))$ 
Construct a sub-memplex of s frogs selected randomly according to the assigned probabilities;
    {
For the sub - memplex;
Let  $P_b$  and  $P_w$  be the position of the best and worst frog in this sub-memplex;
Calculate the new position of the worst frog in each sub-memplex by Eqs. (1) and (2);
Set it = 0;
For (it < SFLA_N)
If new position  $P_n$  is not better than  $P_w$ 
Calculate the new position of the worst frog by Eqs. (3) and (4);
If another new position  $P_{na}$  is not better than  $P_w$ 
Randomly generate a new frog to instead the old one;
Else
 $P_w = P_{na}$ ;
Else
 $P_w = P_n$ ;
it++;
shuffle all the memplex;
    }
Repeat for  $E_i$  iterations;
End
    
```

FIGURE 1. Pseudo code of SFLA algorithm

The new position P_{na} is best as old one; replace P_w with P_{na} else worst frog is replaced by a randomly generated frog. After number of iterations, frogs are shuffled and divided into m new memplexes. Such that frog leaping

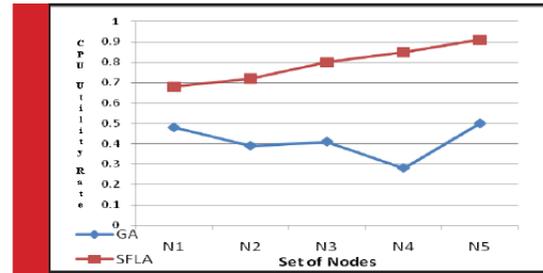


FIGURE 2. Comparison of CPU Utilization Rate

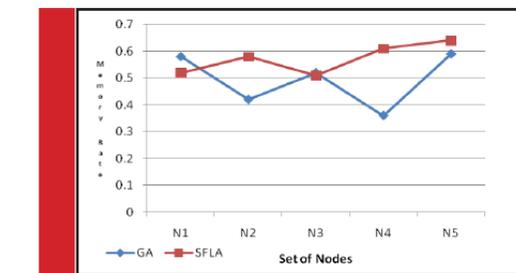


FIGURE 3. Comparison of Memory Rate

and shuffling processes are repeated until the stopping condition is satisfied. The following Figure 1 represents the pseudo code of Shuffled Frog Leaping Algorithm.

EXPERIMENTAL RESULTS

The proposed work is implemented using SFLA algorithm in cloud computing network, the interaction between the cloud layer and client layer is Fog computing. The load balancing problem is to allocate the jobs to cloud and Fog is balanced by SFLA algorithm. The code is implemented in MatLab 9.1 and system running on core i5 processor.

The Fig 2 and Fig 3 show the comparative result analysis of the proposed system with existing technologies. The existing load balanced technique Xin et al. results are compared in this proposed work. The Fig 2 represents, CPU utilization rate is more efficient when considering load balanced approach compared to existing technique.

CONCLUSION

The load balancing problem is implemented in SFLA algorithm to balance the jobs between cloud layer and client layer communication through Fog layer is introduced the efficiency of cloud environment by utilization of bandwidth is reduced from client. After implementing the proposed approach it satisfies CPU Utilization Rate and maintains Memory Rate.

REFERENCES

- Ahиландесwari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 2(2): 1-5.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Bhujbal, A., Jakate, P., Wagh, M., Pise, M. and Marathe, M.V., 2015. Load Balancing Model in Cloud Computing. *International Journal*, 1.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- El-Sayed, T., El-kenawy, A.I.E.D. and Mohamed, F., 2012. Al-rahamawy "Extended Max-Min Scheduling Using Petri Net and Load Balancing" *International Journal of Soft Computing and Engineering (IJSCE)* ISSN: 2231-2307. Volume-2, Issue-4, September.
- Florence, A.P. and Shanthi, V., 2014. A load balancing model using firefly algorithm in cloud computing. *Journal of Computer Science*, 10(7), p.1156.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Lu, X., Zhou, J. and Liu, D., 2012. A method of cloud resource load balancing scheduling based on improved adaptive genetic algorithm. *Journal of Information & Computational Science*, 9(16), pp.4801-4809.
- Marphatia, A., Muhnot, A., Sachdeva, T., Shukla, E. and Kurup, L., 2013. Optimization of FCFS based resource provisioning algorithm for cloud computing. *Journal of Computer Engineering (IOSRJCE)*, pp.2278-0661.
- Pawar, C.S. and Wagh, R.B., 2012, December. Priority based dynamic resource allocation in cloud computing. In *Cloud and Services Computing (ISCOS), 2012 International Symposium on* (pp. 1-6). IEEE.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Verma, M., Bhardwaj, N. and Yadav, A.K., 2016. Real Time Efficient Scheduling Algorithm for Load Balancing in Fog Computing Environment. *International Journal of Information Technology and Computer Science (IJITCS)*, 8(4), p.1.

Solving Virtual Machine placement in Cloud data centre based on Novel Firefly algorithm

T. KalaiPriyan*¹, J. Amudhavel² and Sujatha Pothula³

^{1,3}Department of Computer Science, Pondicherry University, Puducherry, India

²Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

ABSTRACT

Allocation of Virtual machines (VM) into optimal physical machines (holds the capability of serving for more than one VM) with less power consumption and wastage is the key issue in cloud service centers. In this paper, novel Firefly algorithm is used for solving VM placement in cloud data centers. The objectives addressed in this paper are wastage modelling and power consumption. Simulation is made for different data instances and the results are tabulated.

KEYWORDS: CLOUD COMPUTING, EVOLUTIONARY ALGORITHM, FIREFLY ALGORITHM, GREEN COMPUTING, VIRTUAL MACHINE PLACEMENT

INTRODUCTION

Cloud computing provide access to remote servers and is steadily evolving in the era of computer science. Cloud provides resource scalability and utilization, hence it requires more number of data centers which increases energy consumption, network usage, bandwidth allocation and operational cost. The process of selecting virtual machine to be placed in physical machine is Virtual Machine Placement (VMP). In recent, years green computing is an another paradigm which plays a vital role in utilizing the resources of cloud computing in an efficient manner. The process of VM placement turned out to be an optimization problem as the number and size of VM

requests varies with respect to user's utilization. In the past few decades when it comes to be an optimization problem there comes the bio-inspired algorithms which plays a significant role in solving optimization problems with less computational time. Firefly Algorithm is one among the evolutionary algorithm which is inspired from the attracting nature of fireflies. In this paper, this Firefly algorithm has been used to solve VM placement in cloud centers. The rest of the paper is organized as follows: section 2 addresses the related work with respect to VM placement. Section 3 holds the System model and problem definition. Section 4 explains the firefly algorithm. Section 5 holds experimental evaluation and section 6 concludes the work.

ARTICLE INFORMATION:

*Corresponding Author: kalaip27@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

LITERATURE WORK

The authors proposed methodology to reduce network traffic and consumption of power in data center by placing VM. For scheduling tasks, the author designed service in PaaS. They proposed modified Particle Swarm Optimization algorithm to ensure guaranteed QoS tasks by migrating VM (Dashti and Rahmani 2016). The migration and energy consumption of VM are minimized using First-Last algorithm. They treated this problem as bin packing algorithm and tested on various heterogeneous VM. First fit algorithm is used to place VM in host node, best fit is used to select the minimum resource wastage. Next fit algorithm is to place VM in active host (Kumar et al., 2016).

To develop SLA aware VM placement with minimized energy based on utility functions. Genetic algorithm will maximize the search operation in utility function (Mosa and Paton 2016). VM machine are responsible in placing different types of jobs within completion time. The author proposed methodology to minimize the job completion time using off-line VM placement method. Migration algorithm is used to migrate from VM to best PM directly (Li et al., 2015). Pauling Florence and Shanthi proposed cloud computing environment to balance load from clients (Florence and Shanthi 2014). They adapted Firefly algorithm (Lukasik and Zak 2009) to schedule job in cloud network. The proposed algorithm does population generation, indexing and priority based node selection. Yang and Xin-She (Yang 2010) proposed fuzzy based Firefly optimization algorithm to solve load balancing in Cloud environment. They created load balancing in partitioned environment to efficiently utilize the resources. Some of the other applications of recent advancements in these environments include (Anusha et al. 2015, Dharshini et al. 2016, Govindharaj et al. 2016, Ahilandeswari, et al. 2016, Gayathri et al. 2016, Shanmugam et al. 2017, Rajadurai et al. 2017 and Sasidevi et al, 2017)

SYSTEM DESIGN

Let PM denotes the physical machines and N denotes number of PM in cloud and the complete set can be represented as $PM = \{pm_1, pm_2, \dots, pm_i, \dots, pm_N\}$. VM denotes virtual machines requested by the users can be represented as $VM = \{vm_1, vm_2, \dots, vm_j, \dots, vm_M | M \leq N\}$ where M denotes the total number of VM's.

Let us assume $\forall i \in N$ the given resources of CPU and Memory are same and it can be represented as $R(pm_i) = \begin{bmatrix} CPU \\ MEM \end{bmatrix} \forall i \in N$. The resources requested by each VM differs in terms of CPU utilization.

Memory demand and this can be represented as $R(vm_j) = \begin{bmatrix} CPU_j \\ mem_j \end{bmatrix} \forall j \in M$. The total VM's that a PM can hold

can be represented as $\sum_{j=1}^g R(vm_j) \leq R(pm) | g \subseteq |VM|$ where $||$ represents the cardinality measure. $(S_{ij})_{N \times M}$ is the matrix with size $N \times M$ which denotes the allocation of virtual machines to respective physical machine. S_{ij} holds the value 1 if j^{th} virtual machine is allocated to i^{th} physical machine else 0.

Problem definition

The task is to assign N virtual machines within M physical machines with respect to the objectives and the objectives are listed below

Resource Wastage

$$RW_i = \frac{Ut_i^{CPU} - Ut_i^{RAM}}{Max_i^{CPU} + Max_i^{RAM}}$$

where Ut denotes demand of CPU and RAM requested by the virtual machine which has to be allocated to physical machine i and Max denotes the maximum CPU and RAM resources that can be allocated to VM.

The formulated objective for resource wastage is

$$\min \sum_{i=1}^M RW_i$$

where N denotes the number of physical machines.

Power Consumption

The consumed power by a machine can be calculated as

$$PW_i = (PW_i^{awake} - PW_i^{idle}) \times Max_i^{CPU} + PW_i^{idle}$$

The formulated objective for consumed power by all the machines is

$$\min \sum_{i=1}^N PW_i$$

NOVEL FIREFLY ALGORITHM

The attractiveness of the firefly is calculated based on the brightness of light. Thus light intensity and attractiveness are monotonically decreasing function. Gaussian function is used to calculate square law and absorption using

$$I(r) = I_0 e^{-\gamma r^2} \quad (1)$$

where I and I_0 are the light intensity and original light intensity, γ is light absorption coefficient which is constant. The attractiveness of the light depends on the light intensity using

$$\beta(r) = \beta_0 e^{-\gamma r^2} \quad (2)$$

where β_0 is constant and provide attractiveness with constant $r = 0$. Cartesian distance is used to calculate the distance between two fireflies a and b by

$$r_{i,j} = |x_i - x_j| \quad (3)$$

The firefly i movement is attracted to firefly j , attractiveness is calculated by

$$\Delta x_i = \beta_0 e^{-\gamma r_{ij}} (x_j^t - x_i^t) + \alpha \varepsilon_i \quad (4)$$

$$x_i^{t+1} = x_i^t + \Delta x_i \quad (5)$$

where a is the randomization parameter with ε as constant vector from Gaussian distribution. Yang [8] replaced randomized parameter with Levy flight and the step size for levy flight is calculated by

$$L(s) = A s^{-(1+\lambda)} \quad (6)$$

$$A = \lambda \Gamma(\lambda) \sin\left(\frac{\lambda\pi}{2}\right) \frac{1}{\pi} \quad (7)$$

where $\Gamma(\lambda)$ is Gamma function with λ as exponent of the distribution. The annealing scheduling for initial α_0 is

$$\alpha = \alpha_0 \theta^t \quad (8)$$

where $0 < \theta < 1$ is random constant.

EXPERIMENTAL EVALUATION

The given system model has been implemented in MATLAB 9.1 with the configured system Intel core i7 with the processor speed of 3.2GHz and 4GB RAM. For generating the sample values of CPU and Memory demand, a correlation coefficient factor for the reference values of 0, 0.25, 0.5, 0.75 and 1. The memory utilization for CPU and Memory by any VM's are restricted to 25% and 45% in the simulation setup. Parameter settings for the

Table I. Results for 100 node instances			
CPU & MEM Ut.	Corr.	Power (W)	Wastage
25%	-0.748	6310.33	20.48
	-0.359	6289.88	17.73
	-0.051	8353.34	19.01
	0.38	7235.91	18.84
	0.748	7748.61	9.86
45%	-0.754	11314.86	27.12
	-0.379	11931.8	26.58
	-0.053	13227.02	32.17
	0.389	12407.08	28.94
	0.751	11002.29	23.75

Table II. Results for 100 node instances			
CPU & MEM Ut.	Corr.	Power (W)	Wastage
25%	-0.748	11049.46	6.76
	-0.359	11542.19	4.57
	-0.051	13539.21	7.66
	0.38	12305.52	2.25
	0.748	12249.57	5.19
45%	-0.754	21733.15	13.02
	-0.379	22867.22	8.78
	-0.053	24482.46	10.36
	0.389	22893.83	9.55
	0.751	21848.95	12.29

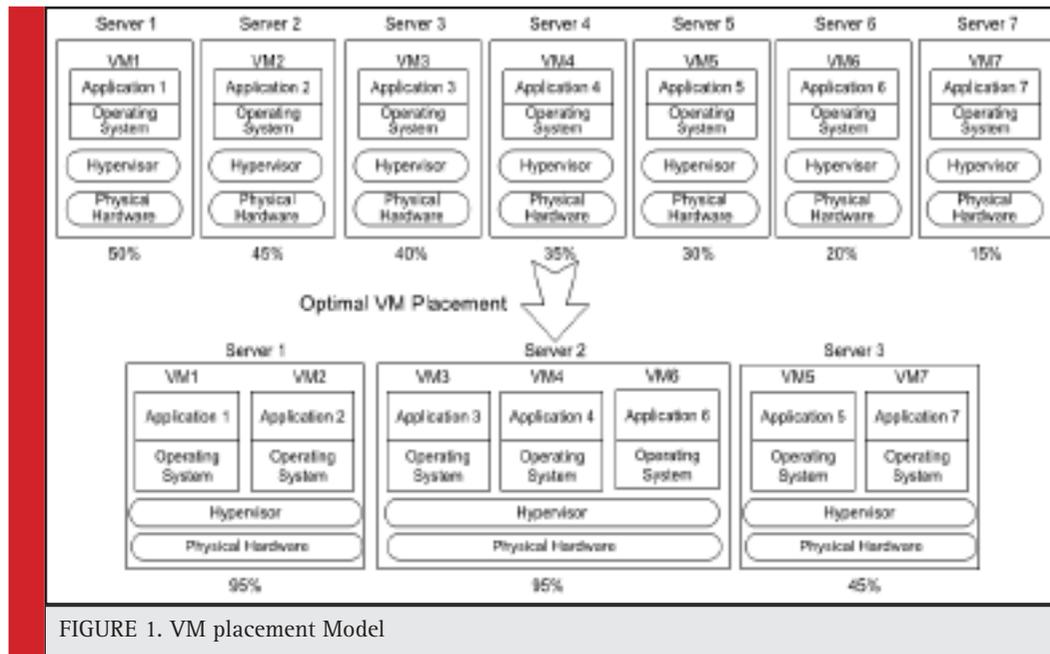


FIGURE 1. VM placement Model

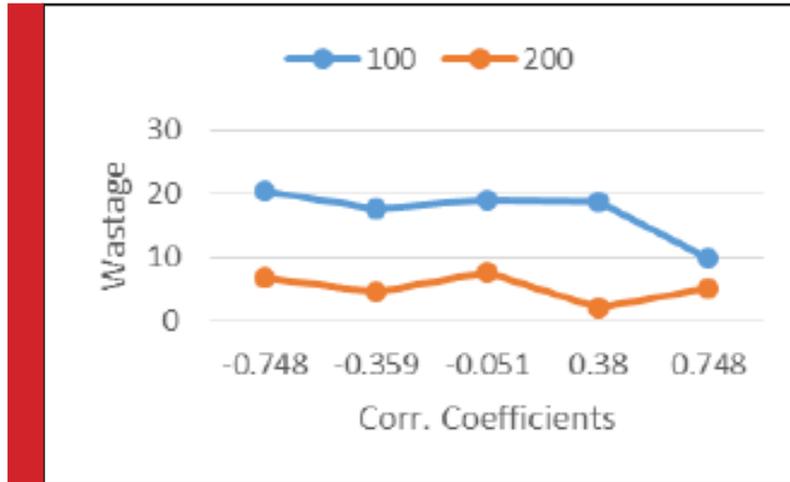


FIGURE 2. Comparison of wastage for CPU=RAM=25%

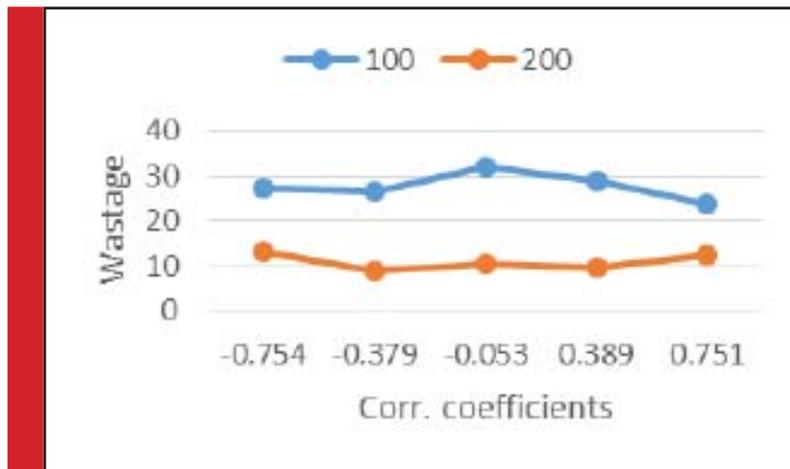


FIGURE 3. Comparison of wastage for CPU=RAM=45%

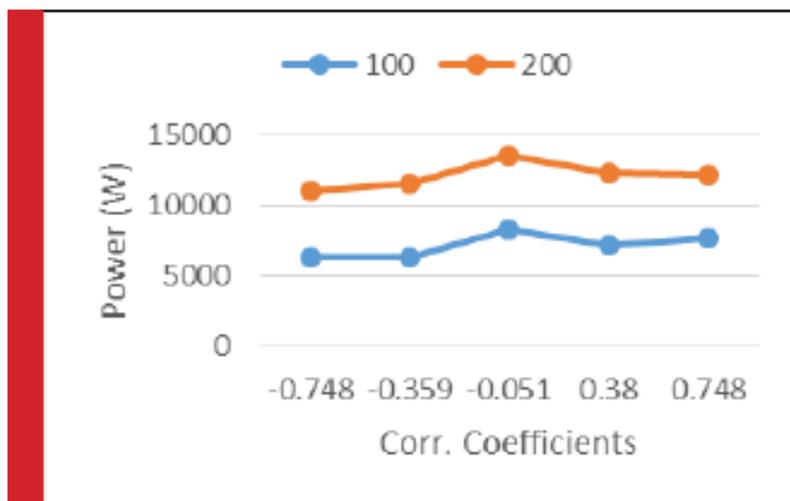


FIGURE 4. Comparison of power consumption for CPU=RAM=25%

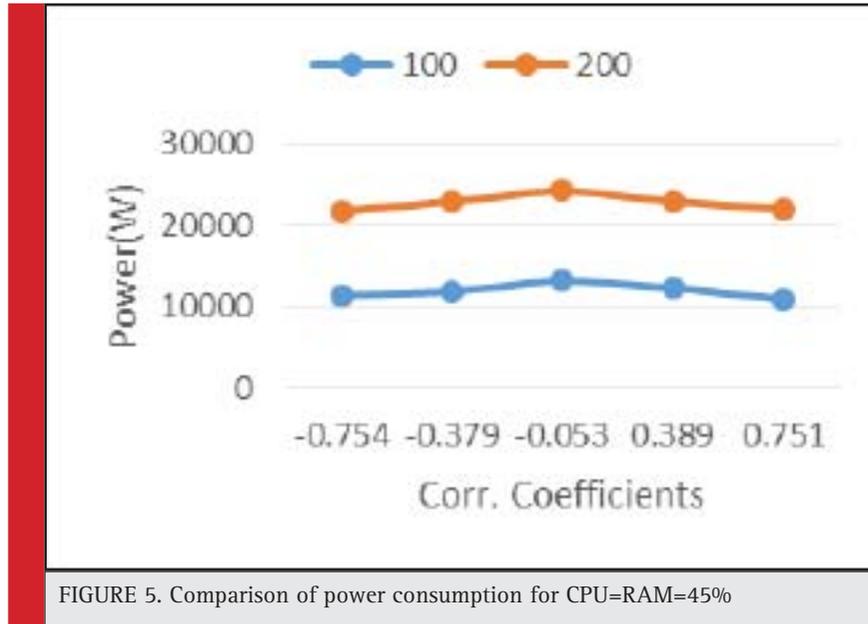


FIGURE 5. Comparison of power consumption for CPU=RAM=45%

<p>Algorithm 1: Pseudo code for Firefly Algorithm</p> <p>Pseudocode For Firefly Algorithm</p> <p>Input: $R(PM)$, $R(vm_j) \forall j \in M$, objective function $f()$</p> <p>Initialize the population $Pop_t = \{X_1, X_2, \dots, X_s\}$ where $X_i = [X_{i,1}, X_{i,2}, \dots, X_{i,M}]$ where t is the generation number, s is the population size.</p> <p>Define the fitness function $f(X_i) \forall i \in N$</p> <p>$\forall i \in Pop_t$ do $I_i \leftarrow f(X_i)$ end for</p> <p>Initialize γ, β_0, λ</p> <p>while ($t < MaxIT$) do $\forall i \in Pop_t$ do $\forall j \in Pop_t$ do if ($I_i > I_j$) then Compute $r_{i,j} \leftarrow \sqrt{(X_{i,k} - X_{j,k})^2}$ Compute ΔX using Eq. (4) Update the position of firefly i using Eq. (5) end if Evaluate $f(X_i)$ end for end for $(S_{i,j})_{N \times M} \leftarrow \min f(Pop_t)$ $t \leftarrow t + 1$ end while</p> <p>Output: $(S_{i,j})_{N \times M}$</p>
--

given algorithm are $\gamma = \frac{1}{D}, \beta_0 = 1, \lambda = 3/2$ and Υ represents gamma function.

Table I refers the resultant values for 100 node instances and Table II holds the results of 200 node instances for the CPU and RAM utilization of 25% and 45% respectively. The results are calculated in terms of the power consumed by all nodes and total wastage that the dataset made. Figure 2-5 compares the results of

Table I and II with respect to different correlation coefficients.

Table I refers the resultant values for 100 node instances and Table II holds the results of 200 node instances for the CPU and RAM utilization of 25% and 45% respectively. The results are calculated in terms of the power consumed by all nodes and total wastage that the dataset made. Figure 2-5 compares the results of Table I and II with respect to different correlation coefficients.

CONCLUSION

VM placement in cloud data centers are the major concern which comes under combinatorial optimization. In this paper in order to address the VM placement in the physical machines with minimal power consumption and wastage, novel Firefly algorithm has been used. The results tabulated in table I and II holds for different correlation coefficients. The future enhancement of this work can be extended to solve more number of objectives in this addressed problem.

REFERENCES

- Dashti, S.E. and Rahmani, A.M., 2016. Dynamic VMs placement for energy efficiency by PSO in cloud computing. Journal of Experimental & Theoretical Artificial Intelligence, 28(1-2), pp.97-112.
- Florence, A.P. and Shanthi, V., 2014. A load balancing model using firefly algorithm in cloud computing. Journal of Computer Science, 10(7), p.1156.
- Kumar, A., Sathasivam, C. and Periyasamy, P., 2016. Virtual machine placement in cloud computing. Indian Journal of Science and Technology, 9(29).

- Li, K., Zheng, H., Wu, J. and Du, X., 2015. Virtual machine placement in cloud systems through migration process. *International Journal of Parallel, Emergent and Distributed Systems*, 30(5), pp.393-410.
- Lukasik, S. and Zak, S., 2009, October. Firefly Algorithm for Continuous Constrained Optimization Tasks. In *ICCCI* (pp. 97-106).
- Mosa, A. and Paton, N.W., 2016. Optimizing virtual machine placement for energy and SLA in clouds using utility functions. *Journal of Cloud Computing*, 5(1), p.17.
- Yang, X.S., 2010. Firefly algorithm, Levy flights and global optimization. *Research and development in intelligent systems XXVI*, pp.209-218.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Ahilandeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 1-5.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 22-29.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 3(1): 7-21.

A novel NM-ABC algorithm to solve Economic Power Dispatch Problem with Transmission Loss and Valve Points

M. Rajeswari^{1*}, J. Amudhavel² and P. Dhavachelvan³

^{1,3}Department of Computer Science, Pondicherry University, Puducherry, India

²Department of Computer science and Engineering, KL University, Andhra Pradesh, India

ABSTRACT

This work presents an investigation on meta-heuristic approach which integrates Artificial Bee Colony (ABC) algorithm with Nelder-Mead method (NM-ABC) to solve Economic Power Dispatch (EPD) problem. EPD is a optimization problem to provide shared demand among power generating units by minimizing generating cost without violating constraints. Nelder-Mead (NM) method is used as local search in the onlooker bee phase of ABC algorithm to enhance the process of intensification in search strategy. Incorporation of NM method at onlooker bee phase improve solution strategy of the proposed algorithm NM-ABC. To validate the efficiency of the proposed algorithm NM-ABC, large scale bench mark test unit forty-unit system is taken for simulation. To show the superiority of the proposed algorithm, the simulation results are compared with other meta-heuristic algorithm.

KEY WORDS: ARTIFICIAL BEE COLONY ALGORITHM, ECONOMIC POWER DISPATCH PROBLEM, EVOLUTIONARY ALGORITHM, NELDER-MEAD METHOD, POWER SYSTEM

INTRODUCTION

Economic power dispatch (EPD) problem of power generating unit plays significant role in electrical power system industry. Economic dispatch (ED) is an optimization problem and provides optimal distribution of total power demand over generating units by minimizing generation cost and operational constraints. ED is one

of the fundamental optimization problems in electrical power system which is characterized by reliability, security, stability and economy (Yoshida et al., 2000; Aravindhababu and Nayar 2002; Chen et al., 1993; Dodu et al., 1972). EPD is complex and non-linear non-convex optimization problem with ramp rate limits, prohibited operating zones, valve point effects and transmission fuels characteristics with equality and inequality

ARTICLE INFORMATION:

*Corresponding Author: raji.rajeswari18@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

constraints makes challenging one (Anusha et al. 2015; Shanmugam et al. 2017; Rajadurai et al. 2017).

EPD is the process of allocating output power generation to the generation units so as to distribute power load which will minimize the total generation fuel cost while satisfying constraints (Selvakumar et al., 2008). Each generation unit in power system produces unique generation cost influenced by its fuel cost coefficients (Gayathri et al. 2016; Sasidevi et al. 2017). ED in supplying total power load will determine the optimal power generation of the generating units and thus the total power generation and total power demand at the station should be equal (Sinha et al., 2003).

Artificial Bee Colony algorithm is inspired from the foraging behavior of the honey bees and is used to solve complex non-convex functions (Karaboga et al., 2007; Basturk et al., 2006). Hemamalini et al. (2006) solved ELD problem using ABC algorithm by considering non-smooth functions, valve effects, ramp rate limit, multiple fuel and prohibited zones. The performance of the algorithm is tested using 10, 13, 15 generator units with non-linear cost functions. The algorithm is compared with other population based optimization algorithms like Artificial Immune system (Vanaja et al., 2008), Genetic Algorithm and PSO to prove its efficiency. Directed Bee colony optimization (Kumar et al., 2012) is proposed and used to solve multi-objective emission constrained EPD problem. The algorithm is composed of multi-agent system, deterministic search and bee decision making process. This hybridization helps to obtain unique and fast convergence towards optimal solution (Dharshini et al. 2016). The algorithm is tested on a standard IEEE 30 bus six generator and the results are compared with existing algorithms like multi-objective stochastic search and linear programming and non-dominated sorting genetic algorithm (NSGA).

Labbi et al. (2014) proposed a variant of ABC algorithm to solve ED problem in power systems with valve points. The new ABC variant is applied to ED problem with incremental fuel cost function and valve point effects for 3, 13 and 40 generator units (Govindharaj et al. 2016). The simulation result reveals the superior performance of ABC algorithm to solve ED problem with valve point effects (Govindharaj et al. 2016). ABC algorithm is applied to solve EPD problem with pollution control effects (Slimani et al., 2013). The authors considered emission control with minimizing the fuel cost of the generation unit. ED problem is solved with the hybridization of Lamda method and bee colony optimization algorithm (Khamseen et al., 2016). The authors have considered multiple fuel cost factors and power loss in solving ED problem. The performance of the algorithm is evaluated in terms of computational time, quality of the solution and convergence rate (Rajeswari et al., 2017).

ABC algorithm is flexible, provide memory management and it utilizes search space for multiple decision making process. It comprises of balanced operator which helps to improve the solution and speed up the searching process. ABC algorithm requires high computational time however with the help of operators used and hybridization with local search algorithm results fast computation (Zhao et al., 2010; Ozturk et al., 2015; Abbas et al., 2014).

To increase the convergence speed and overcome premature convergence, the authors propose a hybrid algorithm which integrates the advantage of ABC and local search method NM. The proposed algorithm is named as Nelder-Mead Artificial Bee Colony algorithm (NM-ABC). The local search strategy is incorporated in the onlooker bee phase of ABC to obtain optimal solution in solving EPD. However, ABC follows same strategy of neighborhood search on employed bee and onlooker bee phase. Improving neighborhood search strategy on individuals increase the probability to obtain global optimum compared with random process. Thus in our work, we introduced NM as local search on onlooker bee phase to enhance the process of neighborhood search. NM-ABC is formulated to solve EPD problem and tested on three different test systems a ten-unit problem, a thirteen-unit problem and forty-unit problem. The simulation results are compared with recent existing algorithms.

ECONOMIC POWER DISPATCH PROBLEM

EPD problem in power system is used to calculate the optimal combination of generating units for power outputs. The objective of the EPD is to calculate the minimum total fuel cost in power system by satisfying the various equality constraints, inequality constraints and power balance (Yuan et al., 2009). EPD problem is either static or dynamic, static allows constant power with minimum cost and the generation cost depends on regular evaluation of scheduled period and it varies according to customer demand. While dynamic EPD are based on the constraints like ramp rate to minimize the generating cost (Xia, X et al., 2010). The objective of EPD is to minimize the generating cost of the power unit. The cost includes the summation of N generating unit cost and it is formulated as

$$\text{Minimize } X_c = \sum_{i=1}^N X_i(U_i) \quad (1)$$

where X_c is the cost of the generating power unit, N is the total number of generator and U_i is the power of the i^{th} generator.

The cost of the generating unit can be represented in the form of polynomial function with α , β , γ as coefficient of the generator.

$$X_i(U_i) = \alpha_i U_i^2 + \beta_i U + \gamma_i \quad (2)$$

The search space of the power unit becomes non-smooth and higher-order non-linear sinusoidal function and it is added with valve point effects represented as

$$X_i(U_i) = \alpha_i U_i^2 + \beta_i U + \gamma_i + |x_i \sin(\gamma_i (U_i^{\min} - U_i))| \quad (3)$$

where V_i is the higher-order nonlinearity and non-smooth sinusoidal function caused by value point effects and x_i, γ_i are the coefficients of the i^{th} generator, U_i^{\min} is the minimum limit of the generating unit.

The mathematical representation of the objective function of EPD with valve points is represented as

$$\text{Minimize } X_G = \sum_{i=1}^N \alpha_i U_i^2 + \beta_i U + \gamma_i + |x_i \sin(\gamma_i (U_i^{\min} - U_i))| \quad (4)$$

To solve EPD, both equality and inequality constraints should be satisfied for generating unit and power balance,

Power balance

The total generating power can be calculated with the summation of total power demand and transmission loss in power unit as

$$\sum_{i=1}^N U_i = U_D + U_{TL} \quad (5)$$

B coefficient matrix is used to calculate transmission loss U_{TL} and it is represented mathematically by

$$U_{TL} = \sum_{i=1}^N \sum_{j=1}^N U_i B_{ij} U_j + \sum_{i=1}^N B_{oi} U_i + B_{oo} \quad (6)$$

where B_{oo}, B_{ij} and B_{oi} are the loss coefficient constant, ij^{th} loss coefficient of the matrix B, and i^{th} element of loss coefficient vector respectively.

Generating capacity

The power unit should satisfy minimum and maximum generating capacity and it is formulated as

$$U_i^{\min} \leq U \leq U_i^{\max} \quad (7)$$

where U_i^{\min} is the minimum limit and U_i^{\max} is the maximum limit of generator i .

PROPOSED ALGORITHM

I. Artificial Bee Colony Algorithm

ABC is a population-based optimization algorithm inspired by the foraging behavior of the honey bees (Karaboga et al., 2007). The intelligent behavior of honey bees helps to find the global optimal solution for the optimization problem. ABC consists of three intelligent groups of honey bees namely employed bee, onlooker bee, and scout bee. The entire colony of honey bee consists of an equal number of employed and onlooker bees and each solution are held by one employed bee in the population. Employed are responsible for searching in

the direction of the food particle and share the information about the food particle position through waggle dance. Onlooker bee selects the higher quality food particle based on the probability calculation and continues for further search in search space. The remaining low-quality food particle is abandoned and converted to scout bee. The scout bee starts its search towards newer food particle.

The initialization of solution in population is generated with food particle FP and represented as $X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,n}\}$. The food particles are randomly distributed to employed bees. The fitness calculation for the solution is evaluated accordingly

$$x_{i,j} = x_{\min,j} + \text{rand}(0,1) * (x_{\max,j} - x_{\min,j}) \quad (8)$$

The candidate solution is generated in the employed bee phase and the positions of food particle are monitored. The candidate solution can be generated using

$$v_{i,j} = x_{i,j} + \phi_{i,j} (x_{i,j} - x_{k,j}) \quad (9)$$

Where $j = 1, 2, \dots, S$ and we chose the value of k different from $i, k = 1, 2, \dots, FP$, value of ϕ ranges $[-1, 1]$. The greedy selection to select among candidate solution v_i and x_i is based on the fitness calculation. If the fitness value of v_i is higher than x_i , the solution x_i is replaced by v_i . After the fitness calculation, the employed bees share the information about the direction of food particle with onlooker bees. The onlooker bees evaluate the fitness of the employed bees solution using probability value p_i . The probability calculation can be generated using

$$P_i = \frac{fit_i}{\sum_{j=1}^{FP} fit_j}$$

$$fit_i = \begin{cases} \frac{1}{1 + f_i} & f_i \geq 0 \\ 1 + abs(f_i) & f_i < 0 \end{cases}$$

Where fit_i is the fitness of the solution x_i , f_i denotes the objective function value of the solution x_i . The onlooker bee randomly chooses food particle x_i based on the probability value p_i . The onlooker bee computes candidate solution using Eq. (9) and to select the best solution among x_i and v_i greedy method is used. If the solution is cannot be further improved and exhausted for the predefined number of iterations, the food particle is abandoned. The corresponding employed bee associated with the solution becomes scout bee and started to search for new food particle using Eq. (8).

II. Nelder-Mead Method

Nelder-Mead is a local search method and it is represented in the simplex triangle with three vertices (Nelder et al., 1965). For each iteration, the poor performing vertex is found and replaced with the new vertex in the

simplex triangle. The search is traced to obtain the best vertex and the size of the triangle converges towards the optimal solution. The vertices are ordered based on the fitness value: best vertex (B), good vertex (G) and worst vertex (W). Let $f(x,y)$ is the function to be minimized and NM works on four process namely reflection, expansion, contraction and shrinkage.

Midpoint (M)

The midpoint between line joining two best and good vertices are calculated using

$$M = \frac{B+G}{2} \quad (10)$$

Reflection (R)

The reflection point R is chosen along the side of \overline{BG} . The point R is found away from the worst vertex using

$$R = 2M - K \quad (11)$$

Expansion (E)

The fitness calculation at vertex R is calculated and if it is less the worst vertex then move towards minimum value. Extend the line segment from R to E by the distance d. The fitness value at vertex E is less than vertex R, then it is moved towards optimal value. The expansion point can be found using

$$E = 2R - M \quad (12)$$

Contraction (C)

When the fitness value for R and W are not minimum and need another vertex to continue the process. Then contraction towards midpoint without replacement is done. The contraction points C_1 and C_2 are drawn along the line joining \overline{WM} and \overline{MR} for the length of $d/2$. The formula to calculate C is

$$C = R + M \quad (13)$$

Shrinkage (S)

The fitness value at C is calculated and if it not less than, then the vertices G and W should be shrunk towards best vertex B. The vertex G is replaced with M and worst vertex W is replaced with S. The point S is the midpoint of and W. The process is continued until the minimum value is found.

III. ABC algorithm with NM method (NM-ABC)

ABC algorithm is flexible and the complexity of the algorithm is reduced since it requires fewer control parameters. To obtain the global solution with less computational time, the algorithm should be good at both exploration and exploitation process. However, ABC is

good at exploration process but fails in exploitation process. The local search strategy is imposed on ABC to maintain the balance of exploration and exploitation process. The NM method is population local search algorithm which requires fewer functions when compared to other meta-heuristic algorithms. But, this method gets entrapped in local optima and shows poor at diversification process. To deal exploitation problem in ABC, NM method is incorporated as local search in our proposed algorithm. The NM method is embodied in onlooker bee phase of ABC algorithm. The local search strategy is incorporated in onlooker bee phase than employee bee phase since the individuals in onlooker bee phase are selected based on probability calculation. Thus intensive search results in global optima rather than using random individuals. The work flow of our proposed NM-ABC algorithm is shown in Figure 1.

EXPERIMENTAL DESIGN AND ANALYSIS

I. Experimental Environment setup

The performance of the proposed algorithm on solving EPD problem are tested using large scale benchmark test systems forty-unit system (Sinha et al., 2003) is considered. The experiments are implemented to solve EPD in MATLAB 8.3 in the system configuration of Intel Core i7 Processor with 3.2 GHz speed and 4GB RAM. The parameter settings for implementation are shown in Table I. The empirical evaluations will set the parameters of the proposed system. Appropriate parameter values are determined based on the preliminary experiments.

II. Case Study

Test case contains forty-unit system with valve point effects and the data have been obtained from (Sinha et al., 2003). The overall power demand for the system is 10500MW. Transmission loss for 40 generation power systems is neglected for comparison purpose of the proposed algorithm NM-ABC. The performance of NM-ABC to solve forty-unit system are compared with the variants of ABC algorithm like IABC (Tsai et al., 2009), CABABC (Gao et al., 2013), LABABC (Rajasekhar et al., 2011), MABC (Gao et al., 2012) and standard ABC algorithm. On comparing the simulation results of the proposed NM-ABC algorithm outperforms other existing variants of ABC in terms of fuel cost on solving the problem are shown in Table II.

CONCLUSION

This paper solves non-convex economic power dispatch problem using NM-ABC algorithm. To solve EPD efficiently ABC algorithm is chosen for global search and

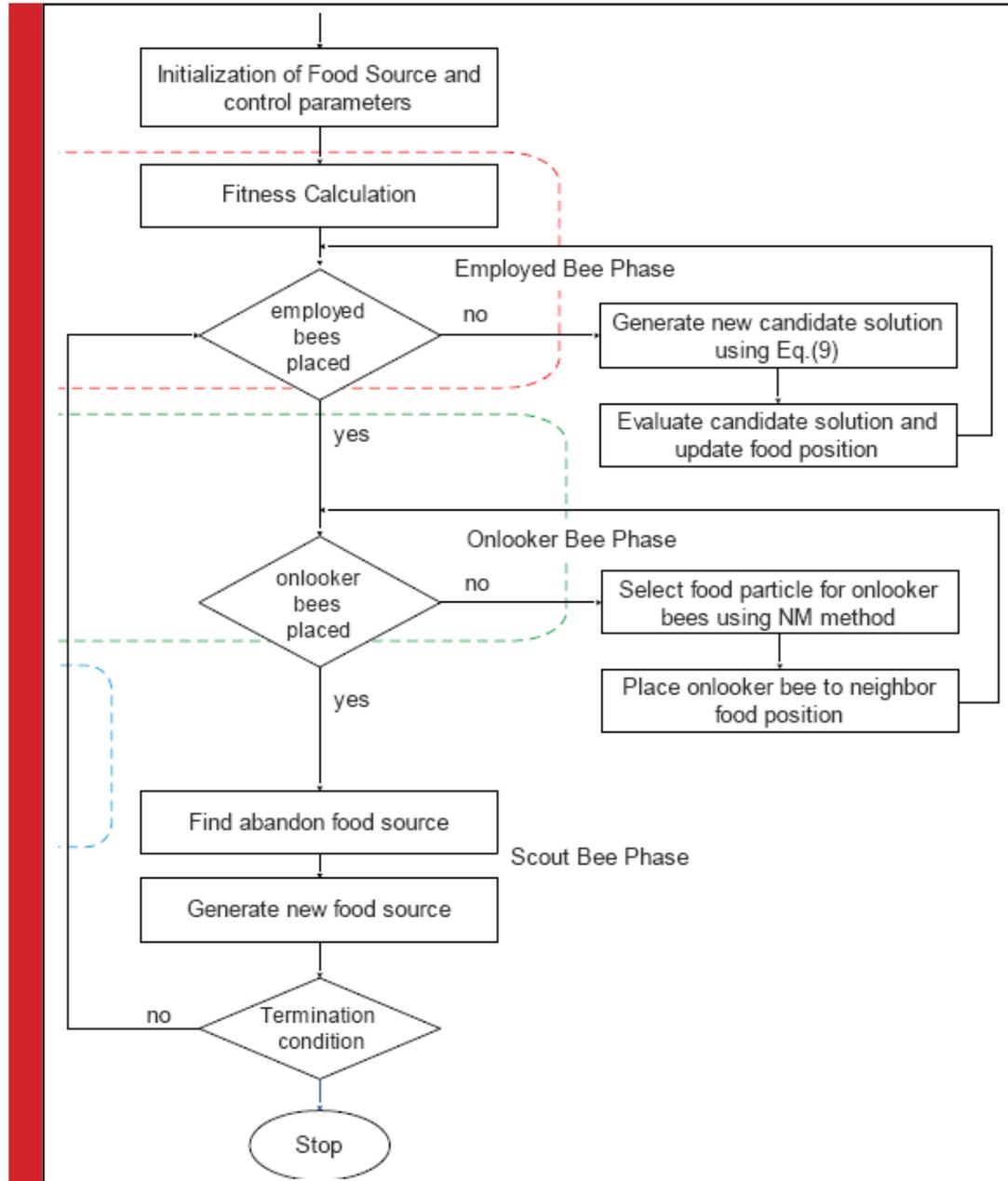


FIGURE 1. Work flow of proposed NM-ABC algorithm

TABLE I. Parameter settings for Experimental Evaluation

Type	Method
Number of Bees	100
Maximum Iterations	1000
Heuristic Used	Modified Nelder-Mead Method
Termination Condition	Maximum Iterations
Run	20
∅	[1,1]

NM method is used as local search. To assess the performance of the proposed NM-ABC on solving EPD large sized benchmark unit systems forty-unit system is chosen. From the simulation result, our proposed algorithm shows significant performance in succeeding the goal of minimizing the cost of generating unit. NM-ABC produces minimized generation cost with other nature inspired algorithms. Thus our proposed algorithm provides an effective heuristic approach to solve EPD for large sized test case.

TABLE II. Unit output of different methods for test case 3						
Unit	NM-ABC	ABC	CABC	IABC	LABC	MABC
P1	48.10337	48.78228	37.31549	49.63385	36	48.31595
P2	114	113.4788	63.50122	71.83804	113.138	113.2195
P3	120	102.1727	108.9332	63.98893	60	102.2052
P4	92.75032	104.364	190	176.0576	190	180.5617
P5	97	97	97	97	97	97
P6	140	96.34776	68	68	121.1038	112.4644
P7	297.2447	110	300	281.8259	219.4286	203.4315
P8	267.5719	300	182.2293	288.4604	250.2106	292.0738
P9	176.3421	297.1281	185.8387	273.9423	267.2882	293.3526
P10	210.4669	162.5414	173.0284	216.2571	270.9595	196.6467
P11	136.9714	324.6816	375	240.3223	364.4466	151.3181
P12	167.0008	140.049	150.4608	144.6459	254.6637	207.5963
P13	414.7065	181.1742	438.9648	240.1508	423.8642	481.1929
P14	306.333	499.45	500	223.7075	394.5674	225.3509
P15	500	383.4203	453.2914	500	334.9145	493.6452
P16	273.6516	442.6797	500	422.8685	364.5887	257.3414
P17	453.2891	468.7174	395.6553	452.2047	491.8115	476.2226
P18	497.2624	447.726	339.8796	361.8167	499.9959	497.4329
P19	479.8874	373.7093	550	550	318.5355	298.5769
P20	549.6337	550	550	501.0201	421.3289	543.669
P21	550	547.8926	550	549.6811	550	494.0024
P22	549.0776	489.3099	329.2822	364.1384	504.7502	546.4916
P23	547.366	550	550	538.1944	550	538.2078
P24	550	547.5793	531.7071	550	465.8048	541.9934
P25	550	497.6082	475.0315	550	537.5953	513.1214
P26	550	526.0304	373.7858	550	333.748	550
P27	10	46.3273	37.13355	30.52611	10	40.65403
P28	40.56873	10	10	34.40069	10	48.53446
P29	31.724	10	10	42.60481	25.81941	10
P30	97	96.99925	86.86509	97	97	97
P31	190	190	190	190	190	173.7211
P32	190	180.8876	190	190	190	190
P33	190	190	134.054	190	190	190
P34	106.8925	199.301	200	200	196.4184	120.1845
P35	183.6122	176.5349	144.9811	200	188.8609	113.8885
P36	191.417	196.2276	172.9326	188.4513	90	198.2997
P37	76.24618	102.5088	110	109.0241	83.34546	78.04431
P38	81.25063	107.9787	67.16313	67.33025	101.205	105.0594
P39	75.29182	101.6068	51.47514	87.36818	100.5231	73.66182
P40	331.5849	424.7498	550	483.6308	513.3216	539.8704
Total Cost	123169.9	123370.7	126645.4	123703.9	124402.2	124379.9

ACKNOWLEDGMENT

This work is a part of the Research Projects sponsored by the Major Project Scheme, UGC, India, Reference Nos: F.No./2014-15/NFO-2014-15-OBC-PON-3843/ (SA-III/WEBSITE), dated March 2015. The authors would like to express their thanks for the financial supports offered by the Sponsored Agency.

REFERENCES

- Abbas, N.H. and Aftan, H.S., 2014. Quantum artificial bee colony algorithm for numerical function optimization. *International Journal of Computer Applications*, 93(9).
- Ahिलandeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 1-5.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Aravindhbabu, P. and Nayar, K.R., 2002. Economic dispatch based on optimal lambda using radial basis function network. *International journal of electrical power & energy systems*, 24(7), pp.551-556.
- Basturk, B. and Karaboga, D., 2006, May. An artificial bee colony (ABC) algorithm for numeric function optimization. In *IEEE swarm intelligence symposium* (Vol. 8, No. 1, pp. 687-697).
- Chen, C.L. and Wang, S.C., 1993. Branch-and-bound scheduling for thermal generating units. *IEEE Transactions on Energy Conversion*, 8(2), pp.184-189.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Dodu, J.C., Martin, P., Merlin, A. and Pouget, J., 1972. An optimal formulation and solution of short-range operating problems for a power system with flow constraints. *Proceedings of the IEEE*, 60(1), pp.54-63.
- Gao, W.F. and Liu, S.Y., 2012. A modified artificial bee colony algorithm. *Computers & Operations Research*, 39(3), pp.687-697.
- Gao, W.F., Liu, S.Y. and Huang, L.L., 2013. A novel artificial bee colony algorithm based on modified search equation and orthogonal learning. *IEEE Transactions on Cybernetics*, 43(3), pp.1011-1024.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Hemamalini, S. and Simon, S.P., 2010. Artificial bee colony algorithm for economic load dispatch problem with non-smooth cost functions. *Electric Power Components and Systems*, 38(7), pp.786-803.
- Karaboga, D. and Basturk, B., 2007. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*, 39(3), pp.459-471.
- Karaboga, D. and Basturk, B., 2007. Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. *Foundations of fuzzy logic and soft computing*, pp.789-798.
- Khamsen, W. And Takeang, C., 2016. Hybrid of lambda and bee colony optimization for solving economic dispatch. *Przeglad Elektrotechniczny*, 92(9), pp.220-223.
- Kumar, R., Sadu, A., Kumar, R. and Panda, S.K., 2012. A novel multi-objective directed bee colony optimization algorithm for multi-objective emission constrained economic power dispatch. *International Journal of Electrical Power & Energy Systems*, 43(1), pp.1241-1250.
- Labbi, Y., Attous, D.B. and Mahdad, B., 2014. Artificial bee colony optimization for economic dispatch with valve point effect. *Frontiers in Energy*, 8(4), pp.449-458.
- Nelder, J.A. and Mead, R., 1965. A simplex method for function minimization. *The computer journal*, 7(4), pp.308-313.
- Ozturk, C., Hancer, E. and Karaboga, D., 2015. Dynamic clustering with improved binary artificial bee colony algorithm. *Applied Soft Computing*, 28, pp.69-80.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Rajasekhar, A., Abraham, A. and Pant, M., 2011, October. Levy mutated artificial bee colony algorithm for global optimization. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on* (pp. 655-662). IEEE.
- Rajeswari, M., Amudhavel, J., Pothula, S. and Dhavachelvan, P., 2017. Directed Bee Colony Optimization Algorithm to Solve the Nurse Rostering Problem. *Computational intelligence and neuroscience*, 2017.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 7-21.
- Selvakumar, A.I. and Thanushkodi, K., 2008. Anti-predatory particle swarm optimization: solution to nonconvex economic dispatch problems. *Electric Power Systems Research*, 78(1), pp.2-10.

- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 1-8.
- Sinha, N., Chakrabarti, R. and Chattopadhyay, P.K., 2003. Evolutionary programming techniques for economic load dispatch. IEEE Transactions on evolutionary computation, 7(1), pp.83-94.
- Slimani, L. and Bouktir, T., 2013. Economic power dispatch of power systems with pollution control using artificial bee colony optimization. Turkish Journal of Electrical Engineering & Computer Sciences, 21(6), pp.1515-1527.
- Sai, P.W., Pan, J.S., Liao, B.Y. and Chu, S.C., 2009. Enhanced artificial bee colony optimization. International Journal of Innovative Computing, Information and Control, 5(12), pp.5081-5092.
- Vanaja, B., Hemamalini, S. and Simon, S.P., 2008, November. Artificial Immune based Economic Load Dispatch with valve-point effect. In TENCON 2008-2008 IEEE Region 10 Conference (pp. 1-5). IEEE.
- Xia, X. and Elaiw, A.M., 2010. Optimal dynamic economic dispatch of generation: A review. Electric Power Systems Research, 80(8), pp.975-986.
- Yoshida, H., Kawata, K., Fukuyama, Y., Takayama, S. and Nakanishi, Y., 2000. A particle swarm optimization for reactive power and voltage control considering voltage security assessment. IEEE Transactions on power systems, 15(4), pp.1232-1239.
- Yuan, X., Wang, L., Zhang, Y. and Yuan, Y., 2009. A hybrid differential evolution method for dynamic economic dispatch with valve-point effects. Expert systems with applications, 36(2), pp.4042-4048.
- Zhao, H., Pei, Z., Jiang, J., Guan, R., Wang, C. and Shi, X., 2010. A hybrid swarm intelligent method based on genetic algorithm and artificial bee colony. Advances in Swarm Intelligence, pp.558-565.

Enhanced Hybrid QoS Framework for Intelligent Decision Making (IDM) in Fog Computing

Gnanasekaran Sambasivam¹, J. Amudhavel² and Raju Govindaraj Babukarthik³

^{1,2,3}*Department of Computer Science and Engineering, KL University, Andhra Pradesh, India*

ABSTRACT

The Internet of Things (IoT) technology has been the recent emerging fields of computing today. This technology is used in many applications such as health, banking and E-commerce. The Fog computing is mainly used with the cloud computing technology, end to end connection increases the efficiency for end users as well by reducing the workloads performed in the cloud computing. The fog computing research needs to be improved in various aspects to increase the efficiency like data transfer, security and Quality of Service(QoS). In this paper, we proposed a novel framework for fog computing model (FOG_QoS) incorporating the QoS parameters, to achieve better performance in Fog computing.

KEY WORDS: FOG COMPUTING, IOT, MATCHMAKING, QOS AND SERVICE PROVIDER

INTRODUCTION

The cloud computing environment usually provide services for software as service, platform as a service and infrastructure as a service (Bonomi et al. 2012, Rahman et al. 2016). All the services are accessed by various users from various locations. While accessing the services from the cloud the data transfer is enormous so that the data load will be very high with high traffic. Due to the over load and heavy traffic the user sometime may not get their services efficiently so that they efficiency will be low (Teing et al. 2016, Vaquero et al. 2014). For that purpose, a new paradigm called Fog computing has been evolved and it works to the edge of the network

and mainly to reduces the latency and congestions in the network. The Fog computing is usually managing the large distributed and it supports services in the virtual environment. So, in this work we have proposed a Quality of service based novel model for fog computing. Which we have added number of QoS parameters for the model and makes the model more efficient. The output of every model in the computing requires a high quality to be delivered to the end user with many number of quality aspects (Yi et al. 2015 Yan Sun et al. 2017).

Based on the QoS only the user will be satisfied with the delivery of the model. So, if the model provides a good QoS with user satisfaction in terms of the generalized QoS parameters or specific parameters. The fog

ARTICLE INFORMATION:

*Corresponding Author: gsambu@gmail.com

Received 27th April, 2017

Accepted after revision 29th July, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

computing will have several benefits with relevant to the domains such as wearable computing, big data analytics, used in reducing the latency, increased security, Internet of Things (IoT), and privacy (Yi et al. 2015). In the field of IoT, fog computing plays a major role because it will provide a unique interface and for the heterogenous computations the resources will be available in flexible manner. The proposed model QoS based fog model will provide quicker response to the user and data delivery to the requested destination will be faster than ever. So, with the proposed QoS model it makes the fog computing model more efficient in terms of access and quality of services.

The rest of this paper is organized as following sections like in section 2 tells about Related works section 3 presents the Proposed model explains our novel framework of Fog Computing, section 4 presents the Algorithm for the proposed framework, section 5 discuss about the various QoS parameters for proposed framework. In section 6, the paper concludes with the summary of the framework proposed.

MATERIAL AND METHODS

There is no novel framework for QoS in the fog computing architecture. QoS is the most important in the fog framework, as the requirement of the consumers increased by day-by-day then there arises the quality related problems. So incorporating the QoS parameters in the fog architecture is very important. If the QoS is added with the framework then it will be much more efficient. Due to the workload in cloud computing archi-

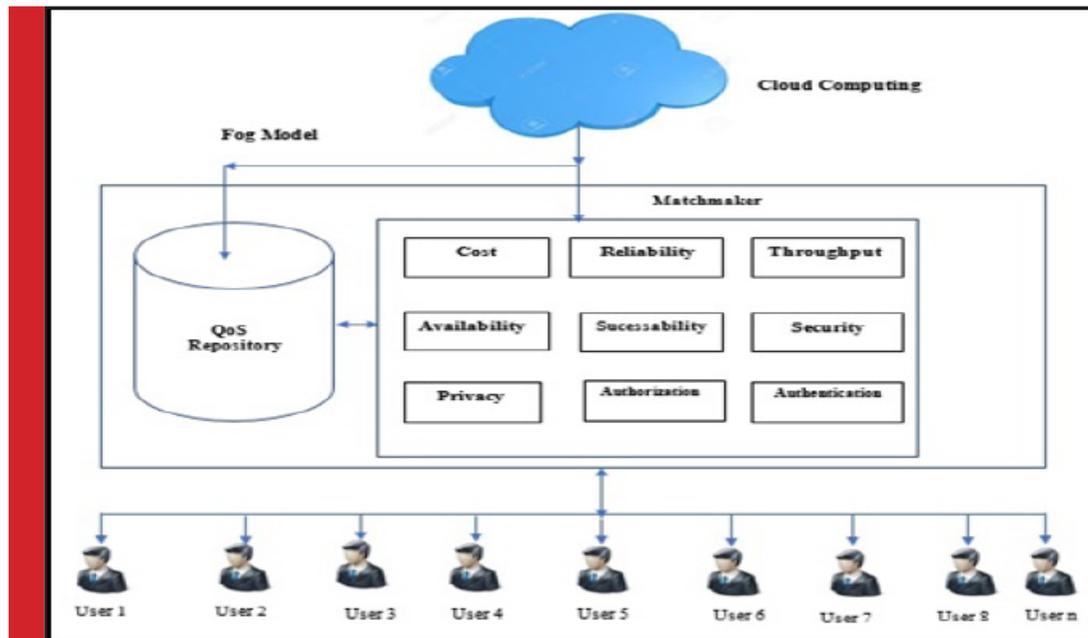
tecture then some of the constraints may fail due to overload. Because the cloud handles large amount of data in the cloud and it is so tedious process to handle all these procedures. So for these purpose the fog computing has been evolved to make the cloud should be easily accessible. The fog computing will be available and accessible to the end users easily since it is present in the edge of the networks. The proposed system shown in fig.b consists of matchmaker and QoS relational database. The components of the proposed are explained below.

CLOUD COMPUTING: The cloud computing offers many services like Platform as a SERVICE (PaaS), Infrastructure as a service(IaaS) and Software as a Service (SaaS) for the consumers. Usually, the cloud computing stores data, manage and process the data in the remote servers. In the proposed architecture, the data will be accessed efficiently with the QoS in the fog framework.

MATCHMAKER: The matchmaker will match the user requests with the cloud data inside the fog model. The matchmaker will use the matching with the available QoS and the data will be presented to the user.

QOS RELATIONAL DATABASE: The Quality of Service(QoS) values will be present in the QoS relational databases. The matchmaker will query the QoS database for the QoS values and based upon the results the user gets the service.

The Fog_QoS algorithm shown in fig.a starts with the user requests to the matchmaker. The matchmaking algorithm will match the user requests with the matchmaker with the retrieved QoS values from the user. After the matchmaking phase gets over the data will be presented to the user efficiently.



QOS PARAMETER

Cost: The cost is the value of the service which will be paid to the service provider for using the service. So as like as cloud computing user should pay for the services which they are accessing. The repository will have the detailed information about the cost services.

Reliability: In the real world, most of the networks and their channels are not providing a reliable service. In QoS the reliability is very important for the customers otherwise they cannot use the services. If the services provided by the service provider not a reliable there will be major loss in the businesses.

Throughput: It is defined as the total time taken by the service per unit of time. The amount of work done per unit of time.

Availability: It is measured using the services which is in operational status. This refers to the ratio of time which the server is up and running. If the server is up then the reliability of the service will be increase.

ALGORITHM FOR HYBRID QOS FRAMEWORK FOR FOG

