International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP—22-27

Security issues related to query phasing using metaheuristic algorithm

Mohanty Anita[1], Mishra Sambit Kumar[2]
1(Department of MCA, Ajay Binay Institute of Technology, Cuttack)
Email: anitamohanty56@gmail.com
2 (Department of Computer Sc.&Engg., Gandhi Institute for Education and Technology, Baniatangi)
Email: sambitmishra@gietbbsr.com

ABSTRACT *: Big Data applies to data that cannot be prepared or broken down utilizing customary procedures or devices. For decades, organizations have been settling on business choices in light of value based information put away into relational databases. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets.*
Cloud computing is the development of parallel computing, distributed computing, grid computing and virtualization technologies which define the shape of a new era. Cloud computing is an emerging model of business computing. Cloud computing is one of the emergent domains in which remote resources are used on the basis of demand, even without the physical infrastructure at the client end. In cloud computing, the actual resources are installed and deployed at remote locations. Cloud computing can expand and shrink as per the need of storage. Cloud computing is a kind of internet based computing where the users can store and process their data provided by a third party centre. The resources and data are provided to the users on demand. So it is also known as On -demand computing. Firefly Algorithm is one of the recent evolutionary computing models which is inspired by fireflies behaviour in nature. The attractiveness is proportional to brightness and they both decrease as their distance increases. While analysing and processing of Big data, it has been observed through a number of phases starting from Acquisition, Extraction, Integration , Analysis and Interpretation.In this paper, firefly optimization technique may be applied to avoid the security issues in big data.

KEYWORDS - *Big Data, Cloud Computing, Security Issues, FA algorithm*

I. INTRODUCTION

As there is rapid evolution of new technologies, devices and communication like social sites, blogs etc., the amount of data produced are growing rapidly now-a-days. There is a big difference between the data we know before and the data we are dealing now. The data is changed not only in size but also in structure too. So to deal with such kind of data new concepts take place in information technology world. Big data can store all types of data like structured, semi-structured and unstructured. Now a days the data stored in big data can be operated on clusters of computers at the same time. The user always thinks that he is operating on a single system which is actually working through a number of computers. Big data is always associated with cloud computing. Big data is implemented on cloud computing platform to do all the necessary operations on data.The cloud makes it possible for users to access information from anywhere anytime. It removes the need for users to be in the same location as the hardware that stores data. Once the internet connection is established either with wireless or broadband, user can access services of cloud computing through various hardwares. This hardware could be a desktop, laptop, tablet or phone. Cloud provides a reliable online storage space. It is the way to store your software or data in Internet (server) and you simply use this either free or sometimes paid. Quantifying the performance of scheduling and allocation policies in a real Cloud environment for different application and service models under different conditions is extremely challenging because: (i) Clouds exhibit varying demand, supply patterns, and system size; and (ii) users have heterogenous and competing quality of service requirements. Many organizations collect, store and analyze huge amounts of data. This data is commonly known as "big data" because of its volume, the velocity with which it arrives and the variety of data it stores. It can be better defined as • High volume—means amount of data • High velocity—the rate at which data created • High variety—the different types of data(both homogeneous and heterogeneous) As big data has all the above three characteristics so new technologies and techniques are required to capture, store and analyze it. Big data is

captured from many sources. The cloud makes it possible for users to access information from anywhere anytime. It removes the need for users to be in the same location as the hardware that stores data. Once the internet connection is established either with wireless or broadband, user can access services of cloud computing through various hardwares. This hardware could be a desktop, laptop, tablet or phone. Cloud provides a reliable online storage space. It is the way to store your software or data in Internet (server) and you simply use this either free or sometimes paid.

II. REVIEW OF LITERATURE

Lina Zhang et.al. [1] discussed that global optimization is challenging to solve due to its nonlinearity and multimodality. Traditional algorithms such as the gradient-based methods often struggle to deal with such problems and one of the current trends is to use metaheuristic algorithms. In this paper, a novel hybrid population-based global optimization algorithm, called hybrid firefly algorithm (HFA), is proposed by combining the advantages of both the firefly algorithm (FA) and differential evolution (DE). FA and DE are executed in parallel to promote information sharing among the population and thus enhance searching efficiency. In order to evaluate the performance and efficiency of the proposed algorithm, a diverse set of selected benchmark functions are employed and these functions fall into two groups: unimodal and multimodal. The experimental results show better performance of the proposed algorithm compared to the original version of the firefly algorithm (FA), differential evolution (DE) and particle swarm optimization (PSO) in the sense of avoiding local minima and increasing the convergence rate.

