ISSN 2278-3091
Volume 8, No.2, March - April 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse36822019.pdf
https://doi.org/10.30534/ijatcse/2019/36822019

Security Enhancement of Genome Sequence Data in Health Care Cloud

G.Sunil[1], Srinivas Aluvala [2], Nagendhar Yamsani[3], Kanegonda Ravi Chythanya[4], Srikanth Yalabaka[5]

[1,2,3,4] Department of CSE, SR Engineering College, Warangal, Telangana State, India
[5]Department of ECE, SR Engineering College, Warangal, Telangana State, India

ABSTRACT

In the recent years, rapid enhancement of genome data sequence analysis is playing vital role in the health care field. It gives tools in support to process the genome raw data. It takes huge amount of time for processing and analysis. To minimize consumption of time, we are applying the big data techniques for genome sequence data. These techniques are more helpful in genome sequence analysis and speedup the delivery of results. The adoption rate is increased in the processing of genomic data from large volumes of data. However we achieved amazing results by applying of big data analytic technology in genome data analysis. The interpretation of personal genomic data is needed to find personalized medicine in health care. However many procedures are introduced for sharing the genomic data in cloud, but there is no technology for prediction of threat model in cloud data sharing. In this paper we introduced a novel threat detection model for prediction of threats in genomic cloud computing. It is a multitier security model, using this we achieved high throughput, low latency and also provides data safety and integrity in genomic cloud. Experimental results shows better performance when compared with existing security model in cloud computing.

Key words: prediction, Genome data, cloud

1. INTRODUCTION

The most essential information will comes out by the study of human genome for every individual. The genomic study is very useful for human individual for diagnosis of diseases and research on personalized medicine[1-3]. In modern days a huge amount of genomic data is generated. The maintenance, processing and analysis of genome data is very complicated, and risk level is high. Many processing tools and frameworks are introduced to overcome the problem in processing of genome sequence patterns. The novel genome sequencing technologies are useful to reduce the process time from days to hours. This is helpful to health care and it can lead to innovate efficient e-health applications[4]. It also support the genome sequences data interpretation model to research and find personalized medicine. The main aim of research on personalized medicine is treating patients specifically on genome data. The big data frameworks also proven the advancement in dynamic memory allocation and analyze the patient data.

Cloud computing is a phenomenon of large pool of storage servers connected as a single network. The internet based computing provided by cloud computing. Where virtually shared servers are provided storage, infrastructure and platform. Cloud Computing provides many services such as create new applications, storage and recovery of data services, blogs and website hosting and data analysis. The figure1 shows the main services of cloud. The cloud customers get the services of cloud computing in the fashion of Pay-as-You-Use model [5]. The cloud service providers classified the cloud computing services in three different ways.

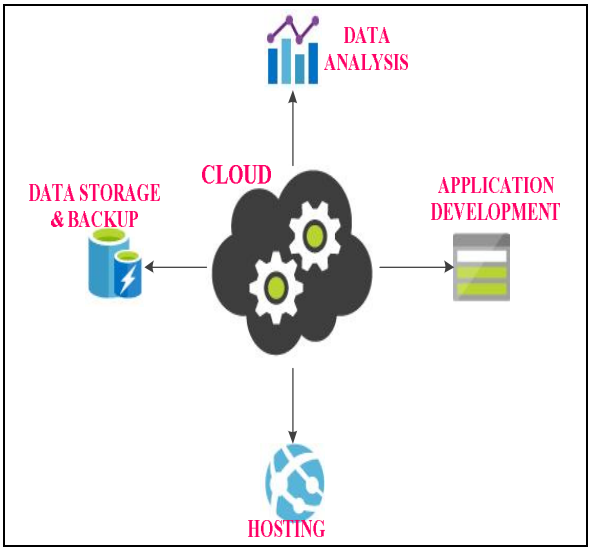

Figure 1: Shows the service of cloud in real time

Platform as a Service
Using the PAAS all the resources shared to build applications and services completely. There is no need of downloading and installation. PAAS also provides designing, developing and deploying an application. Here other premier services also provided such as database integration, security and scalability.

Software as a Service
Using the SAAS cloud, customer get facility to host an application in cloud. It also provides the services like different software's, operating system an

Infrastructure as a Service
IAAS provide the services to its users to build their own virtual infrastructure. IAAS as different type of tools for virtualization. IAAS transfer the physical resources into virtual resources; it is very helpful to give the better service by providing the resources to the customer on urgent basis.
So the cloud computing services are very useful in health care services. By predefined services of cloud like storage, availability and scalability[4]. Most of the people are interested to use cloud, and hence It is a best platform to store the genomic sequence data in cloud. The adaptation rate of cloud computing services is rapidly increased due to offered low rate infrastructure and high availability and scalability. But the open access of cloud computing services is also leads to data threat. So it is much important to safe-guard the genomic data.