```

Hybrid QoS framework for fog Algorithm
Input: No. of Services (S), No. of QoS.
Output: Optimal Selection(QoS).
Step 1: Parameter Initialization
        Number of Services (s), Number of QoS (qos), User
        Requirements (UR)
Step 2: Execution
        Request Service from the data repository (CR) with
        matchmaker
        Store the services in QoS repository in fog repository
        For (i=0; i< data. length;i++)
        {
            fog repository []={fQoS1,fQoS2,fQoS3 .....fQoSn
            //}w.r.t data requested by user
        }
        If (requested data == Qos value)
        {
            Store data with Qos values
        }
        Fog_repository []={QoS1,QoS2,QoS3 .....QoSn} //w.r.t
        keyword search
Step 3: Combine both the search results
        User_data_QoS []=fog_repository{fQoS1,fQoS2,fQoS3 .....
        fQoSn}n = data _repository [ ]={CQoS1,CQoS2,CQoS
        ....CQoSn}
        Go to step 2.
Step 4: Output
        Return (User_data_QoS)
    
```

Successability

Successability will be calculated by the services which the services are processed successfully. It is also calculated by number of messages got through responses after the request will be processed completely.

Security: The security is the most important QoS factors for the fog computing because enormous data will pass through the fog layer to consumer from the cloud computing. The various threats are confidentiality, availability, integrity.

Privacy: The privacy is also the important QoS factors which it will protects the sensitive information. The privacy uses with law and regulations with privacy. It can also be measured with privacy information how it will be protected.

Authorization: The authorization is the particular person will be able to access the particular files for the support confidentiality and integrity. There are predefined polices, access methods and many security levels for the support of authorization.

Authentication: The authentication is used to identify the user credentials in many ways using the username and password.

RESULTS AND DISCUSSION

The fog computing provides services from the edge of the cloud computing. It is the extension of the cloud computing. In this paper, we have presented a novel framework for Fog computing based on Quality of Services, identified nine QoS parameters for optimal service and thereby increasing the performance of proposed framework.

REFERENCES

Bonomi.F (2012),Fog Computing and Its Role in the Internet of Things, Proc. 1st MCC Workshop Mobile Cloud Computing, 13–16.

Rahman N.H.A, W. B. Glisson,Yang.Y (2016), Choo,Forensic-by-design framework for cyber-physical cloud systems",IEEE Cloud Computing,50–59.

Teing.Y.Y, Dehghantanha.A, Choo.K..K. R.,Yang.L. T. (2016), 1000 Forensic investigation of {P2P} cloud storage services and backbone for iot networks: Bittorrent sync as a case study, Computers & Electrical Engineering,1–14.

Vaquero L.M.,Rodero-Merino.L(2014), Finding Your Way in the Fog: Towards a Comprehensive Definition of Fog Computing, Sigcomm Computer Comm. vol.44, 27–32.

Yan Sun, Nan Zhang, (2017) A resource-sharing model based on a repeated game in fog computing,vol. 24,Issue 3,687–694.

Yi.S, C. Li, and Li.Q(2015), Survey of Fog Computing: Concepts Applications and Issues, Proc. Workshop on Mobile Big Data,37–42.

Yi.S, Qin.Z, and Li.Q (2015), Security and Privacy Issues of Fog Computing: A Survey. Wireless Algorithms, Systems, and Applications, LNCS 9204.Springer, 685–695.

Securing SQL Injection flaw

R. Rajadurai*¹, V. Bhadrinath², B. Anandhavel³, N. Danapaquame⁴ and J. Amudhavel⁵

^{1,2,3,4}Department of CSE, Sri ManakulaVinayagar Engineering College, Puducherry, India

⁵Department CSE, KL University, Andhra Pradesh, India

ABSTRACT

SQL injection flaw is predominant in last few years of network security. By exploiting the input validation, the attacker can inject the SQL injection query from the front end to explore the back end data stored in the database. Securing of data from the attacker who involved in injection flaw is becoming more challenging. We proposed an approach to detect the injection attack query and retrieving back the access control from where it last. We are utilizing pattern matching technique and token management technique. KMP (Knuth Morris Pratt) pattern searching algorithm is used for validation of input queries and token management technique is used to retrieve back the access control from where it last. Since provision of both prevention and recovery process makes the system more secure than the existing one.

KEY WORDS: NETWORK SECURITY, WEB SECURITY, PREDICTIVE ANALYSIS, INJECTION FLAW, VULNERABILITY AND SQL DATABASE

INTRODUCTION

The internet is most integrated part of our life. People have become more close to the internet. A network is a series of nodes or links interconnected by communication links. Security is one of the most crucial parts of the internet. Network security is used in the protection of all important documents while allowing access to only those who has access to it. Network security secures the activity such as trade secrets, medical reports, etc.

A. Vulnerability

The vulnerability is cyber security term that refers to foibles in a system by making itself open to ambush. Computer users and network providers protect computer servers by providing software security packages up to date.

Top 10 OWASP network vulnerabilities

- Injection flaw
- Broken authentication
- Cross site script

ARTICLE INFORMATION:

*Corresponding Author: rajadurai@smvec.ac.in

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

- In secured direct object references
- Security misconfiguration
- Sensitive data exposure
- Missing function level access control
- Cross site request forger
- Using components with known vulnerability
- Undirected redirect and forward

B. Authentication

Authentication is a process, by which verifying a person is a valid user or not by checking their necessary security credentials which are given in their database. If credentials matched, the authentication process starts. The authentication process always needs authorization process.

C. Authorization

Authorization is a function specifying access rights to a particular user to perform a particular action. Authorization helps us to control access rights by granting and denying permissions. In the multi-admin server, the super user will give permission for all users to access the database.

D. Injection flaw

Injection flaws are a class of security vulnerability that allows an attacker to steal web application details. It is ranked first in top 10 OWASP network vulnerabilities. An attacker can exploit injection flaw by flaring-up of necessary details and by add additional details and offer unintended the functionality. By allowing injection flaw you allow an attacker to make changes to any arbitrary data available in the application. It can also cause sever

E. Threats in injection flaw

- Injection flaw causes repudiation issues such as changing account balances or voiding transactions.
- Injection flaw causes complete disorder or disclosure of data available in the system.
- Injection flaw may destroy data or make it unavailable and become administrator of the database server.
- Any business affected by injection flaw will cause severe damage to entire system since there is a loss of personal data, financial information, and other aspects.

F. SQL Injection

SQL injection commands is a technique where an attacker creates or alters existing SQL commands to reveal concealed data or to cause damage to SQL data or to override valuable ones. When an application uses internal SQL commands and user input credentials (like login screen), then the SQL data be injected that is it

can create, read, update, or delete any data available in the system. This is mostly accomplished by the application taking the user input and combining with preferred parameters to attain needed SQL query. Owing to the lack of input validation, the attacker can connect to the database on behalf of a super user (super user are those who creates other users). Here the attacker may create a super user in your database. The need for network security is the lack of awareness of risks and awareness.

Small business owners are not so aware of thinking that the hackers use automation tools or can inject SQL query to exploit their data within their own network. The network security will build trust with customers. Not only small scale business even large scare

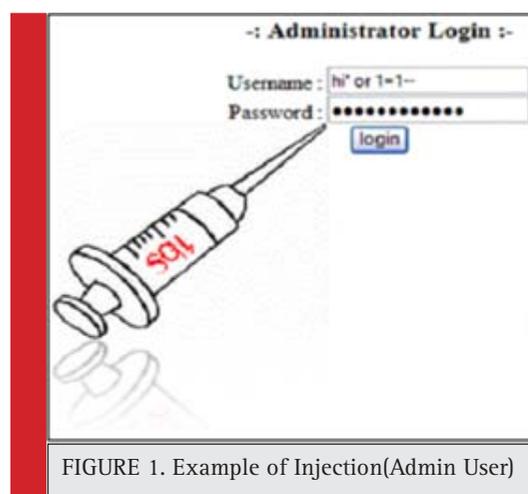


FIGURE 1. Example of Injection(Admin User)

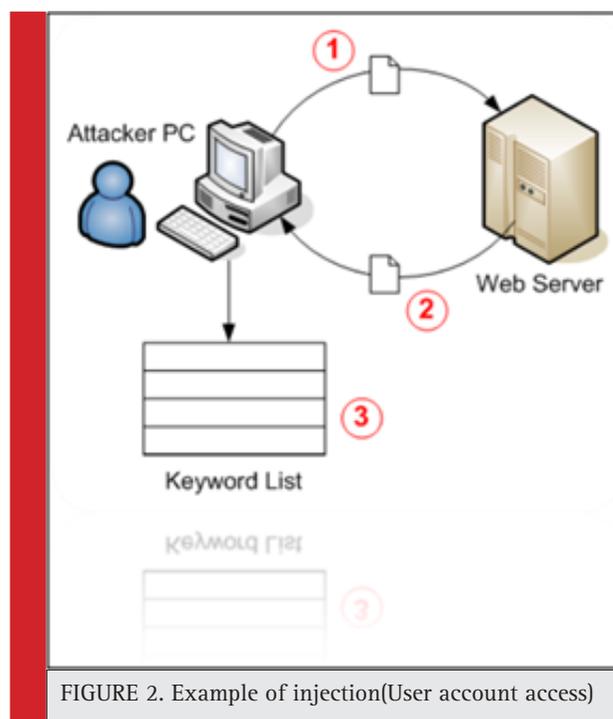


FIGURE 2. Example of injection(User account access)

business need to be careful about securing their website. The hackers not just steal the data alone. They are also creating a watering hole to malware which can cause severe problems in future such as denial of service, data theft, etc. The injection flaw occurs mainly in web pages where the developer uses open source packages or third party resource. The security is important for any business growth. The malicious attacks are increasing day by day. The security needs to match the existing term of attacks today.

The main consequences of SQL injection flaw are

- Confidentiality
- Authentication
- Authorization
- Integrity

G. SQL Injection types

SQL injection is classified based on data extraction. It is classified as

1. In band SQL injections
2. Out of band SQL injections

In-band SQL injections

SQL injections that use the same communication channel as input to dump the information back and are called in-band or inline SQL Injections. This is one of the most common methods, readily explained on the internet in different ways. For example, a query parameter, if injectable, leads to the dumping of info on the web page.

Out of band SQL injections

Injections that use a secondary or different communication channel to dump the output of queries performed via the input channel are referred to as out-of-band SQL injections. For example, the injection is made to a web application and a secondary channel such as DNS queries is used to dump the data back to the attacker domain.

RELATED WORK

Panigrahi et al. proposed the conceptual design and also this paper is titled as SQLiGoT means predicting SQL injection attacks using the graph of tokens and Support Vector Machine (SVM). SQL injection attacks detection has become challenging due to dynamic changes in attack vendors. They thus proposed a novel approach to detect injection attacks by considering SQL queries as the graph of tokens and Support Vector Machine (SVM). They propose an alternative design for the creation of graph tokens. They primarily focused on web applications developed using MySQL and PHP. Their approach can be easily transferred to another platform. Their technique

effectively identifies injection SQL queries with negligible performance. The core approach consists of converting SQL query into nodes of the sequence of tokens and interacting between them with weighted edges. By training an SVM using by the centrality measure of nodes and identify injection queries may occur during runtime. They drag in the issues of high time complexity (Due to the generation of the graph of tokens), the system lags in accuracy and system cover only detection module, and they didn't talk about retrieving of data once it has lasted.

Ala' Shakhathreh et al. derived an SQL-injection vulnerability scanning tool for automatic creation of SQL-injection attacks. They had proposed a tool that can create SQL injection attacks using different attacking patterns automatically to detect different network vulnerability. They designed a web scanning tool named as MySQL Injector with enhanced features of providing penetration test on the website. They designed their tool only for PHP based website. MySQL Injector will automate penetration test process. This tool can be easily handled by anyone even for those who aren't familiar with hacking. They proposed nearly 10 attacking patterns. Even though if one pattern fails the other pattern survive in the checking vulnerability. The MySQL Injector tool includes several levels where the first level will verify web path against SQL injection attack and the second level is the confirmation level to make sure that the web page is injected. This tool falls in issues of working only in PHP based websites, it is limited towards only contained patterns, this tool finds more difficulty in detecting most of the network vulnerabilities and this tool is not so useful since any tool need to withstand existence of attacks. We have conducted extensive experiments to validate the accuracy, privacy, and efficiency of our solutions. For future work, we plan to focus on designing a host-assisted mechanism for the complete data-leak detection for large-scale organizations.

Thilagam et al. proposed an approach titled as Securing web applications from injection and logic vulnerabilities: Approaches and challenges. Injection flaw will allow the user to steal sensitive data and can perform adversary actions. Hence there is a necessity to secure your web application from attackers and hackers. Implementation of flaws in the application will allow attacker or hacker to steal sensitive data and perform unrelated actions. Their objective is to securing web applications against injection attacks. Their articles lie in categorizing based on detection of flaws and prevention against attacks. Since there are various papers are available, this paper contributes to the field of securing web applications by discussing various types of vulnerabilities, illustrating the existing review articles, illustrating the approach on various phases of Software Development

Life Cycle (SDLC), identifying research gaps involved and future work direction.

The limitation of vulnerability involved in scanners and also in spotting out some applications that are adopted to secure web application. Xiaokui Shu et al. designed Privacy- Preserving Detection of Sensitive Data Exposure. Among various data leak cases, the human mistake is the main cause of data loss. The solution exists in term of detecting inadvertent sensitive data leaks caused by human mistakes and to provide alert to a firm. The usual approach is to screen content of exposed sensitive information in storage and transmission. It requires detection operation to conduct secrecy. The secrecy requirements are so challenging in these applications. They had presented some privacy for preserving data loss detection (DLD) solution to solve the issue of secrecy requirements. The advantage of the method is that it enables the user to detect the operation of semi-honest-user without revealing the sensitive data to provider involved in usage. This may cause the data to be exposed before acknowledging the user about the action involved. These privacies are only limited to some extends.

The prevention technique that is used in those existing systems is more time consuming. The proposed system consumes less time when compared to the existing ones. Proposed system doesn't lags in accuracy. Instead of only covering prevention model, we proposed a system which covers both prevention and retraction of defected data.

EXISTING SYSTEM

A. Objective

SQL injection attacks detection has become challenging due to the heterogeneity of the attack vendors. They presented a novel approach to detect injection attacks by modelling SQL queries as a graph of tokens and Support Vector Machine (SVM). They propose an alternative design for the creation of graph tokens

B. Their proposal

They presented a novel approach to detect injection attacks by modelling SQL queries as a graph and Support Vector Machine (SVM). They propose an alternative design for the creation of graph tokens. They primarily focused on web applications developed using MySQL and PHP. Their approach can be easily transferred to another platform. Their queries can perform with negligible performance. The core approach consists of converting SQL query into a sequence of tokens as the node and interaction between them as weighted edges.

By training an SVM classifier using the centrality measure of nodes and identify malicious queries during runtime.

C. Privacy- Preserving Detection of Sensitive Data Exposure

Among various data leak cases, the human mistake is the main cause of data loss. The solution exists in term of detecting inadvertent sensitive data leaks caused by human mistakes and to provide alert to a firm (Deepa G et al. 2016). To solve the issue of sensitive data digits during detection privacy preserving data-leak detection (DTD) is used. With small number of false alarms the data-leak can be preserved. The detection accuracy is measured by using the detection module (Kar et al. 2016).

D. SQL injection vulnerability scanning tool

They proposed a powerful tool that can automatically create SQL- injection attack using different attacking patterns to detect various website network vulnerabilities. They designed a web scanning tool (MySQL Injector) tool with enhanced features that will provide various penetration tests (Ali et al. 2010). To reduce the injection attacks this tool can be used and it can maximize the security of the application and provide protection of the data over the network.

E. Implementation of flaw in an application

Their objective of this review is to summarize the current state of securing web applications from injection flaw. Their articles are categorized based on detecting the flaws and preventing the attacks (SQL-injection attacks). The core of the existing system approach consists of

- Converting an SQL query into a sequence of tokens by preserving its structural composition (Shu et al.2015).
- Generating a graph consisting of tokens as the nodes and interaction between them as weighted edges.
- Training a Support Vector Machine (SVM) classifier using the centrality measure of nodes.
- Using the classifier to identify malicious queries at runtime.

They are:

- Database firewall
- Database server
- Web application
- Web application firewall
- Web clients

F. Benefits of existing system

- Graph of tokens has been proposed to process SQL injection at larger extent.
- The performance impact plays nearly 2.5% of the average response time of web application using experiment (Kumar et al.2015).

- This approach can protect multiple web applications which hosted on a shared web server.
- This approach is practically useful in real time environments.

G. Issues in existing system

- Inappropriate to retrieve back control
- High time complexity
- It is designed to work only in database firewall layer
- It is only applicable to PostgreSQL
- Lags in accuracy

PROPOSED SYSTEM

To reduce the effort of injection flaw we need some changes at the database level. They changes can be performed by limiting super user access, initial injection flaw prevention process, database account retrieval and incrementing the security credentials. Super user has the authorization to create and delete the users of the database (Anusha et al. 2015). Since the super user is very dangerous in the database as any super user can delete or protect another super user to access the database. In initial injection flaw prevention process Knuth Morris Pratt (KMP) algorithm is

used to find the injected query which is given by the user. In database account retrieval the revocation and save-point technique is used and the user request the super user to revoke the account access (Dharshini et al. 2016). These solutions will maximize the security of the database account. The approach is discussed briefly as follow using a architecture diagram Fig. 3.

A. Super user protectivity

Super user has the authorization to create and delete the users of the database. Since the super user is very dangerous in the database as any super user can delete or protect another super user to access the database. So, we are limiting super user authorization only once per database server (Govindharaj et al. 2016). If really needed the user can contact the network provider to provide new super user access with proper verification. Which interns restricts the attacker to deny the authorized user to access data.

As shown in above fig. 4, Existing web architecture consists of mutple super users to access the database. Super users are one who has authentication to create or delete other super users from accessing the database.

At the same time, super users are dangerous to database because it can stop the authorized user from

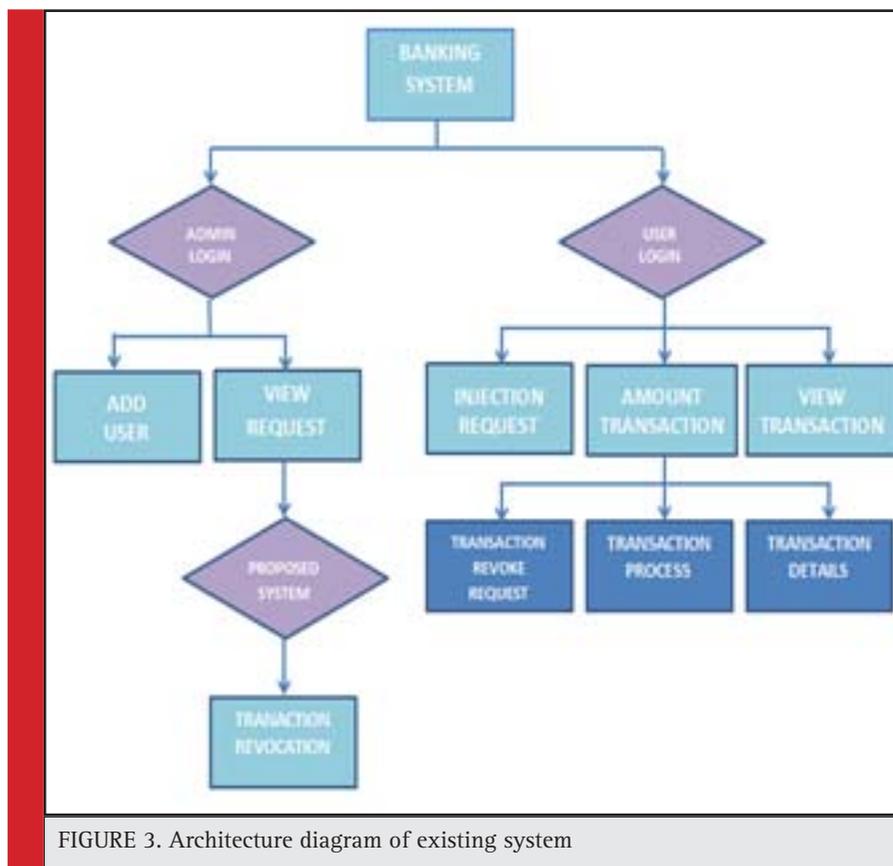


FIGURE 3. Architecture diagram of existing system

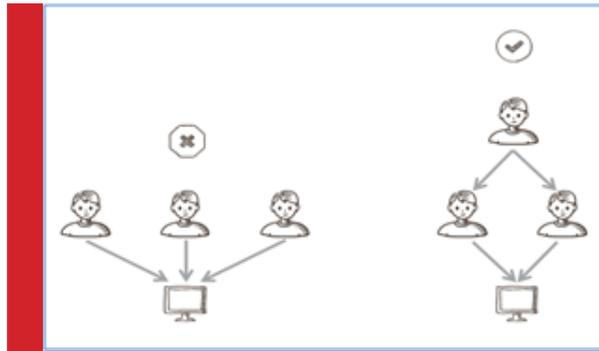


FIGURE 4. Super user protectivity diagram

accessing the database (Ahilandeswari et al. 2016). So in fig. 4, we are limiting the super users by creating only once per database. Thus in our proposed system, we had demonstrated this module through administrator login. The administrator login defines the super user in our system. The administrator login (i.e. super user) consists of functionalities such as adding other users and view injection request as shown in fig. 3.

B. Initial injection flow prevention process

As initial security, the pattern matching technique continuously acknowledges the input credentials with existing injection attack queries as shown in figure fig. 4. Knuth Morris Pratt (KMP) pattern searching algorithm is used for acknowledging the input credentials. Knuth Morris Pratt (KMP) is the best pattern searching algorithm found because of its less processing time (Sasidevi et al. 2017). It searches the pattern along with resource with is assigned within the string to find the needed injected query. The time along which the redundancy control can be possible by which they reduce the consistency towards the query. The time complexity of Knuth Morris Pratt (KMP) algorithm is $O(n)$ in the worst case (Gayathri R et al. 2016). The advantage of using Knuth Morris Pratt (KMP) algorithm is it will avoid checking for matching characters in a given string. KMP algorithm utilizes lps[] array which indicates longest proper prefix in the given string. The lps[] string is utilized to skip matching characters.

As compared to existing system Support vector machine method (SVM), Knuth Morris Pratt algorithm is found more efficient. Time complexity of Knuth Morris Pratt (KMP) algorithm is $O(n)$ in the worst case which is very much lesser when comparing to existing Support vector machine method (SVM) which is of $O(m(n-m+1))$. The utilized control can be forced back along with the parameters which are possible to be injected along with the query to identify the injected query (Shanmugam et al. 2017). These can be performed along within the

injection prevention process with in the query set and the processing over the network. According to other set of conditions through the condition (Rajadurai et al. 2017). In KMP algorithm we take the advantage of degeneration property which will avoid matching characters while comparing.

KMP algorithm is utilized as

```

AlgKMPSearch(char *pat, char *txt)
Begin
    M = strlen(pat)
    N = strlen(txt)
    LPSArray(pat,M,lps)
    I=0
    J=0
    Do
        If pat [j] == txt[i]
            J++
            I++
        End if
        If j==m
            //Found pattern Return Injection Flaw Found
            J= lps [j-1]
        Else if
            //Do not match lps [], they will match anyway
            If j != 0
                J = lps [j-1]
            Else
                I=i+1
            End if
        Until i>n
    End
    