Shuhao Yu et.al. [2] mentioned that Firefly algorithm is a bio-inspired optimization algorithm which has been empirically demonstrated to perform well on many optimization problems. However, it can easily get trapped in the local optima and causes low precision. Therefore, improvement of this disadvantage is the very important. In this paper, we propose a wise strategy for step setting, which considers the information of firefly's personal and the global best positions. The results show that the modified algorithm enhances the performance of the basic firefly algorithm.

Guangchen Ruan et. al. [3] discussed that many scientific investigations require data-intensive research where big data are collected and analyzed. To get big insights from big data, we need to first develop our initial hypotheses from the data and then test and validate our hypotheses about the data. Visualization is often considered a good means to suggest hypotheses from a given dataset. Computational algorithms, coupled with scalable computing, can perform hypothesis testing with big data. Furthermore, interactive visual interfaces can allow domain experts to directly interact with data and participate in the loop to refine their research questions and redirect their research directions. In this paper we discuss a framework that integrates information visualization, scalable computing, and user interfaces to explore large-scale multi-modal data streams. Discovering new knowledge from the data requires the means to exploratively analyze datasets of this scale—allowing us to freely "wander" around the data, and make discoveries by combining bottom-up pattern discovery and top-down human knowledge to leverage the power of the human perceptual system. We start with a novel interactive temporal data mining method that allows us to discover reliable sequential patterns and precise timing information of multivariate time series. We then proceed to a parallelized solution that can fulfill the task of extracting reliable patterns from large-scale time series using iterative MapReduce tasks.

Drew Schmidt et.al. [4] presented a tutorial overview showing how one can achieve scalable performance with R. We do so by utilizing several package extensions, including those from the pbdR project. These packages consist of high performance, high-level interfaces to and extensions of MPI, PBLAS, ScaLAPACK, I/O libraries, profiling libraries, and more. While these libraries shine brightest on large distributed platforms, they also work rather well on small clusters and often, surprisingly, even on a laptop with only two cores.

Elena Geanina Ularu et. al. [5] mentioned that nowadays companies are starting to realize the importance of using more data in order to support decision for their strategies. It was said and proved through study cases that "More data usually beats better algorithms". With this statement companies started to realize that they can chose to invest more in processing larger sets of data rather than investing in expensive algorithms.

Satoshi et. al.[6] have introduced two technologies i.e. distributed database and complex event processing and work flow description in distributed environment. It is an application promoting technology for processing Big Data in Cloud environment. It is a technology to change the system configuration dynamically. The paper presented distributed data store and complex event processing which are basic technologies for big data processing in cloud environments.

Divyakant Agrawal et.al.[7] have discussed design choices made by some successful systems large scale database management systems, analyze the application demands and access patterns, and enumerate the desiderata for a cloud-bound DBMS. It provides an organized picture of the challenges faced by application developers and DBMS designers in developing and deploying internet scale applications. The background study encompasses both classes of systems: (i) for supporting update heavy applications, and (ii) for ad-hoc analytics and decision support. The focus is on providing an in-depth analysis of systems for supporting update intensive web-applications and provide a survey of the state-of-the art in this domain.

Chandramouli et.al.[8] proposed a new progressive analytics system based on a progress model called Prism that allows users to communicate progressive samples to the system, allows efficient and deterministic query processing over samples; and provides repeatable semantics and provenance to data scientists. It works with streaming engines to support progressive SQL over big data because the increasing quantity of data stored in the Cloud has become very expensive, particularly due to the pay-as-you-go Cloud computation model. It is shown that one can realize this model for a temporal relational queries using an unmodified temporal streaming engine, by re-interpreting temporal event fields to denote progress. Based on Prism, a progressive data-parallel computation framework was built for Windows Azure, where progress was understood as a first-class citizen in the framework.

