International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Dynamic File Storage System with DNA Crypto System and Clustering Techniques

Manjunath M[1], I. Manimozhi[2]

[1]*Student, Dept. Of CSE, MVJCE, Bangalore, Karnataka, India*
[2] *Assistant professor, Dept. Of CSE, MVJCE, Bangalore, Karnataka, India*

Abstract - *Storing large user data into the cloud storage servers containing sensitive information may raise privacy issues. Clustering the user data with a guarantee of no private data misuse or the sensitive data loss forms a major security overhead. Storing the clustered data into the cloud storage helps in reduction of cost invested for system infrastructures. In this paper, we propose a dynamic file storage system with the DNA encryption for the information stored in the public cloud storage. In our implementation we provide weightage for the each keys used in the user File. Based on the weightage we cluster the user data into the different blocks and also make use of KNN algorithm in case of clustering issues. We succeed in de duplication concept by not storing the existing blocks into the cloud which in turn will increase the performance and reduce storage. DNA encryption is implemented to make sure that the user data is secured and threat free. In our design we can reduce the storage cost and increase the security with high performance.*

Key Words: **Cloud Computing, KNN Classification, Clustering, Privacy preserving, DNA Encryption.**

1. INTRODUCTION

Data mining and data analysis is the most dynamic and the ever growing with the need of the current trend and used in the most of the domains like healthcare, social network, pattern recognition .clustering plays a major role in the data mining [9] and data analysis, but the increase in the volume and variety and the velocity of the big data in today's trend makes clustering difficult. Using public cloud is the soul part in the reduction of cost for the IT infrastructure which considers with the performance and the investments.

Most of data is associated with the third party [1] storages and contains the sensitive financial and the health information, Hence the protection of their data without a leakage to any outside user is the most important security and the privacy concern. And the use the proper privacy protection mechanism is much require. The privacy-preserving K-means clustering has a problem investigated under the multi-party secure computation model, in which owners of distributed datasets interact for clustering without disclosing their own datasets to each other. In the multi-party setting, each party has a collection of data and wishes to collaborate with others in a privacy preserving [7] manner to improve clustering accuracy.

In this paper we have made use of the KNN clustering over the data sets effectively which can be resolving the problem of the clustering and can be efficiently outsourced to the public cloud servers. Our goal is to provide the privacy with the accuracy of the user data. We make use of the DNA encryption for the blocks stored in the clusters which will be in the numerical format and hence providing the security to the user information in the cloud and any data hacks can result in the zero data misuse and hence meeting the needs of the privacy of the user contents and increase the efficiency by effectively maintaining the clusters in the public cloud. When compared to the k means privacy preserving [6] with the unencrypted datasets, the encrypted datasets achieve accuracy and hence maintaining the scalability and performance. The rest of this paper is organized as follows: In Section II, we discussed about objectives of module Section III describes review of related work. Section IV describes about system model. Section V describes about detailed construction of scheme and section VI concludes the work.

2. OBJECTIVES

The objective of the system is to cluster the similar data objects into a one set and storing the data blocks in the cloud storage .we use KNN clustering over the large dataset to cluster the similar datasets are classified to the particular group to which they belong, then the trained dataset for every different cluster with medical information is loaded, the file is divided into different chunks with DNA encrypted values and stored. Deduplication is used while uploading the file which saves the storage space and reduces the conflicts

of the redundant data storage and hence increasing the performance and security of the system.

3. LITERATURE SURVEY

a) Packed Ciphertext in LWE-based Homomorphic Encryption

In this paper [2] work they are using the Peikert Vaikuntanathan waters (PVW) method for converting the contents into the ciphertext and by making use of the polynomial CRT methods can make the alternative ciphertext storage. Light weight encryption techniques using the PVW techniques along with the light weight encryption schemes leads to worse asymptotic efficiency than the earlier schemes used.

b) Privacy Preserving Back-Propagation Neural Network Learning Made Practical with Cloud Computing

In this paper [3] they have tried to improve the accuracy of the learning result, in general back propagation techniques help in the multi parties where in no user is ready to disclose their data with the other user.

c) Notes on Two Fully Homomorphic Encryption Schemes Without Bootstrapping

Two fully homomorphic encryption is used in the IACR without the bootstrapping. But here the dataset stored are really insecure in the trivial way but the insecure one.

d) Privacy of Outsourced k-Means Clustering

Privacy and the performance forms the main skeleton of this work where in the k means classification techniques have been used in the paper for the improvement of the privacy and the clustering techniques was used to overcome the large data stored into user storage.

4. METHODOLOGY

Clustering techniques plays the major role in the data classification techniques. We implement the KNN classification on the data set which has to be loaded into the cloud storage by the user with the predefined key weightages given to the keys in the data set and the clustering [10] will happen based on the key weightages. Then each of this clusters are divided into the blocks with the prefixed size into the cloud.

The blocks undergo the DNA encryption which converts the user data into the binary numbers where in the user data of the characters will be converted to ascii value of that particular character and then the compliment is applied to the binary value for the obtained output and then DNA strand is applied and the data is replaced with the string of zeros and ones later which this sequence of data is stored in the block contents hence giving the high protection to the user data. The decryption process will be carried out for the file to be downloaded where the files indexes are gathered for files and then decoded using the strand values of the DNA techniques are used to help the decoding of the blocks the group of blocks for the particular files are required to output the file in the same format what the user is uploaded, by then secret of the file where and what blocks belong to the particular file cannot be obtained. The cloud data usage can be reduced considerably in the cloud by not storing the same content twice with the help of Deduplication concept on the storage blocks.