```

Kmp algorithm follows two steps

1. Prefix function
2. String matching

Prefix function

Lps[] – longest proper prefix

P	C	O	C	A	C	O	L	A
Lps[]	0	0	1	0	1	2	0	1

String matching

P	a	b	a	b	c	a	b	c	d
S	a	b	c	d					

In above table P specify the password and S specify the search string

P	a	b	a	b	c	a	b	c	d
Cmp	a								

In above table P specify the password and Cmp specify the injection query compare string

P	a	b	a	b	c	a	b	c	d
Cmp		b							

P	a	b	a	b	c	a	b	c	d
Cmp			c						

P	a	b	a	b	c	a	b	c	d
Cmp			a						

P	a	b	a	b	c	a	b	c	d
Cmp				b					

It differs in case of

P	a	a	a	a	a	b	a	a	a
S	a	a	a	a					

P	a	a	a	a	a	b	a	a	a
Cmp	a	a	a	a	a				

P	a	a	a	a	a	b	a	a	a
Cmp						a			

P	a	a	a	a	a	b	a	a	a
Cmp							a		

B. Pattern matching scheme

The Knuth Morris Pratt (KMP) algorithm is used for pattern matching. Knuth Morris Pratt (KMP) pattern searching algorithm is used for acknowledging the input credentials. Knuth Morris Pratt (KMP) is the best pattern searching algorithm found because of its less processing time. It searches the pattern along with resource with is assigned within the string to find the needed injected query. The time along which the redundancy control can be possible by which they reduce the consistency towards the query. Injection queries will be look like result of the query should returns true. So that the condition of the password matches with the needed criteria and the system is accessed. (ie., Result = true)

Let us consider the inputs of the password of following combinations.

1. Anything or true
Eg. Ab or 1 = '1'
2. True or anything
Eg. 1 = '1' or Ab
3. True or True
Eg. 1='1' or 1='1'
Eg. 1='1'
4. True and true
Eg. 1='1' and 1='1'

From the above mentioned examples the number of repeating numbers are always in even in such case we make use of least proper suffix (lps[]) array used in KMP algorithm.

Working mechanism for comparison is

1. Pattern function

P	a	b	o	r	l	=	'	l	'
lps[]	0	0	0	0	0	0	0	1	0

lps[] array returns one for repeating value

Consider another example

P	l	=	'	l	'	o	r	l	=	'	l	'
lps[]	0	0	0	1	0	0	0	1	0	1	1	2

Repeating value changes one in all cases of injection queries. We can exhibit the injection query here itself. For further verification we are utilizing string matching technique.

2. String matching

P	a	b	o	r	l	=	'	l	'
S	1	=	'	1	'				

P	a	b	o	r	l	=	'	l	'
Cmp	1	1	1	1	1				

P	a	b	o	r	l	=	'	l	'
Cmp						=			

P	a	b	o	r	l	=	'	l	'
Cmp							'		

P	a	b	o	r	l	=	'	l	'
Cmp								1	

P	a	b	o	r	l	=	'	l	'
Cmp									'

Thus the injection query is found inside the system. Now the web page violates the user to acknowledge the database.

C. Retrieving back the access control

The commit is the command used to save changes performed in the database. The roll back is the command used to undo changes until the last commit or rollback command. The save point is the point in the transaction where we can roll back to the certain point without rolling back to the entire transaction. The release save point

command will remove a save point. The auto commit is the default transaction management mode in SQL database. Every SQL transaction is committed and rolled back when it is completed. If the database is attacked by the attacker and there is an emergency for the user to retrieve back the control before it fully lasts.

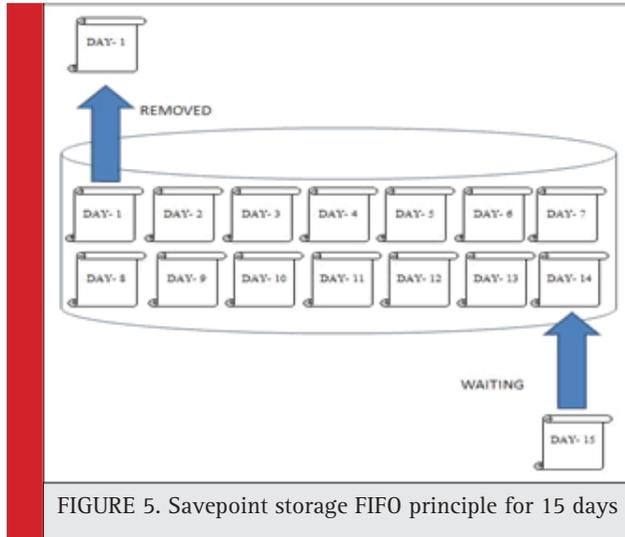


FIGURE 5. Savepoint storage FIFO principle for 15 days

In above fig. 8. The transaction which was happened on 27-02-2017 is considered as hacked after losing the control of one's own system. Since the transaction has happened beyond one's own knowledge it needed to revoke. So the user can request the system admin to retrieve back the system control from where it lost. The admin can able to revolt the control upon doing the proper verification. Now the user account will be revolted until where the user is asked as shown in fig. 9. Steps performed to retrieve back the control can be proposed as

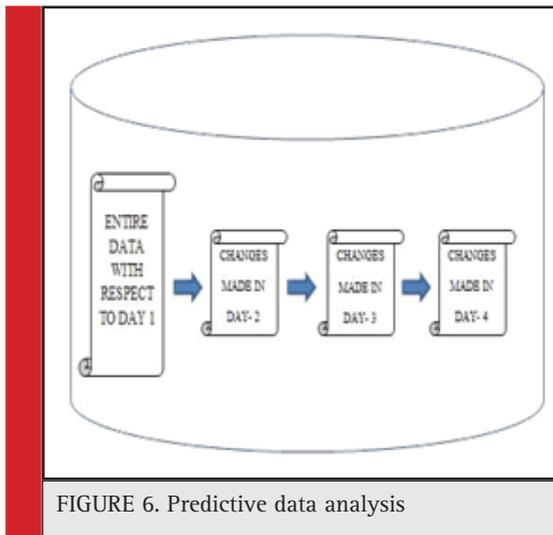


FIGURE 6. Predictive data analysis

- Auto commit is performed at the end of each process and it is stored in the database as a separate save point entry.
- The save point entries are stored for every process.
- The save point entries are stored as FIFO queue in the database.
- Those entries are effectively maintained using big data analysis.
- The data can be retrieved by the user from network provider if in case the data are attacked or vanished as shown in fig. 7.

The save point are usually assigned in SQL as save point name followed by keyword "save point". Since auto save point is not available in SQL. We can set the auto commit and can assign the stored commit data in save point.

The fig 10 illustrates the predictive data analysis of the database. Once the data is stored in the query it will store the previous day information in a small cache memory and store it into the table along with the current date status information.

The big data analysis is a process of gaining, standardizing and analysis of large volumes of data. The predictive analysis is a process of extracting information from existing data. Therefore using predictive analysis we can extract the information from the top of the queue stored in the server. All the other save points contains only the changes performed in existing data. If a roll back is made with request from a user by network provided the sequence of operation performed from day one to corresponding day. Once the user found the injection flaw in the system he can request the network provider for rolling back of data up to specified date as shown in fig. 6. The network provider will perform roll back function up to specified date and he will give back access control to the user. The user can alter the data and can include necessary security and can secure the data for future.

D. Incrementing credentials process

To control the injection flaw, it is necessary to increment the security credentials. After detecting that the crime has been done and data has been recovered. There is an urgency to lock that injection query. The cyber will track the details of "how the culprit attacked the system? If the culprit tries to utilize the same procedure to attack the same network again then the culprit instead of gaining the database access, his acknowledgment gets violated. Those details of user track records are also tracked and recovered using predictive analysis which is used for storing data in big data analysis.

The cyber cam will find the culprit by using the pattern matching technique and will when he tries to access with the same technique they will find him and arrest him for the crime which he committed.

E. Time complexity

Time complexity of the algorithms is calculated using empirical and theoretical analysis and tabulated as shown in table 1. Where time complexity is calculated using the formula eq. 1.

$$T = C / (n - m + 1)$$

Where,

- T - Time taken to compare the string of lengthm
- C - Expected number of comparisons for random text
- M - Length of the pattern
- N - Size of the password

For example,

Let the sample pattern be

X =

a	b	o	r	l	=	'	l	'
1	2	3	4	5	6	7	8	9

Using SVM method,

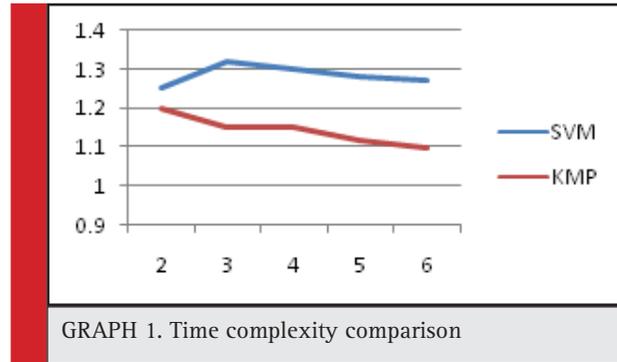
- Expected number of comparison(C) = 7.5
- Length of the pattern (N) = 9
- Size of the comparing string (N) = 5
- Time taken (T) = C / (n - m + 1) = 7.5 / (10 - 5 + 1) = 1.25

Using KMP method,

- Expected number of comparison(C) = 6
- Length of the pattern (N) = 9
- Size of the comparing string (N) = 5
- Time taken (T) = C / (n - m + 1) = 6 / (10 - 5 + 1) = 1.2

Table 1. Time complexity		
Length of the patterns	Time complexity	
	SVM method	KMP algorithm
2	1.25	1.2
3	1.32	1.15
4	1.30	1.15
5	1.28	1.12
6	1.27	1.1

From graph 1, we predicted that the time complexity for KMP algorithm is very much lesser than SVM method. The predicted graphs shows that the complexity of the Support Vector Machine (SVM) method. In this the time taken to in SVM method take 1.25 seconds to



execute and the size of the comparing string can be optimized through the KMP algorithm with 1.2 seconds with the same size.

CONCLUSION

Life doesn't about waiting for the storm to pass; it is about learning to dance in rain. The effort caused by injection flaw is immense but it can be controlled to larger extend. The pre-prevention, detection, and solution are added to the existing system which will secure the system to the maximum. We can't assure any system to be secured so one needs some procedure to retrieve back the access control in terms some flaw has happened. To prevent the same kind of action from happening again in future, we need some kind of prevention technique as well. Those techniques and precautions excellently utilized in this procedure of solution building. Super user accessibility is very much important to provide authorization for users in the database. But it, in turn, can cause a negative result to security. The limitation of super user authorization is very much helping in saving table data in SQL. The queries can be filtered with the same sample of queries along .If in case the data is attacked; it can be easily recovered using this solution. The backup of each save point of data stored in the database will not consume more space. This solution in turns secures the data by nearly major percentage from injection flaw.

REFERENCES

Abdul Bashah Mat Ali, Ala' Yaseen Ibrahim Shakhatrehb, Moh dSyazwanAbdullahc, JasemAlostadd (2010), SQL-injection vulnerability scanning tool for automatic creation of

Ahilandeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(2): 1-5.

Amin Zammouri, Abdelaziz Ait Moussa (2016), SafeBrowse: A new tool for strengthening and monitoring the security con-

- figuration of web browsers, IEEE Information Technology for Organizations Development (IT4OD) International Conference.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 1(1): 1-7.
- DebabrataKar, SuvasiniPanigrahi, Srikanth Sundararajan (2016), SQLiGoT: Detecting SQL injection attacks using graph of tokens and SVM, Computers & Security, 60:206-225.
- Deepa G, SanthiThilagam P (2016), Securing Web Applications from Injection and Logic Vulnerabilities: Approaches and Challenges, Elsevier journal of Information and Software Technology, 2(4), 1051-7.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). Secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(1): 1-6.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(1): 7-13.
- Kumar A., TK, Hong Liu, Johnson P. Thomas, GoutamMylavarapu (2015), Identifying Sensitive Data Items within Hadoop, High Performance Computing and Communications (HPCC), IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 24-26.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 9-16.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 1-8.
- SQL-injection attacks, Procedia Computer Science 3:453-458.
- Xiaokui Shu, Danfeng Yao (2015), Privacy-Preserving Detection of Sensitive Data Exposure, IEEE Transactions on information forensics and Security, 10,1092-1103.

Secure Data Process in Distributed Cloud Computing

P. Sharmila*¹, Dr. N. Danapaquiame², R. Subhapriya³, A. Janakiram⁴ and J. Amudhavel⁵

^{1,2,3,4}Department of CSE, Sri Manakula Vinayagar Engg.College, Pondicherry, India

⁵Department of CSE, KL University, Andhra Pradesh, India

ABSTRACT

Cloud is display for empowering administrations, clients' requirement, convenient and on-request N/W access to a common pool of configurable processing assets. Cloud is for the most part transfer and download the information. Transferring information's with separate to the information transmissions. The different instruments and gathered at an exceptional scale. Such huge volume of information produced is alluded to as large information, which now are altering all parts of our life going from endeavour's to people, from science groups to governments, as they show incredible possibilities to enhance effectiveness of ventures and the personal satisfaction. To get nontrivial designs and get important data from enormous information, an essential issue is the way to legitimately put the gathered information by various clients to circulated mists. Proficiently dissect the gathered information to spare client costs in information stockpiling and preparing, especially the cost reserve funds of clients who share information. Thu sly, it needs the nearby joint efforts among the clients, by sharing and using the enormous information in appropriated mists because of the multifaceted nature and volume of huge information.

KEY WORDS: CLIENT REQUEST, DISTRIBUTED CLOUD RESOURCES, DOWNLOADING DATA, UPLOADING DATA

INTRODUCTION

Cloud is a model for empowering administrations and client's requirement, convenient and on-request arrange access to a mutual pool of configurable registering resources. Cloud is creating innovation to encourage advancements of vast scale and on-request benefits. Distributed computing with utilizing our information on anyplace at whatever time on the planet. Cloud is utilizing for security reason in view of a few devices

and methodologies. Hadoop is an open-source system that permits storing and process huge information in an appropriated domain crosswise over bunches of PCs utilizing basic programming models. It is intended to scale up from single servers to a huge number of machines, each offering nearby calculation and capacity. Huge information implies truly a major information; Cloud is a model for empowering administrations and client's ubiqites, convenient and on-request arrange access to a mutual pool of configurable registering resources.

ARTICLE INFORMATION:

*Corresponding Author: Sharmiksd93@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

Cloud is creating innovation to encourage advancements of vast scale and on-request benefits. Distributed computing with utilizing our information on anyplace at whatever time on the planet. Cloud is utilizing for security reason in view of a few devices and methodologies. Hadoop is an open-source system that permits storing and process huge information in an appropriated domain crosswise over bunches of PCs utilizing basic programming models. It is intended to scale up from single servers to a huge number of machines, each offering nearby calculation and capacity. Huge information implies truly a major information.

In Computer science, distributed computing portrays a sort of outsourcing of PC administrations, like the route in which power supply is outsourced. Clients can basically utilize it. They don't have to stress where the power is from, how it is made, or transported. Consistently, they pay for what they devoured. The thought behind distributed computing is comparable: The client can essentially utilize capacity, figuring power, or extraordinarily created advancement situations, without worrying how these work inside. Distributed computing is generally Internet-based registering. The cloud is a similitude for the Internet in view of how the web is depicted in PC organize graphs; which implies it is a deliberation concealing the perplexing foundation of the web. It is a style of figuring in which IT-related abilities are given "as an administration", permitting clients to get to innovation empowered administrations from the Internet" in the cloud "without learning of, or control over the advancements behind these servers.

As indicated by a paper distributed by IEEE Internet Computing in 2008 "Distributed computing is a world view in which data is for all time put away in servers on the Internet and reserved briefly on customers that incorporate PCs, tablets, handholds, sensors, and so on." Cloud processing is a general idea that uses programming as an administration (SaaS, for example, Web 2.0 and other innovation drifts, all of which rely on upon the Internet for fulfilling clients' needs). For instance, Goggle Apps gives regular business applications online that are got to from a web program, while the product and information are put away on the Internet servers. Distributed computing is regularly mistaken for lattice registering (a type of circulated processing whereby a "super and virtual PC" is made out of a bunch of organized, approximately coupled PCs, cooperating to perform substantial errands), utility figuring (the bundling of registering assets, for example, calculation and capacity are given as a deliberate administration that must be paid like a conventional open utility, for example, power) and autonomic registering (PC frameworks equipped for self-administration).

Many distributed computing organizations are fuelled by matrices, have autonomic qualities and are charged like utilities, however distributed computing can be viewed as a characteristic next stride from the lattice utility model. Some effective cloud models have next to zero concentrated foundation or charging frameworks at all including shared systems like Bit Torrent and Skype (Khan *et al.* 2013). The greater part of distributed computing foundation as of now comprises solid administrations conveyed through server farms that are based on PC and capacity virtualization advancements. The administrations are available any place on the planet, with The Cloud showing up as a solitary purpose of access for all the processing needs of buyers. Business offerings need to meet the nature of administration prerequisites of clients and normally offer administration level assertions. Open measures and open source programming are additionally basic to the development of distributed computing.

As clients for the most part don't claim the foundation or know all insights about it, for the most part they are getting to or leasing, so they can expend assets as an administration, and might pay for what they don't require, rather than what they really do need to utilize. Many distributed computing suppliers utilize the utility figuring model which is comparable to how conventional open utilities like power are expended, while others are charged on a membership premise (Loukopoulos *et al.* 2004). By sharing consumable and "impalpable" figuring power between numerous "occupants", use rates can be enhanced (as servers are not left sit without moving) which can lessen costs fundamentally while expanding the speed of utilization improvement. A symptom of this approach is that "PC limit rises drastically" as clients don't need to build for pinnacle loads. Appropriation has been empowered by "expanded fast data transfer capacity" which makes it conceivable to get a similar reaction times from brought together foundation at different destinations.

Distributed computing is being driven by suppliers including Goggle, Amazon.com, and Yahoo! And in addition conventional merchants including IBM, Intel, Microsoft and SAP. It can embrace by a wide range of clients, be they people or huge ventures. Most web clients are at present utilizing cloud administrations, regardless of the possibility that they don't understand it. Web-mail for instance is a cloud benefit, as are Casebook and Wikipedia and contact list synchronization and online information backups. The basic idea goes back to 1960 when John McCarthy communicated his supposition that "calculation may some time or another be sorted out as an open utility" and the term Cloud was at that point in business use in the mid-1990s to allude to vast ATM systems. By the turn of the 21st century, distributed

computing arrangements had begun to show up available, however the vast majority of the concentration as of now was on Software as a service. Amazon.com assumed a key part in the advancement of distributed computing while redesigning their server farms after the website bubble and giving access to their frameworks by method for Amazon Web Services in 2002 on a utility registering premise.

RELATED WORK

Distributed computing is a rising world view that gives registering, correspondence and capacity assets as an administration over a system. Correspondence assets frequently turn into a bottleneck in administration provisioning applications. Consequently, information replication which brings information (e.g., databases) nearer to information shoppers (e.g., cloud applications) is viewed as a promising arrangement. It permits limiting system deferrals and data transfer capacity use. In this paper we think about information replication in distributed computing server farms. Dissimilar to different methodologies accessible in the writing, we consider both vitality productivity and data transfer capacity utilization of the framework (Gu *et al.* 2014).

This is notwithstanding the enhanced nature of administration QoS got subsequently of the decreased correspondence delays. The assessment comes about, got from both numerical model and broad recreations, help to uncover execution and vitality proficiency trade-offs and in addition control the plan of future information replication arrangements (Jun Liu *et al.* 2014). To address this hole, we propose an information replication procedure for distributed computing server farms which advances vitality utilization, arrange transfer speed and correspondence postpone both between geologically dispersed server farms and inside each datacenter. In particular, our commitments can be condensed as takes after.

Demonstrating of vitality utilization attributes of server farm IT foundations. Improvement of an information replication approach for joint streamlining of vitality utilization and transfer speed limit of server farms. Enhancement of correspondence deferral to give nature of client experience to cloud applications. Execution assessment of the created replication methodology through scientific demonstrating and utilizing a bundle level distributed computing test system, Green Cloud. Examination of the trade-off between execution, serviceability, unwavering quality and vitality utilization (Marshall *et al.* 2012).

One of the fundamental objectives of replication is to guarantee information accessibility. A sudden disappointment away framework or data centre power out-

age could bring about inaccessibility. To oppose these impacts, duplicates of information articles are kept up on excess foundations and in a few geologically appropriated data centres (Limam *et al.* 2010). In this way, the accessibility is generally measured by a likelihood of disappointments in server farm segments or capacity framework. In the proposed replication approach, we expect that each information protest is permanently stored in DB and expansion, contingent upon the get to example, it is recreated in Datacenter DB and Rack DB. Any disappointments in Datacenter DB can be recuperated from Central DB and the other way around. Additionally, not at all like in a few different strategies, the proposed approach actualizes a dynamic replication to just keep up ideal number of reproductions to guarantee both accessibility and the QoS of cloud applications.

Another vital purpose behind information replication is in the decreased information get to reaction time for cloud applications (Anusha *et al.* 2015). Conveying and keeping up the information nearer to the servers where applications are executed essentially diminish access time for this information and significantly enhances general framework execution. Be that as it may, on the opposite side, the number and area of copies ought to be chosen precisely as over the top replication may build the related costs and traffic stack in the server farm arrange required for copy upgrades. The proposed replication approach considers the trade-off between information measure, information get to and upgrade rates, accessible system transmission capacity and properties of the actualized server farm topology to settle on ideal choices. In the first place, information items are reproduced in Rack DB nearer to processing hubs and henceforth reaction time is lessened. Second, information questions that are often got to are recreated which diminished aggregate number of reproductions. Keeping up ideal number of imitated information objects limits arrange stack required to stay up with the latest and system reaction time.

This paper audits the subject of information replication in geologically circulated distributed computing server farms and proposes a novel replication arrangement which notwithstanding customary execution measurements, for example, accessibility of system transfer speed, upgrades vitality effectiveness of the framework (Dharshini *et al.* 2016). What's more, streamlining of correspondence defers prompts to changes in nature of client experience of cloud applications. It augments a preparatory adaptation of this work which has been distributed.

The assessment of the proposed replication arrangement depends on the created numerical model and re-enactments utilizing Green Cloud, the test system concentrating on vitality proficiency and correspondence

forms in cloud computing server farms. The acquired outcomes affirm that reproducing information nearer to information customers, i.e., cloud applications, can diminish vitality utilization, data transmission use and correspondence delays generously. Distributed computing has increased noteworthy footing for late years. It is a type of conveyed figuring whereby assets and application stage are shared over the web through on request and pay on usage premise. A few organizations have officially manufactured Internet customer administrations, for example, web crawler, utilization of a few sites to speak with other client in sites, E-mail administrations, and administrations to buy things online that utilization distributed computing framework. However, this innovation experiences dangers and vulnerabilities that keep the clients from putting stock in it. The event of these dangers may come about into harming of private information in cloud condition. This overview paper means to examine the different uncertain security dangers in distributed computing which are influencing the different partners connected to it. It likewise depicts the upsides and downsides of the current security procedure and furthermore presents the current issues in distributed computing, for example, information uprightness, information isolation, and security etc.

Distributed computing is a general term for anything that includes conveying facilitated benefits over the Internet. It is a rising figuring innovation that uses the web and focal remote servers to look after information. This framework is extremely useful for various clients so they can undoubtedly utilize the framework with no outside support to programming and equipment. They can likewise get to their own documents at any PC on web. This innovation takes into account significantly more effective registering by concentrating stockpiling, memory, handling and data transfer capacity. Additionally, Users can ordinarily interface with mists by means of various sorts of online administrations or web programs. Cloud framework gives parcel of advantages and disadvantages to the customers (Govindharaj *et al.* 2016). This paper examines the different uncertain security concerns and the security hazard related with big business distributed computing including its dangers, hazard and weakness. Despite the fact that there is many preferred standpoint worried in distributed computing, the association are moderate in tolerating it because of security issues related with it. Security is one of the essential issues in cloud condition. It is the defilement of Domain Name System (DNS) server.

A DNS server plays out the interpretation of a space name to an IP address. For this situation, the client ask for one IP address yet it is diverted to some other unapproved cloud. Counter measure for this assault is that Domain Name System Security Extensions (DNSSEC).

However, this safety efforts end up being insufficient one. With a specific end goal to secure cloud against different security dangers, diverse cloud specialist organizations embrace distinctive methods. The best answer for enhance the security is that to build up the secured structure which has extreme security design. With the goal that we will secure client's information, message and data against different assault. The secured structure must utilize solid verification and solid get to control components. So it will give greater security to information of clients from that are right now display inside the distributed computing administrations (Ahilandeswari *et al.* 2016).

The secured structure must utilize solid encryption calculation keeping in mind the end goal to ensure the delicate information before entering into the cloud. There are a few encryption procedures are accessible in cryptography. Among all Gentry portrays homomorphic encryption calculation which is utilized to secure the information in cloud condition. A stand out amongst the most utilized encryption procedures is Homomorphic encryption method, which permits particular sorts of calculations to be done on figure message and get an encoded result which unscrambled matches the after effect of operations performed on plain content. Distributed computing offers awesome potential to enhance efficiency and lessens costs. It likewise postures numerous new security dangers. This paper portrays the review of the different uncertain security dangers in distributed computing which are influencing the different partners connected to it. More than ten papers were additionally reviewed with regard to the distributed computing, benefits of distributed computing, hazards in distributed computing and different ways to deal with settle those dangers each with their advantages and disadvantages. We trust that because of the many-sided quality of cloud system, it is extremely hard to accomplish security (Gayathri *et al.* 2016). New security procedures should be produced and more seasoned security systems should have been profoundly wound to have the capacity to work with the mists design.

Cloud frameworks are progressively utilized for a wide scope of computational needs in private and open associations. We call inhabitant an autonomous association that is client of the arranged administrations offered by a cloud supplier. Get to control over the assets of a multitenant domain is a testing issue due to the gigantic number of end clients included and the required seclusion of the security organization crosswise over various associations. Disseminated approval has as of now been widely contemplated with regards to arranged administrations, e.g., circulated document frameworks. Be that as it may, a cloud situation presents extraordinary attributes that warrant re-examination of the presumptions

and arrangement properties. We are especially intrigued to exploit benefit co-area in the datacenter to better combine the capacity framework utilized by normal information records at the application (e.g., coordinated effort archives) or framework level (e.g., root pictures). Secure capacity combination at the file system level is progressively upheld as the Multitenancy world view for cloud conditions. Albeit virtual plates are appealing for their forming, segregation and movement properties, a document based interface can moreover bolster fine-grained controlled sharing, simple asset organization, and record level execution enhancements.

Existing record based arrangements confront versatility constraints since they either need bolster for various visitor occupants, depend on worldwide to-nearby personality mapping to deal with the clients of various inhabitants, or have the visitors and a brought together file system (or intermediary) running at a similar host. In the proposed arrangement we depend on a protest based, appropriated file system to deal with the capacity prerequisites of customers (e.g., virtual machines) having a place with various occupants. In our plan we require that every customer straightforwardly mounts the file system as opposed to having the file system mounted by a middle of the road intermediary (Shanmugam *et al.* 2017).

The file system locally deals with the get to control metadata of each inhabitant, and guarantees that each occupant can just get to its own name space. Controlled document sharing is moderately direct accordingly of the record based access to a typical file system with document granularity get to control. We give model usage of the above approach in the generation grade, distributed file system. With miniaturized scale benchmarks and application-level investigations we quantitatively exhibit the restricted execution overhead of our plan. We can condense our commitments as takes after: Analysis of get to control prerequisites in solidified stockpiling for virtualization. Engineering outline of local get to control in a multi-tenant file system with backwards compatibility to protest based capacity. The customers and servers of the file system all keep running in one datacenter that is physically secured and worked by a free supplier.

A safe co-processor affirms the product stack on each physical host (e.g., hash chain created by a Trusted Platform Module). A focal screen builds up the trust of the framework from the trustworthiness of the taking part hubs. Open keys (or hashes thereof) particularly distinguish occupants, principals and administrations. The hubs safely impart over transitory symmetric keys progressively settled upon by means of open key cryptography.

The private keys of principals and administrations are for all time put away in encoded frame and just show

up in clear content shape at the unpredictable memory of approved hubs. Prior to the reallocation of host memory crosswise over various hubs, the memory substance are cleaned to avoid data spillage. The file system secures the classification and uprightness of put away information and metadata by limiting gets to approved principals.

We accept that the supplier has no malevolent expectation to bargain the framework security. In any case, there might be different reasons (e.g., poor practices) for which the supplier is not trusted for some applications. All things considered the occupant may remotely apply methods of encryption, hashing and evaluating to accomplish end-to-end classification, uprightness and freshness. We target file system get to control with no express endeavour to give answers for open key conveyance, substance verification, refusal of administration and activity examination. At long last, we don't address general appropriated handling, which includes multi-tenant sharing of assets other than capacity (e.g., calculation). In a file system mount demand to a MDS, a customer needs to safely distinguish the inhabitant of the vital. Keeping in mind the end goal to infer the occupant identifier (TID) we apply the RIPEMD-160 cryptographic hash work on general society key of the inhabitant. At that point we incorporate the TID into an extended Client Session ask for and send it to the MDS over a protected session. For validation purposes the demand ought to moreover convey an inhabitant issued accreditation (not upheld yet in our prototype) (Rajadurai *et al.* 2017). The MDS removes the TID from the M Client Session message and stores it as a field of the session class.

QUALITY OF SERVICES

Synergistic analysts at various geographic areas can share the information by getting to and dissecting it. A credulous situation for such substantial volume of enormous information may bring about immense expenses on information transmission among the datacentre's as well as the worked together scientists. Furthermore, the information produced at various areas must be genuinely put to the circulated cloud. Enormous information administration issue is testing, which lies in a few viewpoints. The coordinated effort mindful clients/huge information applications progressively and ceaselessly produce information from various land areas, the high cost will be brought about when overseeing such information due to their geo-dispersions and huge volumes. Users ordinarily have Quality of Service (QoS) necessities. Reasonable use of cloud administrations is essential, generally, one-sided allotment of cloud assets may bring about that unsatisfied clients no longer

utilize the administration, the specialist organization may fall into unsavoriness, and its income will be altogether decreased. Preparing and investigating the set enormous information require huge registering asset. In any case, the registering asset in each data centre normally is constrained. Enormous information administration issue is testing, which lies in a few perspectives. The joint effort mindful clients/enormous information applications powerfully and constantly produce information from various land areas, the high cost will be caused when overseeing such information due to their geo-dispersions and substantial volumes (Sasidevi *et al.* 2017).

Clients commonly have Quality of Service (QoS) prerequisites. Reasonable use of cloud administrations is pivotal, something else, one-sided designation of cloud assets may bring about that unsatisfied clients no longer utilize the administration, the specialist organization may fall into offensiveness, and its income will be fundamentally diminished. Preparing and breaking down the set enormous information require huge processing asset. Notwithstanding, the registering asset in each data centre ordinarily is constrained.

SYSTEM MODEL

We consider a dispersed cloud $G = (V [FE;E])$, consisting of various Datacentre's situated at various geological areas and interconnected by Internet joins, where V and FE are the arrangements of datacentre's and front-end servers, and E is the arrangement of correspondence connections amongst datacentre's and amongst datacentre's and front-end servers. Give v_i a chance to be a datacenter in V and E_iC a connection in E between data centre's v_i and v_j . The capacity and figuring assets of each Datacenter v_i are utilized to store and process information, and the transfer speed asset of each connection is utilized for information transmission between datacentre's. For comfort, we here just consider processing and system assets of G as capacity asset provisioning is like that of registering asset.

Mean by $B_c(v_i)$ the limit of registering asset at datacenter V_i in V , and $B_b(E_iC)$ the transfer speed asset limit of connection E_{ij} in E . Expecting that time is isolated into equivalent schedule vacancies, the measure of accessible processing asset of datacenter V_i is spoken to by $A_c(v_i; t)$, and the measure of accessible transfer speed asset of connection E_iC is spoken to by $A_b(E_iC; t)$ at availability t . Give F_{em} a chance to be a front-end server in FE , where $1 \leq m \leq jFE_j$. Each front-end server F_{em} fills in as an entryway hub to the circulated cloud, and has a specific stockpiling ability to cushion information from its close-by clients. Without loss of sweeping statement, we accept that the quantity of front-end servers is cor-

responding to the quantity of data centres, i.e., $jFE_j = O(jV_j)$. Clients, for example, undertakings, associations and foundations, these days are outsourcing their huge information to the appropriated cloud G .

COORDINATED EFFORTS IN BIG DATA MANAGEMENT

Two community clients in a similar gathering need to share, prepare and break down the middle of the road after effects of each other with a specific end goal to infer their own last outcomes on a long haul premise. We along these lines utilize a coordinated effort gathering to demonstrate an arrangement of communitarian clients that make utilization of the middle consequences of each other in the disseminated cloud. Expect that there are K joint effort assembles in the framework. Mean by g_k , the k Th coordinated effort amass for every k with $1 \leq k \leq K$. For transfer speed asset investment funds, we accept that the middle consequence of every client u_j in gathering g_k will be multicast to all individuals in the gathering. In this paper, we consider information clients like ventures, associations, and organizations, which have a tendency to have more joint efforts with their companions to fabricate more riches and enhance day by day lives of individuals. Notwithstanding, in all actuality, vast coordinated effort bunches with many gathering numbers are difficult to frame and oversee, on the grounds that many entangled duties should be consulted among the individuals. We in this manner accept that the quantity of individuals in each gathering is given as from the earlier and don't change after some time.

1. Taken a toll demonstrate

Overseeing datasets brings about the operational cost of a cloud specialist organization, where the operational cost incorporates the information stockpiling cost, the information handling cost, and the correspondence cost of exchanging datasets and middle outcomes between data centres. These expenses are corresponding to the volume of information put away, handled and exchanged.

PROPOSED SYSTEM

We then propose an enhancement system for the issue, under which we devise a guess calculation with an ensured estimate proportion, by diminishing the issue to the base cost multi ware stream issue. We at last assess the execution of the proposed calculation through exploratory reproductions. Thusly, it needs the nearby joint efforts among the clients, by sharing and using the enormous information in conveyed mists because of the many-sided quality and volume of huge information. Since registering, stockpiling and transmission capacity

assets in a disseminated cloud ordinarily are restricted, and such asset provisioning normally is costly, the collective clients require making utilization of the assets decently. Cloud figuring has increased huge footing for late years. It is a type of appropriated figuring whereby assets and application stage are shared over the web through on request and pay on usage premise. Even however there is many preferred standpoint worried in distributed computing, the association are moderate in tolerating it because of security issues related with it. Security is one of the essential issues in cloud condition.

SYSTEM ARCHITECTURE

The user want to access their file in cloud resources or server they need some security process to access them. Due to security process user's data will be secure in cloud and it avoid the duplication, hacking the data ,unauthorised person .the new user register their details in the process and get the username and password. The process to upload the files and request the file whenever they need. User enter the username and password with security key for process. For user uploading Time password (OTP) for each time. OTP is generated dynamic for each time user enter into the process. Then, User upload the file into server. Now the Cloud resources is created for the process. We created the Virtual machines under the

cloud which consist of speed, RA, Bandwidth and VM id. Then, we category the Virtual machines by the capacity. Then the depending upon the capacity of file and virtual machine we allocate the file to the virtual machines in cloud infrastructure. Now, user send the request with security key for file to the CSP. For user Downloading, we verify the user with security key and generate one time password (OTP) for each time.

Then request received by the server, they maintain the database information. Request forward to the CSP and CSP forward the request to the Owner of the File's send the request to owner of the file. Then the owner give response as access or denied. After that request forward to CSP and CSP verify the response. Then allow the request user to access the file and download to him/her. In this system we use collaboration framework for fairness aware among the distributed cloud server to provide security to user data.so it data only provide to correct authentication user. For task allocation use the task selection algorithms to divide the file based on some condition such as size, bandwidth. For interconnection of multi cloud use Atmosphere.

It is a cloud-based software Platform as a Service used by customers that want to integrate their various cloud-based tools with each other and with on-premise applications. New functionality provided as part of an upgrade last summer allows usage of multiple data

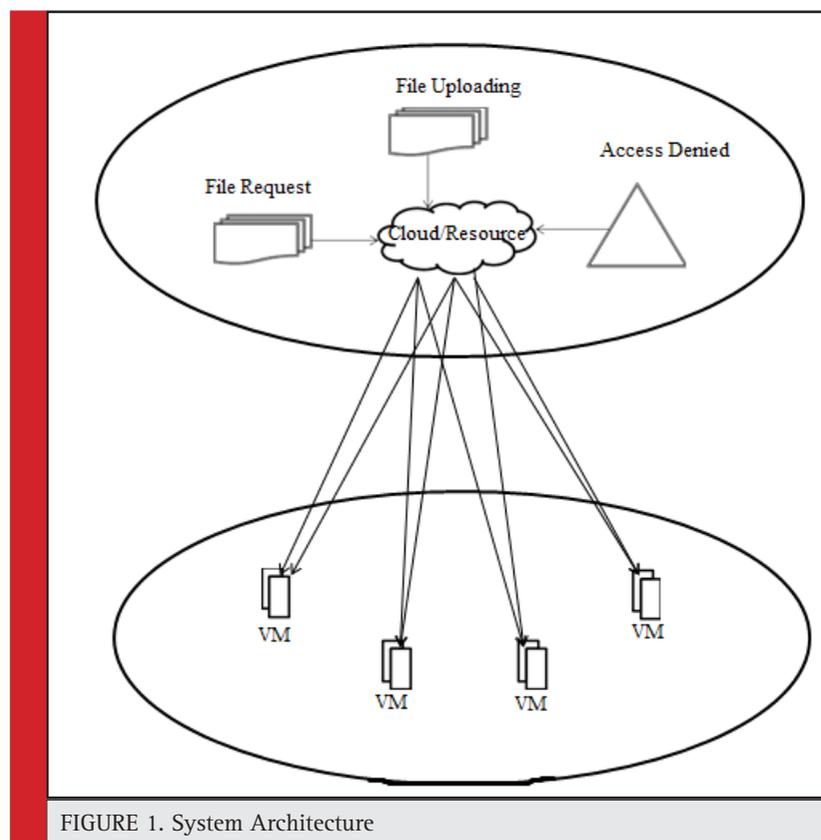


FIGURE 1. System Architecture

Table 1. data values while uploading and downloading in cloud.

Primary	+	Elasticity	CPU and Computation	Data Bandwidth	Date Proximity	Governance and Jurisdiction
Workload Responsiveness	Public	Public and Hybrid	Public and Hybrid	Public and Hybrid	Hybrid and Dadedicated	Private and Hybrid
CPU and Computation	Public and Dedicated	Hybrid	Hybrid	Public and Hybrid	Hybrid and Dadedicated	Private and Hybrid
Data Volume	Public and Dedicated	Hybrid and Dadedicated	Hybrid and Dadedicated	Hybrid	Hybrid and Dadedicated	Private and Hybrid
Data Bandwidth	Public and Dedicated	Hybrid and Dadedicated	Hybrid and Dadedicated	Hybrid and Dadedicated	Hybrid and Dadedicated	Private and Hybrid
Data Proximity	Hybrid and Dadedicated	Private and Hybrid				
Governance and Jurisdiction	Private and Hybrid					

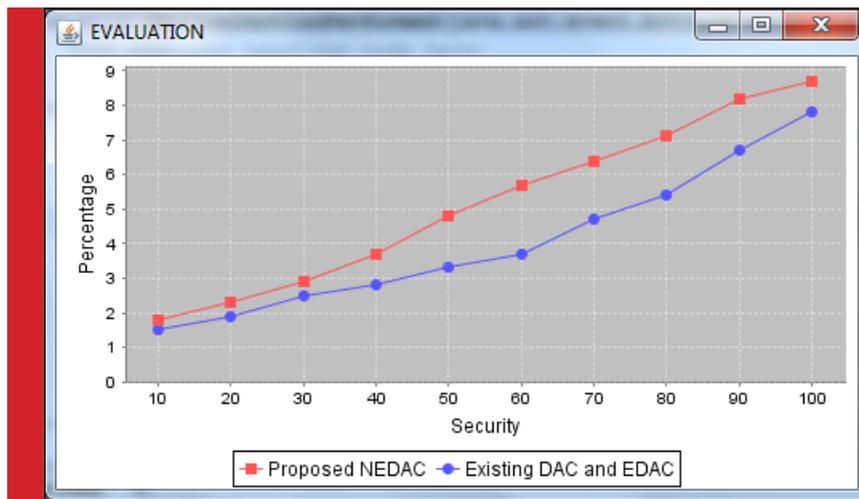


FIGURE 2. Security Performance.

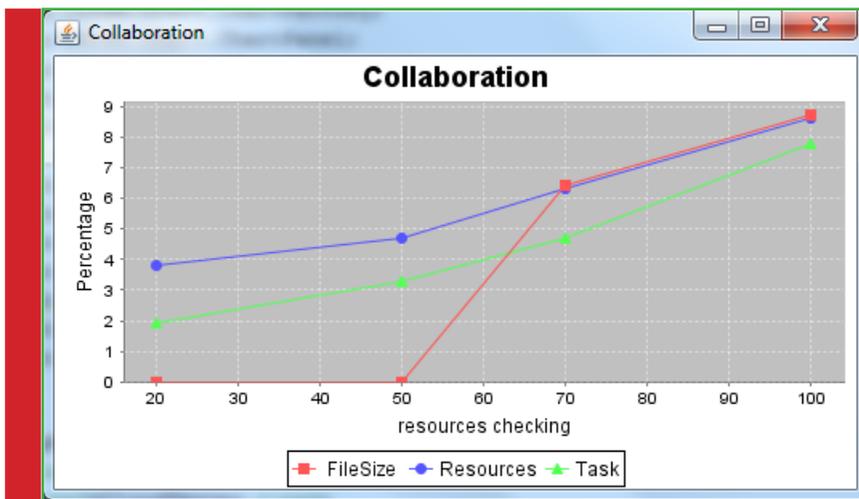


FIGURE 3. Resources checking

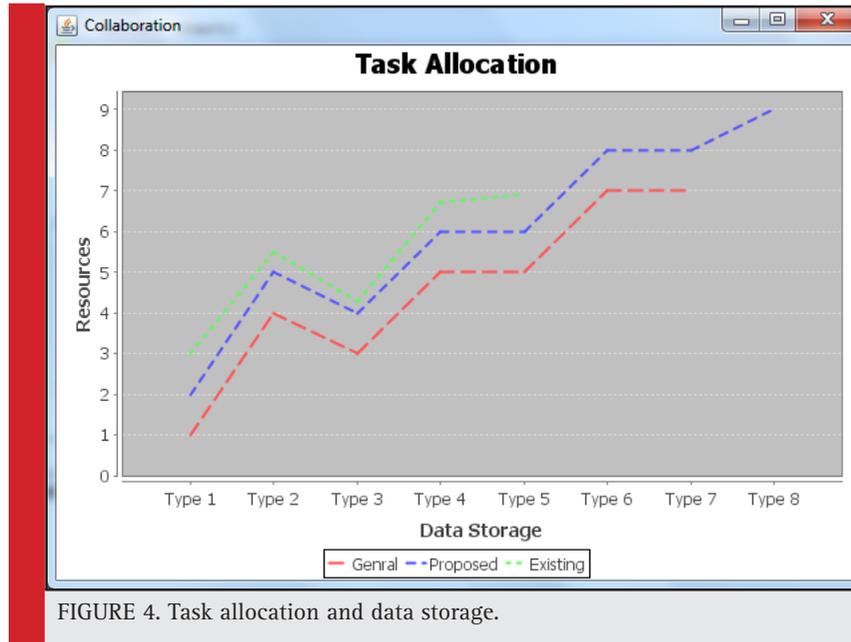


FIGURE 4. Task allocation and data storage.

sources and reservation of cloud resources for real-time data transfer in integration

In distributed cloud computing placing the data is major issues because where data will be stored in distributed cloud server. Another problem is uploading and downloading the data from cloud server. so overcome this problems use collaboration and aware between the distributed cloud to provide trustworthy security for user data in cloud. The data will be stored in distributed server based on data bandwidth. Where the server has place frustrated data in cloud. Interconnection between cloud server the data upload and download of user data will be efficient.

Overseeing datasets brings about the operational cost of a cloud specialist organization, where the operational cost incorporates the information stockpiling cost, the information handling cost, and the correspondence cost of exchanging datasets and middle outcomes between data centres. These expenses are corresponding to the volume of information put away, handled and exchanged.

PERFORMANCE ANALYSIS

Performance analysis for collaboration and fairness aware big data management in clouds for storing and retrieving data efficiently. But generally data is stored on cloud and the allocation for which resources is to be checked. We describe a framework for data and operation security in IaaS, TPM-based certification of a Remote Resource (TCRR) and “Verify My VM” With TCRR a tenant can verify the integrity of a remote host and establish a trusted channel for further communication. In “Verify My VM”, the hyper visor running on an

attested host uses an emulated TPM to verify on-demand the integrity of running VMs. Proposed a “Trusted Cloud Compute Plat-form” (TCCP) to ensure VMs are running on a trusted hardware and software stack on a remote and initially untrusted host. To enable this, a trusted coordinator stores the list of attested hosts that run a “trusted virtual machine monitor” which can securely run the client’s VM. The collect

The collaboration overcome the major problem of big data management in distributed cloud is how to properly place data by different users to distributed cloud. While file size of user data will place correctly and collaboration allocate the task. Due to these performances will be increased.

In Task allocation for collaboration and fairness aware cloud data for available resources to be shared for multiple virtual machines to be stored. In existing with performance analysis for resources allocation on cloud, through agents and describe the implementation of the proposed protocols on an open-source cloud platform and present extensive experimental results that demonstrate their practicality and efficiency. The non-cryptographic nature of the proposed scheme makes it faster to perform the required operations on the data. It improved security. The simulation results show that the performance of the proposed algorithm is used, which can reduce the operational cost of the distributed cloud significantly.

CONCLUSION

We considered a collaboration and fairness-aware big data management problem in distributed cloud environ-

ments. We developed a novel optimization framework, under which we then devised a fast approximation algorithm for the problem. We also analysed the time complexity and approximation ratio of the proposed algorithm. We finally conducted extensive experiments by simulations to evaluate the performance of the proposed algorithm. Experimental results demonstrated that the proposed algorithm is used, and outperforms other two mentioned heuristics.

FUTURE ENHANCEMENTS

Here we will develop with big data as; data is stored on HDFS successfully. Hadoop distributed file system is always to ensure the data has been analysed. Another point with Auto updated mechanism using our data is easily knows the available resources on the cloud computing. Different algorithms using encrypt the data to be performed.

REFERENCES

Ahиландесwari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 2(2): 1-5.

Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.

Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.

Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 2(2): 22-29.

Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.

Gu L, Zeng D, Li P, Guo S. (2014). Cost minimization for big data processing in geo-distributed data centres. *IEEE Trans. on Emerging Topics in Computing*, 2(3): 314-323.

Jun Liu, Feng Liu, Ansari N. (2014). Monitoring and analyzing big traffic data of a large-scale cellular network with Hadoop," *Network*, 28(4).

Khan N, M. Kiah M.L, Madani S. A, and Ali M. (2013). Enhanced dynamic credential generation scheme for protection of user identity in mobile-cloud computing. *The Journal of Supercomputing*, 66(3): 1687-1706, 2013.

Limam N, R. Boutaba. (2010). Assessing Software Service Quality and Trust worthiness at Selection Time. *IEEE Transactions on Software Engineering*, 36(2): 559-574.

Loukopoulos T, Ahmad I. (2004). Static and adaptive distributed data replication using genetic algorithms *Journal of Parallel and Distributed Computing*, 64(11): 1270-1285.

Marshall M. S. (2012). Emerging practices for mapping and linking life sciences data using RDF—A case series," *J.Web Semantics*, 14: 2-13.

Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.

Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 3(1): 7-21.

Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.

A Novel Clustering Algorithm for Big Data: K-Means-Fuzzy C Means

A. Manikandan¹, N. Danapaquiame², R. Gayathri³, E. Kodhai⁴ and J. Amudhavel⁵

^{1,2,3,4}Department of CSE, Sri Manakula Vinayagar Engineering College, Pondicherry, India

⁵Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

ABSTRACT

Data mining is the infusion of data from a large network. It plays an important role in identifying the objects covered data patterns and gives guessing information which applies for business. The function performed in data mining is classification class description, prediction, association, and clustering. More number of research works were carried out related to clustering in data mining in the past, but which has still issued. In this paper, KF algorithm has been proposed for data mining and clustering techniques. Clustering analysis is used to find clusters impacted in the data and KF is the combination of k-means and Fuzzy C means. The proposed KF algorithm gives more valuable performance than the existing in terms of measuring purity, entropy, recall and precision metrics.

KEY WORDS: DATA MINING, CLUSTERING, K-MEANS, FUZZY C MEANS, KF ALGORITHM

INTRODUCTION

Data mining, the extraction of concealed prescient data from huge databases, is an effective new innovation with extraordinary potential to help organizations concentrate on the most critical data in their information distribution centers. Information mining devices foresee future patterns and practices, enabling organizations to make proactive, learning driven choices. The mechanized, forthcoming examinations offered by information mining move past the investigations of past occa-

sions given by review devices run of the mill of choice emotionally supportive networks. Information mining instruments can answer business addresses that customarily were excessively tedious, making it impossible to determine. They scour databases for concealed examples, finding prescient data that specialists may miss since it lies outside their expectations. Most organizations effectively gather and refine monstrous amounts of information. Information mining procedures can be executed quickly on existing programming and equipment stages to improve the benefit of existing data assets, and can be

ARTICLE INFORMATION:

*Corresponding Author: maniraghu22@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

incorporated with new items and frameworks as they are brought on-line. At the point when actualized on superior customer server or parallel preparing PCs, information mining devices can dissect monstrous databases to convey.

Data mining includes to “removing or mining” learning from a lot of information. There are numerous different terms conveying a comparative or marginally unique intending to information mining, for example, learning mining from databases, information extraction, information/design examination, information archaic exploration, and information digging. Many individuals regard information digging as an equivalent word for another prominently utilized term, Knowledge Discovery in Databases”, or KDD. On the other hand, others see information mining as basically a basic stride during the time spent learning revelation in databases.

DATA MINING

Data mining is the abstraction of hidden data from a large data sets which is used to develop the business readiness. Expanded availability of huge amounts of data are possible for turn such information and knowledge into a useful and understandable data. The data mining is also has the development of data cleaning, pattern evaluation, data presentation and data transformation. It plays an important role in identifying the objects covered data patterns and produce the detailed information in the business.

Bunching is information mining system of collection articles or information into groups in which protests inside the group have high similitude, yet are extremely not at all like questions in alternate bunches. Likenesses and Dissimilarities are measured on the trait values which depict the items. Bunching strategies are utilized to plan and pigeonhole the information, for information pressure and model development, for discovery of anomalies and so forth Normal approach of all bunching techniques is to discover groups focus which speak to each bunch.

In view of the likeness metric and information vector bunch focus helps in figuring out which bunch is closest or most comparable one. Data mining is used to specify the patterns to perform data mining task. Data mining task are classified into two types they are descriptive and predictive. Descriptive is used to find the general properties of data from database. Predictive is used to find the attributes using the value of other attributes.

APPLICATION OF DATA MINING

The appliace of data mining is: Market analysis and management, fraud detection, and Corporate Analysis and Risk Management. Market analysis: Data mining as

various fields: Customer profiling: Data mining helps to verify what kind of people buy what kind of products-finding the customer requirements: For different customer they provide the perfect product that may attract the customer.

Cross market analysis:

It performs correlation between the products

Target marketing: it is used for model of clusters

Fraud detection: It is used to detect frauds of telephone and credit card service. In telephone fraud it is used to find the destination, time and date.

Finance analysis:

It involves in cash flow analysis

Resource planning: It involves in comparing the resources and spending.

Competition: It is used to monitor the competitor

CLUSTERING

Clustering is the descriptive data mining. Clustering is the aggregation of objects that are related to same class. Clustering is one by one learning skill that follows to divide cases into clusters which partition same quantities. The technic used in clustering are portioning, distance based, probabilistic, spectral and hierarchical. The application of clustering in datamining is pattern recognition and economic science. The application of clustering in data mining is pattern recognition and economic science. Clustering can be viewed as the most critical unsupervised learning issue; thus, as each other issue of this kind, it manages finding a structure in a gathering of unlabeled data. A free meaning of Clustering could be the way toward sorting out items into gatherings whose individuals are comparative group is consequently an accumulation of articles which are comparative amongst them and are “divergent” to the items having a place with different bunches. The principle thought is to characterize k means, one for each bunch. These k-means bought to be set slyly due to various area causes diverse outcome.

