Scientific Journal of Impact Factor (SJIF): 4.14
e-ISSN : 2348-4470
p-ISSN : 2348-6406

International Journal of Advance Engineering and Research Development

Volume 3,Issue 5,May -2016

Digital Signature to Secure Data in Cloud Using Homomorphic Encryption Technique

Tulsi Snehi[1]

[1]Department of Computer Engineering
Alpha College of Engineering and Technology,
Khatraj, Gujarat, India
tulsi_snehi@yahoo.com

Abstract —*Cloud Computing has been the most promising innovation in the computing world in past decade. With the advancement in technology, industry and research a large amount of complex and pervasive digital data is being generated which is increasing at an exponential rate and often termed as big data. Traditional Data Storage systems are not able to handle Big Data and also analyzing the Big Data becomes a challenge and thus it cannot be handled by traditional analytic tools. Cloud Computing can resolve the problem of handling, storage and analyzing the Big Data as it distributes the big data within the cloudlets. In spite of its numerous advantages in both technical and business aspects, cloud computing still poses new challenges particularly in security of Big Data storage. Data Privacy is one of the major issues while storing the Big Data in a Cloud environment. Data Mining based attacks, a major threat to the data, allows an adversary or an unauthorized user to infer valuable and sensitive information by analyzing the results generated from computation performed on the raw data. In this thesis digital signature and homomorphic encryption algorithm are used to protect confidentiality of data stored in cloud by using a secure k-means data mining approach assuming that the data to be distributed among different hosts maintaining the privacy of the data. The approach is able to maintain the correctness and validity of the existing k-means to generate the final results even in the distributed environment.*

Keywords*-cloud computing, data mining, homomorphic encryption, Security, Digital Signature*

I. INTRODUCTION

Cloud computing is the apt technology for the decade. Cloud computing refers to the web-based computing, providing users or devices with shared pool of resources, information or software on demand and pay per-use basis. It allows user to store large amount of data in cloud storage and use as and when required, from any part of the world, via any terminal equipment. It frees a user from the concerns about the expertise in the technological infrastructure of the service. It allows end user and small companies to make use of various computational resources like storage, software and processing capabilities provided by other companies. The cloud services can be divided into three categories: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS)[1]. Despite all the above powerful functionalities provided by the cloud computing techniques, a lot of perspective customers and users lack interest for cloud services. Since cloud computing is rest on internet, security issues like privacy, data security, confidentiality and authentication is encountered. The users on cloud who uses the service are not always the trusted person so privacy preservation of the data-owner's data is required or it may be possible that any adversary brake the security and hake the original data. Any attacker or adversary having an unauthorized access to the storage on cloud can mine the data and retrieve large amount of confidential data. So security in the cloud is nothing but hide the original data from the service-provider or the user who may be an adversary. Thus security in the cloud is current research topic and in this work research is done to provide the privacy to the data-owner's data from the any attacker or user. Various data analysis techniques or algorithm are available today which can be used successfully to mine valuable information from the large datasets by analyzing the behavioral and statistical data.

II. RELATED WORK

Preserving the privacy of the data mining algorithm has been a concern of researchers for long and a number of algorithms have been proposed for the same. Focuses on improving the security of two-party k-means while maintaining the correctness of algorithm. K-anonymity [10], noise transformation and multiplicative transformation are some PPDM (privacy preserving data mining) methods. Compared to PPDM secure cloud mining is a relatively newer field. The attacks in a Cloud Data Mining system can be listed as Denial of Service attack, Distributed Denial of Service, *Sniffing, DNS attack, Man in the Middle attack* etc. [22] gives a detailed survey on the security issues in cloud and a description of the types of attacks possible in a Cloud Data mining environments with their impact and possible solution to some of them. According to [23] data mining attacks in cloud falls in three classes: network-level, application level and virtualization level. The Network level attacks of the Cloud system and propose a solution for these type of attacks which is deployed on IBM SCE in the form of "Security-as-Service". This application prevents the high-level security attacks. Application level security is discussed in [24]. This discusses various issues regarding the deployment, moving a service on cloud in detail. It mainly focuses on building transparent cloud application using loosely coupled services. Virtualization is the key concept of cloud computing these days but it too act as a loophole in the security of the Cloud. The security of the virtual network residing in a virtual environment. They first discuss the security issues in the virtual machines and network and then propose a solution in the form of a framework to control these security issues but cryptography alone cannot prevent the attacks on the cloud mining systems and some other form of security must also be