2. RELATED WORK

Atallah et al., [4] has proposed a privacy-preserving protocol; it computes two sequences based on dynamic programming. The protocol takes two servers which are to process two input genome sequences in engage model. The computation results are exchanged in each iteration. However it achieves better results but the communication overhead is increased by exchanging computation results in each iteration.
Jha et al. [6] has proposed computation efficiency method. But it used same methodology like iterative protocol. So the problem communication overhead is not resolved.
Wang et al.[7] has proposed a genomic computing privacy framework, according to this partitioning the genomic data possess different sensitivity levels. Security applies for only higher level sensitivity data. However it gain better results, but it leads to leakage of low sensitivity genome data.
Troncoso-Pastoriza et al. [5] has proposed a FSM(Finite State Machine) method to calculate, edit distance with an encrypted input sequence. According to this scheme Client FSM is input of query sequences and servers have FSM for DNA sequence. In this process client and servers are participated in interactive model. In this way it is having heavy communication overhead. Huge amount of memory is required for generated sequence FSM.

3. IMPLEMENTATION

The genomic sequence data is stored in cloud due to its huge amount of data and to make it available, scalable. The security model (figure 2) is when we placed genomic data in public clouds. The patient genomic data is placed in cloud. Here mainly four trusted parties such as Genome sequence generate lab, patient, concern doctor and cloud service provider. These parties are associated with proposed security model. According this model the lab generated genome sequence for each of the patient is stored in cloud.
When storing the genome data in cloud, the system model encrypts the data and generates an alpha numeric secret key. The secret key will be used for further to view genomic data [8]. Excepted Cloud other three parties are requested to security model to view/read the genomic data[9, 10]. But the view of genomic data is available with the combination of two parties.
For each party, we check the authentication and validates the authorization. The party which have authorization can get the permission from the authorized patient[11]. At each level we authenticate the person/party based on character. We generate dynamic random alpha numeric secret key for storing and viewing the data. The secret key is unaware of each person/party. The major responsibility is taken by proposed security model.

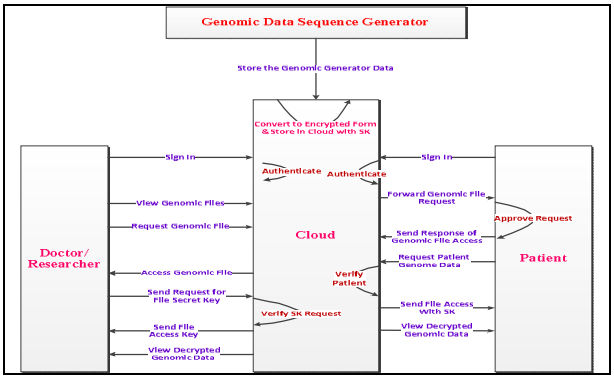

Figure 2: Working model of proposed security

The generation of secret keys and validation of keys is done by security model. Doctor gets the genomic sequence data for testing purpose. The patient genomic data is verified by researcher for diagnosis

the diseases[12]. The generated sequence pattern is very useful to find personalized medicine (figure 3).

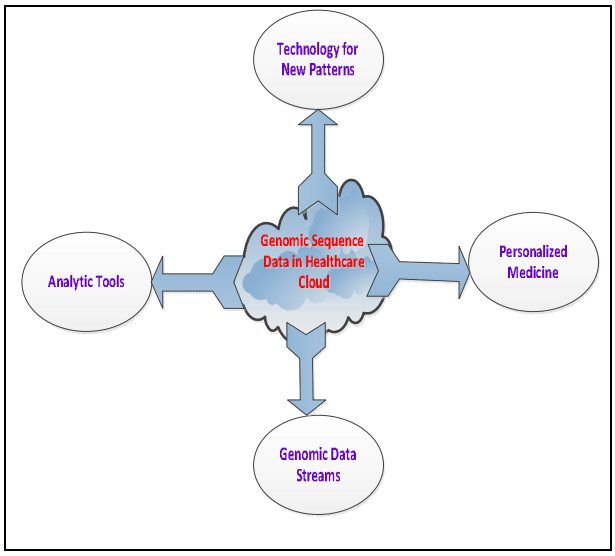

Figure 3: Healthcare cloud for personalized medicine

4. RESULTS AND DISCUSSION

A novel security model is configured in genomic cloud data for giving security and predicts the threats [13]. The security model gives authentication and authorization in every stage of accessing genome sequence data[14]. In this discussion we compare the results of proposed model with existing system model.