In Fig 1 the clustering algorithm would frequently merge smaller to larger until the result. The result will

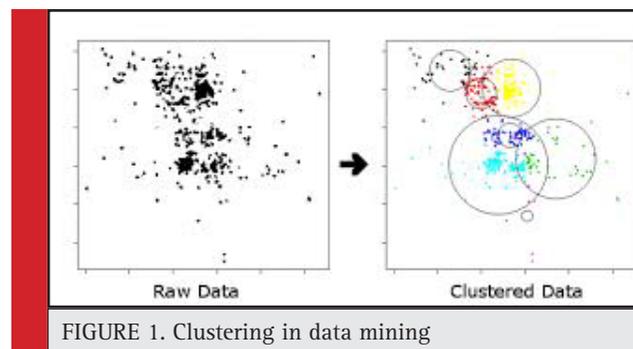


FIGURE 1. Clustering in data mining

be smaller set of combined datas. The algorithm used in clustering techniques are K-means, Fuzzy C means, Kernel based, Graph based algorithm.

K-means is centroid based cluster. Various clusters is illustrated in the center vector. This algorithm is mainly used for partitioning method. The K mean algorithm determine the arithmetic means of all cluster from data-set. In the cluster K means has an individual records.

Fuzzy C means is hard clustering. It is an individual data elements that correspond accurately to one cluster.

In this paper KF algorithm is proposed from K-means and Fuzzy C means clustering algorithm. Though number of research work is related to data mining there is some issues. KF algorithm gives more valuable performance than the existing in terms of measuring purity, entropy, recall and precision metrics.

RELATED WORK

Kavita Nagar et al. 2015 analysed the various kind of data. The data are scientific data, mathematical data, financial data and environmental data. The information sharing is more in networks so the data is increasing as the age network grows. The clustering are based on minimizing the intra class similarity and maximizing the intra class similarity. The overview of different clustering methods is described with an algorithm. The clustering methods are Hierarchical method, grid based, density based, portioning method and model based clustering. Ahilandeswari et al. discussed the Hierarchical method use the data sets available according to the top down and bottom up approach. Grid based cluster use data representation. Density based use arbitrary shape cluster. Partition method use clustering. Model based cluster use the concept of assumption for features of distribution.

Aastha Joshi et al analyzed and compared about the six techniques in the data mining of clustering. The six techniques are K-means, Hierarchical clustering, OPTICS, STING, DBSCAN clustering and Fuzzy C means clustering. They have been described briefly about the techniques, algorithms, types and its process in data mining of clustering. Concluded has K-mean has biggest advantage in clustering of data mining. It increases the number of cluster but decreases the values of numeric. So Agglomerative and Divisive Hierarchical algorithm was created for categorical data due to some complexity they developed K-means for categorical attribute. It is done numeric by assigning rank. OPTIC and DBSCAN clustering are designed in arbitrary shape and portioning and hierarchical are designed in spherical shaped cluster so it facilitates the fast processing of clustering in data mining.

Narander Kumar et al is mainly developed for insurance organization. Unidentified and covered data pattern

from huge data of insurance organization. It requires the strong customer base with huge database. Cluster analysis is a great arithmetic tool for remote and multivariable database. The K-mean method is used to develop the relationship in database. The data of LIC customer, compute the distance of cluster using the attributes of customer with the help of cluster with K-mean. It allocates S_1 to S_n with cluster C_1 . It will increase the revenue and profit of the organization with the customer satisfaction. Johanner Grabmeier and Andreas Rudolph described about what is clustering in data mining with an example and described briefly about the techniques of cluster in data mining. It is accomplished by various similarity, distance measures and optimization criteria for clustering. Each technique has an advantage and disadvantage. In the second focus cluster algorithm discuss about the demographic cluster algorithm.

Yang Zaihua Fuzzy principle in fuzzy C means convention of the c-means algorithm. The convention Fuzzy C means algorithm is used for gene sequence. The convention is taking pseudo F statistics into the method. To verify the validity they use nodes instead of gene to verify the efficiency. The result is high processing speed and stable performance. A. Banumathi, A. Pethalakshmid described about the clustering of data mining and the detailed description about the algorithm and how it works. K-means and Fuzzy C means are used to analysis the quality of cluster it will select randomly or sequentially. In real time large database difficult to predict so they proposed two algorithm UCAM and Fuzzy-UCAM clustering algorithms Unique clustering is obtained with the help of affinity measures.

Sandeep Panda, Sanat Sahu, Pradeep Jene It is an iterative technique that is used to partition an image into K clusters. K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid

CLUSTERING OF DATA MINING ALGORITHM

Clustering is the distribution of data into clusters of identical entities. If the clusters are interior then the data has high identity but if it is exterior then data has variance discussed in Jeff Z. Pan and Ian Horrocks. Clustering is valuable in data mining and data analysis. It is placed

on minimizing inter- class similarity and maximizing intra- class similarity.

The data reporting of less clusters certainly loss the details, but execute simplification. The goal of the clustering is understand the clusters. The result should take “natural” structural information. Clustering is a standalone data mining tool. Cluster analysis is used in information retrieval, pattern recognition, machine language, marketing, database applications, CRM, web analysis and statistics listed in Bo Liu et al. The advantage of clustering is adoptable i.e., easy to change. K-means and Fuzzy algorithm study of the resultant cluster is depending on the initial stage where it is chosen either randomly or sequentially.

NECESSITY IN CLUSTERING OF DATA MINING

Scalability: It deals with huge database for high scalable clustering algorithm.

Dissimilar kind of attributes: Algorithm should be able to perform any form of data such as numerical data, categorical data and binary data.

High dimensional: It not only consists of low dimensional but also high dimensional.

Qualification of noisy data: Due to booming data loss occurs. Some algorithm is sensory so it leads to poor quality clusters.

Understandability: The result of clustering should be useful, comprehensible and interpret.

K-MEANS

Partitioning clustering is a technique of K means clustering. It symbolizes the algorithm for dividing the data

into K number of clusters. This performance is dependon the center point of cluster so it is calledas center based clustering technique.It is of spherical shape. It specifies the number of disjoint and non-hierarchical data. Clusters appoint the rank values to the cluster categorical cluster used in mathematical method. Categorical data is transformed into numeric by appointing the rank values.

Clustering design to division of n perception into n clusters and each perception concern that the clusters has mean value. The main concept is to specify k center, for individual clusters. Centers must locate in a delicious way variant location purpose variant result. So the best way is to place them by a long way from each other. The coming step is to hold all points acceptance is like data sets and it has companion with close centroid.

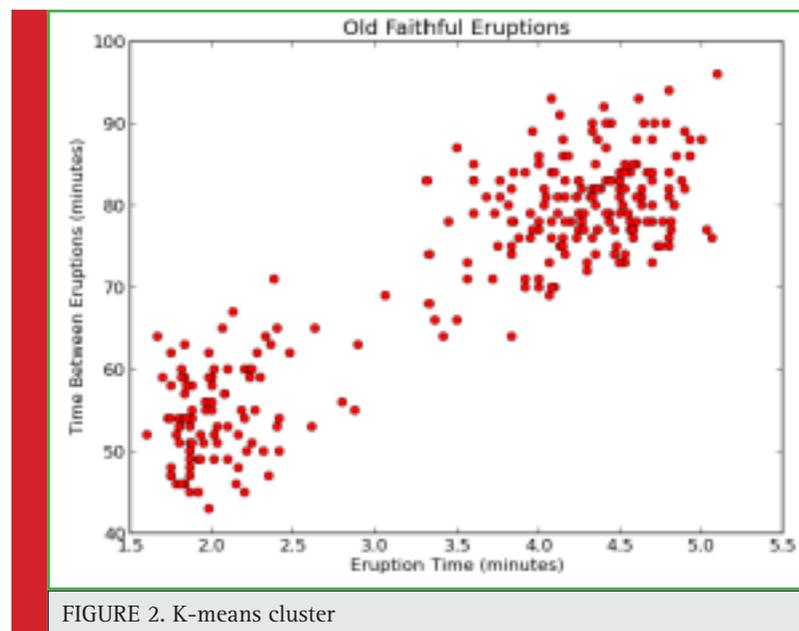
The first step is finished when no points is waiting. Create the loop. At the moment we need to consider K new centroid until no more changes done in the clustering. Iterative algorithm is use in K-mean clustering data mining that minimize the distance of clustering. The time complexity is $O(n*k*i*d)$ Where, K number of centroid, n number of objects, I number of iteration and d number of attributes are discussed Jun Liu et al.

The function is;

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

Where, $\|X_i(j)-C_j\|_2$ the distance between data point X_i and clusters C_j .

In Fig the clusters appers in large amount to make it into simple method they divide clusters by separation of the data points. In Fig b the centroid is proposed-for splitting of data into two clusters. The center based partitioning is simple method that all the datas around



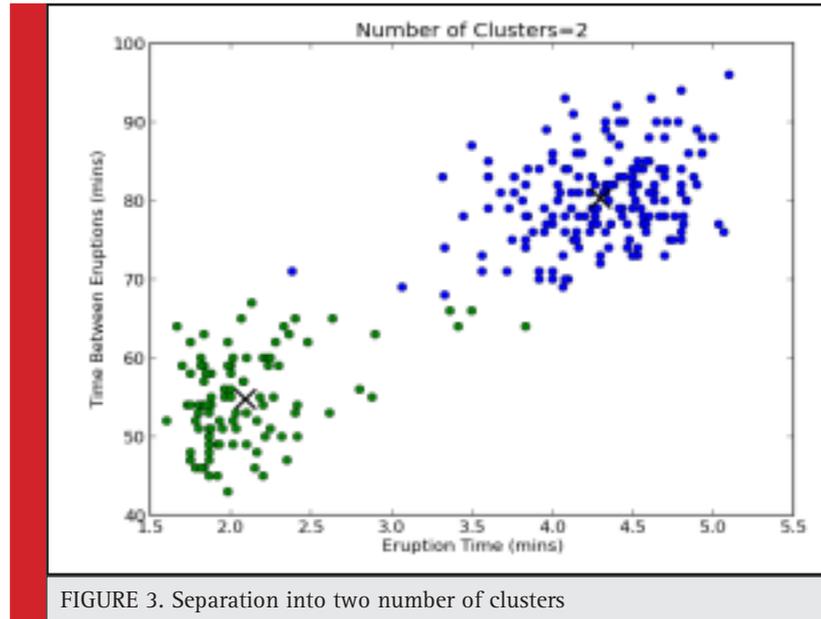


FIGURE 3. Separation into two number of clusters

centroid is clusters. In Fig c the center based is applied for partioing into six clusters the result of K-means is that it cannot partition the data sets used Cheng Chen et al.

ALGORITHM OF K-MEANS

Let $X = \{x_1, x_2, x_3, \dots, x_n\}$ set of data points and $C = \{c_1, c_2, c_3, \dots, c_n\}$ is set of cluster

1. Randomly choose the initial value of clusters. μ_i where $i=1, 2, \dots, k$.
2. Measure of each clusters and data points.
3. From the center of cluster the data point must be minimum of all the clusters.

4. Depend on cluster locate the mean value for all data point.

$$\mu_i = \frac{1}{|c_i|} \sum_{j \in c_i} x_j, \forall i$$

5. Recalculate the data point and the cluster until no data points occurs.

ADVANTAGES

1. Fast, understandable, useful, and robust
2. It is very efficient.
3. If it is well separated it gives the best result.

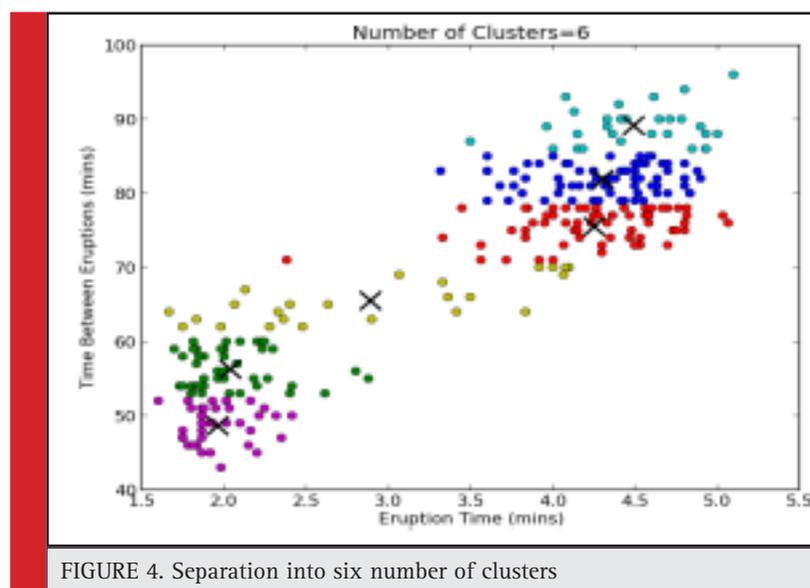


FIGURE 4. Separation into six number of clusters

DISADVANTAGES

1. Unable to handle noisy data
2. It is only applicable when the centroid is determined
3. It will not measure equal weight of each cluster
4. If there is overlapping it could not resolve that there are two clusters.

FUZZY C MEANS

Fuzzy C means is the extended version of Hard C means clustering methods. The basic C Mean optimizing the clustering using Fuzzy clustering VivianaMascardi et al. Fuzzy C means works by selecting membership to every data points compliment to every cluster center it support the distance within cluster center and data point. The data is closet the cluster center more is it membership near to the exact cluster center PavelShvaiko and Jerome Euzenat.

Summation of membership of several data point should be alike. It is unsupervised algorithm that is used expansive range. The complexity of Fuzzy C means is $O(n*d*c^2*i)$.

The edge of the cluster may has less degree of points in center of the clusters. the center based of each points weighted has the degree of the clusters:

$$C_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}$$

ADVANTAGES

Conclusion gives the most excellent result for overlapping of data set and relatively better then K means

Unlike k-means where data point must particularly be part of one cluster center here data point is selected as membership to every cluster center as a conclusion of which data point may belong to more than one clusters.

DISADVANTAGES

1. A priori requirements of the number of clusters
2. With β value may lower so we get the better effect but at the expense of more number of iteration.
3. Euclidean distance scope cannot equally weighted fundamental factors.

DIFFERENCE OF K-MEANS AND FUZZY C MEANS

K means and Fuzzy C means are same way of partitioning clustering. The major alteration is that fuzzy C means clustering weighted must be pointed in every point associate with specific clusters. Since a security model is a theoretical portrayal of the security operations in a working framework, we have to comprehend the semantics of the security operations exactly to build a precise security demonstrate in Valencia-Garcia et al. This, be that as it may, is frequently troublesome on the grounds that the semantics of security operations is unobtrusive and shifts among various working frameworks (such as various kinds of the Unix framework).

In addition, their documentation is now and then deficient or wrong. We advocate depending on the bit code for the development of security models, since the bit code decides the semantics of the security operations Vajargah and M. Shoghi. We receive a two stage prepare: in the initial step, discover all the piece factors that influence the security operations and after that decide the states in the FSA in view of these piece factors; in the second step, decide the moves among these states in the FSA. The initial step should more often than not be possible by hand, however physically doing the second step is regularly relentless and blunder inclined in view of the expansive number of moves discussed Mabu et al. We handle this issue by composing a state-space pilgrim that thoroughly executes all the security operations on

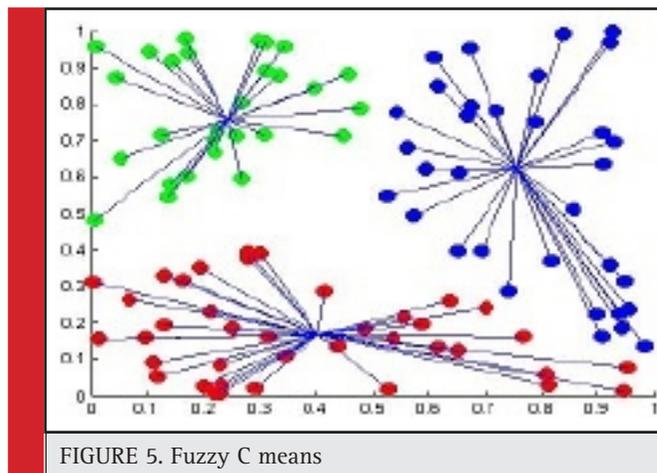


FIGURE 5. Fuzzy C means

the working framework and consequently makes every one of the moves in the security display.

Keeping in mind the end goal to clarify how the directions are permuted, we ought to depict first the structure of the MIPS ISA we utilize. Each direction begins with a 6-bit operation code that incorporates up to three 5-bit registers and, perhaps, a 16-bit prompt esteem. The contention of the SWITCH guideline characterizes some change more than 26 numbers and another stage τ more than 25 numbers. τ is utilized to permute the operation code, and τ is utilized to permute the registers. Quick values can be encoded either by registering them, as depicted by Roles, or by scrambling them utilizing. A wide range of standard family unit contraptions can be adjusted to work in an Data framework. Wi-Fi organize connectors, movement sensors, cameras, receivers and other instrumentation can be implanted in these gadgets to empower them for work in the Internet of Things.

Home mechanization frameworks as of now actualize primitive adaptations of this idea for things like lights, in addition to different gadgets like remote scales and remote circulatory strain screens that each speak to early cases of Data contraptions. Wearable figuring gadgets like watches and glasses are likewise imagined to be enter segments in future Data frameworks. A similar remote correspondence convention like Wi-Fi and Bluetooth normally reach out to the Internet of Things too. As late, the processing situations are changing to the Data, however, the Data administrations have high-security issues, for example, hacking and abusing on the grounds that practically gadgets of Data frameworks associated on the web and transmit information over the system. Data sensors or gadgets are presented to generally high-security danger than the conventional server framework inside firewall or IDS if such terminal gadgets are under assault, all things considered, the entire Data-based administrations can't work typically or work with irregular practices Prof. Neha Soni¹, Dr. Amit Ganatra.

Hence, another pattern proposing coding advisers for comprehending programming shortcomings at the coding stage has risen. Subsequently, if shortcomings are hindered from the product advancement arrange, the critical expenses put resources into perceiving and modifying the product at the execution stage can be spared. What's more, this can contribute extraordinarily to developing programming which is protected from programmers. Current savvy gadgets, Google's Android stage in light of Java, Apple's iOS utilizing Objective-C have been discharged and have developed an overwhelming situation. Additionally, as of late, Microsoft has declared a stage Windows Phone utilizing C# as the principle dialect. These past keen gadget stages have free application improvement condition - create dialect, instruments - and singular attributes per stage. Every

application must experience a characterized improvement handle so as to be executed in a specific stage. The protest code likewise must be made independently relying upon the stage. In this way, with a specific end goal to give administration to contrast stages regardless of the possibility that the substance is the same, contingent upon the question gadget, independent.

The field of virtualization is the utilization of virtual machine (VM) based computerized rights and duplicate security. The two objectives of acquainting VM with computerized rights security are to encode and muddle the program. Compelling the programmers to relocate from a well-known x86 condition to a new also, jumbled virtual condition is proposed to represent a more prominent test in breaking the product duplicate assurance. Our exploration group is attempting to take care of this issue with the point of creating high dependable Data benefits by creating innovation for Data secure programming improvement and execution in view of a compiler and virtual machine.

In this paper, we propose the utilization of a safe compiler and virtual machine with a stack observing technique to create secure Data applications and ensure strange conduct in figuring conditions containing different Data gadgets. A safe compiler was intended for forestalling programming shortcomings in the source code amid the application improvement stage, and it is joined with a conventional compiler and shortcoming analyser to produce the objective code and expel the shortcomings. The safe compiler is actualized in conjunction with a virtual machine, which screens strange conduct, for example, cradle flood assaults or entrusted input information dealing with to secure the framework while the applications are running.

Emerging pattern in the field of virtualization is the utilization of virtual machine (VM) based computerized rights and duplicate security. The two objectives of acquainting VMs with computerized rights security are to encode and muddle the program. Driving the programmers to relocate from a well-known x86 condition to a new furthermore, muddled virtual condition is proposed to represent a more noteworthy test in breaking the product duplicate security. A nonexclusive and self-loader strategy for breaking VM based assurance. It accept that the VM is, comprehensively, an unending circle with a substantial switch explanation called the operation code dispatcher. Each case in this switch explanation is a handler of a specific operation code.

The interpreter is a product device that maps the program directions from the VM dialect to some other dialect picked by the designer, for instance, x86 get together. The VM might be stack based or enrol based. The figure out work is comparative in both cases. Roles calls a dialect that the interpreters decipher the code it

peruses into, i.e., a moderate portrayal (IR). The first step is done once for a specific VM-based assurance, despite what number of programming frameworks are ensured utilizing the same VM. In the second step, the technique removes the VM operation code dispatcher and the muddled directions from the executable code. The operation codes of these directions, be that as it may, do not need to compare to those of the interpreter: the operation codes for each program occasion can be permuted in an unexpected way.

In the third step, the strategy inspects the dispatcher of the VM and uncovers the significance of each operation code from the code executed by its handler. At last, the muddled guidelines. Now, the program is not secured any longer since it can be executed without the VM. Roles additionally applies an arrangement of enhancements to accomplish a program, which is near the first one. Indeed, even by utilizing Roles' suppositions, we contend in such a VM. The product of today trades information in the web condition making it hard to secure the legitimacy of the information and yield. There exists the likelihood of being vindictively assaulted by an arbitrary intruder. This shortcoming has been the immediate reason for programming security occurrences which create huge financial misfortunes.

Security frameworks introduced to keep security occurrences from happening, for the most part, consist of firewalls, client verification framework and so on. In any case, agreeing Gartner's report 75% of programming security episodes happen because of use projects including shortcomings. Along these lines as opposed to making security frameworks for the outer condition all the more firm, developers making programming codes all the more firm is the more central and successful technique for expanding the security levels. Be that as it may, endeavours to lessen the shortcomings of a PC framework are still for the most part one-sided to network servers.

As late, there has been an acknowledgment of this issue and along these lines investigate on secure coding, composing secure codes from the advancement stage, is being done effectively. Particularly, CWE discussed M.Vijayalakshmi, M.Renuka Devi (Common Weakness Enumeration), an association which investigates the shortcomings that can emerge from programming dialect, has investigated and determined the different shortcomings that can happen in the source code creation organize by the distinctive dialects. Likewise, CERT are discussed by Deepti Sisodia et al (Computer Emergency Response Team) characterizes secure coding rules to guarantee secure source code creation.

Fortify SCA is a shortcoming location device created by Fortify. Sustain SCA distinguishes shortcoming in source codes by utilizing dynamic and static investiga-

tion techniques and it bolsters 12 dialects including C/ C++ and Java. Recognized shortcoming data is informed to the clients alongside measurable information. Coverity prevents is a static investigation instrument for source codes. Covertly Prevent shows shortcomings recognized through a whole source code with records. Each rundown contains the area of a shortcoming's event and its cause. MOPS are listed by Krotzsch et al. is a model monitor created in Berkeley. MOPS characterized security powerlessness components as properties, what's more, formalized them utilizing limited robot.

In this manner, all displayed vulnerabilities can be recognized at low examination costs. In any case, there is a roof to the helplessness examination on the grounds that no examination is performed on information streams. Find bugs, created at the University of Maryland as an open source static examination instrument to investigation shortcoming in Java source code which utilized by Overshadow module. Find bugs utilizing the predefined 401 bug examples and it has the trademark that doesn't require the source since it plays out the objective Java program breaking down on byte code level. Also, the bug examples can grow effectively when the new shortcomings are included.

Table 1. Differences of K-means and Fuzzy c means

K-MEANS	FUZZY C MEANS
It entirely concern with one cluster	Each point concern with every cluster.
Leads to fast then fuzzy C means	Leads to slow then K-means
Use distance calculation	Use inverse distance weighting
Coverage epsilon=0	Coverage does not exit

CONCLUSION

In this paper we tends to study two wide used algorithm for data processing functions such as K means and Fuzzy C means due to the combination of Fuzzy C means and K mean algorithm we portioning the data or cluster in useful manner and easy understanding. the main purpose of KF algorithm is gives more valuable performance than the existing in terms of measuring purity, entropy, recall and precision metrics.

REFERENCES

- Ahilandeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455-8907, 2(2): 1-5.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. Interna-

- tional Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 1(1): 1-7.
- Barak S, Modarres M, (2015), Developing an Approach to Evaluate Stocks by Forecasting Effective Features with Data Mining Methods. *Journal of Expert Systems with Applications*, 11: 1325-1339.
- Bo Liu, Keman Huang, Jianqiang Li, MengChu Zhou, (2015), An Incremental and Distributed Inference Method for Large-Scale Ontologies Based on MapReduce Paradigm. *International Journal of Cybernetics*, 45(1).
- Cheng Chen, Zhong Liu, Wei-Hua Lin, Shuang Huang Li, Kai Wang, (2013), Distributed Modeling in a Map Reduce Framework for Data-Driven Traffic Flow Forecasting. *International Journal of Intelligent Transportation Systems*, 14(1).
- Deepti Sisodia, Lokesh Singh, Sheetal Sisodia, Khushboo Saxena, (2012), Clustering Techniques: A Brief Survey of Different Clustering Algorithms. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*. 1(3).
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). Secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455-8907, 2(1): 7-13.
- Jeff Z. Pan, Ian Horrocks, (2007), RDFS (FA): Connecting RDF(S) and OWL DL," *IEEE Transactions on Knowledge and Data Engineering*, 19(2).
- Jun Liu, Feng Liu, N. Ansari, (2014), Monitoring and analyzing big traffic data of a large-scale cellular network with Hadoop, *International Journal of Network*, 28(4), 32-39.
- Kavita Nagar, (2015). Data Mining clustering methods: A reviews. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5(4): 575-579.
- Krotzsch M, Simancik F, Horrocks I, (2014), Description Logics. *International Journal of Intelligent System*, 29(1).
- Mabu S, Obayashi M, Kuremoto T, (2015), Ensemble learning of rule-based evolutionary algorithm using multi-layer perceptron for supporting decisions in stock trading problems. *International Journal of Applied soft computing*, 36: 357-367.
- Neha Soni, Amit Ganatra, (2012). Comparative study of several Clustering Algorithms. *International Journal of Advanced Computer Research*, 2 No.6.
- Pan. J. Z, (2007), A Flexible Ontology Reasoning Architecture for the Semantic Web. *Journal of Knowledge and Data Eng*, 19(2).
- Pavel Shvaiko, Jerome Euzenat, (2013), Ontology Matching: State of the Art and Future Challenges," *Knowledge and Data Eng*, 25(1).
- Pei J, Han J, MortazaviAsl B, Wang J, Pinto H, Chen Q, Dayal U, Hsu M. C, (2004) Mining sequential patterns by pattern-growth: The Prefix Span approach. *IEEE Transactions on Knowledge and Data Engineering*, 16(11): 1424-1440.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Vajargah K.F, Shoghi M, (2015), Simulation of Stochastic Differential Equation of Geometric Brownian Motion by Quasi-Monte Carlo Method and its Application in Prediction of Total Index of Stock Market and Value at Risk. *Mathematical Sciences*, 9(3): 115-125.
- Valencia-Garcia R, Garcia-Sanchez F, Castellanos-Nieves D, Fernandez-Breis J. T, (2011) OWL Path: An OWL Ontology-Guided Query Editor. *Journal of Systems, Man and Cybernetics*, 41(1): 121 - 136.
- Vijayalakshmi M, Renuka Devi M, (2012). A Survey of Different Issue of Different clustering Algorithms Used in Large Datasets," *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(3).
- Viviana Mascardi, Angela Locoro, Paolo Rosso, (2010), Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation. *Journal of Knowledge and Data Eng*, 22(5).

Blocking Abusive and Analysis of Tweets in Twitter Social Network Using NLP in Real-Time

R. Rahin Batcha¹, K. Prem Kumar², N. Danapaquiame³, J. Arumugam^{*4} and D. Saravanan⁵

^{1,2,3,4}Dept of CSE, Sri Manakula Vinayagar Engineering College, Puducherry, India

⁵Dept of CSE, KL University, Andhra Pradesh, India

ABSTRACT

In day-to-day life, everyone need to share their thoughts, opinion, feelings and information etc., between their friends, relatives and society in fast and rapid speed of data exchange range rate for every seconds. But we cannot measure the originality of the message and message is having any false, positive or negative, castigating words or vulgar words what they share. In existing social network, Twitter Account the linguistic communication process is employed to move with the Human words and machine language to spot and to rate the comments shows whether or not the announce comment is positive or negative/neutral. During this paper, the work is predicated on sentimental analysis for rating and blocks the castigating words used by the twitter followers from their Bag of Words. Finally, Dashboard visualizing tools is employed to represent the restricted words from the announce comments of twitter followers.

KEY WORDS: SOCIAL NETWORK, NATURAL LANGUAGE PROCESSING, MACHINE LANGUAGE, CASTIGATING WORDS, BAG OF WORDS AND DASHBOARD

INTRODUCTION

There are a unit variety of social networking services and a lot of usage today condition. Social networking provides each benefits and downsides. Folks use social network just for communication purpose. Twitter is one in all the usually used social media between short times it gained a worldwide quality, quite 350 countless posts area unit tweeted per day. The main downside of existing and the tweets may be visited by anyone via com-

mented. Thus twitter doesn't offer privacy and security. The networking problems became a matter of great concern. The user feels insecure since victimization twitters because it shares all the non-public data in common with third parties. In our planned system we tend to use tongue process (NLP) to spot malicious feedback ratings. IP acts as a sleuthing technique that detects the negative or malicious comments and conjointly blocks the abused or negative comments stopping it from sharing the non-public details to the public journal (Chen et al. 2010).

ARTICLE INFORMATION:

*Corresponding Author: batcha35@gmail.com

Received 27th April, 2017

Accepted after revision 29th July, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF : 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

Twitter is employed as an excellent tool for twitting messages and sharing some universal connected topic or discussions it's more used for expressing feelings. Twitter once cantered in business purpose it's terribly helpful. However once it deals with confidentiality it lacks stability. Twitter whereas thought of as an excellent business platform consumes more help in official usage. By mistreatment twitter an oversized cluster of communication is obtainable and target markets area unit insured that is helpful for the business purchasers to develop or exaggerate their business. In different aspect twitter is that the first website that enables spammers. Every sender ought to be known, remove and filtered from the list time to time. The sender within the twitter might enter the site and misuse the shared information of different users. They'll conjointly pass negative comments and unfold malicious feedback (Zheng *et al.* 2007)

Twitter, Facebook and other social media encourage frequent user expressions of their thoughts, opinions and random details of their lives. Tweets and status updates range from important events to inane comments. Most messages contain little informational value but the aggregation of millions of messages can generate important knowledge. Several Twitter studies have demonstrated that aggregating millions of messages can provide valuable insights into a population. Analyzing user messages in social media will live totally different population characteristics, as well as public health measures. As an example, recent work has correlative Twitter messages with respiratory disease rates within the United States; however this has for the most part been the extent of mining Twitter for public health. With the ascension of social media, users particularly adolescents are defrayal vital quantity of your time on numerous social networking sites to attach with others, to share data, and to pursue common interests (Maximilien *et al.* 2012).

In 2011, seventieth of teens use social media sites on a day to day and nearly one in four teens hit their favourite social-media sites ten or additional times on a daily basis. Whereas adolescents enjoy their use of social media by interacting with and learning from others, they're conjointly at the chance of being exposed to massive amounts of offensive on-line contents. Scan Safe's monthly "Global Threat Report" found that up to eightieth of blogs contained offensive contents and seventy four enclosed smut within the format of image, video, or offensive languages. Additionally, cyber-bullying happens via offensive messages denote on social media. It has been found that nineteen of teens report that somebody has written or announce mean or embarrassing things concerning them on social networking sites. As adolescents are additional probably to be negatively tormented by biased and harmful contents than

adults, detection on-line offensive contents to shield adolescent on-line safety become associate pressing task (Malik *et al.* 2007).

RELATED WORK

A balanced stage ought to be maintained with obtaining friendly along with your followers. Once this exceeds could cause major drawback and ends in negative commenting and abused words. This happens typically in twitter. As an example, think about A and B WHO shares posts or tweet some info concerning politics or post tweets commenting movies. Therein case a quarrel or fight could occur between the 2. This causes disturbance to different user WHO read the comments. There are ninetieth of possibilities for different user to publish the fight and comment concerning it in comment. This promotes a theoretical state and therefore the state is inescapable. In existing system bloom filter wherever accustomed notice malicious feedback rates and forestall it from abusing remark and attacks. The abusing remark are measured and rated in success magnitude relation recommending the service. This checks the comments of the online service that successively is utilized metric and investigates whether or not to be counselled to the user or not.

Then, the metric sends Associate in nursing authentication that the comments delivery for you is abused. Then the user might tend to erase the comments (Xu *et al.* 2007). This technique might not facilitate all told times. If the authentication delivered the ends up in the delay or corrupted this could have an effect on the user details. ACS (abusing comment system depends on previous info establishing bond among unknown user. ACS of internet services could be an issue that no history of recent comers is gift. In twitter there could also be pretend tweets and pretend addresses that allot malicious feedback to different users. Malicious feedback in twitters could be a desirable assessment of associate attribute delineate on single entity relating observations inflicting issues. Quite one supply square measure simulated etymologizing abused comments. A respect to collective perception is ensured permitting the service requesters for providing abusing comments (Ardagna *et al.* 2007). Web Service recommendation systems may be utilized to recommend the best internet service for satisfying user's needs.

Service recommendations are useful for users once 2 or additional net services have identical practicality however totally different tongue process (NLP) performance. Information science is outlined as a collection of non-functional properties, together with Abusing comments, latency, reliability, etc. once multiple net services developed AC provides identical practicality, and then a

feedback rating demand is used as a secondary criterion for service choice. Language process could be a set of non-functional attributes like service latency, throughput, reliability, and handiness. Service computing are used with multiple and separate systems adopting many business domains as a package practicality cloth routines (Conner *et al.* 2009). Whereas exploitation twitters the privacy is in peril solely attributable to the social media that affects the pride of an individual.

B. Literature Audit

A major concern once incorporating giant sets of various n-gram options for sentiment classification is that the presence of clatter, irrelevant, and redundant attributes. These considerations will typically create it tough to harness the increased discriminatory potential of extended feature sets. (Abbasi *et al.*, 2011) proposed a rule-based variable text feature choice methodology referred to as Feature Relation Network (FRN) that considers linguistics data and additionally leverages

Furthermore, by incorporating grammar data concerning n-gram relations, FRN is {in a position is ready} to pick out options in a lot of computationally economical manner than several variable and hybrid techniques. Twitter sentiment analysis has become wide in style. However, stable Twitter sentiment classification performance remains elusive because of many issues: significant class imbalance during a multi-class drawback, delineative richness problems for sentiment cues, and therefore the use of numerous conversational linguistic patterns. These problems square measure problematic since several varieties of social media analytics consider correct underlying Twitter sentiments. Consequently, a text analytics framework is planned by (Ammar Hassan, and Ahmed Abbasi, Daniel Zeng 2013) for Twitter sentiment analysis. The framework uses associate degree elaborate bootstrapping ensemble to quell category imbalance and delineative richness problems. Consequently, it's able to build sentiment statistic that square measure higher able to mirror events eliciting robust positive and negative sentiments from users.

Fast algorithms capable of process huge volumes of knowledge square measure currently needed within the field of power systems. It presents a completely unique parallel detrended fluctuation analysis (PDFFA) approach for quick event detection on huge volumes of PMU information. The algorithmic rule is evaluated mistreatment information from put in PMUs on the transmission of nice kingdom from the aspects of hurrying, quantifiability, and accuracy. A revision to the law is then proposed by (Zheng and Dagnino, 2014) suggesting enhancements to its capability to analyse the performance gain in computation once parallelizing information intensive applications in a very cluster computing surroundings.

The overwhelming majority of existing approaches to opinion feature extraction have confidence mining patterns solely from one review corpus, ignoring the non-trivial disparities in word spatial arrangement characteristics of opinion options across completely different corporates, (Hai *et al.* 2014) planned a completely unique technique to spot opinion options from on-line reviews by exploiting the distinction in opinion feature statistics across 2 corpora, one domain-specific corpus (*i.e.*, the given review corpus) and one domain-independent corpus (*i.e.*, the different corpus).

Experimental results on 2 real-world review domains show the planned IEDR approach to outdo many different well-established ways in distinctive opinion options. The current Analytics tools and models that square measure accessible within the market square measure terribly pricey, unable to handle huge knowledge and fewer secure. The standard Analytics systems take a protracted time to come back up with results, thus it's not useful to use for Real Time Analytics. So, (Rajurkar and Goudar, 2015) planned the work resolve of these issues by combining the Apache Open supply platform that solves the problems of Real Time Analytics mistreatment HADOOP. The HANDOOP is versatile and ascendible design. The planned work I - based mostly upon the development of combination of open supply package in conjunction with artefact hardware which will increase the profit of IT trade.

To address issues on children's access to offensive content over web, directors of social media typically manually review on-line contents to find and delete offensive materials. However, the manual review tasks of distinctive offensive contents square measure labour intensive, time intense, and so not property and scalable truly. Some automatic content filtering code packages, like Append and web Security Suite, are developed to find and filter on-line offensive contents. Most of them merely blocked WebPages and paragraphs that contained dirty words. These word-based approaches not solely have an effect on the readability and usefulness of internet sites; however conjointly fail to spot delicate offensive messages. For instance, underneath these standard approaches, the sentence "you area unit such a crying baby" won't be known as offensive content, as a result of none of its words is enclosed generally offensive lexicons.

In addition, the false positive rate of those word-based detection approaches is usually high, because of the word ambiguity downside, *i.e.*, identical word will have terribly totally different meanings in numerous contexts. Moreover, existing strategies treat every message as a freelance instance while not tracing the supply of offensive contents. Sentiment analysis or opinion mining are often defined as a specific application of information mining,

that aims to mixture and extract emotions and feelings from differing types of documents (Nepal *et al.* 2009).

The number information of knowledge of information out there on the net is growing exponentially; these data, however, square measure chiefly in an unstructured format and thence aren't machine method in a position and machine explicable. Thus graph-mining and linguistic communication process (NLP) techniques might contribute to the distillation of data and opinions from the massive quantity of data gift within the internet. Sentiment analysis will enhance the capabilities of client relationship management and recommendation systems by permitting one, as an example, to search out that options customers are notably inquisitive about or to exclude from ads things that have received unfavourable feedback. Likewise, it can be used in social communication to enhance anti spam systems. Business intelligence can even like sentiment analysis (Jurca *et al.* 2007). Since predicting the angle of the general public toward a complete or a product has become of crucial importance for corporations, Associate in nursing increasing quantity of cash has been endowed in promoting ways involving opinion and sentiment mining. That state of affairs crystal rectifier to sentic computing, that tackles those crucial problems by exploiting emotive reasonable reasoning (i.e., the per se human capability to interpret the psychological feature and emotive data related to natural language). Above all, sentic computing leverages a reasonable knowledge domain designed through crowd sourcing.

Common sense is helpful in many various applied science applications, together with knowledge mental image, text recognition, and human-computer interaction. During this context, logic is employed to bridge the linguistics gap between word-level linguistic communication knowledge and therefore the concept-level opinions sent by these (N Limam *et al.* 2010). To perform affection commonsensible reasoning, information info is needed for storage and extraction of the affection and emotional data related to word and multiword expressions.

Graph-mining and spatiality reduction techniques are used on a knowledge domain obtained by the mixing of idea internet, a directed graph illustration of commonsensible information, with Word Net-Affect, a linguistic resource for the lexical illustration of feelings. Not like Word Net-Affect, Sentic internet associated its extended versions exploit an ensemble of common and commonsensible information to travel on the far side word-level opinion mining.

With the increase of the Social internet, there are unit currently variant humans providing their information on-line, which implies that the data is hold on, searchable, and simply shared. This trend has created

associate degree maintained a system of participation, wherever the worth is formed by the aggregation of the many individual user contributions. Such contributions, however, area unit meant for human consumption and thus area unit hardly accessible and method in a position by computers (Anusha *et al.* 2015). Creating sense of the massive quantity of social information accessible on the online needs the adoption of novel approaches to linguistic communication understanding that may provides a structure to such information in a very approach that they'll be a lot of simply aggregative and analyzed.

FEASIBILITY STUDY

Feasibility is the determination of whether or not a project is worth doing. The processes is followed in making this determination is called a feasibility study. Feasibility study is the test of system proposal according to its workability, Impact on the organisation ability to meet user's needs and effective use of resources. The result of feasibility study is a formal proposal. This is simply a report – a formal document detailing the nature and scope of the proposed solution. The main objective of a feasibility study is to test the technical, social and economic feasibility of developing a computer system. This is done by investigation the existing system in the area under investigation and generating ideas about a new system. On studying the feasibility of the system, three major considerations are dealt with, to find whether the automation of the system is feasible. They are discussed as follows.

A. Technical Feasibility Study

A system threat can be developed technically and that will be used if installed must still be a good invested for the organisation. The assessment of technical feasibility must be based on an outline design on system requirements in terms of inputs, outputs, files, programs, procedures. Technical feasibility centres on the existing computer system and to what extent it can support the proposed system provides adequate response to inquiries and whether it can be expanded if developed (Dharshini *et al.* 2016). The current project is to design so as to fit to the expectations of various categories of people concerned with it. Besides some technical experts who also have the computer knowledge are to be trained over the project enabling them to take care of the technical problems. The system is developed to meet the demands of the existing. The system is also reliable and easy to use. So it is found that this project is technically practicable keeping the clients requirements in mind.

B. The Java Platform

A platform is the hardware or software environment in which a program runs. The java platform differs from

most other platforms in that it's a software-only platform that runs on top of other, hardware-based platforms. Most other platforms are described as a combination of hardware and operating system.

C. MVC Framework of Java

MVC framework is used to separate the data access layer, business logic code and the graphical user interface that has to be defined and designed to let the user interact with the application. This application has three parts,

- Model - this part of the framework is to store the data of the application, such as databases, text data, files and/or other web resources.
- View - this is the graphical user interface of the application. That would contain different buttons, text boxes and other controls to let the user interact with the application to complete his projects depending on sort of the software he is using (Govindharaj *et al.* 2016).
- Controller - the actual back-end code constitutes the controller of the framework. A controller controls the data coming from the users, or going to the user from a model.

D. NETBEANS IDE

The java code has been written in an easy to use "Net Beans IDE": which is a reusable framework for simplifying the development of other desktop applications (Ahi-landeswari *et al.* 2016). When an application based on the Net Beans Platform is run, the platform's Main class is executed. Available modules are located, placed in an in-memory registry, and the modules' start up tasks is executed. Generally, a module's code is loaded into memory only as it is needed. Applications can install modules dynamically. Any application can include the Update Centre module to allow users of the application to download digitally signed upgrades and new features directly into the running application. Reinstalling an upgrade or a new release does not force users to download the entire application again (Gayathri *et al.* 2016).

E. SERVLETS

Servlets are Java programming language classes that dynamically process requests and construct responses. *JSP pages* are text-based documents that execute as servlets but allow a more natural approach to creating static content. Although servlets and JSP pages can be used interchangeably, each has its own strengths. Servlets are best suited for service-oriented applications (web service endpoints are implemented as servlets) and the control functions of a presentation-oriented application, such as dispatching requests and handling non-textual data. JSP pages are more appropriate for gener-

ating text-based mark up such as HTML, Scalable Vector Graphics (SVG), Wireless Mark up Language (WML), and XML.

F. JSP

Java Server Pages (JSP) is a server side Java Technology that allows software developers to create dynamically generated web pages, with HTML, XML, or other document types, in response to web client request to a Java Web Application container (server) (Shanmugam *et al.* 2017).

G. Oracle 12 c (server class)

Oracle Database (commonly referred to as Oracle RDBMS or simply as Oracle) is an object-relational database management system produced and marketed by Oracle Corporation. The Oracle RDBMS stores data logically in the form of table spaces and physically in the form of data files ("data files"). Table spaces can contain various types of memory segments, such as Data Segments, Index Segments, etc. Segments in turn comprise one or more extents. Extents comprise groups of contiguous data blocks. Data blocks from the basic units of data storage.

H. Economic Feasibility Study

The technique of cost benefit analysis is often used as a basis for assessing economic feasibility. Economic feasibility deals with the analysis of costs against benefits (i.e.) whether the benefits to the enjoyed due to the new system are worthy when compared to the costs to be spent on the system. Economic analysis is the nose frequently used technique for evaluating the cost effectiveness of the proposed project (Rajadurai R *et al.* 2017). More commonly known as cost / benefit analysis, the procedure is to determine whether the project have the benefits and savings. Further compared with the existing-costs in the manual procedure, the current projects involve less investment.

I. Operational Feasibility Study

Proposed projects are beneficial only if they can be turned into information systems that will meet the company's operating requirements. Simply stated, this test of feasibility asks if the system will work when it is developed and installed. There are questions that will help to test the operational feasibility of a project. Will the proposed system cause harm? The project is made such that a man with ordinance prudence will be able to operate the system without much difficulty if he is trained properly. The system is completely Menu driven and a warning has been set so as to inform the user that there is a flaw in the operation. The operational feasibility depends on the people who operate it. Due to the interactive nature of this project, the users would find it very

easy to work with it. Keeping all this in mind the project is found to be operationally feasible. Several interrelated managerial and financial factors are considered while evaluating the feasibility of this project (Sasidevi et al. 2017). It is mainly related to human organisational and political aspects.

MODULE DESCRIPTION

A. Fetching raw twitter data

First of all, we need to fetch the twitter information from the first twitter API. Real time Streaming information is obtained from twitter Public API. Association to twitter is established exploitation Twitter 4j API gift in spark streaming. Twitter API uses client Key, client Secret, Access Key, Access Token to attach and attest twitter that is provided to code at run time.

B. Cleansing raw twitter data

Hash tags and @mentions are declared within the code as filter so as to obtained target supply knowledge from twitter. Solely necessary fields like SCREEN NAME, TEXT, CREATED_DATE, LANGUAGE, and RETWEET_COUNT are extracted from the raw twitter knowledge obtained from the general public API. All junk character and alternative unessential symbols aside from English words are removed so model coaching may be done expeditiously.

C. Pre - processing the data

Pre-processing is that the main step in methodological processes. Usually information pre-processing is reworking of information into a lucid format. During this project, made-to-order Schema with some further fields that square measure necessary for later parts square measure engineered.

D. Analytics

In this module, three processes are needed to be done.

- (i) Creation, (ii) Training, (iii) Prediction.

Model Creation

All the Analytical models are designed victimization MLIB library gift in artificial language .Since Sentiment Analysis is classification downside, classification model is constructed with Naïve mathematician Model because the alternative of algorithmic program. Addition libraries like Lucien English, Hashing TF, Stemmer are wont to additional take away words sort of a, the, or and some perennial words that has no weight age is removed.

Model training

Using around one million historical records of clean twitter information, that is comparable to the \$64000

time information obtained from twitter area unit accustomed train the model created within the on top of step. Once the model is trained and once potency is evaluated with many various classification algorithms best ensuing model like Naïve Thomas Bayes Model is saved to the disk. During this part we have a tendency to store solely Meta data which may be employed in later stages.

Prediction module

In this module the important time information is cleaned and sent to stemmer for removing words that has no weight age. Prediction operate is written and known as that successively can used the already trained and saved model for prediction in real time. All the output of prediction operates with some extra info is saved in to MYSQL tables for information visualisation and report-age.

E. Data visualization

This module is to check the date that area gathered in the on top of method. The gathered data are then visualized in form of graph or charts.

The two kind of development used in this module are Front End Development

- Front End Development
- Back End Development

Front end development: Front End View is further divided into 2 parts

- Archival Stats
- Real Time Tweets with sentiment

The languages employed in this project for forepart are hypertext mark-up language, CSS and JSP. Dashboard skeletal layouts are created. Alternative Navigation links to and between home and real time tweets with sentiment page is formed. Charts and Dashboards are set to automotive vehicle refresh to point out the \$64000 time information changes with stripped delay. Dashboards are chiefly wont to visualize the info in a very linear manner.

Back end development: Data is sourced from 2 different MYSQL tables

- Real Time Table
- Archive Table

Servlet with JDBC affiliation code is written to supply knowledge from MYSQL tables. knowledge is classified and drill down is formed with the assistance of DC intrinsic functionalities Charts and alternative metrics area unit created exploitation D3 and Dc JavaScript libraries. Necessary operate that area unit to be known as through frontend is formed to point out dashboard seamlessly .

F. Data integration

The main purpose of this module is to integrate the info from the information. Servlet is employed because the Integration layer to drag the info from the MYSQL tables to dashboards sporadically with smallest delay.

Real-time tables are truncated for each new knowledge force from twitter .Stored Procedure is formed to manoeuvre the info from real agenda to deposit table. Stored Procedure pulls the info before it gets truncated. All the backend knowledge movement is automated/ managed by writing schedule script and line of work the keep procedure sporadically in MYSQL.

All the side elements are sporadically rested primarily based out on the predefined calculated interval set to show/populate the dashboard with terribly smallest delay. Integration is otherwise termed as programming.

IMPLEMENTATION AND RESULTS

Malicious feedbacks are rated consequently with ACS (abusing computing system). ACS acts as a police investigation sector during which the comments are rated and measured. The comments tweeted are separated as division.

The abused and malicious feedback will be classified by measuring:

- Positive feedbacks
- Negative feedbacks

Bar char based on the tweet score is illustrated in figure 1 in twitter the feedback ratings are calculated by distinctive the indications. The indications are at the same time recruited and verified. Abusing comments are boots

trappers that access recently deployed services. The rating defines a theoretical analysis during which measurements are profiled. The name feedback is additionally measured other ways. If it shows positive indication then the comments are positive and also the feedback is in stable condition. There's no want for any malicious hindrance during this stage of indication. If it indicates negative indication then the comments are negative and also the feedback isn't in a very stable condition. During this stage there's want of security and privacy protection. Previous studies on email spam found that click-through and conversion rates considerably varied. The estimate click-through rates (i.e., the number of people who arrive at the website having clicked the link in the email) ranged from 0.003% to 0.02%.The 2010 study on Twitter spam estimated the click-through rate at 0.13%, which suggests that the click-through rate for Twitter spam was two times higher in magnitude higher than for email spam.

