Secure and Efficient Data Mining in Cloud Using Homomorphic Encryption

[1]Hrushikesh Zanzad, [2]Prashant Shinde, [3]Swapnil Thakur, [4]Akash Mane ,[5]Prof. M. Lingayat

Department of Computer Engineering Zeal College of Engineering & Research, University of Pune.

Abstract- In the advancement in technology, industry, ecommerce analysis a pervasive digital knowledge great deal of complicated is being generated that is increasing rate usually termed as huge knowledge. ancient knowledge Storage systems isn't able to handle large knowledge additionally analyze the large knowledge becomes a challenge so it can not be handled by ancient analysing tools. Cloud Computing will resolve drawback of handling, storage & analyze the large knowledge because it distributes the large knowledge among the cloudlets. Their isn't any doubt, the Cloud Computing is that the best answer on the market to the matter of giant knowledge storage its analyses however having aforesaid that, there's forever a possible risk to the safety of giant knowledge storage in Cloud Computing, that is got to be addressed . knowledge Privacy is in an exceeding one amongst one in every of} the key problems whereas storing the large knowledge in a Cloud surroundings. data processing based mostly attacks, a significant threat to the info, it permits Associate in Nursing human or Associate in Nursing unauthorized user to infer valuable sensitive data by analyzes results generated from computation performed on the information. This proposes a secure k-means data processing approach assumptive the info to be distributed among completely different hosts protective the privacy of the info. The approach is usually able to maintain the correctness validity of the prevailing k-means to get the ultimate results even within the distributed surroundings.

KEYWORDS- Data Mining, Security, Cloud Computing, Homomorphic Encryption, Distributed Computing.

1. INTRODUCTION

Cloud computing is representing to the web-based computing, providing users or devices with shared pool of resources, information or software on demand & pay per-use basis. It frees a user from the concerns about the expertise in the technological infrastructure of the services. It allows end user and small companies to make use of various computational resources like storage, software & processing capabilities provided by other companies. The cloud services can be divided into three types:- Software as a Service (SaaS), Platform as a Service (PaaS) & Infrastructure as a Service (IaaS) [2]. Amazon, Microsoft, Google these are some of the major cloud service providers. Google App Engine (GAE) is a type of PaaS provided by Google which allows web application hosting. Windows Azure, SQL Azure is some of the services offered by Microsoft providing processing and storage capabilities for large datasets [3]. Amazon Web Services (AWS) including Simple Storage Service (S3), SQS, EC2 are cloud services provided by the Amazon [1]. Thus convenience, on demand measured access, shared easily configurable computational resources, rapid provisioning, location independence and self-service are some of the major characteristics of a cloud environment [2]. Despite all the above powerful functionalities provided by the cloud computing techniques, a lot of perspective customers and users lack interest for cloud services. Reason being the cloud issues which includes: availability or business continuity, Data confidentiality, data transfer bottlenecks, performance unpredictability, scalable storage, bugs in large distributed systems, scalability, reputation fate sharing and software licensing [3]. Of all the above issues Security comprising of data privacy issue or confidentiality of data is one of the major. As all the data resides with the cloud provider, a serious data privacy issue arises if the provider misuses the data or the information. Also any attacker or adversary having an unauthorized access to the storage on cloud can mine the data and retrieve large amount of confidential data. 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 behaviour and statistical data. Many cloud providers offer these data mining facilities to users which can be used by an adversary. Google also uses some data mining technique to predict search results by analyzing the user behaviors [4]. So, data mining can be a serious threat to the cloud security. Specially, to the organizations dealing with the financial, governmental, education or legal issues of people, leaking of which can sometime result in national catastrophes for e.g. collection of financial, health etc information by Total Information Awareness in 2002 [5] and analysis of phone records of people gathered from phone companies by NSA for identifying the possible terrorists in june 2006[5]. Also, according to a survey conducted by Rexer analytics, 7% of the data miners analyzing data using the cloud [6], due to the cheap and elastic computing powers offered by the cloud computing. So, maintenance of client privacy goes in parallel with data privacy in cloud and is a major area of concern for the cloud provider as well as cloud user. This paper presents an approach to mine the data securing using k-means algorithm from the cloud even in the presence of adversaries. This approach considered that the data is not stored in a centralized location but is distributed to different hosts. This proposed approach prevents any intermediate

data leakage in the process of computation while maintaining the correctness & validity of data mining process and the end results.