imposed. Fragmentation technique or partitioning of the database into chunks is another method for security which suggests that keeping the data with different cloud service provider or nodes will prevent an adversary from having the access to complete data and thus will not be able to infer correct results. [25] Discusses the k-anonymity and k-anonymity noise taxonomy in a multi-cloud environment to perform frequent pattern mining. It proves that distributed data or a multi-cloud environment prevents the attacker from getting hold of the complete data thus cannot infer valuable information from the data. A one-time pass key mechanism can be used to preserve the privacy of the user as well as the service provider. This approach is based on the terminology of the authentication of both user and the provider. This paper proposes practical scheme as most of the schemes assumes the models to be semi-honest adversary model. It presents a case study of knn (k-nearest neighbor), SVM (Support Vector Machine) and k-means in the above mentioned outsourced collaborative environment. A lot of Privacy Preserving Data Mining techniques exist today. It finally concludes that most of the existing techniques are an approximation and need to be perfected further if efficiency and accuracy is required as most of the algorithms compromise one for the other and to get a balance between them more robust, dedicated and perfect PPDMs are required. Along with this PPDMs authentication between cloudlets is also required.

III. PROPOSED APPROACH

Let D= { $d_1, d_2, \ldots . d_n$ } be a multivariate database, where n is the number of attributes, which holds the user's data. The Database is horizontally partitioned and stored at two locations .i.e. Host A and Host B. Host A has {$d1_A \ldots dn_A$} and Host B has {$d1_B, \ldots, dn_B$ }. We want to perform data mining on the given data using k-means clustering approach while maintaining the privacy of the content at both the host and also preventing the intermediate values to be leaked to the adversary. It is desired that the hosts know their inputs, the final outputs and no intermediate values.

1 Encryption Formulas

To preserve the privacy of the data of each host and the intermediate results which are communicated to and fro we need an encryption system in which if any specific operation is performed on encrypted data or cipher text, the results generated matches the operation performed on plaintext when decrypted. This system of encryption is known as Homomorphic encryption system. For this purpose we use the Pallier cryptosystem which satisfies the need of the approach. We use E (a).E (b) =E (a+b) and E (a)^b=E (a*c) in this approach, where E is the required encryption scheme.

2 Assumptions

- A semi-honest model of adversary is assumed by the proposed approach in which a host can reveal other host's data, if not secured, while maintaining the privacy of its own.
- This approach assumes that the data input by client is stored as chunks at different locations instead of storing whole of the data centrally, as, the centrally stored data is more vulnerable to the attacker. Thus the client's data is stored in a decentralized manner by partitioning the database horizontally. Horizontal partitioning is referred to the partitioning scheme where each site has different records which contain same or equal set of attributes.

3 *Data Distribution*

A multivariate relational database depicted as D= {d_1, d_2..., dn} in which Host A has D_A= {$d_1^A \ldots d_1^A$} and Host B D_B= {$d_1^B \ldots d_1^B$}. As the database is multivariate, each data object is denoted by a vector set $d_i=x_{i,1},\ldots\ldots, x_{i,m}$ where m is the number of attributes. Now, let Host A have a set of private clustering centers $H1^A$, $H2^A$,……., HK^A while Host B has H_1^B, H_2^B,……., H_K^B and (C_1,C_2,……,C_k)={ $H_1^A + H_1^B$,……, $H_k^A + H_k^B$ } as the joint cluster centers[12]. Here, k is the number of clusters.