The Trend Micro Web Reputation Technology has a component that allows users to obtain malicious anonymized feedback if they wish to. We examined the feedback data to determine which malicious URLs embedded in tweets were clicked. However, without access to the platform's backend infrastructure, it was difficult to determine the absolute

Twitter spam click-through rate. However, we were able to sensibly compare the relative effectiveness of malicious campaigns and determined that there was great variability across campaigns.

We classified the groups and domains we analyzed in Section V into the following categories:

Malware: Tweets with embedded links that led to malware-distribution websites.

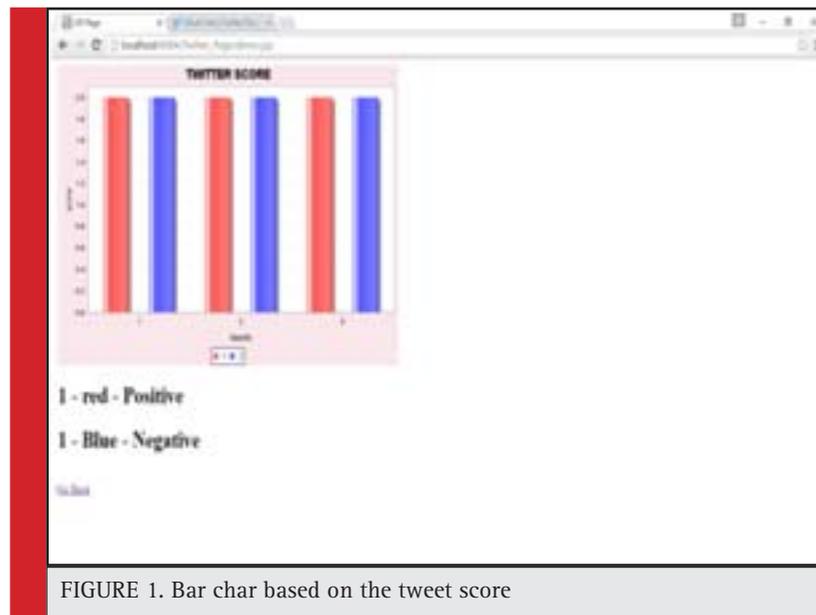


FIGURE 1. Bar char based on the tweet score

Traditional phishing: Tweets with embedded links that led to phishing websites.

Twitter-specific scam: Tweets that led to the Twitter follower scam described in Section VI.

Spam: Tweets that were sent by groups or domains involved in spam distribution. We split this category into three subcategories because the different spam flavours had distinct characteristics. The subcategories include the following:

- Traditional spam
- Spam with shortened URLs
- Russian spam, including the most prolific type, Russian-138 spam, described in Section VI
- Spam related to a viral Japanese campaign

There were enormous variations in the effectiveness of the different approaches to twitter spamming. For example, the viral Japanese campaign was approximately 5,000 times more effective than the Russian spam campaign. Public Tweets PT: (the default setting) noble metal could be a default setting that are visible to anyone, whether or not or not they need an account. They are doing not think about any followers or don't calculate the user desires. Publically tweets the comments flow also are published. Protected Tweets: may solely be visible to your friends and approves solely Twitter followers. Solely specific users are granted permission to follow the tweets we have a tendency to post.

If others attempt to interrupt a sign to admin are going to be delivered and also the admin might block that individual user tweet ID and details. The projected solutions use totally different techniques measurement internet service reputations supported user feedback ratings relating to abusing words or comments. We have a tendency to validate our projected malicious feedback rating hindrance theme through theoretical analysis, and conjointly appraise our projected measure.

A name derivation model had conjointly been projected to mixture feedbacks into a name price that higher reflects the behaviour of the service at choice time. The projected methodology reduces the abnormality of the name measure. The success quantitative relation of the net service recommendation will be improved.

In service-oriented environments wherever honest and malicious service suppliers co-exist, finding the precise balance between fairness and accuracy for abusing comments bootstrapping is non-trivial. For example, a malicious service supplier might decide to clear its (negative) Abusing comments history by discarding its original identity and getting into the system with a replacement one. In distinction, a service supplier could also be getting into the system for the primary time with none malicious motives. This may even be avoided and protection will be provided.

NATURAL LANGUAGE PROCESSING (NLP) AS MALICIOUS FEEDBACK DETECTOR

Natural language processing (NLP) is a field of computer science, artificial intelligence and computational linguistics concerned with the interactions between computers and human (natural) languages, and, in particular, concerned with programming computers to fruitfully process large natural language corpora. Challenges in natural language processing frequently involve natural language understanding, natural language generation (frequently from formal, machine-readable logical forms), connecting language and machine perception, managing human-computer dialog systems, or some combination thereof. Since the so-called "statistical revolution" in the late 1980s and mid 1990s, much Natural Language Processing research has relied heavily on learning. Formerly, many language-processing tasks typically involved the direct hand coding of rules, which is not in general robust to natural language variation.

The machine-learning paradigm calls instead for using statistical inference to automatically learn such rules through the analysis of large corpora of typical real-world examples (a corpus (plural, «corpora») is a set of documents, possibly with human or computer annotations). Many different classes of machine learning algorithms have been applied to NLP tasks. These algorithms take as input a large set of «features» that are generated from the input data. Increasingly, however, research has focused on statistical models, which make soft, probabilistic decisions based on attaching real-valued weights to each input feature. Such models have the advantage that they can express the relative certainty of many different possible answers rather than only one, producing more reliable results when such a model is included as a component of a larger system.

Natural Language process is outlined as a language detector otherwise called malicious feedback detector. The detector is employed to spot the abusing comments provided by alternative batty. NLP Natural Language Process NLP human language technology information science informatics information processing IP acts as a language processing detector. Natural language processing initially trace for harsh or abusing comments it validate through indication meter. In indication meter if the indication points to positive then the comments valid square measure positive therefore the server permits the comments to be announce. If the indication points towards negative sign then there's detection of abused word usage. in this case the server are blocked the actual comment is unbroken for verification within the main server. If the admin allows/permit then the verification is completed once more and also the specific comment is reevaluated. However during this method it ne'er permits

the comment to submit to or to be announced. Steps followed in detection of malicious feedback/abusing word.

- Indication check: Check for the indication positive permits user to post comment. Negative indication then don't allow user to post the malicious feedback.
- Admin verification: promptly once the malicious feedback is detected pin purpose the comment. Picks out the abusing comment sent to main server for Admin verification.
- Provision to permit: the abusing comment is detected through NLP natural language process NLP human language technology information science informatics information processing IP and also the language processing transfers mechanically to admin for verification not permitting the server to tweet the abusing comment/post within the twitter.
- Block the post: if the doubt is processed and also the comment is abused then promptly the admin block the post/malicious feedback from the user. The admin conjointly perceive and block or mark list the actual user from twitter.

The development of NLP applications is challenging because computers traditionally require humans to "speak" to them in a programming language that is precise, unambiguous and highly structured or, perhaps through a limited number of clearly-enunciated voice commands. Human speech, however, is not always precise - it is often ambiguous and the linguistic structure can depend on many complex variables, including slang, regional dialects and social context. Current approaches to NLP are based on machine learning, a type of artificial intelligence that examines and uses patterns in data to improve a program's own understanding. Most of the research being done on natural language processing revolves around search, especially enterprise search. Common NLP tasks in software programs today include:

- Sentence segmentation, part-of-speech tagging and parsing.
- Deep analytics.
- Named entity extraction.
- Co-reference resolution.

The advantage of natural language processing can be seen when considering the following two statements: "Cloud computing insurance should be part of every service level agreement" and "A good SLA ensures an easier night's sleep even in the cloud." If you use natural language processing for search, the program will recognise that cloud computing is an entity, that cloud is an abbreviated form of cloud computing and that SLA is an industry acronym for service level agreement.

FUTURE ENHANCEMENT

By this the malicious feedback won't be announce. This protects the opposite user from quarrel and fights. Once the abusing comment or malicious feedback is blocked there won't be any drawback incidence relating the feedback. Sentimental Analysis may also be used for different social networks like Facebook, WhatsApp, thus it helps to cut back the abusive comments on social media thereby increasing the peace and healthy setting.

It can also be may be half-tracked for on-line product and services to review the genuineness of the tweets in order that different on-line websites can build use of social mediate tweets area unit extracted as a individual task within the planned, whereas we are able to enhance it to review the tweet whenever it's denote. Language process is completed just for English, we are able to enhance for different languages reference to the situation of the user. It may also depict mood of the user with whom one is chatting and once the chat exceeds the morality, the involved user is warned.

CONCLUSION

The service comment (positive & negative comments) score is typically calculated victimisation feedback ratings provided by users. Though the name measuring of net service has been studied within the recent history, current malicious and subjective user reaction ratings usually result in a bias that degrades the performance of the service recommendation system. During this paper, we have a tendency to propose a safer comment passing for twitter by victimisation language process (NLP) measuring approach for net service recommendations. During this projected system the feedback measuring within the twitter approach utilizes malicious feedback rating detection and additionally feedback similarity computation to live the name and harmful quarrel ob net services in common. The hindrance theme also can establish the informatics address with abusing/offending comment ratings and block them victimisation the human language technology [12]. Human language technology as a sleuthing technique finds the incorrect comment and interact the comment to admin for verification. And blocks the abusing feedback ratings within the user net suggested system and defend the user.

REFERENCES

Ahilandeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. International Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 2(2): 1-5.

- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Ardagna D, Pernici B. (2007). Adaptive service composition inflexible processes. *IEEE Transactions on Software Engineering*, 33(6): 369-384.
- Chen X, Liu X, Huang Z, Sun H, (2010), Region KNN: A scalable hybrid collaborative filtering algorithm for personalized web service recommendation. In: Proc. of the 8th IEEE International Conference on Web Services (ICWS'10): 9- 16.
- Conner W, Iyengar A, Mikalsen T, Rouvellou I, KNahrstedt. (2009), A trust management framework for service oriented environments. In: Proc. of the 18th international conference on World Wide Web (WWW'09): 891-900.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Jurca R, Faltings B, Binder W. (2007), Reliable QoS monitoring based on client feedback. In: Proc. of the 16th international conference on World Wide Web (WWW'07): 1003-1012.
- Malik Z, Bouguettaya A. (2007). Evaluatingrater credibility for reputation assessment of web services. In: Proc. of the 8th International Conference on Web Information Systems Engineering (WISE'07): 38-49
- Maximilien E M, Singh M P, (2012), Conceptual model of web service reputation. *ACM SIGMOD Record*, 31(4): 36-41.
- N Limam, R Boutaba. (2010). Assessing Software Service Quality and Trustworthiness at Selection Time. *IEEE Transactions on Software Engineering*, 36(4): 559-574.
- Nepal S, Malik Z, Bouguettaya A. (2009). Reputation Propagation in Composite Services. In: Proc. of the IEEE International Conference on Web Services (ICWS'09): 295-302.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Xu Z, Martin P, Powley W, Zulkernine F. (2007). Reputation enhanced QoS-based web services discovery. In: Proc. of the IEEE International Conference on Web Services (ICWS'07): 249-256.
- Zheng Z, Lyu M R, (2007) Collaborative reliability prediction of service-oriented systems, In: Proc. of the 32nd ACM / IEEE International Conference on Software Engineering (ICSE): 35-44.

Frequent Item set Using Abundant Data on Hadoop Clusters in Big Data

N. Danapaquiame¹, V. Balaji², R. Gayathri³, E. Kodhai⁴ and G. Sambasivam⁵

^{1,2,3,4}Department of CSE, Sri Manakula Vinayagar Engineering College, Pondicherry, India

⁵Department of CSE, KL University, Andhra Pradesh, India

ABSTRACT

Big data mining faces lots of challenges within the huge information generation. The rule mining is an important place of studies within the field of information mining. Association rule mining set of rules isn't always sufficient to procedure massive data sets. Apriori algorithm has limitations just like the excessive I/O load and coffee overall performance. The FP-growth set of rules additionally has certain obstacles like less inner memory. Mining the frequent item set in the dynamic eventualities is a difficult mission. To overcome these issues a parallelized method the usage of the map reduce framework has been used. The mining set of rules has been applied the usage of the Hadoop clusters.

KEY WORDS: FP-GROWTH ALGORITHM, BIG DATA, DATA MINING, FREQUENT ITEM SET MINING, HADOOP CLUSTERS

INTRODUCTION

Data mining faces a number of demanding situations in this big data era. The term big data refers back to the voluminous amount of records which is tough to shop, examine and process. The massive information includes various technologies to obtain beneficial information from the huge quantity of statistics. The mining of huge facts is a hard technique. The records mining and the huge statistics have lead to the emergence of the business intelligence. The statistics is collected and analyzed to improve the earnings of the groups through adopting

many techniques. The huge statistics has positive unique characteristics just like the volume, veracity and cost [8]. The big information has packages in many fields just like the healthcare, governmental companies, manufacturing industries, media, retail banking and studies. The large data calls for many technology to obtain useful data. Big records have emerge as the essential location of research nowadays. With using the social media like the fb, twitter and plenty of other social systems the records is developing in a speedy way. The fast growing data turns into the premise for the large facts. The association rule mining is a vital location of research.

ARTICLE INFORMATION:

*Corresponding Author: n.danapaquiame@gmail.com

Received 27th April, 2017

Accepted after revision 29th July, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF : 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

The association rule mining turns into a tedious system within the case of the huge statistics. The algorithms used for acquiring the frequent item sets aren't green in the case of the dynamic eventualities. The databases are updated dynamically.

The threshold values like the minimum aid depend that are used for mining purposes also are updated dynamically. In such cases the mining procedure turns into a hard undertaking. The mining procedure must be repeated on every occasion the values are updated in the databases or the threshold values are reset. To overcome these problems a parallelized algorithm has been proposed. The large volume of World Wide Web Consortium (W3C) data and their vast growth, diverse applications have emerged in a variety of domains such as social networks, public sectors, healthcare and life sciences, business process management, expert systems e-marketplace Web services composition and cloud system management. The Semantic Web was estimated to contain 7.7 billion triples in 2010 and has now reached over 40 billion triples. Its growth rate is still increasing. As it is evolving into a global knowledge-based framework to promise a kind of swarm intelligence, supporting knowledge searching over such a big and increasing dataset has become an important issue. But the Big data means the data sets whose size is vast when compared to the ability of current technology method and theory to capture, manage, and process these data within an endurable lapsed time. Today, Big Data management provides viewpoints as a challenge for all IT companies. The solution to such a challenge is shifting increasingly from providing hardware to provisioning more manageable software solutions. Big data also brings new opportunities and critical challenges to industry and academic world.

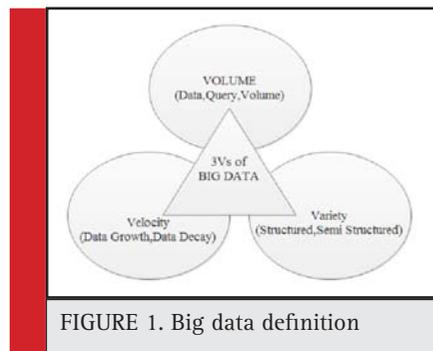


FIGURE 1. Big data definition

LITERATURE SURVEY

JW.Han, J. Pei and YW.Yin have proposed a brand new method called as the frequent sample tree method. The frequent sample tree stores the compressed records in an extended prefix tree shape. The common styles are

saved in a compressed shape. A FP-tree based totally mining method known as the FP-growth is developed. The proposed set of rules facilitates in mining the frequent item sets without the candidate set technology. The Three techniques had been employed to gain the performance of mining. A massive database is converted right into a small facts shape to keep away from the repeated database scans which is said to be high priced. It adopts a pattern frequent growth method to avoid generating huge candidate sets which may be very luxurious. The mining tasks are divided into smaller mission which is very beneficial in decreasing the quest space. The FP- tree primarily based mining additionally has many research issues like the square-based FP-tree shape with excessive scalability, mining common patterns with constraints and using FP-tree structure for mining sequential styles.

H. Li, Y. Wang, D. Zhang, M. Zhang, and E. Chang have proposed a parallel FP-growth set of rules. In parallel FP- increase set of rules the mining challenge is split into some of partitions. Every of the walls is supplied to the one of a kind machines and every partition is computed independently. To triumph over the demanding situations confronted through the FP-boom algorithm just like the garage, distribution of computation and surprisingly highly-priced computation parallel FP-increase algorithm is proposed. The PFP set of rules consists of five steps. Inside the first step, the database is divided into small elements. In the third step the frequent objects are grouped. Inside the Fourth step the FP-tree is built and the common item sets are mined. In the fifth step the neighborhood frequent item sets are aggregated. The PFP set of rules is powerful in mining tag-tag associations and website associations which might be used in query recommendation or another seek.

Zhigang Zhang, Genlin Ji, Mengmeng Tang has proposed a parallel algorithm MREclat primarily based on Map/reduce framework. Inside the vertical layout algorithm the common patterns are mined the use of the set of rules Eclat. The algorithms for mining frequent patterns in horizontal layout databases are unique from the algorithms for mining vertical databases like the Eclat. A parallel set of rules MREclat which makes use of a mapreduce framework has been proposed to obtain the common itemsets from the large datasets. Set of rules MREclat includes 3 steps. Within the preliminary step, all frequent 2-itemsets and their tid-lists are received from transaction database. The second one step is the balanced organization step, wherein common 1-itemsets are partitioned into businesses. The third step is the parallel mining step, where the records were given in step one is redistributed to unique computing nodes. Every node runs an improved Eclat to mine common item sets. Finally, MREclat collects all of the output from every

computing node and formats the very last result. MREclat uses the advanced Eclat to process statistics with the equal prefix. It has been proved that MREclat has high scalability and excellent speedup ratio.

Hui Chen, Tsau younger Lin, Zhibing Zhan and Jie Zhong have proposed a parallel algorithm for mining frequent sample in huge transactional data. It makes use of a prolonged Map Reduce Framework. Some of sub files are received by way of splitting the mass facts report. The bitmap computation is executed on each sub file to acquire the frequent styles. The frequent sample of the overall mass records file is obtained through integrating the consequences of all sub files. A statistic evaluation technique is used to prune the insignificant styles while processing every sub file. Has been proved that the method is scalable and efficient in mining common patterns in massive facts.

Xinhao Zhou and Yongfeng Huang have proposed an improved parallel Apriori algorithm. The map reduce framework is used to find the and the candidate set generation. The proposed algorithm is compared with the prevailing conventional apriori algorithm. It has been proved that the proposed set of rules is extra efficient compared to the traditional algorithm.

Jingui Liao, Yuelong Zhao and Saiqin lengthy have proposed an Pre Post set of rules. It's far a parallel set of rules that's implemented the usage of the Hadoop platform. The MRPrePost is a progressed PrePost algorithm which uses the mapreduce framework. The MRPrePost algorithm is used to locate the affiliation rules by mining the massive datasets. The MRPrePost set of rules has 3 steps. Inside the first step the database is divided into the records blocks referred to as the shards which can be allotted to each worker node. In the 2d step the FP-tree is constructed. Inside the very last step the FP-tree is mined to attain the common itemsets. Experimental results have proved that the MRPrePost algorithm is the quickest.

Sheela Gole and Bharat Tidke have proposed a new approach, Clust Big FIM. Huge datasets are mined the usage of the Map reduce framework in the proposed set of rules. Big FIM algorithm is changed to obtain the Clust Big FIM set of rules. Clust Big FIM algorithm affords scalability and velocity which might be used to achieve useful data from big datasets. The useful statistics may be used to make higher choices in the enterprise pastime. The proposed Clust Big FIM algorithm has four essential steps. Within the first step the proposed set of rules uses k-approach set of rules to generate the clusters. Within the 2nd step the frequent item sets are mined from the clusters. By building the prefix tree the global TID list are received. The sub trees of the prefix tree are mined to acquire the common item sets. The proposed Clust Big FIM algorithm is proved to be more as compared to the BigFIM set of rules.

Surendar Natarajan and Sountharajan Sehar have proposed a new algorithm named affiliation rule mining primarily based on Hadoop (ARMH). The proposed algorithm utilizes the clusters correctly and facilitates in mining frequent sample from big databases. The workload some of the clusters is managed the use of the hadoop disbursed framework. The hadoop allotted file gadget stores the huge database. Map reduce jobs have been used to mine the frequent styles. The proposed ARMH algorithm obtains the common sample from large databases. Semantic implication has attracted much helpfulness from both academic and industry present days. Many implication engines have been developed to support the reasoning over Semantic Web, structured web, unstructured web. For Example Paulheim and Bizer studied the problem of extrapolation on blaring data and presented the SDType method based on statistical distribution of types in RDF datasets to deal with blaring data. Grau et al. Existing an incremental reasoning approach based on modules that can reuse the information obtained from the earlier versions of ontology. This method is used for OWL but not suitable to the increasing RDF data. Besides, since no distributed reasoning method is adopted, the reasoning speed is a huge problem when dealing with a large ontology base. To speed up the updating process with newly-arrived data and fulfill the requirements of end-users for online queries, this paper presents a method ant colony and swam intelligence based on Map Reduce and Hadoop, which can well leverage the old and new data to minimize the updating time and reduce the reasoning time when facing big RDF datasets

Jayavardhana Gubbi, Rajkumar Buyya, proposed system is to deploy large-scale, platform-independent, wireless sensor network infrastructure that includes data management and processing, actuation and analytics. It is often quite important to take advantage of the benefits of metadata for transferring the information from a database to the user via the Internet.

Marisol García-Valls, Pablo Basanta-Val, proposed were the level of hardware, networked embedded systems (NES) are becoming a cloud of hundreds, and even thousands, of heterogeneous nodes connected by means of heterogeneous networks as well; they are now used in various domains such as cloud or grid.

Keling DA, Marc DALMAU, Philippe ROOSE, proposed system is context collector first collects information on the operating environment from the operating system and the user context.

PROPOSED WORK

The unique database is known as the D. the new datasets which can be inserted into the database at a later degree

is called the D'. The unique minimal guide count number is referred as T that is the edge value. The up to date minimum help be counted is referred to as the T'. The entire process is divided into the 2 stages.

Within the First segment the frequent item sets are obtained from the database D underneath the brink cost T. the first segment includes map reduce responsibilities. The database D is divided into small chunks using a process called as the input split. Those chunks are sent to the mapper and the reducer to gain the assist depend values. The use of the help be counted value and primarily based at the minimum aid remember the common listing is formed.

The group listing shaped the mapper and the reducer to form the local FP-bushes. From the local FP-timber the local frequent item sets are extracted. The nearby common item sets are included to attain the common item sets of the database D.

In the second phase, the new datasets are inserted into the database. The minimum assist be counted is modified as T'. Based on the brand new assist count number T' the common listing is up to date. The new updated group list is acquired. The brand new updated group list is given to the mapper and the reducer to obtain the nearby FP-timber.

From the up to date nearby FP-bushes the nearby frequent item sets are acquired. The local common item sets are integrated to obtain the final result. In this proposed the new algorithm i.e. Ant Colony Swarm Intelligence Optimization has been proposed for optimizing the unstructured big data to structured data. The proposed algorithm gives better performance in terms of scalability, less time and efficiency compared to Traditional centralized reasoning methods.

Semantic Web is based on RDF, which integrates a variety of applications by using extensible markup language (XML) for syntax and universal resource identifier (URI) for naming. RDF is an assertion language envisioned to be recycled to express propositions via precise formal vocabularies. The fundamental unit of RDF is a triple that is used to describe the relationship between two things. Its formal definition is <subject, predicate, object>, in which subject denotes a resource, and predicate denotes properties or aspects the resource and expresses in relationship between the resource and the object.

FREQUENT ITEM SET MINING

It is mainly used for market basket analysis. The regularities in the shopping behavior of the customers can be found using the frequent item set mining. The products which are bought together can be found using the frequent item set mining.

DYNAMIC THRESHOLD VALUE

The threshold value refers to the values like the minimum support count which changes dynamically in the incremental databases. The minimum support count is used for obtaining the frequent item sets. The threshold values play an important role in the pattern mining.

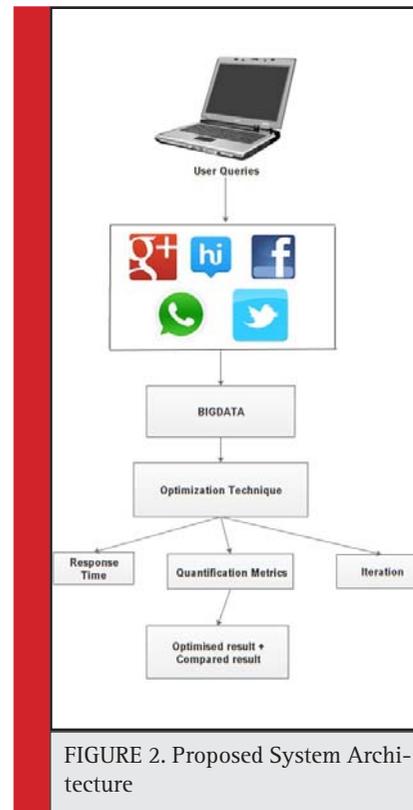


FIGURE 2. Proposed System Architecture

The scalable and cost effective processing technique to deal with growing amount of semantic web data. A framework that builds using Hadoop to store and retrieve large number of RDF triples by exploiting the cloud computing paradigm. A scheme is described to store RDF data as plain text files in Hadoop Distributed File System. An algorithm is developed to generate the best possible query plan to answer a SPARQL query and RDF query language based on the cost model. This cost model is derived from summary statistics that collect from RDF Graphs for query processing plans.

In this approach, data is not stored in a single file because, in Hadoop, a file is the smallest unit to a Map Reduce job and therefore divides the data into multiple files. Summary statistics is a proven tool which helps to generate the best possible query processing plan. It is widely used in relational databases and effective to handle enormous amount of data. The outcome of this

approach is Hadoop framework is capable of handling enormous amount of RDF data, which is a distributed and highly fault tolerant system and inherits those two properties automatically. The cost model is used in algorithm improves best processing plan to answer a SPARQL query.

An efficient RDF query engine to evaluate SPARQL queries, where the inverted index structure is used for indexing the RDF triples. A SPARQL query graph model is defined that can effectively expresses the query semantics. Then a set of operators is implemented on the indexed triples for transforming the SPARQL query graph into an optimal execution tree for evaluation. To achieve this goal, a main-tree-shaped optimization algorithm was developed to identify the optimal execution plan by effectively reducing the search space to determine the optimal joining order. The optimization uses a set of RDF statistics to estimate the execution cost of the query plan. The algorithm includes an IR based solution for indexing triples and a set of highly-efficient operators for query optimization and evaluation, a set of RDF statistics for estimating the execution cost of the query plan, and a main-tree shaped optimization algorithm for identifying the optimal query plan. Tests show that this approach is very efficient and scalable for querying large-scale RDF triples. Current work focuses on the query optimization problem on SPARQL basic graph patterns, and union and optional patterns. Map Reduce is a programming model and an associated implementation for processing and generating large data sets.

Users specify a map function that processes a key/value pair to generate a set of intermediate key/value pairs, and a reduce function that merges all intermediate values associated with the same intermediate key. Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. The run-time system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine communication.

This allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a large distributed system. Our implementation of Map Reduce runs on a large cluster of commodity machines and is highly scalable: a typical Map Reduce computation processes many terabytes of data on thousands of machines. Programmers find the system easy to use: hundreds of Map Reduce programs have been implemented and upwards of one thousand Map Reduce jobs are executed on Google's clusters every day. Therefore, the implementation makes efficient use of these machine resources.

FP-GROWTH ALGORITHM

FP-Growth algorithm is mainly used to find the frequent item set without candidate item set generation. Two steps are followed in the FP-growth algorithm. In the first step, the FP-tree is constructed. In the second step the frequent item set are extracted from the FP-tree. It is an important data mining model used to find the interesting relationship between the data in the database. It is mainly used for the Market basket analysis to help improve the business activity. Association rules are obtained using two main criteria the support and the confidence. The support indicates how frequently the items appear in the database. The confidence refers to the number of times the if/then statements have found to be true. Map Reduce is used to execute the Mongo db query and provide the parallel processing over a large number of nodes to simplify the data. Finally, the Mango db query language is created for the data is retrieved from the HDFS. The scalable keyword search on big RDF data for generating the summary based solution for query processing. Keyword search is a useful tool for exploring large RDF data sets and it is an important tool for exploring and searching large data corpuses whose structure is either unknown, or constantly changing. So, keyword search has already been studied in the context of relational databases, XML documents, and more recently over graphs, and RDF data.

Existing techniques either rely on constructing a distance matrix for pruning the search space or building summaries from the RDF graphs for query processing. It have serious limitations in dealing with realistic, large RDF data with tens of millions of triples. Furthermore, the existing summarization techniques may lead to incorrect/incomplete results.

ASSOCIATION RULE MINING

To address these issues, an effective summarization algorithm to summarize the RDF data. Given a keyword query, the summaries lend significant pruning powers to exploratory keyword search and result in much better efficiency compared to previous works [18]. A keyword query against an RDF graph looks for (smallest) sub graphs that contain all the keywords. An RDF data set is a graph (RDF graph) composed by triples, where a triple is formed by subject, predicate and object in that order. A typical approach used here for keyword-search is backward search. In this approach, a schema to represent the relations among entities of distinct types is summarized from the RDF data set. Backward search is first applied on the schema/summary of the data to identify promising relations which could have all the keywords being queried. Then, by translating these rela-

tions into search patterns in SPARQL queries and executing them against the RDF data, the actual sub graphs are retrieved. The BACKWARD method will initiate many random accesses to the data on disk, and has to construct numerous search paths in order to complete the search. However, the majority of them will not lead to any answers. Intuitively, we would like to reduce the input size to BACKWARD and apply BACKWARD only on the most promising sub graphs. The proposed summarization process in has a limitation: it bundles all the entities of the same type into one node in its summary, which loses too much information in data as to how one type of entities are connected to other types of entities. As a result, this approach could generate erroneous results (both false positives and false negatives).

To alleviate this problem, a type based summarization approach on the RDF data is proposed. The idea is that, by operating our keyword search initially on the summary, we can navigate and prune large portions of the graph that are irrelevant to the query, and only apply BACKWARD method on the smaller sub graphs that guarantee to find the optimal answers.

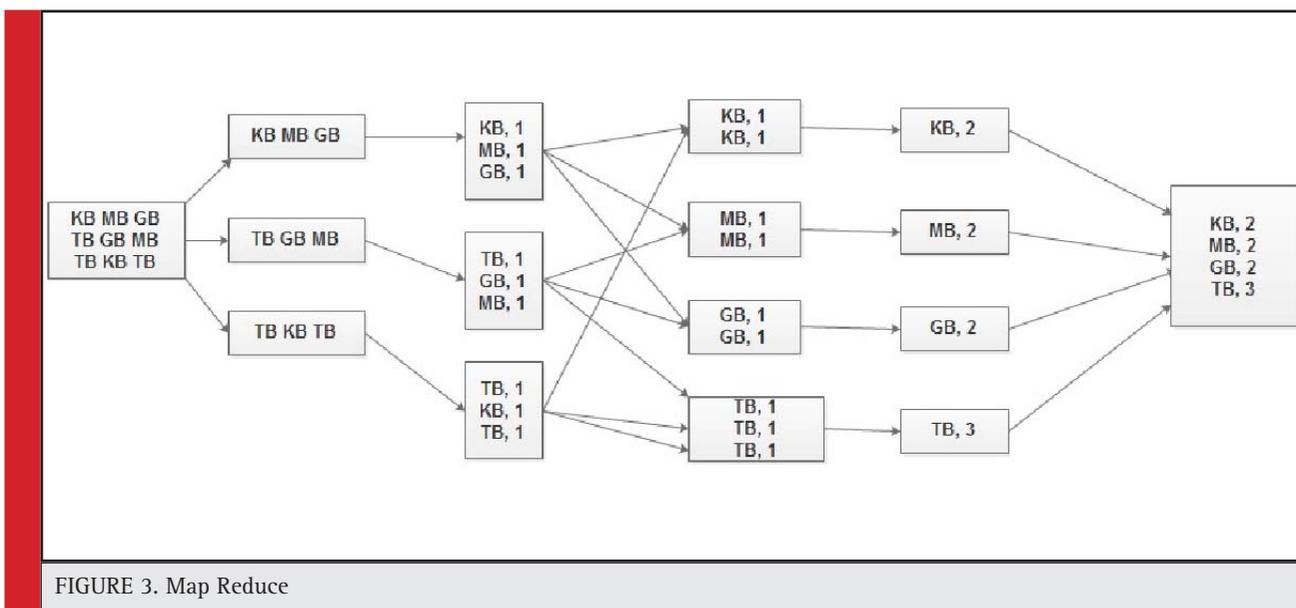
In general, the keyword search can benefit from the summary in two perspectives. With the summary, we can obtain the upper and lower bounds for the distance traversed in any backward expansion without constructing the actual path and we can efficiently retrieve every partition from the data by collaboratively using SPARQL query and any RDF store without explicitly storing the partition.

In this approach, the algorithm requires sub graph isomorphism for two graphs and the testing is hard for RDF related graph.

A different order to partition the data could lead to different performance in evaluating a keyword query. The summarization process starts with splitting the data graph into smaller but semantically similar and edge-disjoint sub graphs. However, finding an optimal order to partition the RDF data set is not efficient. In this approach, Keyword search with summary is a scalable and exact search algorithm to overcome the problems in previous method. The SUMM algorithm uses the summary of the RDF data to reduce the amount of data accessed in the BACKWARD method. For the algorithm to be effective, we should be able to efficiently identify and retrieve the instance data from selected partitions.

It performs a two-level backward search: one backward search at the summary-level, and one at the data-level. Only for identified connected partitions that are found to contain all the distinct keywords at the summary-level and whose score could enter the top-k answers. The path-length computation is at the heart of backward search and pruning. While working at the summary-level, exact path lengths are not available. Any shortest path in the underlying RDF graph must go through a number of partitions, and for each intermediate partition the path connects two of its portals.

Therefore, this approach explains about how to estimate the path length of the actual data. In this method, first define a mapping function Join. It takes as input a graph g , a list of disjoint sets of vertexes and outputs a new graph g_0 . The function Join itself is a homomorphism mapping, which constructs a homomorphism from g to g_0 . This approach stores the RDF data in an RDF store and works by dynamically identifying the data of a partition using appropriately constructed SPARQL



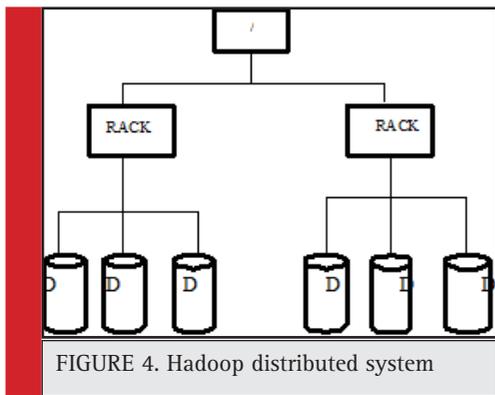


FIGURE 4. Hadoop distributed system

queries that retrieve only the data for that partition. In contrast to BACKWARD, the difference in time for the SUMM method is relatively moderate as the size of the data changes. This is keywords that are far apart can be pruned in the first level of the backward search.

Therefore, SUMM finds the top-k answers for any top-k keyword search query q on an RDF graph and the summaries we built can be upd incrementally and efficiently.

HDFS were planned to be accessible, distributed more storage system that works carefully with Map Reduce. HDFS will “impartial work “under a structural of carnal and system atmospheres. By using distributed large storage and Computations across to many servers [17].

Big data is the data sets whose size is vast when compared to the ability of current technology method and theory to capture, manage, and process these data within an endurable lapsed time.

The structural and unstructured data are the fast growth in ontology based challenges in performance more scalability and efficiently. The ant-colony algorithm is a probabilistic technique for solving large data sets problems which can be reduced time to find particular data sets. The issues in a traditional centralized reasoning methods is using in small datasets (TIF/EAT)

the performance is less. In this paper new algorithm i.e. Ant Colony Swarm Intelligence Optimization has been proposed for optimizing the unstructured big data to structured data.

The proposed algorithm gives better performance in terms of scalability, less time and efficiency compared to Traditional centralized reasoning methods. The system model is included two or more data owners and a cloud. Every data owners have a private database, and the data owners encrypt their private databases prior to outsourcing the encrypted databases to the cloud. Data owners can also request the cloud to mine association rules or frequent item sets from the joint database for their behalf. The (honest but curious) cloud is tasked with the compiling and storing of databases got from various data owners, the mining of association rules or frequent item sets for data owners, and the sending of the mining result to relevant data owners.

PERFORMANCE ANALYSIS

The novel incremental FP-growth mining set of rules is implemented at the hadoop framework. The classical dataset T10I4D100K is used to enforce the proposed set of rules. The T10I4D100K dataset include approximately one hundred, transactions and approximately 870 objects. The proposed algorithm is determined to value the least amount of time when in comparison with the apriori set of rules. The proposed algorithm is located to be more optimized.

CONCLUSION

The Traditional association rule mining set of rules is not green in mining the big statistics. The existing algorithms cannot be implemented in the dynamic eventualities. The database is up to date periodically and while the threshold values changes the mining process turns into a tedious undertaking. Data Analysis is a very vast

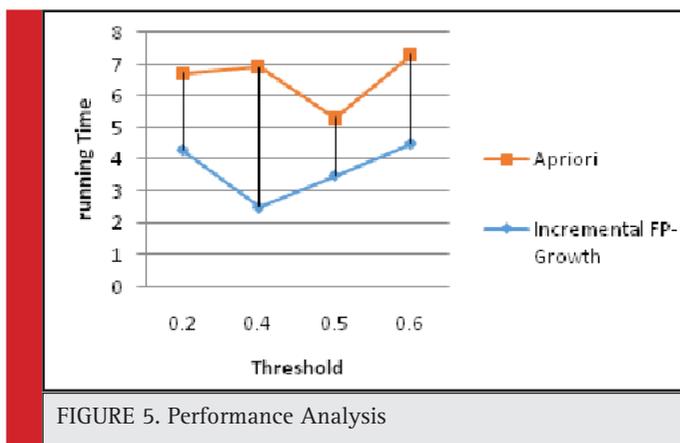


FIGURE 5. Performance Analysis

and interesting topic which can be used along with different technologies to analyze the data efficiently and accurately. Data mining brings a lot of benefits to businesses, society, governments as well as the individual. However, when huge data comes into picture, it is more efficient to use Big Data analysis using a tool that is very effective to analyze. Out of all the method in which data mining can take place, we found R to be very easy, user friendly and effective when large data has to be analyzed. Thus, we will be using R for implementation in our further analysis. Those troubles are overcome by the proposed novel incremental FP-boom set of rules which is applied using the hadoop. So the proposed set of rules is discovered to be greater effective. In the big data, reasoning on a World Wide Web Consortium (W3c) becomes progressively stimulating because of the large volume of data involved and the time and space complexity of the task. This paper for the first time proposes an (ACOSIO) to deal with large scale incremental in RDF datasets to our knowledge. User will excited their query more effectively. Our method is implemented based on map reduce and Hadoop. It will reduce the time complexity and space complexity. It will be update the datasets at simultaneously process.

REFERENCES

- Ahиландесwari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 1-5.
- Anagnostopoulos C, Hadjiefthymiades S, (2009), Advanced inference insituation-aware computing. *IEEE Trans. Syst, Man, Cybern. A, Syst, Humans*, 39(5): 1108–1115.
- Anish C. M, Majhi B, (2016), Hybrid Nonlinear Adaptive Scheme for Stock Market Prediction using Feedback FLANN and Factor Analysis. *Journal of the Korean Statistical Society*: 45: 64-76.
- Anusha B, Noah, Sivaranjani C, Priyanka S. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Chang F, Dean J, Ghemawat S, Hsieh WC, (2008) Big table: A Distributed Storage System for Structured Data. *ACM Trans. Computer Systems*. 26(2), article 4.
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Guo J, Xu L, Gong Z, Che C-P, Chaudhry S S, (2012), Semantic inference on heterogeneous e-marketplace activities. *IEEE Trans. Syst, Man, Cybern. A, Syst, Humans*, 42(2): 316–330.
- Han JW, Pei J, Yin YW, (2004), Mining frequent patterns without candidate generation. In: *Proc.s of the Springer International Conference on Management of Data*, 29(2): 1-12.
- Jun Liu, Feng Liu, Ansari N, (2014), Monitoring and analyzing big traffic data of a large-scale cellular network with Hadoop Network. 28(4), 2014.
- Kim Y, Enke D, (2016), Developing a Rule Change Trading System for the Futures Market using Rough Set Analysis. *Journal of Expert Systems with Applications*, 5(9): 65-173.
- Li H, Wang Y, Zhang D, Zhang M, Chang E, (2008) PFP: Parallel FP-growth for query recommendation. In: *Proc.of the ACM Conference on Recommender Systems*: 107-114.
- Limam N, Boutaba R, (2010) Assessing Software Service Quality and Trust worthiness at Selection Time. *IEEE Transactions on Software Engineering*, 36(2): 559-574.
- Marshall M S, (2012), Emerging practices for mapping and linking life sciences data using RDF—A case series. *Journal of Web Semantics*, 14: 2-13.
- Paranjape-Voditel P, Deshpande U, (2011), An Association Rule Mining Based Stock Market Recommender System. In: *Proc. of the Emerging Applications of Information Technology (EAIT), Second International Conference*: 21-24.
- Paulheim H, Bizer C, (2013), Type inference on noisy RDF data. In *Proc. ISWC, Sydney, NSW, Australia*: 510-525.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Rout M, Majhi B, Mohapatra U M, Mahapatra R, (2012) An Artificial Bee Colony Algorithm Based Efficient Prediction Model for Stock Market Indices. In: *Proc. of Information and Communication Technologies (WICT), World Congress*: 750-754.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 1-8.
- Siddique Ibrahim S P, (2014), Extract data in large database with hadoop. *International Journal of Advances in Engineering and Scientific Research*. 1: 1- 5.

N. Danapaquiame *et al.*

Suthar M, Ranavadiya A, (2015), An Insight into Various Techniques on Association Rule Mining using Multi Objective Genetic Algorithm. *International Journal of Advent Research in Computer and Electronics*, 2(12): 1-8.

Xu Z, Martin P, Powley W, Zulkernine F, (2007), Reputation enhanced QoS-based web services discovery. In: *Proc. of the IEEE International Conference on Web Services (ICWS)*: 249-256.

Yang T, Qian K, Lo D. C, Xie Y, Shi Y, Tao L, (2016) Improve the Prediction Accuracy of Naïve Bayes Classifier with Association Rule Mining, In: *Proc.of International Conference on IEEE 2nd International Conference on Big Data Security on Cloud (Big-DataSecurity)*: 129-133.

Zheng Z, Lyu M R, (2010), Collaborative reliability prediction of service-oriented systems. In: *Proc. of the 32nd ACM / IEEE International Conference on Software Engineering (ICSE)*: 35-44.

Topmost-K Frequent Itemsets mining using compressed POC Tree and Nodesets

Boina Nataraj¹, Pothula Sujatha*² and J. Amudhavel³

¹PG Student, Department of Computer Science, Pondicherry University, Puducherry, India

²Assistant Professor, Department of Computer Science, Pondicherry University, Puducherry, India

³Department of CSE, KL University, Andhra Pradesh, India

ABSTRACT

The Frequent Itemsets Mining (FIM) is the core part of any Association Rule Mining (ARM) algorithm for unearths hidden rules in any transactional database. There are a lot of techniques developed for mining out frequent item sets which can be broadly categorized as i) Itemsets that use candidate generation ii) Itemsets does not use candidate generation. Late systems basically make utilization of novel data structures for the purpose of generating candidate sets instead of storing all the generated intermediate candidate itemsets for memory efficiency. Broad assortments of special data structures have been proposed to solve this problem, comprising of the linear prefix trees, Pre-post coding trees, pre-order coding trees, nodelists, nodeSets and so forth. This work employed one such structure named Pre-Order Coding tree (POC). The tree is constructed in such a way that the number of nodes is minimized. This is very much essential and advantageous in order to minimize the memory pre-requisites of the running algorithm. Further, a novel algorithm called Topmost-K Frequent Itemsets mining (TKFIN) using Compressed POC is developed to haul out frequent itemsets by utilizing this pre-order coding tree in an efficient manner. This algorithm is tested with both denser and sparser data sets and the outcomes are compared with FIN algorithm in terms of space and time complexities. The proposed TKFIN using Compressed POC tree algorithm yields better performance when compared to FIN algorithm.

KEY WORDS: COMPRESSED POC TREE, NODE SETS, PREFIX-TREE, VERTICAL FREQUENT ITEM SET ALGORITHM

ARTICLE INFORMATION:

*Corresponding Author: spothula@gmail.com

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

INTRODUCTION

Data mining and knowledge discovery plays a very important role in various fields today. As various techniques are involved in data mining such as classification, clustering, association rule mining, outlier detection, etc., ARM is selected as the problem domain. ARM forms a basic data mining technique which mines out patterns in the item sets which are represented as a set of antecedent and precedent rules. A major component of ARM is the FIM algorithm, FIM is a subset of a more general problem known as frequent pattern mining which encompasses mining itemsets, subsequences or substructures Han et al (2006). The FIM algorithm produces the itemsets that appears most frequently. depends on the support threshold given by the domain expert.

Apriori Algorithm was introduced as a pioneering algorithm for generating association rules Agarwal et al (1993). The problem of mining association rules over quantitative and categorical attributes in large relational tables is defined Srikanth et al (1996). The authors developed an enhanced version of apriori by using a hash-based algorithm, where the size of the candidate-k itemsets were reduced. Instead of maintaining itemset count, bucket counts are maintained which minimized the number of candidate-k Itemsets to be examined Park et al (1995). An efficient method for mining association rules which uses partitioning the data to find candidate Itemsets is described Savasere et al (1995). A new algorithm is introduced to find large Itemsets using very few passes over the database using a technique called item reordering Brin & Sergey (1997). Also, a new way of generating the implication rules has also been exploited. The author targeted the problem of integrating user constraints while mining out association rule to reduce the execution time Srikanth et al (1997). Mining on large data warehouses is considered where in a large range of alternatives have been exploited for conjoining both mining and database systems Sarawagi et al (1998). A sampling approach is proposed and further reducing the number of passes required over the database for mining out association rules Toivonen & Hannu (1996).

An approach of mining frequent itemsets without candidate generation is used Han & Jiawei (2004), where a tree data structure called fp-tree is created using the transactional database. The tree is created using each of the transaction record sorted in L order Han et al (2000). A new approach called ECLAT was introduced which exploited the vertical data format Zaki (2000). A slight but very effective method was proposed to improve the performance of ECLAT Zaki (2003). Maximal Itemsets refers to those Itemsets which satisfy the condition that their support is equal to any of their supersets Burdick et al. (2011). A simple recursive algorithm is proposed to

find frequent Itemsets Borgelt & Christian (2005). The algorithm is proved to be more efficient without using any complicated data structure but not guaranteed to be faster for larger datasets. A new algorithm is developed using N-Lists Deng et al (2012).

A fp-tree like structure is used called PPC-tree which compresses the information about the Itemsets. An algorithm called PrePost is developed which can directly find frequent Itemsets without generating candidate Itemsets. The concept of a prefix tree called PPC-tree is used and a novel data structured called Node-list has been developed Suchahyo et al (2004), which aims to efficiently mine out patterns. This algorithm is shown to be efficient than its counterparts. The overhead of pointer usage is eliminated by using a tree called Linear-Prefix tree or LP-Tree which is composed of array forms. This has been shown to be efficient in terms of memory, scalability and memory. A new algorithm Pre-Order coding tree has been proposed for fast mining of frequent itemsets using Node sets Deng et al (2014). A new technique for pruning generations of candidate Nodesets have been proposed called superset equivalence is also explained in the study. Using this, search space is greatly reduced which in turn reduces the time and fast forwards the mining process.

The rest of the study is organized as follows. The section 2 gives the description of Pre-order coding tree which describes the algorithm for construction of tree and also illustrated it with an example. The section 3 gives the proposed system which introduces the modification to the POC tree such that it becomes memory efficient by further compressing the nodes. The section 4 describes the experiments performed and the results obtained. Section 5 provides a brief discussion on the results obtained from the experiments. Section 6 gives the conclusion of this study and also states the future work that can be done in the same lines.

PRE ORDER CODING (POC) TREE

Nodesets resolve the problems present in the previous structures namely node-lists and N-Lists. Moreover, novel data structure has been proven to be efficient by extensive experiments. This is achieved because the Nodesets, unlike the Node-lists or the N-Lists, are encoded using pre-order code or post-order code. Thus, the benefits of using Nodesets are two way as follows:

1. The tree constructed in the case of Nodesets need not to store both pre-order code and post-order codes at each node. This saves the memory for tree construction.
2. Also the data structure, i.e. Nodesets has only one of such code, the intersection will also take less time.

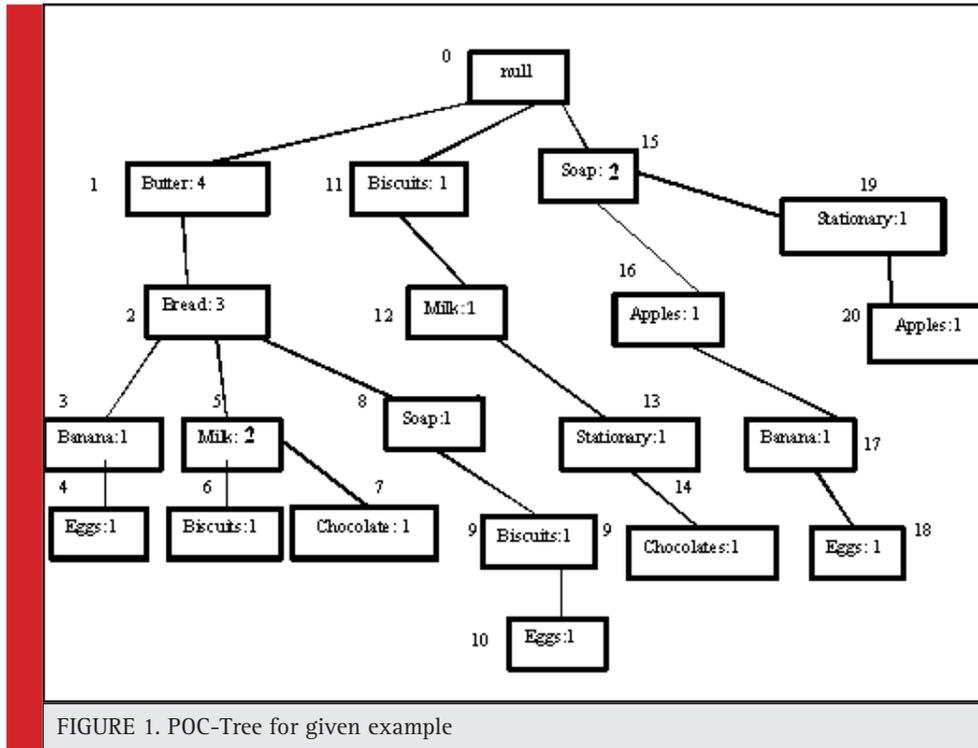


FIGURE 1. POC-Tree for given example

This leads to lesser compression of the tree. The transaction T05 can be arranged in two ways: {Soap, Stationary, and Apples} or {Soap, Apples, Stationary}, because supports for both the items are same. When the POC-tree is constructed with these itemsets, two POC-trees different in the number of nodes is created. When permutation is used {Soap, Stationary, and Apples}, one extra node is added to the POC-tree. This can be avoided by taking the other permutation and hence the tree will be compressed more. But, the problem is to identify which permutation will lead to a more compressed POC-tree. The figure 1 shows the POC-tree obtained when the former permutation is considered.

POC COMPRESSION ALGORITHM

An efficient tree compression algorithm which will compress the tree after it has been constructed by the previous approach is proposed. The algorithm is given as follows: After passing input to algorithm 3 and applies the algorithm 4 recursively on the subtrees of the root nodes and dynamically compresses the tree while iterating over it. Algorithm 1, where each sub-tree of the root node is subjected to the recursive algorithm Compress_Tree.

Algorithm 3: Compression

Input: A POC - Tree obtained from the Algorithm 1

Output: A POC - Tree Compressed POC - Tree

1. root.label:= "null"; // The root label is initialized to null, since

it has no element.

2. root.pre:= 0; // The pre-order code of this root node is set to 0
3. for each child_noderoot.children representing its subtree,
 Call Compress_Tree (child_node, POC - Tree) recursively.
4. Return POC_tree;

This algorithm takes the parent node of the sub-tree and also takes in the POC-tree which it uses as a reference tree for compression, as the tree will get modified. The logic of the Compress_Tree algorithm is simple. It takes in a parent node of a sub-tree and then if it finds out that this is the direct descendant of root or any of its children, it will recursively apply the Compress_Tree over the children of the node. If the nodes are other than the above nodes then the core algorithm will start. First it needs to decide whether at any point there is a need for compression or not. This is achieved by checking its support with its parent's support. If it finds out that the support of its parent is same, then it will conditionally check whether any of its parent's sibling (nodes of the same level) nodes have the same item name as it does. If this is the case then there is a clear need for compression since there is a chance of merging the current node with the parent's sibling who's item-name and thus the algorithm will start the merging procedure. Otherwise the Compress_tree() is recursively called on its children nodes. The merging procedure has four steps. First it copies each sub tree of the node and attaches it to the parent, then removes the node and its sub tree, next it copies the parent's sub tree and attaches it to its sib-

ling node and finally the supports are added since when merging the nodes the partial count of the nodes will be added up for consistency.

Algorithm 5:

```

copySubTree (root, Parent_node, POC - Tree, subtree)
1. subtree := POC - Tree.root.children // All the children are copied
   to a subtree
2. for each node n ∈ POC - Tree.root.children
   a. copySubTree(n, POC - Tree, subtree) // Recursively copy
     subtree of children
   b. end for

```

Algorithm 6

```

attachSubTree (node, root, POC - Tree.subtree)
1. subtree := POC - Tree.root.children // Copy the children into a
   subtree
2. root.child := node // add the current node to the node called root
3. for each ch ∈ subtree.root.children //Recursively build the new
   subtree
   a. attachSubTree(root, ch, subtree, pocTree)
   b. end for

```

Algorithm 4:

```

Compress_Tree (Current_node, POC_Tree)
1. if parent(Current_node).label = "null" || parent(parent(Current_
   node)).label = "null"
   a. for each child_node such that child_node Current_node.
     children
     i. Compress_Tree ( child_node POC - Tree,)
2. else if Current_node.support = parent(Current_node).support
   then for each parent(current_node).siblings do
   ifcurrent_node.label = parent(current_node).sibling.labelthen
   a. // Copy each sub-tree(current_node) and attach it to parent
   b. for each node ch ∈ POC - Tree.current_node.children
     i. subtree.add(ch)
     ii. copySubTree(ch, POC-tree, subtree)
     iii. POC-tree.remove(ch)
   c. attachSubTree(Parent_node, subtree.root, subtree, POC-
     tree) end for // Remove the current_node and its subtree
     POC-tree.remove(node) // Copy the Parent subtree and
     attach it to the sibling node
   d. subtree.add(Parent_node)
   e. copySubTree(Parent_node, POC-tree, subtree)
   f. POC-tree.remove(Parent_Node)
   g. attachSubTree(parent(current_node).sibling, Parent_Node,
     subtree, POC-tree)
   h. // Add up the supports
   i. parent(current_node).sibling.partialSupport = parent
     (current_node).sibling + Current_node.partialSupport
   j. // set compressed flag to "true"
   k. Compressed := true;
   l. end for
3. if compressed = "true" //If compression takes place, then
   recursively call on children of parent node
   a. for each chParent_Node.children
     i. Compress_Tree(ch, POC-tree)
     ii. end for
4. else //If compression does not take place, recursively call on
   children of the current node
   a. for each ch in Current_Node.children
     i. Compress_Tree(ch, POC-Tree)
End for

```

Now, it is explained and compared with the POC-tree in figure 2 and see the difference between the two is explained. The algorithm requires simply taking the output of the above tree and then applying the compression algorithm. The POC-tree obtained after compressing it is given in the figure 2. The POC-tree in the figure 2 may look almost similar to the previous, but the number of nodes in this tree is only 19 whereas in the earlier POC-tree it was 20. In figure 2 the 16th node of the POC-tree is labeled as "Apple" which is the same as the 20th node, the count of 20th node and 19th node were same which also gives us the freedom to swap the order of both the nodes. According to the algorithm 3, the node 20th will be merged with the 16th with their counts added. Thus the resultant POC-tree is shown in Figure 2. However the comparison achieved seems to be very small in terms of the number of nodes present in the tree, but when there are a lot of redundant itemsets with the same support, then this algorithm will give a better POC-tree which will have far less nodes than in the POC-tree obtained from POC-tree construction algorithm. The FIN algorithm for mining top-k frequent itemsets because of its brilliant use of the superset equivalence property that eliminates recursive generation of many nodes in the pattern tree thus constructed while mining. Thus it removes the overhead of creating such nodes and also eliminates the further overhead finding intersection of nodesets which are obviously frequent.

Top-K Frequent Itemset Mining Using Compressed Poc-Tree
Algorithm 7: TKFIN algorithm

```

Input: A transaction database and a minimum support  $\delta$ 
Output: Top-K frequent Itemsets
1. Call Algorithm 1 to construct the POC-tree.
2. Call Algorithm 3 to compress the POC-tree
3. Compute frequent-1 itemsets from the POC-tree
4.  $F_2 = \emptyset$ ; // Initialize the frequent-2 itemset variable
5. For each node  $n \in PreOrderTraversal(POC - tree)$  //do a preorder
   traversal
6.  $I_x$  be the current item in the node  $n$ . //using the label of current
   node
7. For each ancestor of  $n$ ,  $n_A$  do //Combine each node from the
   ancestors
8. Let  $I_y$  be the item in node  $n_A$ .
9. If  $(I_x, I_y) \in F_2$  then // Check if the combination is present in the
   frequent-2 itemset
10.  $(I_x, I_y).support += n.support$ ; // Add up all the supports using the
   Nodesets.
11. Else
12.  $(I_x, I_y).support = n.support$ ; // If not present, then simply assign
   the support
13. Add  $(I_x, I_y)$  to  $F_2$  // Finally add this frequent-2 itemset to  $F_2$ 
14. End If
15. End For
16. Remove  $(I_x, I_y)$  from  $F_2$  where  $(I_x, I_y).support < \delta * |DB|$ 
17. Scan the POC-tree using pre-order traversal and at each node do
18.  $I_y$  be the current item in the node  $n$ .
19. For each ancestor of  $n$ ,  $n_A$  do // search for k-frequent patterns
20. If  $(I_x, I_y) \in F_2$  then //for each nodeset in frequent-2 itemset

```

```

21.  $(I_x, I_y).Nodeset = (I_x, I_y).Nodeset \cup N\_Info(n)$ ;
22.  $max\_F = \emptyset$ ;
23. For each  $(I_x, I_y)$  in  $F_2$  do
24. call construct_pattern_tree ( $(I_x, I_y)$ ,  $\{I \mid I \text{ is a frequent-1 Itemset and Support}(I) > Support(I_x)\}$ ,  $max\_F, F$ )
25. End for
26. Return top_F(F, k);

```

In algorithm 7, a recursive function named `construct_pattern_tree` is called, which is used to build a set enumeration tree. This set enumeration tree will help generate the itemsets which are further evaluated for their frequency. This algorithm employs superset equivalence property which eliminates the generation of nodes for super sets of any frequent itemset which are not so much required since they can be predicted to be frequent. It is shown our compressed POC-tree approach to be memory efficient without much sacrifice of the run time of the algorithm.

Algorithm 8: `top_F`

Input: The List of Nodesets of all the frequent itemsets L_n ,
Cardinality k

Output: Top-K frequent itemsets