Assuncao et.al.[9] discussed approaches and environments for carrying out analytics on Clouds for Big Data applications. Through a detailed survey , the possible gaps in technology are identified and provided recommendations for the research community on future directions on cloud supported Big data computing.

Ms. Rupali S.Khachane et.al.[10] had emphasised Privacy Homomorphism to resolve the security of query processing from client side,cloud with k-Nn on R-tree index query and distance re-coding algorithm. A research had been carried out by many people in cloud computing and its security to preserve query processing data, privacy of data owners and its clients through PH technique.

Khairul Munadi et.al.[11] had proposed a conceptual image trading framework that enables secure storage and retrieval over internet services. Here the target application is the online market, where publishers sell their stock images over the internet using public cloud servers.

Monika et.al.[12] through their experimental evaluation and security analyses demonstrated that robust mechanisms can be deployed with a minimal amount of computational and communicational expense. The security problems with untrusted third parties are
multifaceted in several areas such as privacy, authentication, and recovery. For privacy, the third party should not be able to know what the user's query is since the query itself describes the user's interest. For authentication, the user should be able to verify that the information from the third party is not tampered since the correctness of the query results depends upon the correctness of the information from the third party. For recovery, when the result is found to be forged by an adversary, we should be able to find the adversary and get a correct result by removing the adversary. To address these challenges, they proposed several schemes.

Reza Akbarinia et.al.[13] had proposed the best position algorithm which executes top-k queries more efficiently than Threshold algorithm with a much less execution cost. They had proposed two new algorithms because the TA algorithm incurs a lot of useless accesses to the sorted list.

Venkata Inukollu et.al.[14] discussed security issues for cloud computing, Big data, Map Reduce and Hadoop environment. The main focus was on security issues in cloud computing that were associated with big data. The main focus is on security issues in cloud computing that are associated with big data. Big data applications are a great benefit to organizations, business, companies and many large scale and small scale industries.We also discuss various possible solutions for the issues in cloud computing security and Hadoop. Cloud computing security is developing at a rapid pace which includes computer security, network security, information security, and data privacy.Cloud computing plays a very vital role in protecting data, applications and the related infrastructure with the help of policies, technologies, controls, and big data tools**.** Moreover,cloud computing, big data and its applications, advantages are likely to represent the most promising new frontiers in science.

V.P.Archana et.al.[15] introduced a Firefly algorithm (FA) based record deduplication that discovers or identifies more replica records in data warehouse than the Genetic Programming
approach. The main task of record deduplication is the task of identifying the replica in the records and to find the original data from its data repositories. In Genetic programming approach record Deduplication, works to

find the replica records only in local repository and not in all records, when compared to other optimization it becomes less efficient. This new system introduces a Firefly algorithm (FA) based record deduplication that discovers or identifies more replica records in data warehouse than the GP approach .Firefly algorithm is one of the optimization algorithms and is inspired by fireflies' behaviour in scenery.

III. SECURITY ISSUE IN CLOUD COMPUTING

To users, cloud computing is a Pay-per-Use-On-Demand mode that can conveniently access shared IT resources through the Internet. Where the IT resources include network, server, storage, application, service and so on and they can be deployed with much quick and easy manner and least management and also interactions with service providers. Cloud computing can much improve the availability of IT resources and owns many advantages over other computing techniques.

Overall data breaching was three times more likely to occur for businesses that utilize the cloud than those that don't. The simple conclusion is that the cloud comes with a unique set of characteristics that make it more vulnerable. Attackers now have the ability to use your (or your employees') login information to remotely access sensitive data stored on the cloud; additionally, attackers can falsify and manipulate information through hijacked credentials.

Employees can use their *authorized* access to an organization's cloud-based services to misuse or access information such as customer accounts, financial forms, and other sensitive information.

The expansion of cloud-based services has made it possible for both small and enterprise-level organizations to host vast amounts of data easily. However, the cloud's unprecedented storage capacity has also allowed both hackers and authorized users to easily host and spread malware, illegal software, and other digital properties.

APIs can be a threat to cloud security because of their very nature. Not only do they give companies the ability to customize features of their cloud services to fit business needs, but they also authenticate, provide access, and effect encryption.