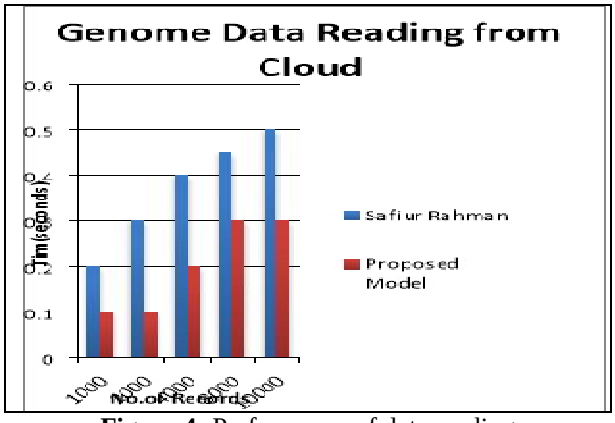

Figure 4: Performance of data reading

As shown in figure 4 the performance genome data reading from cloud of proposed model is better than compare with the existing system [15][16]. It performed better results even though no of records increased. The data reading from cloud is taken very less amount of time.

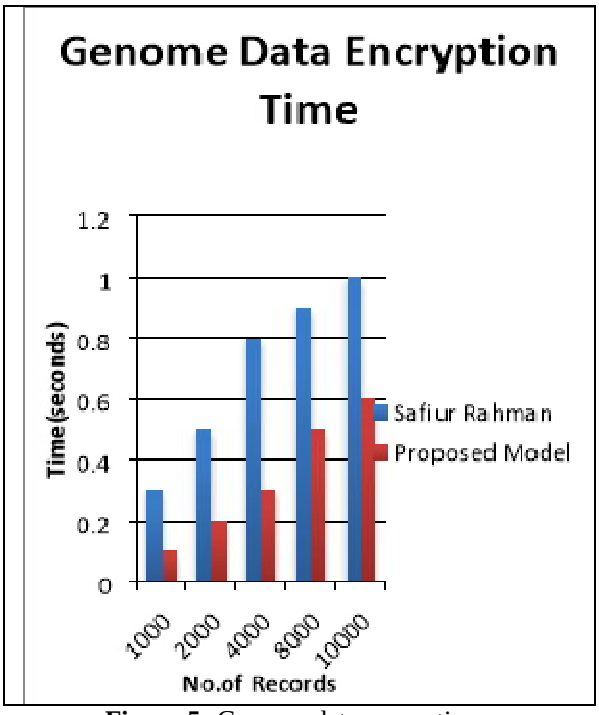

Figure 5: Genome data encryption

Figure 5 shows the genome data encryption time for proposed security model for different no of records. We can conclude that the encryption time of our proposed model is very less when compare with existing model.

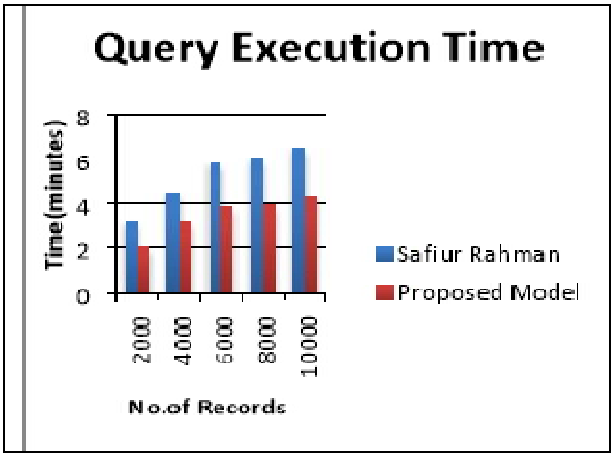

Figure 6: Query execution time

Figure 6 shows the genome query execution time from cloud proposed security model for different no of records. We can conclude that the query execution time of our proposed model is very less when compare with existing model.

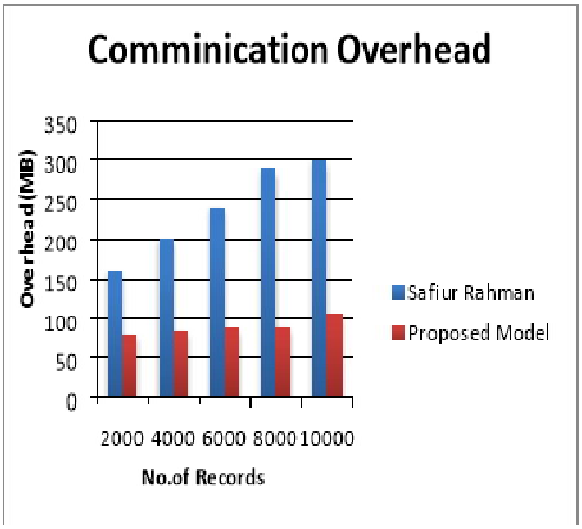

Figure 7: Communication overhead in healthcare cloud

Figure 7 shows the communication overhead healthcare cloud of the proposed model for different no of records. We can conclude that the communication overhead of our proposed model is very less(negligible) when compare with existing model.