```

1.  $lastIndex = \emptyset$ ; // Initialize the lastIndex variable to store the index of last itemset
2.  $maxSize = \emptyset$ ; // Initialize the maxSize variable which will store the maximum cardinality
3.  $top = \emptyset$ ; // Initialize the top variable which will store the top-most k frequent itemsets
4. if  $k = \emptyset$ 
   a.  $maxSize = \text{Number of items corresponding to the last Nodeset}$ 
5. else
   a.  $maxSize = \text{Number of items corresponding to the } k^{\text{th}} \text{Nodeset}$ 
6. if  $(maxSize == 1)$  then return all the frequent-1 itemsets.
7. Else
8. For each nodeset,  $n_s$  in  $L_n$ ,
9. If  $n_s.size == maxSize$  then
10.  $top = top \cup n_s.itemset$  //If the size of itemset is equal to maximum size, then assign to top
11. end if
12. end for
13. end if
14. return top

```

In algorithm 8, there is a provision to select the cardinality of the final frequent itemsets which is supplied in the algorithm 8. After the initialization steps in line 1 to line 3, it is checked if k is supplied or not. If the maximum cardinality is not specified then $maxSize$ in algorithm is set to the number of items available in the last NodeSet which will automatically be of the highest cardinality. Otherwise, if k is specified, then $maxSize$ with the first itemset in the NodeSet list which is of the size k is assigned. In line 6, the $maxSize$ is checked for highest cardinality. If it is one, then return all the frequent 1-itemsets. Otherwise a return all the itemsets which have the Nodeset size as same as the $maxSize$.

EXPERIMENTAL METHODOLOGY

Experimental Setup

Experiment is conducted to compare two major traits of any algorithm: time complexity and space complexity. The time is measured for each run of the algorithm in milliseconds whereas the space complexity is measured in terms of bytes used at runtime by the execution of algorithm. Experiments are performed using a computer with 4 GB memory, Intel core i7-3770 CPU with a clock speed of 3.40 GHz. The operating system used is Windows 7 x 64 Edition. The implementation of the algorithm is done in Java using the integrated development environment of Eclipse which is an open source IDE developed by Apache software foundation. The version of Eclipse used is Kepler 4.3.

Dataset Description

For testing the proposed algorithms, two types of transaction datasets are used: one is a real time dataset and other is a synthetic dataset. The real time datasets are Connect, Chess, mushroom datasets, whereas the synthetic dataset used is T10I4D100K. This dataset is generated using the generator from the IBM Almaden Quest research group (<http://fimi.ua.ac.be/data/>). The datasets includes name of the dataset, number of records in the dataset, the total number of items and an average number of items in each transaction (density of the dataset) are given in table 1. According to the table 1 the Connect, Chess and Mushroom datasets are denser datasets which have a greater number of average transactions per row, whereas density is comparatively less in T10I4D100k datasets. T10I4D100k is the sparser dataset amongst all the considered datasets which is attributed to its very less number of items in each transaction despite having a huge number of transactions and number of items.

Table 1. Dataset Description

Dataset	Number of Transactions	Number of items	Average length of items
Connect	67557	130	43
Chess	3196	74	37
Mushroom	8124	119	23
T10I4D100K	10482	922	10

Clearly, proposed method shows better result on the datasets such as Connect and Chess, and also a slight improvement in the mushroom dataset. But, it may not have any impact on the T10I4D100k dataset. Experiments are conducted using these datasets and applied to the algorithm one by one. The mushroom dataset contains characteristics of various species of mushrooms while connect dataset is derived from game steps. Chess dataset consists of steps moves involved in the chess game.

RESULTS

Table 2-5 shows the memory consumption and runtime of the compared algorithms on connect, chess, mushroom and T10I4D100k with different values of support. The legend used in all Runtime and Memory plots for both the algorithms is shown in figure 3. Graph plots are obtained when the TKFIN and FIN algorithms are run against the “connect”, “chess”, “mushroom” and “T10I4D100K” datasets are depicted in Fig 4 and Fig 5 respectively.

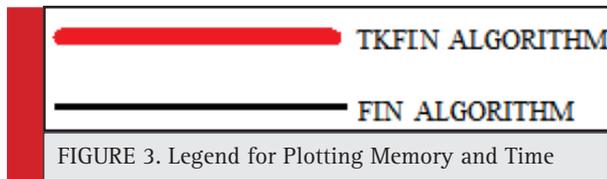


FIGURE 3. Legend for Plotting Memory and Time

Table 2 shows the memory consumption and run time of the compared algorithm on Connect dataset. The support here starts from 40% since below 40% threshold the number of frequent itemsets cannot be accommodated by the java heap. The upper limit is however insignificant since the values will be almost be converging. From fig 4a, the memory consumption by TKFIN algorithm has been stable from 40% to 60% support whereas in the FIN algorithm at 55% support the tree has a sudden increase in the memory attributed to the rise in the POC-tree nodes and similar behavior is achieved at 75% support. From there onwards the memory linearly rises in case of TKFIN and overlaps with the FIN algorithm and significant compression is achieved at 55% and 75% threshold value. The proposed algorithm works efficiently and significant amount of compression is achieved for Connect dataset. The maximum time taken is 15 seconds whereas the minimum was in order of 1 second. Memory consumed is in the orders of 100KBs, where the minimum memory consumed is 300KB and the maximum is consumed at 700KB.

Table 3 shows the memory consumption and run time of the compared algorithm on Chess dataset. The support here starts from 20% and ends up to 95% as shown in Fig. 4 b. The maximum time taken is 39 seconds whereas the minimum was in order of 200 milliseconds. Memory consumed is in the orders of 100KBs, where the minimum memory consumed is 300KB and the maximum is consumed at 700KB. The proposed algorithm works efficiently and significant amount of compression is achieved and run time is reduced for Chess dataset.

Table 4 shows the memory consumption and run time of the compared algorithm on Mushroom dataset and graphically depicted in Figure 4 c. The support here starts from 5% and ends up to 90%. The reason for such a small starting threshold is attributed to the small sizes of the dataset i.e. relatively lesser number of transactions when compared with other datasets. The maximum time taken is 24 seconds whereas the minimum was in order of 200 milliseconds. Memory consumed is in the orders of 100KBs, where the minimum memory consumed is 300KB and the maximum is consumed at 700KB. The proposed method shows slight improvement on Mushroom dataset. Table 5 shows the memory consumption and run time of the compared algorithm on T10I4D100K dataset and graphically depicted in Figure 4d. The support here starts from 20% and ends up to 95% similar to chess dataset. But since the number of transactions is in the range of 100,000s the memories shoot up to a maximum of 4761 KBs, which means almost 4MBs. Thus, the comparison becomes difficult because of two reasons: one is that the dataset is very huge and two the number of items and number of items per transaction is very less. The maximum time taken is 326 seconds whereas the minimum was in order of 300 milliseconds. The proposed method does not have significant improvement in T10I4D100K in terms of both memory and run time. The interesting phenomenon is that the proposed algorithm works well for denser datasets when compared to sparse datasets. In summary, no matter which parameter is

Table 2. Time and Memory Consumption of Connect

	Support	40	45	50	55	60	65	70	75	80	85	90	95
FIN	Time (s)	16.56	9.485	7.66	6.589	5.709	4.203	3.678	3.401	3.073	2.605	2.106	1.701
	Memory (KB)	305.73	305.736	305.73	465.376	305.688	465.328	465.32	465.21	759.768	464.76	610.64	566.08
TKFIN	Time (s)	15.803	9.291	7.82	6.623	5.741	4.072	3.604	3.369	2.979	2.656	2.079	1.685
	Memory (KB)	320.46	305.68	305.68	305.68	305.632	465.272	305.63	465.27	762.264	464.71	612.80	568.05

Table 3. Time and Memory Consumption of Chess

	Support	FIN		TKFIN	
		Time (s)	Memory (KB)	Time (s)	Memory (KB)
	20	37.748	306.488	39.387	306.488
	25	32.376	306.544	32.845	306.544
	30	31.765	306.544	29.907	306.544
	35	21.111	304.632	21.31	304.632
	40	13.47	304.576	13.75	304.576
	45	11.651	464.216	11.706	304.576
	50	8.637	304.312	8.624	463.896
	55	6.395	304.256	6.325	304.256
	60	4.763	463.84	4.752	304.144
	65	1.827	463.608	1.826	463.608
	70	0.687	463.392	0.672	463.392
	75	0.639	558.384	0.624	558.384
	80	0.406	746.008	0.39	746.008
	85	0.296	623.952	0.305	623.952
	90	0.254	473.152	0.25	471.568
	95	0.218	536.576	0.217	536.576

Table 4. Time and Memory Consumption of Mushroom

Support	FIN		TKFIN	
	Time	Memory	Time	Memory
5	23.496	306.496	24.093	306.496
10	9.965	305.256	9.628	305.256
15	4.925	304.528	4.805	304.528
20	2.857	304.184	2.797	304.184
25	1.209	463.528	1.187	463.528
30	0.65	601.688	0.648	601.688
35	0.481	798.672	0.475	695.616
40	0.412	693.512	0.405	693.512
45	0.373	642.4	0.36	642.4
50	0.315	570.568	0.315	570.568
55	0.282	543.144	0.279	543.144
60	0.266	531.024	0.267	531.024
65	0.252	522.584	0.252	522.584
70	0.243	519.488	0.244	519.488
75	0.245	519.488	0.242	519.488
80	0.24	519.448	0.246	519.448
85	0.232	502.064	0.232	516.984
90	0.235	516.776	0.231	501.856

chosen, experimental results suggests that the proposed algorithm is efficient for Chess and Connect dataset than Mushroom and T10I4D100K in terms of memory consumption and run time.

From the above results, the proposed method works well for denser datasets compared to sparse datasets. From fig 5a, the memory consumption by TKFIN algorithm has been stable from 40% to 60% support whereas in the FIN algorithm at 55% support the tree has a sudden increase in the memory attributed to the rise in the POC-tree nodes and similar behavior is achieved at 75% support. From there onwards the memory linearly rises in case of TKFIN and overlaps with the FIN algorithm and significant compression is achieved at 55% and 75% threshold value. The Chess Dataset is dense but not as much as the connect dataset. In chess dataset, memory is stabilized till the support reaches 50% for TKFIN algorithm, whereas the FIM has a spike as the threshold of 45% support. For mushroom dataset, TKFIN shows better results in terms of memory. The curve belonging to TKFIN does not rise up to a memory as much as 800 MB as in case of FIM. The memory consumption in case of TKFIN is not more than 700MB in all support thresholds. Evaluating time complexity between the two algorithms is very difficult since the graphs for all the above datasets render almost the same running times for both the

Table 5. Memory and time consumption of T10I4D100K					
	FIN		TKFIN		
Support	Time	Memory	Support	Time	Memory
25	334.514	4761.808	25	326.876	4761.944
30	121.033	2803.76	30	134.548	2803.76
35	37.678	1517.816	35	38.238	1517.816
40	5.592	749.464	40	5.561	749.464
45	2.233	558.232	45	2.251	558.232
50	1.03	423.8	50	1.04	423.8
55	0.718	946.28	55	0.727	946.28
60	0.527	466.376	60	0.534	481.224
65	0.468	371.016	65	0.46	371.016
70	0.42	321.68	70	0.419	336.736
75	0.375	327.112	75	0.385	327.112
80	0.386	327.112	80	0.386	327.112
85	0.36	332.88	85	0.371	332.88
90	0.364	331.288	90	0.362	331.288
95	0.334	338.656	95	0.337	338.656

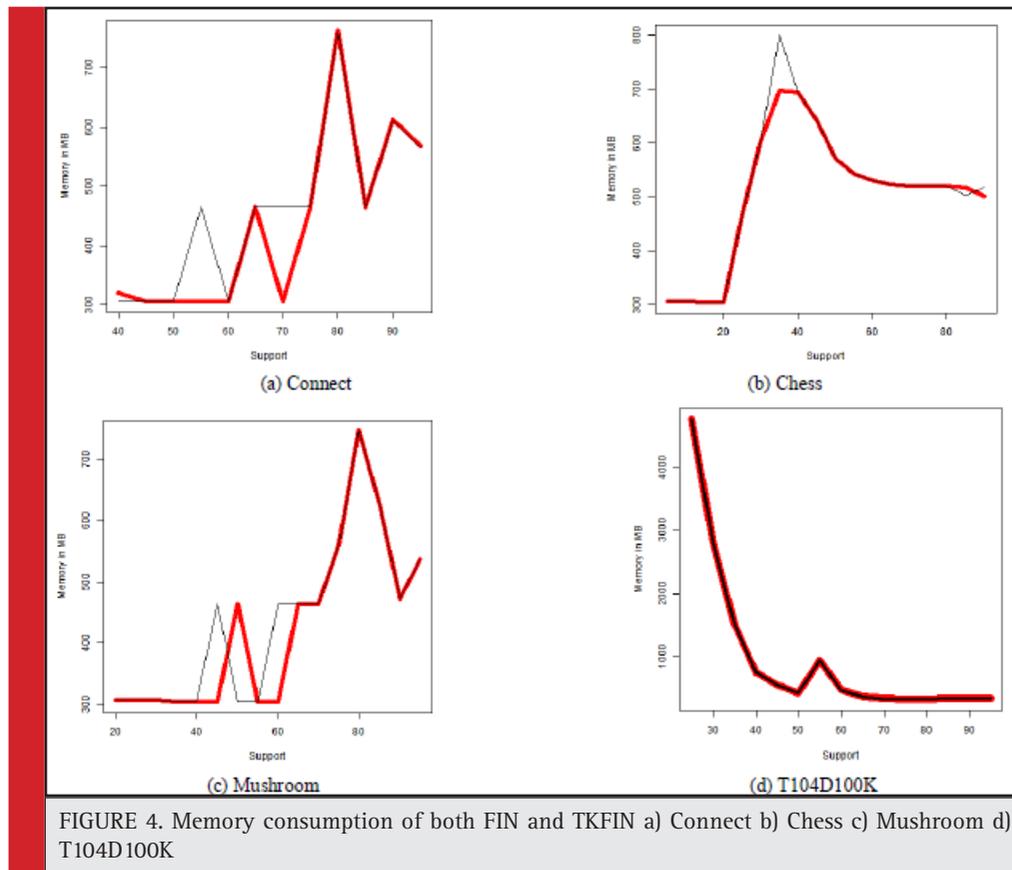
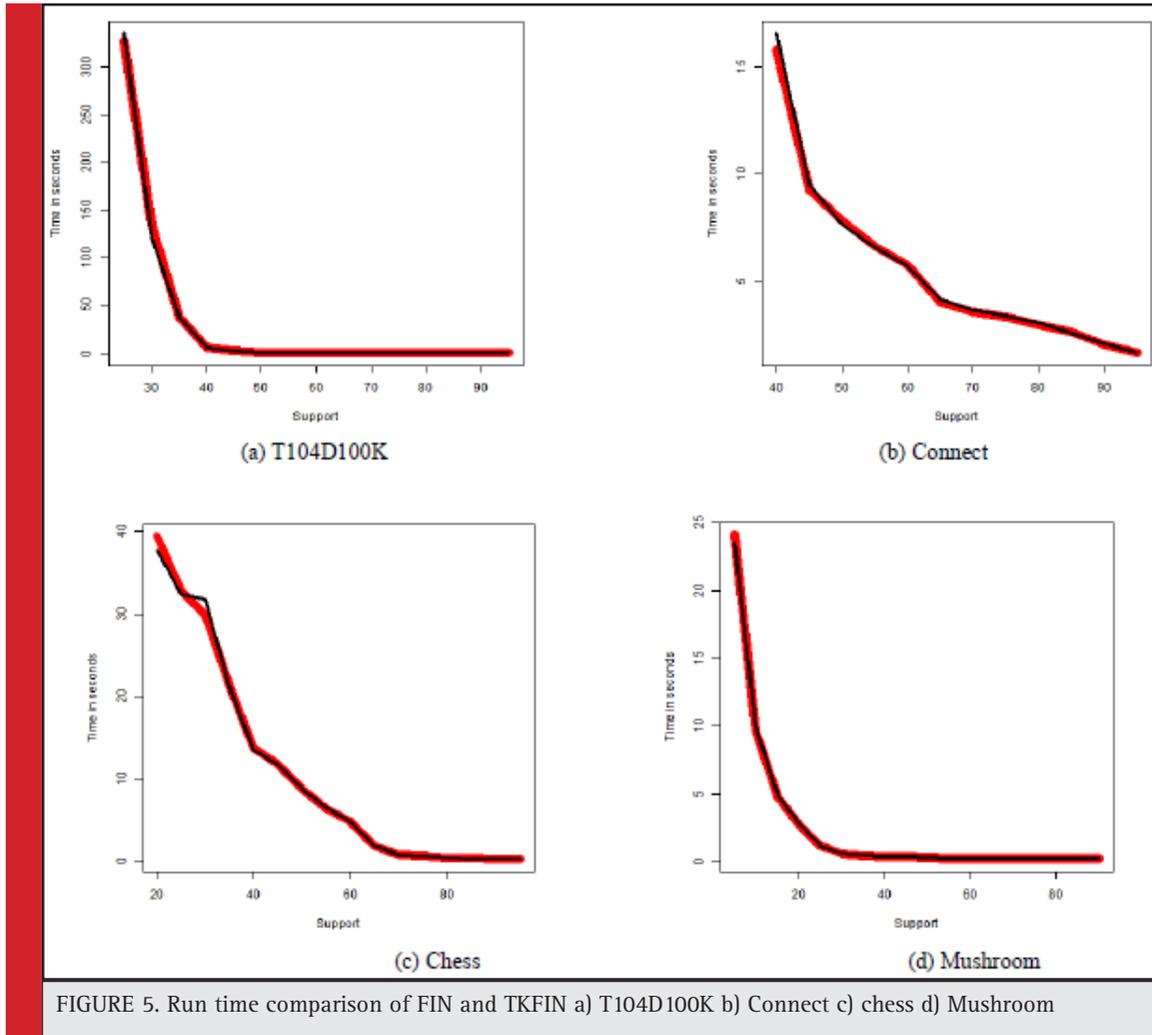


FIGURE 4. Memory consumption of both FIN and TKFIN a) Connect b) Chess c) Mushroom d) T104D100K



algorithms as in Fig 5. The proposed method shows better result on the datasets such as Connect and Chess, and also a slight improvement in the mushroom dataset. But, it may not have any impact on the T10I4D100k dataset.

DISCUSSION

Apriori algorithm provides good results in frequent itemsets to find the largest itemset using a method called

Table 6. Decreasing order of various algorithms in terms of memory consumption and runtime	
Decreasing order of algorithms in terms of memory consumption and runtime	Apriori
	Apriori - hash-based algorithm
	fp (frequent pattern) tree
	ECLAT
	Prepost Coding (PPC) tree
	Linear-Prefix tree
	FIN with Pre-Order Coding tree (POC)
	TKFIN with compressed POC tree

item reordering technique. Apriori algorithm scans the database again and again for finding frequent itemsets which results to slower execution. An enhanced version of apriori by using a hash-based algorithm consumes less memory and runs faster compared to apriori algorithm. Next, a PrePost coding tree algorithm is the fastest in most cases but consumes more memory. Based on Nodesets, an efficient algorithm called FIN to mining frequent itemsets is developed which saves half of memory compared with N-lists and Node-lists. Nodeset structure is efficient and FIN run faster than PrePost and FPgrowth on the whole. Even though FIN algorithm with POC tree is faster, FIN cannot generate topmost k frequent itemsets. Usage of compressed POC tree in TKFIN results to better performance in terms of both memory and run time. The extensive experiments show that the TKFIN algorithm is efficient and runs faster than FIN on the whole. Especially, TKFIN consumes much less memory than FIN on dense datasets. Table 6 shows the Decreasing order of various algorithms in terms of memory consumption and runtime.

CONCLUSION

In this study, the problem of frequent item set mining which forms the essential part of any association rule mining algorithm has been discussed. The problem and have highlighted the pioneering research works done in this field and also have evaluated them in a general manner. The algorithms for frequent itemset mining approach are divided into two categories. One which uses a Horizontal Data Format and other category uses the Vertical Data Format. Since from the past works it the Vertical Data Format has been proven to be the best approach, this field is chosen. The importance of the prefix tree based methods because of the most popular and successful approach known as fp-growth is considered. The work done using the vertical approaches is evaluated in detail, describing the merits and downfalls of each approach and how they can be overcome. An approach is selected that efficiently combines the elegance of prefix-tree based approach and the quick support counting facility of ECLAT approach which uses a tree called POC-tree and a data structure called nodeset. Finally, the algorithm is proposed for compressing the POC-tree and proposes an algorithm for mining top-k cardinality frequent itemsets using the same FIM approach. Furthermore, this algorithm can be used in an association rule miner which will be an efficient approach for mining association rules between the frequent itemsets. Also the compressed POC-tree can be used for mining maximal frequent itemsets and closed frequent itemsets.

REFERENCES

- Agrawal Rakesh Tomasz Imieliński and Arun Swami (1993) Mining association rules between sets of items in large databases ACM SIGMOD Record Vol 22 No 2
- Borgelt and Christian (2005) Keeping things simple: Finding frequent item sets by recursive elimination Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations ACM
- Brin and Sergey (1997) Dynamic itemset counting and implication rules for market basket data ACM SIGMOD Record Vol 26 No 2 ACM
- Burdick Douglas Manuel Calimlim and Johannes Gehrke (2001) MAFIA: A maximal frequent itemset algorithm for relational databases Data Engineering 2001 Proceedings 17th International Conference on IEEE
- Deng et al (2014) Fast mining frequent itemsets using Nodesets Expert Syst Appl Vol 41 No 10 Pages 4505–4512.
- Deng Wang and Jiang (2012) A new algorithm for fast mining frequent itemsets using N-lists Sci China Inf Sci Vol 55 No 9 Pages 2008–2030.
- Han Jiawei (2004) Mining frequent patterns without candidate generation: A frequent-pattern tree a Pagesroach Data mining and knowledge discovery Vol 81 Pages 53–87.
- Han Jiawei and Micheline Kamber (2006) Data Mining Southeast Asia Edition: Concepts and Techniques Morgan kaufmann.
- Han Jiawei Jian Pei and Yiwen Yin (2000) Mining frequent patterns without candidate generation ACM SIGMOD Record Vol 29 No 2 ACM.
- Park Jong Soo Ming-Syan Chen and Philip S Yu (1995) An effective hash-based algorithm for mining association rules Vol 24 No 2 ACM.
- Sarawagi Sunita Shiby Thomas and Rakesh Agrawal (1998) Integrating association rule mining with relational database systems: Alternatives and implications Vol 27 No 2 ACM
- Savasere A Omiecinski E R and Navathe S B (1995) An efficient algorithm for mining association rules in large databases.
- Srikant Ramakrishnan and Rakesh Agrawal (1996) Mining quantitative association rules in large relational tables ACM SIGMOD Record Vol 25 No 2 ACM
- Srikant Ramakrishnan Quoc Vu and Rakesh Agrawal (1997) Mining Association Rules with Item Constraints KDD Vol 97
- Sucahyo Yudho Giri and Raj P Gopalan (2004) CT-PRO: A Bottom-Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tree Data Structure FIMI Vol 4
- Toivonen Hannu (1996) Sampling large databases for association rules VLDB Vol 96
- Zaki (2003) Fast Vertical Mining Using Diffsets KDD '03 Proc ninth ACM SIGKDD Int Conf Knowl Discov data Min Pages 326–335
- Zaki Mohammed Javeed 2000 “Scalable algorithms for association mining” Knowledge and Data Engineering IEEE Transactions on Vol 123 Pages 372–390

An enhancement of data locality in Hadoop distributed file system

A. Siva Krishna Reddy¹, Pothula Sujatha*², Prasad Koti³, P. Dhavachelvan⁴
and J. Amudhavel⁵

^{1,2,4}Department of Computer Science, Pondicherry University, Puducherry, India

³Department of Computer Science, Saradha Gangadaran College, Puducherry India

⁵Department of CSE, KL University, Andhra Pradesh, India

ABSTRACT

The MapReduce system has greater prevalence because of its advantages, for example, programming simplicity, fault tolerance and data distribution. The number of utilizations based on Hadoop is growing because of its robustness and features. Data locality is a critical issue in parallel data applications where the task processing is spending a various amount of time and resource at particular locations. Some methodologies have been proposed to enhance the data locality. In this paper, we identify the DP problem across nodes and improve the data locality. At first, MapReduce system divides the dataset into smaller subsets called data blocks. These data blocks are encoded with erasure coding to achieve the reliability. Then, the Flexible Data Placement (FDP) algorithm applies to the slave nodes (data nodes) which dynamically dispatches the data blocks based on their locality. It will reduce the collision of vulnerability, network traffic and increases the throughput of the Hadoop system. With the help of analytical model, execution time of every task is identified which detects the job with data locality problem. Then, the hash table is built for data blocks to the node. In data locality, a program is transferred to the node where the original data placed. Experiments are conducted on two real-world data sets with different data placement approaches, which show that the proposed methodology diminishes the execution time and upgrades the performance of 42.5%, which is the better performance than the existing methods.

KEY WORDS: DATA PLACEMENT, DISK CONSUMPTION, HADOOP, MAPREDUCE, STORAGE COST

ARTICLE INFORMATION:

*Corresponding Author: prasad.k@pec.edu

Received 30th April, 2017

Accepted after revision 2nd Aug, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF: 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

INTRODUCTION

Large-scale distributed storage systems have been widely deployed in enterprises to provide high storage capacities, for example, Google File System, Amazon Dynamo, and HDFS. Data is striped over different nodes that offer local storage space in such kind of system Li *et al* (2015). In a network environment, nodes are interconnected as either wide area settings or cluster settings Pamies-Juarez *et al* (2013). All data needs to be stored constantly to assure that data failures of few of the storage nodes. The most straightforward approach to present redundancy is to save several copies of each data over the system Wang *et al* (2017).

Data files put away on convenient file system like HDFS which divides data blocks into fixed size. Every data block has three replicas that are set on random nodes. Each job is broken into tasks. For example, there is one map task per data block for MapReduce jobs Gao *et al* (2017). MapReduce is an appealing model for parallel data processing in high-performance cluster computing environments. The scalability of ended up being high, since work in the MapReduce exhibit is partitioned into various little errands running on different machines in a substantial cluster Xie *et al* (2016).

DP choice of HDFS is huge for the data locality which is an essential basis for task scheduling of MapReduce and in the end, influences the application execution Benifa JV and Bibal. Data locality is a critical issue with all data-parallel applications where a data-processing task absorbs various amounts of time and resources in various locations Abad *et al* (2011). Apache Hadoop is a common open source software framework to save and manage large-scale data. Basically, in Hadoop, data blocks are equally placed on Data Nodes Lee *et al* (2014). Hadoop stores three copies of all data in different nodes, without a policy to dynamically replicate this data Eltabakh Mohamed *et al* (2011). Data replication method is designed to solve the placement and replication problems for improved performance of Hadoop Liu *et al* (2016).

Hadoop has attempted a few techniques to improve the accessibility of users running on it, e.g. keeping up various replicas of application and redeploying application tasks in view of failures, yet it does not give high accessibility to itself Wu *et al* (2016). There are several challenges like SPOF (Single Point of Failure) identification, low overhead and flexible configuration to provide high availability for Hadoop Divya and Shunmuganathan (2016). Map and reduce are the functions which are used in Hadoop that process on data and stored in HDFS. It stores expansive records as a progression of pieces distributed over a bunch of data nodes and deals with replication for adaptation to non-critical failure Park *et al* (2016). The DP policy of HDFS tries to adjust the

load by putting blocks randomly; it doesn't consider any information qualities Anna and Theodore (1988).

Guaranteeing information accessibility in distributed storage frameworks is essential, given that hub disappointments are predominant Bui *et al* (2016). Information availability can be accomplished through erasure codes whose principal thought is to encode data fragments into equality portions, with the end goal that a subset of the information and parity sections can adequately reproduce the first information fragments Zhu *et al* (2015). Erasure codes can endure various failures while causing less capacity overhead than replication Li Runhui *et al* (2015). In spite of the fact that erasure codes can possibly lessen capacity costs in appropriated capacity frameworks fundamentally, there are so far two great conditions of using replication to store as of late exhibited information. They are pipelined insertion and data locality Zhu Yunfeng *et al* (2014).

Because of these properties, distributed storage frameworks frequently store recently presented information utilizing replication, and depend on eradication codes to achieve older and infrequently accessed data Weidong *et al* (2013). To adjust stack, Hadoop appropriates information to various hubs given plate space accessibility. Such information arrangement procedure is extremely down to earth and productive for a similar domain where hubs are indistinguishable as far as both processing and disk limit Xiong *et al* (2016).

The data locality is a key element as it significantly affects job completion times and system throughput. Reduce tasks are not influenced by the data locality because the input data for a reduce function is the output of all mapper. The data locality has an impact on map tasks because the input data for a map function is the stored data block in the HDFS. When a map task is performed on the Data node that contains the input data block, the map function can achieve the ideal performance. In particular, this paper makes the accompanying contributions:

1. We demonstrate the FDP which predicts the availability of performing host and assigns data blocks to that host. The extremely local accessible data enhances the execution of the Hadoop framework subsequently.
2. We set the threshold value to data blocks to minimize disk space consumption. So it will improve the Hadoop performance.
3. To improve Map-Reduce regarding saving the time of the execution and the system's resources.
4. To defeat data blocks cross-rack download while the encoding operation.

The remaining part of the paper composed as takes after: the next section describes the recent literature

survey. In Section 3, proposed data placement strategy is explained. Section 4 evaluates the implementation results and Section 5 concludes the work.

RELATED WORK

There is several research works on the performance of DP for MapReduce jobs in Hadoop's cluster. The below literature illustrate the DP strategy. Bui et al (2016) proposed a way to deal with the data file replication in view of predictive examination progressively. The use of every information document can be anticipated with the assistance of probability hypothesis to make a comparable replication methodology. In the long run, the prevalent documents can be in this manner recreated by their access possibilities. An erasure code is connected for the remaining minimal potential data to keep up the reliability. Thus, this methodology enhances the accessibility while keeping the security in contrast with the default conspires.

Wang et al (2014) developed another Data-gRouping-AWAre (DRAW) that address the data grouping impacts to enhance the performance for data-intensive applications with interest locality. DRAW powerfully examines information access from framework log documents. It extracts optimal data clusters and re-composes data layouts to accomplish the greatest parallelism per group subjective to load balance. By testing two certifiable Map Reduce applications with various DP plots on a 40-hub test bed, they presume that DRAW builds the aggregate number of neighbourhood delineates executed up to 59.8%. It diminishes the finish inertness of the guide stage up to 41.7%, and enhances the overall execution by 36.4%, in the examination with Hadoop's default arbitrary placement.

Xiong et al proposed a novel data placement strategy (SLDP) for different Hadoop clusters. Initially, SLDP selects an aware heterogeneity process to partition multiple nodes into many virtual storage tiers (VSTs). Then data blocks places across nodes in each VST indirectly according to the hotness of data. Moreover, SLDP utilizes a hotness corresponding replication to recover disk space and also has an efficient power control function. Experimental outcomes on two real world datasets demonstrate that SLDP is space-saving, energy-efficient and able to enhance MapReduce performance in a different Hadoop cluster well.

Qu et al (2016) proposed a dynamic replica strategy (DRS) in view of enhanced Markov model for HDFS, which incorporates a dynamic replica adjustment strategy (DARS) and a similar replica placement strategy (HRPS). As indicated by the Markov model, which can exchange to another state with an absolute probability value. This homogeneous placement strategy accomplishes uniform appropriation as well as considers the

connection between various types of data to set on the relevant data on the same rack or node to minimize the data transmission time and bandwidth utilization between the nodes or racks.

Tian Bowen et al (2016) concentrated on the preservation issue of data replication in Hadoop by presenting a Secure HDFS (SecHDFS), using a deviation of the SFAS section designation scheme and using it to the HDFS in Hadoop. Subsequent to presenting a file to Hadoop system SecHDFS plan will create an objective rundown of grouped candidate nodes that will be sent into an updated version of Hadoop's DP policy. SecHDFS will attempt to use whatever number of Data Nodes as could be expected under the circumstances to store distinctive parts of the file while keeping replications inside a similar sort of Data Nodes.

Our work is different from some other works on DP where the execution time of the specific task is measured and also the failure-free execution time. It will improve the Hadoop performance when a failure occurs in the execution time. The current DP schemes achieve the overall performance by 36.4%. By using flexible DP algorithm, we can reach the better performance than the existing approaches. Table 1 produces the notations that are used in this paper.

Table 1. Notations

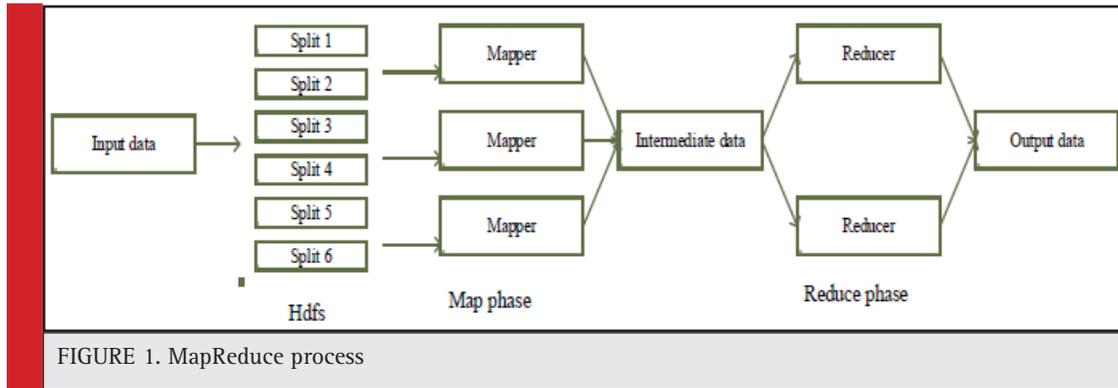
Notation	Definition
m	Block size
n	Node size
λ	Length of one task
γ	Interruption entry rate
t	Time instance
P	Rework cost
Q	System downtime
χ	Failure-free execution time
μ	Interruption mean time

PROBLEM DEFINITION AND FORMALIZATION

The principle point of the recommended mapping framework is to have an efficient distribution of the data blocks in the MapReduce environment. The performance of MapReduce is improved by data locality in heterogeneous clusters. Here, computations are placed in nodes which keep the blocks of data. In the cluster, the nodes parameter always changes because of the character of heterogeneity in Hadoop.

A. MapReduce

Figure 1 demonstrates the process of MapReduce. MapReduce framework is working with a large dataset



which contains two phases, Map and Reduce. The key-value pairs are the input of map tasks, and it will convert them into transitional key-value pairs. These intermediate pairs are used in the reduce tasks to produce the output. Google initially developed MapReduce for web searches. Some other systems like Hadoop and Microsoft Dryad are also designed to enhance the MapReduce framework.

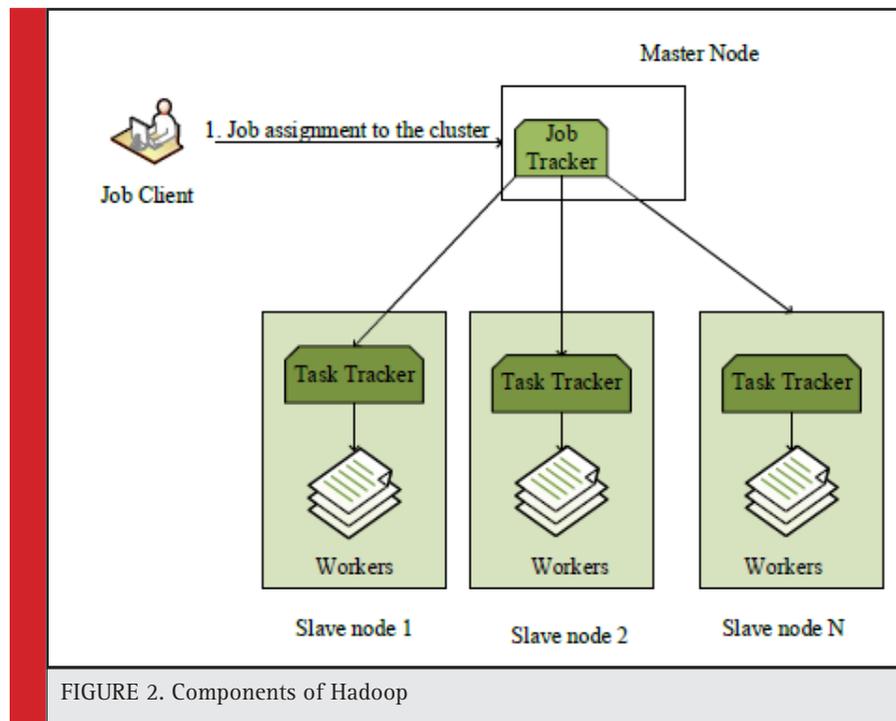
B. Hadoop

Hadoop consists of two subsystems which are Hadoop runtime system and HDFS. The Hadoop runtime system contains one job tracker and multiple task trackers. As shown in Fig. 2, the job tracker gathers the job and distributes the task of every job onto the task tracker. A file is an input which splits into fixed-size data blocks and stored in HDFS. The block of data is replicated and

spreading over the Hadoop cluster for fault tolerance. For processing the input file, job tracker converts the job into small tasks, and these tasks are allocated to an empty server to deal with a data block together. The name node (job tracker) maintains the Metadata management while the nodes store data. The job could not be completed until all the tasks of a particular job completed. Equal size data blocks are created in HDFS files. The blocks are distributed randomly for balanced distribution.

C. Data replication

In HDFS, we can store large files in the more massive cluster. These large files consist of series of blocks, and the blocks in the files are the same size except the last block. Blocks are replicated to enhance the fault tolerance. The placement of replicas is a major thing in HDFS to improve the performance. The rack-aware rep-



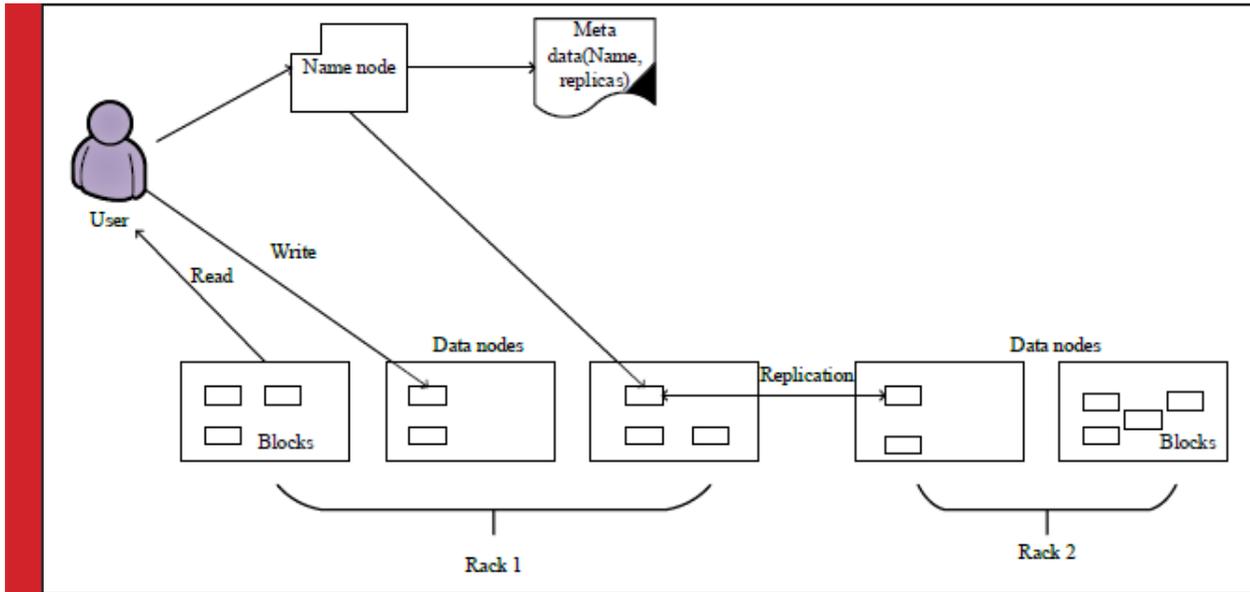


FIGURE 3. HDFS Architecture

lica placement improves the availability, reliability, and bandwidth utilization.

D. HDFS

HDFS is a distributed file system, and it is storage system of Hadoop. HDFS is depended on write-once and read many times arrangement. In contrast with other distributed systems, HDFS had many faults tolerant and created with low-cost asset hardware. The design of HDFS is represented in figure 3. HDFS is processing a large volume of data and grants simple access. The files are stored across

multiple machines to keep such large data. These files are stored repetitively to keep up a strategic separation from possible information misfortunes if there ought to emerge an event of disappointment. HDFS in like manner creates applications accessible for parallel processing.

DATA PLACEMENT APPROACH USING FDP

We propose an efficient DP algorithm for MapReduce to reduce the impact of heterogeneity and unpredictability. Figure 4 demonstrates the overview of proposed

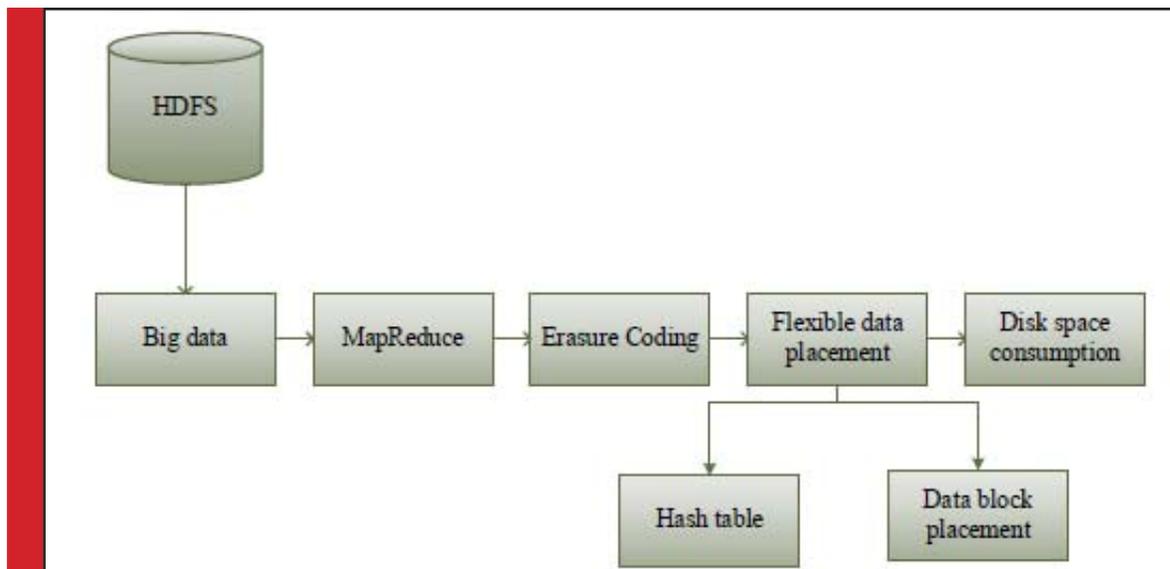
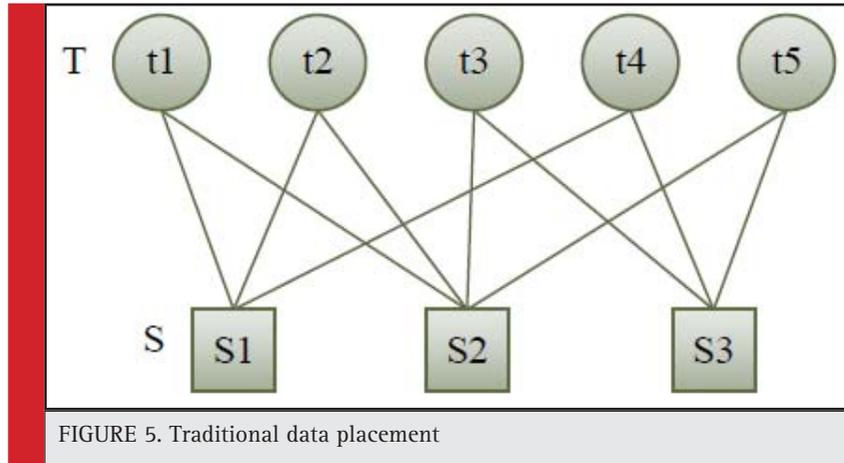


FIGURE 4. Architecture of FDP



work. The data blocks are encoded with erasure coding to protect the data. The FDP dynamically distributes data blocks onto corresponding hosts. Then the disk space consumption will be avoided by setting a threshold value to the nodes. Because of the limited storage limit of every node, it is imperative to have an efficient replication procedure to ensure the overall execution. During the DP process, it prompts to extra disk space utilization for more positive nodes. We fix a threshold to address this issue for the quantity of blocks that can be allotted on each node. The node that achieves the threshold won't be recognized for further placement.

Figure 5 illustrates the graphical view of DP. There are five tasks, and three servers are considered. Each task has two input data replicas.

A. FDP algorithm

DP is critical to the storage system for data accessibility and adaptation to internal failure. An excellent DP method ought to enhance the data availability, reliability and network data transfer capacity. In this manner, we concentrate replication procedure on accomplishing these objectives. In this section, we give details about flexible DP using the hash table. Initially, the substantial dataset are separated into little subsets by MapReduce.

Erasure coding

The coding approaches are the most important part used to reduce the storage overhead. The 'm' size of data is isolated into 'q' blocks. Every storing m/q bytes of data and further parity blocks r are added to it which hold the encoded data of 'q' blocks. These encrypted data is assigned to $n = q + r$ nodes, where 'q' nodes maintain data fragments and the rest of 'r' nodes are parity nodes. Each node store m/q bytes that produces the result in an absolute storage of $n.m/q$ and a rate of the code is q/n . If an exact node fails, its data can be restored by transferring m/q bytes of data from any 'q' cohesion

nodes. Erasure codes can accept up to $n - q$ failures. This approach diminishes storage cost and gives reliability. For that, the nodes need processing capacities to encode and decode the data. Reed-Solomon code is the most useful strategy of erasure coding. When all 'q' data blocks are gotten, the task tracker calculates the equality blocks and keeps in touch with them into HDFS.

B. Execution time estimation of one Task

The task associated with delay constraint is executed for a time it is identified by a newly developed analytical model and based on this the data is placed. The host addressed by $j(0 \leq j \leq n - 1)$ having the autonomous inter-arrival times of delay and accurately assigned arbitrary variables with constant γ (i.e. the opposite of MTBI). The values of γ_j is changed according to γ whereas in the clear situation. The reconstruction time pursues a general assignment with mean μ . While recovering a previous interruption, the disruptions will be occurring between the interruptions and these are managed by First in First out (FIFO) approach. The clustering is done on present interruptions after clearing the previous ones. When we regard as an event in the queue the system will be treat as an M/G/1 queuing system.

Let us consider a particular assignment with failure-free execution time having length χ . On the off chance that no interferences are happened though executing the χ job, then it is known as a pleasant task. If any interference occurs in the period $t < \chi$, then the task will be completed after a failure recovery. In this way, it develops the task completion by checking the expiry articulation.

We characterize the random variable 'D' stand for the period to complete a task of length χ , we have,

$$\begin{aligned}
 D &= \chi + P_1 + P_2 + \dots + P_m + Q_1 + Q_2 + \dots + Q_m \\
 &= \chi + \sum_{j=1}^m P_j + \sum_{j=1}^m Q_j
 \end{aligned}
 \tag{1}$$

Algorithm 1: hash table creation
Input requirements: m data blocks and n nodes, interruption entry rate γ , interruption meantime μ , failure-free execution period of a task χ
Output: hash table $H(0...m - 1)$
 Start
 For every node $j(0 \leq j < n)$
 Compute $E_j(D)$ using eq. 7
 End for

$$\Psi = \sum_{j=0}^{j<n} \frac{1}{E_j(D)}$$
 $g = 0$ // creating hash table keys to node j
 For every node $j(0 \leq j < n)$
 $v_j = m \times rank_j$ // stop creating hash table keys
 $rank_j = \frac{1/E_j(D)}{\Psi}$
 $h = |g + v_j|$
 For $|g| \leq j < |h|$
 Insert j to list $H[j]$
 End
 $g=h$
 End for
 Stop

Algorithm 2: data placement
Input parameters: hash table H built from algorithm 1
Output: node id which hosts the data block
 Start
 Create arbitrary integer $w(0 \leq w < m)$
 If size ($H[w]$)=1
 Return $H[w]$
 Else
 Generate w_1 such that $(0 \leq w_1 < 1)$
 Lower bound=0;
 For every $i \in H[w]$
 $\xi = \sum_{k=0}^{k<size(H[w])=1} rate_i$ // collision handling
 Upper bound=lower bound+ $rank_j/\xi$;
 If lower bound $\leq w_1 < upper\ bound$
 Return i
 End if
 Lower bound=upper bound;
 End for
 End if
 Stop

The hash table is made for each emphasis when the MapReduce application instates its contribution to the HDFS.

DATA BLOCK PLACEMENT

The DP phase takes the decision of placement. At first, we create arbitrary integers $w(0 \leq w < m)$ that utilize the key to finding things in the hash table H. On the off chance that there is a single value relating to the key of w, we just select the relevant node to host the data. When there is an arise in occurrence of crashes on the key, we will deliver another arbitrary number in the scope of 0 and 1, and rehash each an incentive in the crash fasten to pick the host.

Interruptions over job execution affect HDFS, as well as the MapReduce runtime framework. The system is complicated for job processing than data exchange that creates more vulnerable to interruptions. The essential thing in FDP algorithm is disk space utilization. To take care of this issue, we assign a threshold value for the data blocks which can be given to each node. If the node reaches the particular threshold, it does not consider for further DP which helps to tune the DP and keep up the client loyalty. Assume MapReduce system has m blocks, n nodes and 'y' the number of replicas in the MapReduce system. We set the threshold as $m \times (y + 1)/n$ accordingly the blocks allotted to each node do not exceed its normal number with one more replica. Table 2 represents the examination of DP approaches with related methodologies.

Table 2. Comparison of related strategies

Strategies	DRAW	SLDP	Our scheme
Data locality	Medium	Medium	High
Execution time	Medium	High	low
Fault tolerance	Low	Low	High
Load balancing	Medium	Medium	High
Disk space consumption	-	-	High
Cross-rack download	-	-	High
Reliability	Medium	Medium	High
Performance	Medium	Medium	High

EXPERIMENTAL EVALUATION AND ANALYSIS

This part explains the experimental environment used for evaluating proposed DP algorithm. Here we also consider the workload and performance measures used for evaluation. To achieve the better performance, we apply the FDP algorithm in the MATLAB platform on Hadoop with the accompanying configuration.

Operating system: Windows 7
 Processor: Intel Core 2 Quad @ 2.5 GHz
 Mat lab version: R 2016a
 RAM: 3GB

The Apache Hadoop consists of a cache part known as HDFS, and another handling part is called as MapReduce.

Experimental environment

The experiments are conducted with computer simulations. The simulated HDFS consist of one name node and 5 data nodes. The initial usage of data nodes are 0 and the network bandwidth between two nodes is 256 Mb. The master node assigned to run Job Tracker and NameNode, while the slaves run Task Trackers and Data nodes. The master node is designed to show the usefulness of a slave node. Table 3 demonstrates the Hadoop configurable parameters.

Performance evaluation

The data locality determined by HDFSDP strategy is a major factor for MapReduce performance. During the assessment, we compared two DP mechanisms of HDFS such as DRAW and SLDP. The numbers of jobs are set from 100 to 1500 with 100 increments each time. Following performance metrics are used to measure the performance of DP.

Execution time: the job execution time or task execution time details are available in the job tracker database. Strategies that enable jobs to read data locally result in

Parameter	value
Number of nodes	1 name node, 5 data node
CPU	Intel®Core
Block size	64 MB
Bandwidth	8Mb/s
Ratio of interrupted nodes	1/2
Disk	50 GB
Memory	1 GB/ 2 GB/ 4 GB

quicker performance; consequently this metric measures the per-job advantage of data locality. Two types of jobs are used to evaluate the execution of FDP in a Hadoop heterogeneous cluster. The size of input data (job) is 1 GB and 2 GB correspondingly. At first, the execution time of jobs is figuring out for every node to execute various sizes of data. It incorporates the time needed for putting a file to the desired location until the point when the execution procedure finishes.

Figure 6(a) demonstrates the average execution time of input jobs respectively. The execution time of every node is corresponding to the data size. If the size of the data block is set to be larger, the time of executing a local task and non-local task is high. Figure 6(b) represents the average execution time of jobs, and the DP schemes DRAW and SLDP is used for comparison. For example, the job 1 data partitioned into 9 data blocks, and the blocks are applied to erasure coding. Based on erasure coding, data node 1 and two is assigned 2 data blocks, data node three is assigned 3 data blocks and data node 4 and five has assigned one blocks. The comparison with default Hadoop strategy, FDP minimizes the execution time up to 11.5%. Figure 6(c) displays the observation of various DP strategies on job execution period. The proposed FDP strategy provides better results as compared with other procedures.

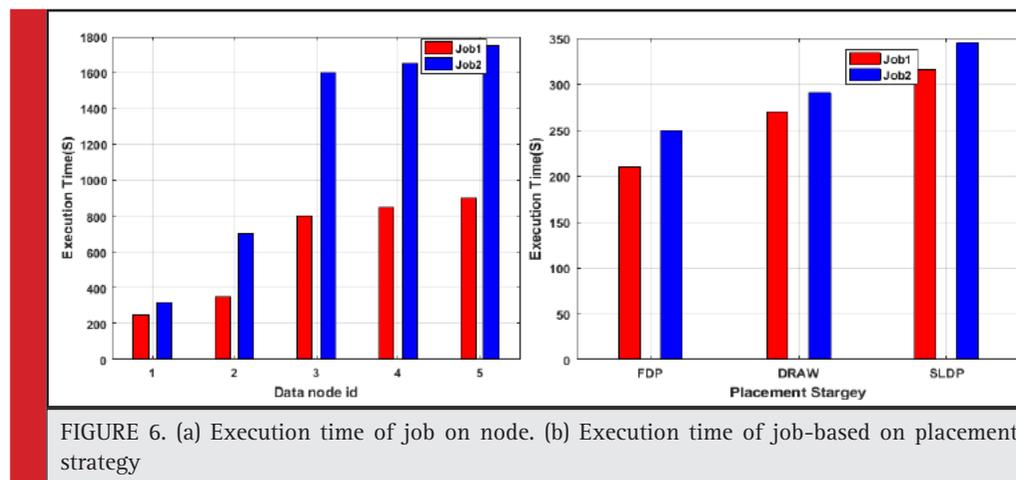


FIGURE 6. (a) Execution time of job on node. (b) Execution time of job-based on placement strategy

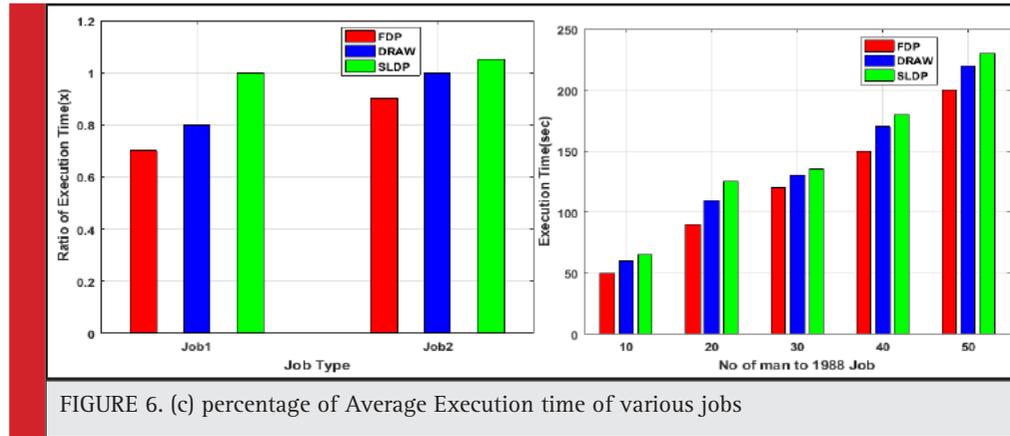


FIGURE 6. (c) percentage of Average Execution time of various jobs

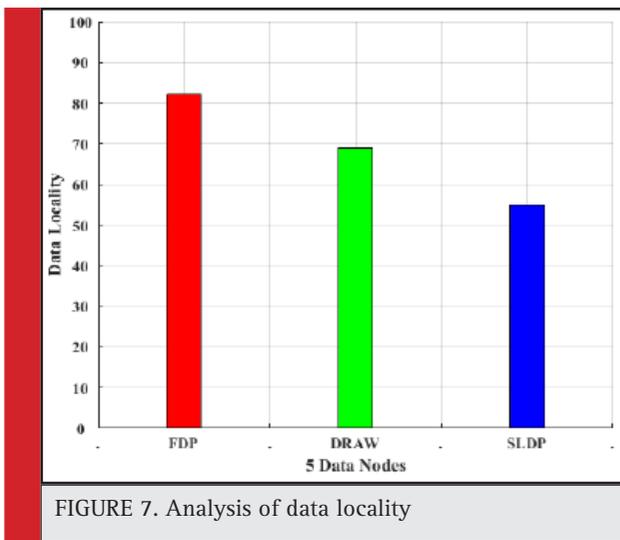


FIGURE 7. Analysis of data locality

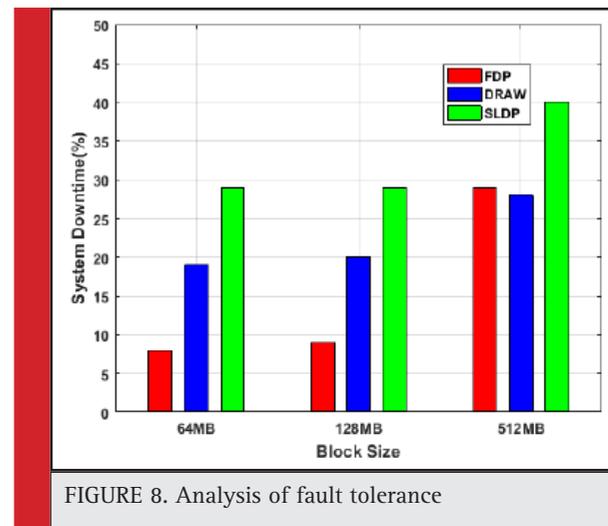


FIGURE 8. Analysis of fault tolerance

Data Locality: The term data locality is characterized as the quantity of tasks running on similar resources where their saved data are found. Running the task on a similar node where the data exists is efficient

and in the event that it is impractical at that point functioning the task in the same rack is embraced. Data locality amplifies the general throughput of the framework.

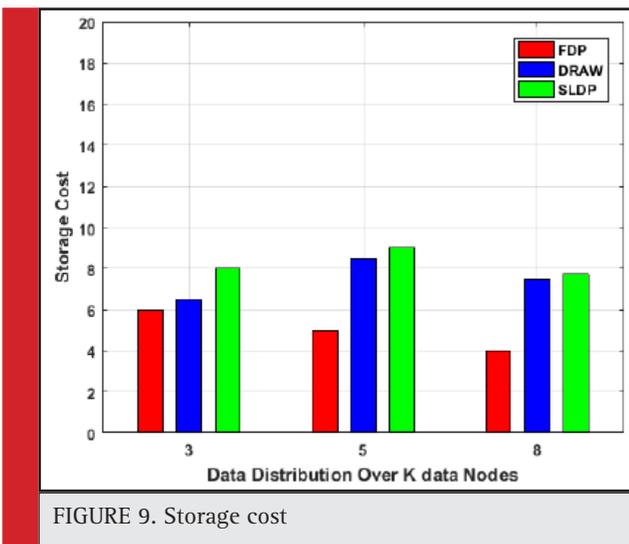


FIGURE 9. Storage cost

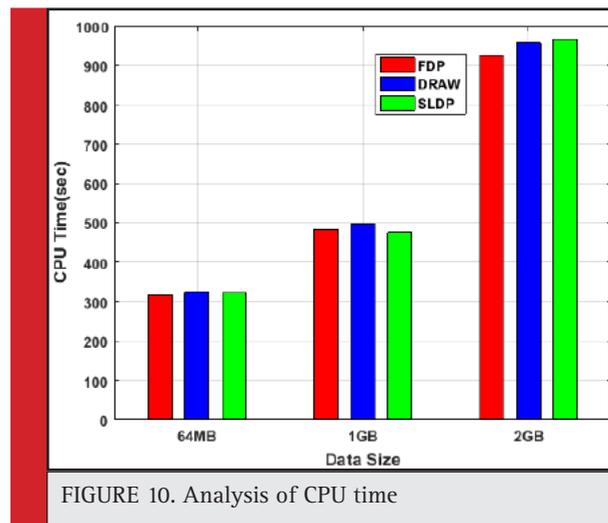


FIGURE 10. Analysis of CPU time

Figure 7 illustrates the data locality of three types of methods. FDP improves data locality by approximately 42.5 % on other techniques. FDP considered data locality view of DP and exchange data to task executions. Despite the fact that different strategies considered critical factors, for example, resource limit and quickest evaluated begin time of VMs, which can build throughput and speed, without feeling data-locality their performance is imperfect. Interestingly, our procedure FDP enhances the general execution time and resource utilization.

Fault tolerance: One primary advantage of MapReduce is the fault tolerance which consequently keeps up the failure and hides the difficulty of the adaptation to non-critical failure from the software engineers. By default, the one replica of each block is stored across the MapReduce in Hadoop, and it's changed only when required. HDFS's fault tolerant can be inspected in such situations when any hub goes down, the information on that hub can be recouped effectively from different hubs. Failures of a particular hub or undertaking are recovered consequently by the structure.

From Figure 8, we report the outcome for HDFS block sizes of 64MB, 128 MB and 512MB. From the graph, we can see the slowdown rate is higher when a block size of 512MB because the measure of halfway outcomes lost because of the disappointment is greater. To discover the adaptation to non-critical failure, we arbitrarily select a hub to execute after half of occupation is finished.

Storage cost: In FDP, the load across the cluster is totally adjusted guarantees least storage cost.

Figure 9 displays the storage cost of various size of data nodes $k=3$, $k=5$ and $k=8$. The x-axis shows three values of k . The y-axis represents the value of storage cost of corresponding values for DP. Minimum storage cost is our requirement, and the cost decreases when the nodes are increases from 3 to 8.

CPU time: Here, the measure of time for which a CPU is utilized for preparing instructions of a program is showed. This is a measure of execution since it measures the effectiveness of the framework. An idle CPU has 0% CPU usage since it offers invalid execution per unit cost. The higher CPU usage accomplishes more noteworthy proficiency of the framework.

Examination of CPU Time utilized by a job is made by fluctuating the span of data provided to the three DP approaches. As appeared in Figure 10, FDP takes less CPU Time.

CONCLUSION

Distributed storage management system is needed to enhance the execution of data intensive applications. In

this paper, we studied the problem of data locality and introduce the FDP algorithm for MapReduce applications in heterogeneous Hadoop environment. This FDP algorithm aims to make proper mapping between data blocks to corresponding data nodes to solve the data locality problem. To maintain the locality, the unsuccessful execution time is measured and hash table is created for every node. Then, the data blocks are placed using the random integer and the collision is handled by rank value. So, the algorithm enhances the throughput and reduces the execution time and CPU time. Various performance measures such as execution time, data locality, storage cost, CPU time, and fault tolerance are compared separately for DRAW and SLDP in a heterogeneous cluster environment. We conduct experimental evaluation using some test bed experiments and analysis, and show that proposed algorithm achieves high data locality when compared to the DRAW and SLDP.

REFERENCES

- Abad M et al (2011) DARE Adaptive data replication for efficient cluster scheduling In Cluster Computing (CLUSTER) 2011 IEEE International Conference on pp 159-168 Ieee 2011
- Benifa JV and Bibal Performance Improvement of MapReduce for Heterogeneous Clusters Based on Efficient Locality and Replica Aware Scheduling (ELRAS) Strategy Wireless Personal Communications Pages 1-25
- Bui et al (2016) Adaptive Replication Management in HDFS based on Supervised Learning IEEE Transactions on Knowledge and Data Engineering Vol 28 No 6 Pages 1369-1382
- Bui et al (2016) Adaptive Replication Management in HDFS based on Supervised Learning IEEE Transactions on Knowledge and Data Engineering 28 no 6 (2016) 1369-1382
- Divya M and Shunmuganathan (2016) A multi-agent based intelligent query processing system for Hadoop with FIPA-OS using cooperating agent in cloud environment In Intelligent Systems and Control (ISCO) 2016 10th International Conference Pages 1-6
- Eltabakh Mohamed et al (2011) CoHadoop flexible data placement and its exploitation in Hadoop Proceedings of the VLDB Endowment Vol 4 No 9 Pages 575-585
- Gao et al (2017) Compact Popularity-Aware and Adaptive Hybrid Data Placement Schemes for Heterogeneous Cloud Storage IEEE Access Vol 5 Pages 1306-1318
- Hac Anna and Theodore J (1988) Dynamic load balancing through process and read-site placement in a distributed system AT&T technical journal Vol 67 No 5 Pages 72-85
- Lee et al (2014) A dynamic data placement strategy for hadoop in heterogeneous environments Big Data Research Vol 1 Pages 14-22
- Li et al (2015) Enabling concurrent failure recovery for regenerating-coding-based storage systems From theory to practice IEEE Transactions on Computers Vol 64 No 7 Pages 1898-1911

- Li Runhui et al (2015) Enabling efficient and reliable transition from replication to erasure coding for clustered file systems In Dependable Systems and Networks (DSN) 2015 45th Annual IEEE/IFIP International Conference on Pages 148-159
- Liu et al (2016) Selective data replication for online social networks with distributed datacenters IEEE Transactions on Parallel and Distributed Systems Vol 27 No 8 Pages 2377-2393
- Pamies-Juarez et al (2013) Rapid RAID Pipelined erasure codes for fast data archival in distributed storage systems In INFOCOM 2013 Proceedings IEEE Pages 1294-1302
- Park et al (2016) An efficient Hadoop data replication method design for heterogeneous clusters In Proceedings of the 31st Annual ACM Symposium on Applied Computing Pages 2182-2184 ACM
- Qu et al (2016) A dynamic replica strategy based on Markov model for hadoop distributed file system (HDFS) In Cloud Computing and Intelligence Systems (CCIS) 2016 4th International Conference on pp Pages 337-342
- Tian Bowen et al (2016) HDFS A Secure Data Allocation Scheme for Heterogenous Hadoop Systems In Networking Architecture and Storage 2016 IEEE International Conference on Pages 1-2
- Wang et al (2014) Yin DRAW A new Data-gRouping-AWAre data placement scheme for data intensive applications with interest locality In Cloud Computing for Data-Intensive Applications Pages 149-174
- Wang et al (2017) A new reliability model in replication-based big data storage systems Journal of Parallel and Distributed Computing
- Weidong et al (2013) Tree-structured parallel regeneration for multiple data losses in distributed storage systems based on erasure codes China Communications Vol 10 No 4 Pages 113-125
- Wu et al (2016) A new data-grouping-aware dynamic data placement method that take into account jobs execute frequency for Hadoop Microprocessors and Microsystems Vol 47 Pages 161-169
- Xie et al (2016) Locality-Aware Scheduling With Stochastic Delay Optimality IEEE/ACM Transactions on Networking
- Xiong et al (2016) An Energy-Aware Workload Balancing Method for Cloud Video Data Storage Management In Advanced Cloud and Big Data (CBD) 2016 International Conference on IEEE 2016 Pages 7-12
- Xiong et al Optimizing data placement in heterogeneous Hadoop clusters Springer Cluster Computing Vol 18 No 4 Pages 1465-1480
- Zhu et al (2015) Boosting degraded reads in heterogeneous erasure-coded storage systems IEEE Transactions on Computers Vol 64 No 8 Pages 2145-2157
- Zhu Yunfeng et al (2014) On the speedup of recovery in large-scale erasure-coded storage systems IEEE Transactions on Parallel and Distributed Systems Vol 25 No 7 Pages 1830-1840

Job Scheduling Optimization using BAT Algorithm in Hadoop Environment

Raghav R.S¹, J. Amudhavel² and P. Dhavachelvan³

^{1,3}Department of Computer Science, Pondicherry University, Puducherry, India

²Department of Computer Science and Engineering, KL University, Andhra Pradesh, India

ABSTRACT

The role of Hadoop in the big data environment is meant to be massive; it needs special attention for processing large data sets in a distributed environment. The role of HDFS is to store the data and Mapreduce is used for processing the data, where the data job is submitted by the master node and the task is assigned to slave node for effective processing. The aim of this article is to reduce the execution time and cost by improving the reliability. Here we use bio inspired algorithm known as BAT algorithm, it identifies the slave node accurately which helps the master node to assign task quickly. It also helps to find the sticky slots and avoid job starvation by reducing the scheduling complexity. The BAT algorithm produces better results by outperforming other basic randomized evolutionary algorithm.

KEY WORDS: BAT ALGORITHM, HADOOP, JOB SCHEDULING, MAP REDUCE AND TASK SCHEDULING

INTRODUCTION

In present world, the growth of data is massive and the gathering of data requires some specific techniques to handle the complexity issues. The big data compresses of 3 V's such as Volume velocity and Variety of data, these data format are classified into different types such as structured data semi structured and unstructured data. In big data environment handling of large data sets is carried by using the Map Reduce . It is defined as distributed computing model. The role Map reduces is to solve the complex task by breaking down them into small chunks which is mentioned to be scalable. It executes parallel processing

combined with the HDFS (Hadoop Distributed file system). The data processing is carried at the location of the data, without migrating data to the execution location. The same physical nodes in the cluster are used for storage and computation of data (Rajadurai et al., 2017).

The huge amount of data can be processed effectively without having any complex issues. The Map Reduce consists of two main phases known as Map and Reduce (Somasundaram and Govindarajan 2014). The role of Map is to accept set of input and it also provides some converting process into intermediate key-value pair. The loading process of files is carried by the mapper to that machine and processes them. The merging of all values

ARTICLE INFORMATION:

*Corresponding Author: vpmrags@gmail.com

Received 27th April, 2017

Accepted after revision 29th July, 2017

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007



Thomson Reuters ISI ESC and Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 Cosmos IF : 4.006

© A Society of Science and Nature Publication, 2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

mapped with the same key and transfers these pair of key-value pairs to Reduce phase. The role of reduce phase is to convert these values to smaller set of values for quick processing. There are varieties of nature-inspired algorithms and it is applied to variety of applications. It is incorporated to different numerical and combinatorial optimization problems (Anusha et al., 2015).

These authors have discussed about the enhancement carried in the Hadoop MapReduce framework, explaining about the variety of algorithms which is mainly suitable for heterogeneous environments. The aim of the algorithm is to provide effective platform for processing the massive data task in heterogeneous environments. Somasundaram and Govindarajan, (2014) explained about the design and evolve a CLOUD Resource Broker (CLOUDRB) for managing cloud resources. It also denotes about the job completion, which is applied on variety of complex environments to handle data within short span of time. Here the authors used deadline concept to complete the task by combining Job Scheduling and Particle Swarm Optimization (PSO)-based Resource Allocation mechanism. In order to reduce the cost and execution time of the process, and this is considered as a fitness function.

Lau and Qiu (2014) have illustrated about consolidate framework for handling generic job shop scheduling issues. The main problem is divided into three main groups such as static based scheduling problems, semi-dynamic scheduling problems and dynamic scheduling problems. An artificial immune systems algorithm is based on human immune system, in order to handle job shop scheduling problems. They provided a high level decision support model for improvising placement of the scheduling scheme. Kaur and Singh (2016) describe about the meta-heuristic method known as BAT algorithm, the feature of this algorithm is to handle multi-objective issues. The theme of the algorithm is to reduce execution time and increase the reliability within low assigned budget, (Mamdapure, et al., , 2016) have introduced an effective way for storing massive amount of data for processing with the help of Mapreduce. The goal of this article is to carry parallelize job execution across multiple nodes. It also helps to reduce execution time of a parallel application for effective processing.

MATERIAL AND METHODS

In order to achieve efficient job and task scheduling with less execution of time and low cost for completing the job (Nagina and Dhingra, 2016). This model helps to fulfil the request generated by the client and provide them a quick response with less execution time (Gayathri et al., 2016). The theme of the work is 1) to reduce the execution time and improvise the reliability. 2) The node which takes more execution time to process the task will be ruled out. 3) To

manage queue jobs and to reduce job starvation 4) to minimize the scheduling complexity and to avoid sticky slots. The number of jobs $J = (j_1, j_2, j_3, \dots, j_n)$, T represents number of tasks $T = (t_1, t_2, t_3, \dots, t_n)$. (Sasidevi et al. 2017).

C_j denotes the cost for each job, C_i completion time of each task. The node in the model to n, which is stored using the set S, S(n) which represents first node and so on. The total number of nodes is described as TN_{nodes} . Then $n \in [1, TN_{nodes}]$. The nodes memory performance is evaluated $P_{NM}(n) = size_{NM}(n)$, where the $size_{NM}$ express about the size of the nodes memory and this parameters is consider to be one of the key factor to evaluate the memory performance. The storage capacity of the node is described using the following notation $size_{nodes}(n)$ for the node n. The used capacity can be denoted as $UC_{nodes}(n)$ for the node n. The reliability of a system depends on certain characteristics of the slave nodes such as memory size, failure rate etc. Here we consider the memory size and it is described by the following equation

$$R(P_{NM}(n)) = \min(r(t_1), r(t_2), \dots, r(t_n)). \quad (1)$$

Each Task is denoted as

$$T_i = \begin{vmatrix} C_t \\ C_j \\ R(P_{NM}(n)) \end{vmatrix} \quad (2)$$

C_j denotes the cost for each job, C_i completion time of each task, $R(P_{NM}(n))$ reliability of a node for each task.