Unlike other kind of cyber attacks, which are typically launched to establish a long-term foothold and hijack sensitive information, denial of service assaults do not attempt to breach your security perimeter. Rather, they attempt to make your website and servers unavailable to legitimate users.

Data on cloud services can be lost through a malicious attack, natural disaster, or a data wipe by the service provider. Losing vital information can be devastating to businesses that don't have a recovery plan.

Insufficient due diligence can pose a security risk when an organization migrates to the cloud quickly without properly anticipating that the services will not match customer's expectation.

IV. PROBLEM FORMULATION

Cloud environment is widely used in industry and research aspects; therefore security is an important aspect for organizations running on these cloud environments. Using proposed approaches, cloud environments can be secured for complex business operations. Big data have various challenges related to security like-computation in distributed programming, security of data storage and transaction log, input filtering from client, scalable data mining and analytics, access control and secure communication. For tackling with such security challenges different security methods are used like Type Based keyword search for security of big data, use of hybrid cloud to provide privacy in big data . With the help of query optimization technique like FA algorithm the security can be tried to improve. Most of the time as a third party is involved so security is hampered in that case. Firefly algorithm is a soft computing technique which helps to improve the security issue we find in cloud database. Firefly algorithm is used to find the replica records not only in local repository but also in all records. Each and every one firefly is associated with their light intensity(I) and their attractiveness varies with distance (d). Each firefly is changing its location iteratively. FA is efficient of all other algorithms because of its advantage that is the whole population can automatically subdivide into subgroups, and each group can swarm around each mode or local optimum. That why FA algorithm can be taken as a way to find the optimum value for more security in cloud database.

V. Algorithm

Begin

1)Objective function:f(x),x=(x1,x2....xn) ;

2) Generate an initial population of fireflies xi where i=1.....n ;.

3) Formulate light intensity I so that it is associated with f(x) (for example, for maximization problems, I=f(x) ;)

4) Define absorption coefficient γ While (t < MaxGeneration) for i = 1 : n (all n fireflies)

for j = 1 : n (n fireflies) if (I(j) > I(i)), move firefly i towards j;

Vary attractiveness with distance r via ;

Evaluate new solutions and update light intensity;

end if

end for j

end for i

Rank fireflies and find the current best;

end while Post-processing the results and visualization;

end

VI. FIREFLY ALGORITHM

The firefly algorithm (FA) is a metaheuristic algorithm, inspired by the flashing behaviour of fireflies. The primary purpose for a firefly's flash is to act as a signal system to attract other fireflies. Firefly algorithm appeared in about five years ago, its literature has expanded dramatically with diverse applications. Here three things are assumed in studying the behaviour of firefly and may be defined as follows.

* Fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex.
* The attractiveness is proportional to the brightness, and they both decrease as their distance increases. Thus for any two flashing fireflies, the less brighter one will move towards the brighter one. If there is no brighter one than a particular firefly, it will move randomly.
*The brightness of a firefly is determined by the landscape of the objective function.

VII. DISCUSSION AND FUTURE WORK

Lina Zhang et.al. [1] discussed that global optimization is challenging to solve due to its nonlinearity and multimodality. Traditional algorithms such as the gradient-based methods often struggle to deal with such problems and one of the current trends is to use metaheuristic algorithms. In this paper, a novel hybrid population-based global optimization algorithm, called hybrid firefly algorithm (HFA), is proposed by combining the advantages of both the firefly algorithm (FA) and differential evolution (DE).

Shuhao Yu et.al. [2] mentioned that Firefly algorithm is a bio-inspired optimization algorithm which has been empirically demonstrated to perform well on many optimization problems. However, it can easily get trapped in the local optima and causes low precision. Therefore, improvement of this disadvantage is the very important. In this paper, we propose a wise strategy for step setting, which considers the information of firefly's personal and the global best positions. The results show that the modified algorithm enhances the performance of the basic firefly algorithm.