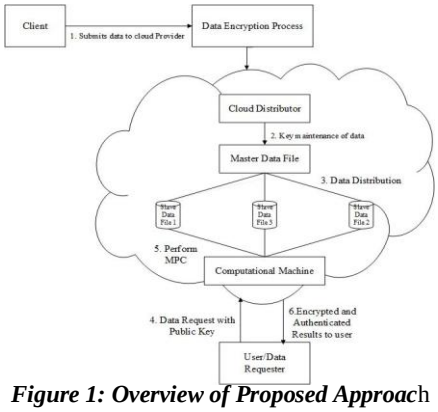

Figure 1: Overview of Proposed Approach

4 Proposed Algorithm

Notations: C_i represents the combined clustering centers which is the sum of Host A and Host B's share i.e. H^A and H^B respectively where C_i= H^A+ H^B.

Input: 1) Database D_A and D_B belonging to Host A and Host B respectively having n data objects.
2) 'k' which is the total number of clusters.

Output: The k cluster which is the combination of D_A and D_B or D.

1) Each party performs Data Normalization on local data.

2) Host A and Host B select their respective k cluster centers $H1_A, H2_{A,\ldots\ldots,} HK_A$ and $H1_B, H2_{B,\ldots\ldots,} HK_B$ randomly.

(C1, C2,…, Ck)={ $H1_A + H1_B, H2_A +, H2_{B,\ldots\ldots,} HK_A + HK_B$ } .

3) Data Private Key maintenance.
4) Job allocation is performed using MapReduce of Hadoop HDFS. (Task tracker).
5) Calculate or perform k-means for Host A and Host B according to the size of data.
6) Save the cluster centers. Hj^{Ai}, Hj^{Bi}.
7) Perform the secure cluster updating and reassign the data objects to their closest clusters locally.
8) Save Hj^{Ai+1}, Hj^{Bi+1}. If the difference between the previous cluster center and the current one is less than or equal to threshold value then stop the iteration else repeat step 7 onwards.
9) Host A and B randomly generates a pair of public/private keys (l_k, p_k).
10) Digital Signature is provided to third party for secure transmission of public key between hosts.
11) Perform Data transmission after user authentication.

IV. EXPERIMENTAL SETUP

1 Specifications:

- **Hardware Specification:**
  - Processor : Intel Core i5-3337U CPU @ 1.80GHz
  - RAM : 4.00 GB
- **Software Specification:**
  - OS: Windows 8 and Ubuntu 14.04 lts
  - System Type : 64 bit OS

2 Parameters Used:

- k –no of the clusters.
- x – No. of iterations .
- dm – distance measure.
- cd – convergence delta or threshold value which is taken as 0.5.

3 Technology used:

- Linux Ubuntu 14.04 lts, 64-bit –40GB hardisk,1.5 GB RAM
- Jdk 1.7.0
- Hadoop-2.6.4
- Mahout-0.9
- Apache Maven-3.2.1
- Eclipse Kepler

V. RESULTS AND ANALYSIS

A. Evaluation Parameters

1 Confidentiality: Confidentiality refers to the degree to which unauthorized parties are prevented from obtaining sensitive information so it is the most important security service for cloud communication.

2 Privacy: Privacy is also considered as maintaining the confidentiality throughout the network to source and destination. User privacy and access privacy are the types of the privacy to hide the identity of the person and give the particular access privilege.

3 Integrity: Integrity is the assurance of the data that either in the transmission process or stored process data is not changed or altered. It is the combination of two parameters like completeness and correctness.

4 Authenticity: Authenticity means all the users must be validated for authenticity before given any access or privilege for communication.

5 Correctness: Correctness means when client fire the query over the cloud he/she get the correct or the genuine result of that query.