The blocks helps in the reduction of the data storage in the cloud by helping us to implement the de duplication concept on the dataset stored. The system architecture of our system with DNA encyprtion on the clusters are demonstrated as shown in the below diagram

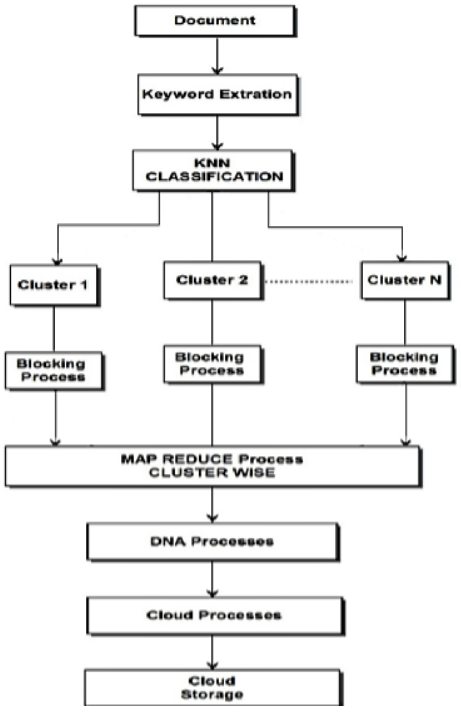

Figure 1: System Architecture

The system helps in preserving the privacy with the high security for the user data cost effectively.

5. IMPLEMENTATION

In this section we determine the implementation of dynamic file storage using clustering Techniques.

a) File classification before upload

The user selects the file which he likes to upload in to the storage and then selects the file upload option in our system design. Wherein we implement the k means clustering on the file we upload and then decide the category or the trained data to which the particular file belongs to with the help of the weightages what we have assigned to the words in the trained data set. In turn it helps in classifying at the earliest and deciding the category of file belongs.

b) Selection of the cluster.

Based on our classification techniques the files classified are divided into the cluster where the dataset of similar data types or the same data contents are stored under one naming or a particular heading. In turn the cluster helps in finding the required files at the ease when required by defining the place where it belongs instead of searching the entire storage space of user.

c) Dividing the file into blocks

By dividing the file into blocks we can able to split the contents of the user to the particular size and the wherein the user data is braked into multiple blocks and stored. By using the blocks it becomes difficult to decide the file to which the particular block belongs to and in case of data loss there is less chance of identifying the file without any detailed information.

d) De duplication

In this process we are ensuring that no two blocks has the same contents or the amount of data stored, in the sense we are avoiding the duplicates in our storage. Which reduces the storage space in the cloud and hence reduction in the cost invested and the maintenance required.

e) DNA encryption and upload

The files are undergone with the DNA encryption before they are divided and stored into the clusters and blocks. The advantage here is the consumer may not worry about their data stored over the network. By using this technique we could able to store the data in the format of binary numbers. In case of any threats the data will be safer ever before.

f) Download the decrypted file

Finally the download of the stored files will happen by grouping all the blocks which was divided by our process before and the content are gathered and then the user can download the file.

6. CONCLUSION

In this paper we proposed a Dynamic File Storage System with DNA Crypto System and Clustering Techniques in cloud computing we could achieve the high data privacy with the faster clustering with exact classification to the datasets. The KNN classifier will define the category or the cluster to which the user data belongs we implement Deduplication concept to overcome the redundant and the repetitive data storage into the cloud. The data stored by the user is made highly secured by using the DNA encryption by which the file content is stored in the form of binary numbers. Hence the high security with the lesser data storage is achieved and also helps in preserving the privacy of the user data.

7. REFERENCES

[1] European Network and Information Security Agency. Cloud computing security risk assessment.https://www.enisa.europa.eu/activities/riskmanagement/files/deliverables/cloud-computing-risk-assessment.

[2] Zvika Brakerski, Craig Gentry, and Shai Halevi. Packed cipher texts in lwe-based homomorphic encryption. In 16th International Conference on Practice and Theory in Public-Key Cryptography (PKC), pages 1–13, February 2013.

[3] Jiawei Yuan and Shucheng Yu. Privacy preserving back-propagation neural network learning made practical with cloud computing. IEEE Transactions on Parallel and Distributed Systems, 25(1):212–221, 2014.

[4] Yongge Wang. Notes on two fully homomorphic encryption schemes without bootstrapping. Cryptology ePrint Archive, Report 2015/519, 2015.

[5] Dongxi Liu, Elisa Bertino, and Xun Yi. Privacy of outsourced k-means clustering. In Proceedings of the 9th ACM Symposium on Information, Computer and

Communications Security, ASIA CCS '14, pages 123– 134, New York, NY, USA, 2014. ACM.

[6] Geetha Jagannathan and Rebecca N. Wright. Privacy-preserving distributed k-means clustering over arbitrarily partitioned data. In Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, KDD '05, pages 593–599, New York, NY, USA, 2005. ACM.

[7] Stanley R. M. Oliveira and Osmar R. Zaane. Privacy preserving clustering by data transformation.In Brazilian Symposiumon Databases, SBBD, Manaus, Amazonas, Brazil, 2003.

[8] Jiawei Yuan and Shucheng Yu. Efficient Privacy-Preserving biometric identification in cloud computing. In 2013 Proceedings IEEEINFOCOM (INFOCOM'2013), pages 2652–2660, Turin, Italy, April 2013.

[9] Jiawei Han and Micheline Kamber. Chapter 7, Data Mining: Concepts and Techniques, Second Edition. Morgan Kaufmann, 2006.

[10] Xiaowei Xu, Jochen J¨ager, and Hans-Peter Kriegel. A fast parallel clustering algorithm for large spatial databases. Data Min. Knowl. Discov. 3(3):263–290, September 1999.