Guangchen Ruan et. al. [3] discussed that many scientific investigations require data-intensive research where big data are collected and analyzed. To get big insights from big data, we need to first develop our initial hypotheses from the data and then test and validate our hypotheses about the data. Visualization is often considered a good means to suggest hypotheses from a given dataset. Computational algorithms, coupled with scalable computing, can perform hypothesis testing with big data. Furthermore, interactive visual interfaces can allow domain experts to directly interact with data and participate in the loop to refine their research questions and redirect their research directions. In this paper we discuss a framework that integrates information visualization, scalable computing, and user interfaces to explore large-scale multi-modal data streams.

Drew Schmidt et.al. [4] presented a tutorial overview showing how one can achieve scalable performance with R. We do so by utilizing several package extensions, including those from the pbdR project. These packages consist of high performance, high-level interfaces to and extensions of MPI, PBLAS, ScaLAPACK, I/O libraries, profiling libraries, and more. While these libraries shine brightest on large distributed platforms, they also work rather well on small clusters and often, surprisingly, even on a laptop with only two cores.

VIII. CONCLUSION

Today big data is in boom and handling such a large volume and variety of data is a big challenge for us. In this paper we have discussed about cloud computing ,its security issues and firefly optimization technique. Cloud is a platform which resides in remote location. Big data are implemented on this platform and uses its tools, softwares and hardwares for the manipulation of its data. There are lots of security issues found in cloud computing which are discussed above.

REFERENCES

[1] Lina Zhang , Liqiang Liu, Xin-she Yang, Yuntao Dai :" A novel Hybrid Firefly Algorithm for Global Optimization " , PLOS ,Vol.11, 2016

[2] Shuhao Yu , Shoubao Su, Quingping Lu, Li Huang :" A Novel Wise Step Strategy for Ffirefly Algorithm " ,IJCM, Vol 91, Issue 12, 2014

[3] Guangchen Ruan, Hui Zhang : "Closed-Loop Big Data Analysis with Visualization and Scalable Computing " , ELSEVIER, Vol 8, July 2017

[4] Drew Schmidt ,Wei-Chen Chen , Michael A. Matheson ,George Ostrouchov :" Programming with Big Data in r: Scaling Analytics from one to thousands of nodes" , ELSEVIER ,Vol 8 , July 2017

[5] Satoshi Tsuchiya, Yoshinori Sakamoto,Yuichi Tsuchimoto,Vivian Lee : "Big Data Processing in Cloud Environments",FUJITSU Sci.Tech.J,Vol.48,No.2(April 2012)

[6] Divyakant Agrawal ,Sudipto Das, Amr El Abbadi : " Big Data and Cloud Computing: Current State and Future Opportunities" , EDBT 2011,Sweden

[7] Badrish Chandramouli ,Jonathan Goldstein, Abdul Quamar : " Scalable Progressive Analytics on Big Data in the Cloud" , VLDB 2013, Vol.6, No.14

[8] Marcos D. Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S.Netto, Rajkumar Buyya: " Big Data computing and clouds: Trends and future directions", ELSEVIER,2014

[9] Ms. Rupali S.Khachane , Dr. Pradeep K. Deshmukh: " Attribute Based Secure Query Processing in Cloud with Privacy Homomorphism", IJCNCS,Vol.3,No.7(July 2015)

[10] Khairul Munadi, Fitri Arnia, Mohd Syaryadhi,Masaaki Fujiyoshi, Hitoshi Kiya : " A secure online image trading system for untrusted cloud environments" , SpringerPlus,2015

[11] Monika Rokade,S.A.Kahate,K.S.Kore : " Secure kNN Query Processing in Untrusted Cloud Environments:an Overview",IJIRCCE,Vol.3,Issue 6(June 2015)

[12] Reza Akbarinia, Esther Pacitti, Patrick Valduriez : " Best Position Algorithms for Top-k Queries ", VLDB,2007

[13] Venkata Narasimha Inukollu , Sailaja Arsi, Srinivasa Rao Ravuri : "Security Issues Associated With Big Data In Cloud Computing", IJNSA, Vol.6, No.3(May 2014)

[14] Kalyani Shirudkar , Dilip Motwani : " Big Data Security " , IJARCSSE, Vol.5, Issue 3 (March 2015)

[15] V.P.Archana Linnet Hailey, N. Sudha : "An Optimization Approach of Firefly Algorithm to Record Deduplication" , IJERT , Vol.2,Issue 9 (September 2013)