6 Availability: Availability means the service of service-provider is all time available for the users in any situation.

B. RESULTS

The proposed approach performs k-means clustering on a dataset which is horizontally partitioned and stored on two different locations. The approach first run locally then performs a joint computation on encrypted intermediate results so as to obtain complete result. It was observed that running secure k-means on the partitioned data with same parameters and computation environment as the original single party k-means, produced the same end results and same inference and validating the correctness of the proposed approach.

Figure 2:Clustering result of K-means

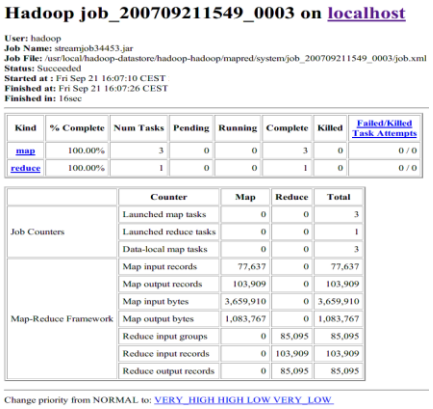

Figure-3: Hadoop Task Result

The above figures show the correctness of the proposed algorithm. It can be seen that the final cluster centers obtained by the merging of the clusters and the clustered points obtained in the final iteration of the two-party k-mean computation is similar to the cluster center obtained by the running of k-means algorithm single time.Thus, it is proved that the algorithm maintains the correctness and validity of the final result and thus can be applied to all situations where a single party k-means can be used.

Coming to the security issue we know that the model uses a partitioned approach to store the large dataset .i.e. the dataset is fragmented horizontally with a certain number of records with n attributes stored on Host A and the other set of record on Host B. Thus, fragmentation is the first step towards the security against data mining based attacks as the intruder which otherwise could, after getting an unauthorized access or entry to the data storage point, easily use the cheap and simple data mining techniques to extract valuable information from the data. But, as the data is fragmented and kept in chunks at different locations getting the correct information from the incomplete data becomes impossible thus fending off the attack by the adversary.

Secondly, the assumed model is that of a semi-honest adversary .i.e. participants try to leak the data of one another while maintaining their privacy. This approach deals with this threat as the intermediate results of both the party goes to a third party, and that too in an encrypted form, and it performs the computation on the encrypted data and returns the encrypted results to each party. Thus, each party only knows their intermediate values and the final value but not the data of the other party.

Lastly, as the data goes to the third party encrypted with a key, if an intruder is able to pick the data in the transition he/she will not be able to decipher the encrypted data to get the original values and to simulate the approach with those values. This prevents Sniffing attack on the data-in transit.

VI. CONCLUSION

Security and privacy is the major issue concerning the clients as well as the providers of cloud services as a lot of confidential and sensitive data is stored in cloud which can provide valuable information to an attacker. The Present work, "Digital Signature to Secure Data in Cloud Using Homomorphic Encryption", basically provides a system to solve the privacy issue of cloud. In this user data is distributed on two hosts and performs a combined k-means clustering using Pallier Homomorphic encryption system to prevent interpretation of intermediate attacker. Along with that digital signature is added to provide authentication to the users. The objective of this research is to implement digital signature to provide authenticity between communicating hosts. After successfully implementation of digital signature it can further be implemented without providing third party authentication (TPA) between two hosts to make it more secure. Also it can be generalized to more number of hosts as per requirement.

REFERENCES

[1] J. Carolan, S. Gaede, J. Baty, G. Brunette, A. Licht, J. Remmell, L. Tucker, and J. Weise, "Introduction to cloud computing architecture. " White Paper, 1st edn. Sun Micro Systems Inc (2009).

[2] H. Dev, T. Sen, M. Basak, and M. E. Ali, "An Approach to Protect the Privacy of Cloud Data from Data Mining Based Attacks" In High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion, pp. 1106-1115. IEEE, 2012.

[3] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes. " In Advances in cryptology-EUROCRYPT'99, pp. 223-238. Springer Berlin Heidelberg, 1999.

[4] Shobha Rajak, Ashok Verma "Secure Data Storage in the Cloud using Digital Signature Mechanism" International Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 4, June 2012.