```

Algorithm1: Pseudo code of BAT algorithm
Objective function  $F(x)$ ,  $x = (x_1, \dots, x_d)^T$ 
Initialization of population of bat  $x_i, v_i$  ( $i=1,2,\dots,n$ )
Pulse frequency is noted as ( $f_i$ ) at  $x_i$ 
Initial pulse rates  $r_i$  & loudness  $A_i$ 
while ( $t < \text{Max no of iterations}$ ) do
     $P_i = P_{min} + (P_{max} - P_{min}) \beta$ 
     $S_i^t = S^{t-1} + (P_o_i^t - P_o_i) P_i$  //The frequency
    adjustment to generate
    new solutions and
    velocity
    updating/location is
    done
if ( $\text{rand} > RP_i$ ) then
    Solution selection among best solution
    local solution generation around the chosen best
end if
Random fly based new solution is generated
if ( $\text{rand} < AL_i$  &  $P(P_o_i) < f(P_o_i)$ ) then
    Acquire new solutions
    Maximize  $RP_i$  and minimize  $AL_i$ 
end if
Order of Ranking the bats to find the current best  $P_o_i$ 
end while
Display of results and visualization

```

ALGORITHM 1: BAT based Task scheduling

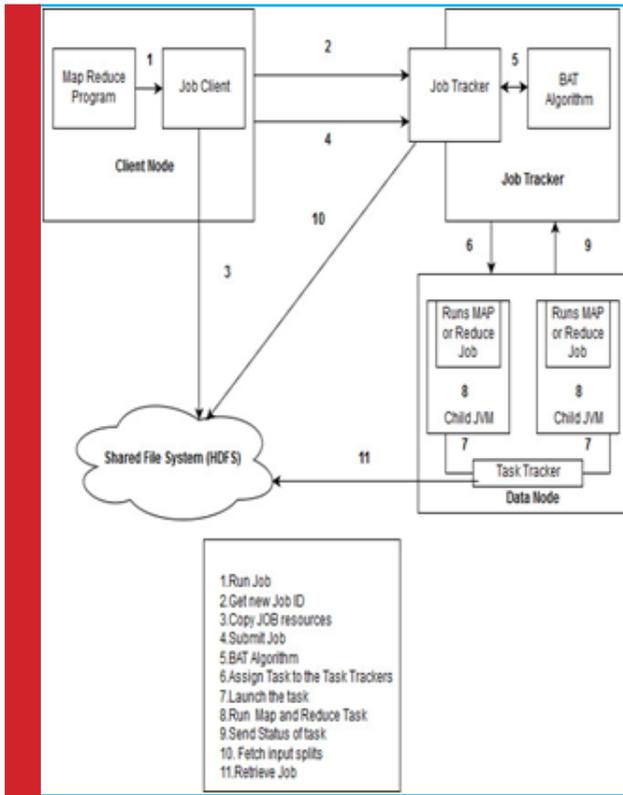


FIGURE 1. System Design BAT based Task scheduling

BAT ALGORITHM

The Bat algorithm is one among the nature inspired algorithms and it is applied to handle numerous optimization problems in multiple fields (Aboubi, et al. 2016), (Shanmugam et al., 2017). It is known as a population-based search algorithm which is completely based on echolocation behavior of bats displayed as Algorithm 1. The Bat algorithm comprises of three main steps 1) Bats can easily identifies the distance between targets of attack, food, and identifies the hurdles by using echolocation. 2) Random fly will be carried by bats at the time of finding of food with speed, with fixed sound pulse frequency; W denotes the different wavelength and loudness L_0 . 3) Loudness can be varied from high value L_0 to low constant value L_{min} . The frequency P varies in a range $[P_{min}, P_{max}]$ and Initially, the location of i th bats is denoted as P_{o_i} , pulse rate is noted as RP_i , L_i indicates loudness, with pulse frequency P_i , S_i is defined as the velocity (Topal and Altun 2016). These factors are declared randomly. The position and the velocity S_i of the bats are mentioned for each time step t is updated as in (2), (3),

$$P_i = P_{min} + (P_{max} - P_{min}) \beta \quad [0, 1] \text{ Defined as a random vector} \quad (2)$$

$$S_i^t = S_i^{t-1} + (P_{o_i}^t - P_{o_i})P_i \quad (3)$$

$$P_{o_i}^t = P_{o_i}^{t-1} + S_i^t \quad (4)$$

The local search is carried by the selected solution among the current best solutions with the help of random walk process a new solution for each bat is generated locally using random walk. $P_{o_{new}} = P_{o_{old}} + \epsilon AL^t \epsilon \in [-1, 1]$ random number $AL^t = \langle AL_i^t \rangle = \frac{1}{N} \sum_{i=1}^N AL_i^t$ is average loudness of all bats, number of bats is N $AL_i^{t+1} = \alpha AL_i^t$, $RP_i^{t+1} = r_i^0 [1 - \exp(-\gamma t)]$ where α and γ are constants (Ahilandeswari et al., 2016)

The system design in figure 1 consists of client node, master node/name node which consists of job tracker node and slave node with task tracker. The Hadoop Distributed File system (HDFS) is used for shared file system for storage system (Kaur and Chhabra 2017). The client node consists of MRP (mapreduce program) for running the job and the job client fetches the new job ID from the name node. Once it fetches the job ID it forwards the ID and copies the job resources to shared file system. Then the client submits the job to the master node. Here the role of BAT algorithm is initiated, where it finds the exact node to complete the job within short span of time (Aboubi et al., 2016). It also helps to find out the nodes which are waiting or starving for jobs. Once this process is done the name node assign task to other data nodes for processing (Arpitha and Sebastian 2017).

The job is reduced into small chunks for quick processing and it is shared to HDFS (Xu and Tang 2015). The role task tracker is to run the task with the help of data node and it also tracks the movement of the task. Once the process is completed the result will be send the status of the task to the master node and to HDFS (Rasooli and Down 2012).

The experimental setup used for processing consists of 5 nodes on local machines. One node was fixed as the JobTracker and NameNode, where the remaining nodes are known as TaskTracker and DataNodes. Each node had a 2.4 Ghz CPU, 4 GB of RAM and 250 GB hard disk, running Ubuntu 10.04 Linux, executing Hadoop 0.20.205.0. The data replication factor was 3. All nodes were on placed in same rack and connected through a 1 Gbps switch. No of iterations = 10000 with 20 runs.

Cost: The cost of the task schedule JS is denoted by C (TS) and is defined as the addition of execution cost of all the tasks of the schedule TS. If TN is the total number of tasks, cost can be calculated as:

$$C(TS) = \sum_{i=1}^{tn} C(LT) \quad (5)$$

Where $C(LT) = (\text{Instruction length of the task } i) * (\text{cost scaling factor})$ and cost scaling factor is high for of higher reliability data nodes and low for less reliable data nodes.

Execution Time: It is defined as addition of execution time for the entire task TS. If n is the total number of task in schedule.

$$ET(TS) = \sum_{i=1}^{tn} T(LT) \quad (6)$$



FIGURE 2. Execution Time comparison of schedulers

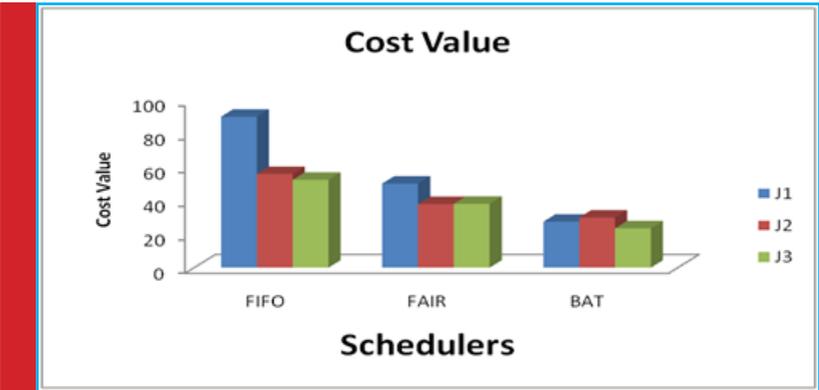


FIGURE 3. Cost Value comparison of schedulers

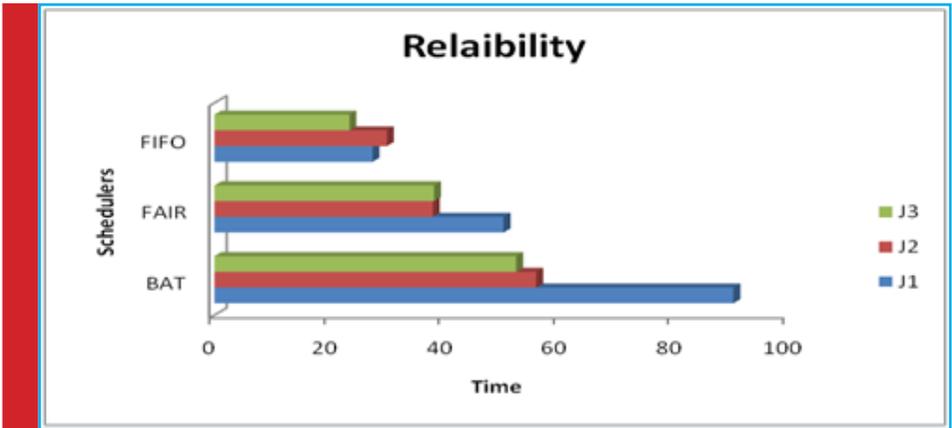


FIGURE 4. Reliability Comparison Of Schedulers

The below figure 2 describes about the execution time of three different jobs (J1, J2, J3) and the execution time is calculated by using the equation 5 (Dharshini et al., 2016). The scheduler doesn't have the ability to find out the appropriate data node processing. It takes more time for searching the correct slave node for each job. The figure2 displays about the execution time for each job and the different scheduler. According to the results the BAT outperforms remaining to scheduler by completing the task within short span of time (Govindharaj et al., 2016).

The figure 3 explains about the cost value of the job. Here we assumed the cost value from 0 as, minimum to 100 as maximum. The BAT algorithm is mainly incorporated for its accuracy, where it identifies the lave node and assigned ask with less time. The cost value will become less once the BAT is incorporated in the environment. The figure 4 illustrates about the reliability of the node. They are meant to be reliable when the number of job gets increased the reliability level also increases. The results clearly display that the BAT algorithm outperforms the FAIR and FIFO in all the scenarios.

CONCLUSION

In his paper, a recently developed meta-heuristic technique called BAT algorithm is incorporated in Hadoop environment for find the proper data node for processing the mapped task. The algorithm not only optimizes the execution time but also fortify the tasks by providing high reliability. It also helps to find the sticky slots and avoid job starvation by reducing the scheduling complexity. The final results show accurately that the BAT algorithm performs better than the other two FIFO and FAIR algorithms.

REFERENCES

- Aboubi, Y., Drias, H. and Kamel, N., 2016. BAT-CLARA: BAT-inspired algorithm for Clustering LARge Applications. *IFAC-PapersOnLine*, 49(12), pp.243-248.
- Ahilandeswari T, Nandhini S, Sivasankari P, Rajalakshmy M. (2016). Intensifying the Generic Middleware for Smart Environment. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 1-5.
- Anjos, J.C., Carrera, I., Kolberg, W., Tibola, A.L., Arantes, L.B. and Geyer, C.R., 2015. MRA++: Scheduling and data placement on MapReduce for heterogeneous environments. *Future Generation Computer Systems*, 42, pp.22-35.
- Anusha B, Noah, SivaranjaniC, PriyankaS. (2015). Predictive analysis of movie reviews using hybrid approach. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.
- Arpitha, H.V. and Sebastian, S., 2017. Comparative study of Job Schedulers in Hadoop Environment. *International Journal of Advanced Research in Computer Science*, 8(3).
- Dharshini G, Subhasri V, Sujitha G, Ganesan M. (2016). Secure Information Retrieval for Decentralised Disruption Tolerant Military Networks using CP-ABE. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 1-6.
- Gayathri R, Indumathi K, Githanjali P, Roobini V. (2016). Securing Multimedia using Data Lineage in Malicious Environment: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(2): 22-29.
- Govindharaj I, Karthiga S, Manishalakshmi R, Mary Silvia Theodore. (2016). Home Power Analyzer with Smart Power Monitoring using IoT. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 2(1): 7-13.
- Kaur, N. and Chhabra, A., 2017. Comparative Analysis of Job Scheduling Algorithms in Parallel and Distributed Computing Environments. *International Journal of Advanced Research in Computer Science*, 8(3).
- Kaur, N. and Singh, S., 2016. A Budget-constrained Time and Reliability Optimization BAT Algorithm for Scheduling Workflow Applications in Clouds. *Procedia Computer Science*, 98, pp.199-204.
- Lau, H.Y. and Qiu, X., 2014. An Artificial Immune Systems (AIS)-based Unified Framework for General Job Shop Scheduling. *IFAC Proceedings Volumes*, 47(3), pp.6186-6191.
- Mamdapure, S., Ginwala, M. and Papat, N., 2016. Task Scheduling in Hadoop. *IJIR*, 2(1).
- Nagina, D. and Dhingra, S., 2016. Scheduling Algorithms in Big Data: A Survey. *Int. J. Eng. Comput. Sci*, 5(8), pp.17737-17743.
- Rajadurai R, Amelia, Aubrey, Anusha A, Danapriya P, Geethashnee D. (2017). Efficient Data Leakage Prevention Strategy using Key Distribution. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 9-16.
- Rasooli, A. and Down, D.G., 2012, November. A hybrid scheduling approach for scalable heterogeneous Hadoop systems. In *High Performance Computing, Networking, Storage and Analysis (SCC)*, 2012 SC Companion: (pp. 1284-1291). IEEE.
- Sasidevi V, Hannah, Sathiyam D, Rajadurai R. (2017). Classification Algorithm: A Survey. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 3(1): 7-21.
- Shanmugam M, Dhivya S, Lavanya B, Keerthana V. (2017). Free Voice Calling in Wi-Fi Network using Android. *International Research Journal of Advanced Engineering Sciences and Technologies*, ISSN: 2455 - 8907, 1(1): 1-7.

tional Research Journal of Advanced Engineering Sciences and Technologies, ISSN: 2455 - 8907, 3(1): 1-8.

Somasundaram, T.S. and Govindarajan, K., 2014. CLOUDRB: A framework for scheduling and managing High-Performance Computing (HPC) applications in science cloud. Future Generation Computer Systems, 34, pp.47-65.

Topal, A.O. and Altun, O., 2016. A novel meta-heuristic algorithm: Dynamic Virtual Bats Algorithm. Information Sciences, 354, pp.222-235.

Xu, J. and Tang, Y., 2015. Improved Particle Optimization Algorithm Solving Hadoop Task Scheduling Problem [C]. In 2nd International Conference on Intelligent Computing and Cognitive Informatics (ICICCI 2015). Atlantis Press.