5. CONCLUSION

In this paper we introduced a novel threat detection model for prediction of threats genomic cloud computing. It is multitier security model using this we achieved high throughput, low latency also provide data safety and integrity in genomic cloud. Experimental results shows the better performance when compared with existing security model in cloud computing.

REFERENCES

1. Mete Akgün, A Osman Bayrak, Bugra Ozer, M Şamil Sağiroğlu, "Privacy preserving processing of genomic data: A survey", *Journal of Biomedical Informatics*, vol. 56, pp. 103-111, 2015

2. Srinivas Aluvala, K. Raja Sekhar, Deepika Vodnala, “Analysis of Security Threats and Issues in MANETs” ISSN (Print): 2319-2526, Volume -4, Issue -5, 2015

3. Paul R Burton, Anna L Hansell, Isabel Fortier, Teri A Manolio, Muin J Khoury, Julian Little, Paul Elliott, "Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology", *International Journal of Epidemiology*, vol. 38, no. 1, pp. 263-273, 2009.

4. Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, Ido Guy, "Make new friends but keep the old: recommending people on social networking sites", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 201-210, 2009.

5. Melissa Gymrek, Amy L McGuire, David Golan, Eran Halperin, Yaniv Erlich, "Identifying personal genomes by surname inference", *Science*, vol. 339, no. 6117, pp. 321-324, 2013.

6. Yaniv Erlich, James B. Williams, David Glazer, Kenneth Yocum, Nita Farahany, Maynard Olson, Arvind Narayanan, Lincoln D. Stein, Jan A. Witkowski, Robert C. Kain, "Redefining Genomic Privacy: Trust and Empowerment", *PLoS Biol*, vol. 12, 11 2014.

7. Tamara Dugan, Xukai Zou, "A Survey of Secure Multiparty Computation Protocols for Privacy Preserving Genetic Tests", *Connected Health: Applications Systems and Engineering Technologies (CHASE) 2016 IEEE First International Conference on*, pp. 173-182, 2016.

8. Astrid Rheinländer, Martin Knobloch, Nicky Hochmuth, Ulf Leser, "Prefix tree indexing for similarity search and similarity joins on genomic data", *International Conference on Scientific and Statistical Database Management*, pp. 519-536, 2010.

9. Cong Wang, Kui Ren, Shucheng Yu, Karthik Mahendra Raje Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data", *INFOCOM 2012 Proceedings IEEE*, pp. 451-459, 2012.

10. Xiao Shaun Wang, Yan Huang, Yongan Zhao, Haixu Tang, XiaoFeng Wang, Diyue Bu, "Efficient genome-wide privacy-preserving similar patient query based on private edit distance", *Proceedings of the 22nd ACM SIGSAC Conference on Computerand Communications Security*, pp. 492-503, 2015.

11. Murat Kantarcioglu, Wei Jiang, Ying Liu, Bradley Malin, "A cryptographic approach to securely share and query genomic sequences", *IEEE Transactions on information technology in biomedicine*, vol. 12, no. 5, pp. 606-617, 2008.

12. Mustafa Canim, Murat Kantarcioglu, Bradley Malin, "Secure management of biomedical data with cryptographic hardware", *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 1, pp. 166-175, 2012.

13. Mohammad Zahidul Hasan, Md Safiur Rahman Mahdi, Noman Mohammed, "Secure Count Query on Encrypted Genomic Data", *Proceedings of the 3rd International Workshop on Genome Privacy and Security (GenoPri)*, 2016.

14. Dina N Paltoo, Laura Lyman Rodriguez, Michael Feolo, Elizabeth Gillanders, Erin M Ramos, Joni

Rutter et al., "Data use under the NIH GWAS Data Sharing Policy and future directions", *Nature genetics*, vol. 46, no. 9, pp. 934-938, 2014.

15. Carmit Hazay, Yehuda Lindell, Efficient secure two-party protocols: Techniques and constructions, Springer Science & Business Media, 2010.

16. Srinivas Aluvala, G.Sunil, Nagendar Yamsani, Bura Vijay kumar, "*An Empirical Study of Issues in Security and Routing of Multicast Routing Protocols in Mobile Ad Hoc Networks*", International Journal of Engineering & Technology, 7 (3.34) (2018) 1015-1018