[5] T. Sivasakthi and Dr. N Prabakaran" Applying Digital Signature with Encryption Algorithm of User Authentication for Data Security in Cloud Computing" International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 2, February 2014.

[6] Wojciech Kinastowski" Digital Signature as a Cloud-based Service"The Fourth International Conference on Cloud Computing, GRIDs, and Virtualization, CLOUD COMPUTING 2013.
[7] Deepti Mittal, Damandeep Kaur, Ashish Aggarwal" Secure Data Mining in Cloud using Homomorphic Encryption " Cloud Computing in Emerging Markets (CCEM), 2014 IEEE International Conference.
[8] ZhenQiWang; HaiLongLi"Research of Massive Web Log Data Mining Based on Cloud Computing"Computational and Information Sciences (ICCIS), 2013 Fifth International Conference.
[9] Ansari, A.S.A.; Devadkar, K.K." Secure cloud mining"Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference.
[10] Jian Li; SicongChen; DanjieSong"Security structure of cloud storage based on homomorphic encryption scheme"Cloud Computing and Intelligent Systems (CCIS), 2012 IEEE 2nd International Conference.
[11] Shashank Bajpai and Padmija Srivastava" A Fully Homomorphic Encryption Implementation on Cloud Computing" International Journal of Information & Computation Technology 2014.
[12] Jaber, A.N.; Bin Zolkipli, M.F.; Binti Abdul Majid, M.; Khan, N.U." A study in data security in cloud computing "Computer, Communications, and Control Technology (I4CT), 2014 International Conference.
[13] Kangavalli,R.; Vagdevi,S."A mixed homomorphic encryption scheme for secure data storage in cloud"Advance Computing Conference (IACC), 2015 IEEE International
[14] Rewagad,P.; Pawar,Y."Use of Digital Signature with Diffie Hellman Key Exchange and AESEncryption Algorithm to Enhance Data Security in Cloud Computing "Communication Systems and Network Technologies (CSNT), 2013 International Conference
[15] ElMakkaoui,K.; Ezzati,A.; Hssane,A.B."Challenges of using homomorphic encryption to secure cloud computing"Cloud Technologies and Applications (CloudTech), 2015 International Conference
[16] Somani,U.; Lakhani,K.; Mundra,M."Implementing digital signature with RSA encryption algorithm to enhancethe Data Security of cloud in Cloud Computing"Parallel Distributed and Grid Computing (PDGC), 2010 1st International Conference [17]Sandha; Durga,M.G."Study on data security mechanism in cloud computing"Current Trends in Engineering and Technology (ICCTET), 2014 2nd International Conference
[18]Gupta,C.P.; Sharma,I."A fully homomorphic encryption scheme with symmetric keys withapplication to private data processing in clouds"Network of the Future (NOF), 2013 Fourth International Conference
[19]Chunhua Su; Feng Bao; Jianying Zhou; Takagi, T.; Sakurai, K." Privacy-Preserving Two-Party K-Means Clustering via SecureApproximation"Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference
[20]S. Owen, A. Robin, T. Dunning, and E. Friedman. Mahout in Action. Manning Publications, 2012.
[21]C. Tai, J. Huang, and M. Chung. "Privacy Preserving Frequent Pattern Mining on Multi-cloud Environment." 2013 International Symposium on Biometrics and Security Technologies (ISBAST), IEEE, pp. 235-240, 2013
[22]R. Bhadauria, R. Borgohain, A. Biswas and S. Sanyal. "Secure Authentication of Cloud Data Mining API " arXiv preprint arXiv:1204.0764, 2012
[23]K. Beaty, A. Kundu, V. Naik, and A. Acharya. "Network-level Access Control Management for the Cloud." 2013 IEEE International Conference on Cloud Engineering (IC2E), IEEE, pp. 98-107, 2013
[24] http://storage.googleapis.com/books/ngrams/books/datasetsv2.html