2. RELATED WORK

Preserving the privacy of the data mining algorithm has been a concern of researchers for long & a number of algorithms have been proposed for the same. [7] Focuses on improving the security of two-party k-means while maintaining the correctness of algorithm. K-anonymity [10], noise transformation & multiplicative transformation [9][17] are some PPDM(privacy preserving data mining) methods. Compared to PPDM secure cloud mining is a relatively newer field. [15] is a detailed survey of the key security challenges faced by the developers while designing the cloud application, privacy risks to the Cloud, the various Privacy Requirements & finally gives the design guidelines for developers to tackle the issue of privacy. [4] proposes an extended solution to the current techniques using trusted computing & various new, modified cryptographic techniques for privacy enhanced business intelligence. The attacks in a Cloud Data Mining system can be listed as DoS (Denial of Service) attack, DDoS (Distributed Denial of Service), Sniffing, DNS attack, Man in the Middle attack etc. [33] gives a detailed survey on the security issues in cloud & a description of the types of attacks possible in a Cloud Data mining environments with their impact & possible solution to some of them. According to [19] data mining attacks in cloud falls in three classes: network-level, application level & virtualization level. [20] discusses about the Network level attacks of the Cloud system& 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 [18]. 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. [21] discusses the security of the virtual network residing in a virtual environment. They first discuss the security issues in the virtual machines & network & then propose a solution in the form of a framework to control these security issues. [22] discusses the recent advances in the cryptographic security mechanism & try to apply those in the cloud environment. But, [11] states that cryptography alone cannot prevent the attacks on the cloud mining systems & some other form of security must also be imposed. Fragmentation technique or partitioning of the database into chunks [12] 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 & thus will not be able to infer correct results. A different approach is proposed in [13] for secure mining. They employs a privacy preserving repository which with a query plan wrapper limits the task of the data sharing & the access to the shared data with the encrypted results thus, maintaining the confidentiality as well. [23] discusses the k-anonymity & 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 onetime pass key mechanism [19] 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 & the provider. [24] proposes a SCM (Secure Cloud Mining) architecture for the generation of secure forecasting reports for an organization by identifying the interesting patterns & links between variables in a multivariate database system using image based encryption for secure forecasting. The secure collaborative outsourcing of data mining is discussed in [25]. This paper proposes practical scheme as most of the schemes assumes the models to be semihonest adversary model. It presents a case study of knn (knearest neighbor), SVM (Support Vector Machine) & kmeans in the above mentioned outsourced collaborative environment. A lot of Privacy Preserving Data Mining (PPDM) techniques exist today. [8] gives a review about all these existing techniques & analyze the representative PPDMs. It finally concludes that most of the existing techniques are an approximation & need to be perfected further if efficiency & accuracy is required as most of the algorithms compromise one for the other and to get a balance between them more robust, dedicated & perfect PPDMs are required.

3. PROPOSED APPROACH

Let D= { d_1,……..,d_a} be a multivariate database, where n is the number of attributes, which holds the user's data. The Database is horizontally partitioned & stored at two locations .i.e. Host A and Host B. Host A has D_A = {d_1^A ,…..,d_a^A} and Host B has = {d_1^B,…..d_a^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 & 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.

A. Encryption Formulas

To preserve the privacy of the data of each host & 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 [16] which satisfies the need of the approach. We use

$E(a).E(b)=E(a+b)$ & $E(a)^b=E(a*c)$ in this approach, where E is the required encryption scheme.

B. *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 [12] 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.

.

C. *Data Distribution*

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

D. *Proposed Algorithm*

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

Input: 1) Database D^A & 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 & D^B or D.

1) Each party performs Data Normalization on local data.

2) Host A & Host B select their respective k cluster centers $H_1^A, H_2^A,\ldots\ldots..., H_K^A$ & $H_1^B, H_2^B,\ldots\ldots..., H_K^B$(locally) randomly.
$(C_1,C_2,\ldots\ldots\ldots,C_k)=\{ H_1^A + H_1^B ,\ldots\ldots, H_k^A + H_k^B \}$

3) Calculate or perform local k-means for Host A & Host B.

4) Save the cluster centers $H_j^{A,i}, H_j^{B,i}$.

5) Perform the secure cluster updation & reassign the data objects to their closest clusters locally 6) Save $H_j^{A,i+1}, H_j^{B,i+1}$. if the difference between the previous cluster center & the current one is less than or equal to threshold value then stop the iteration else repeat step 4 onwards.

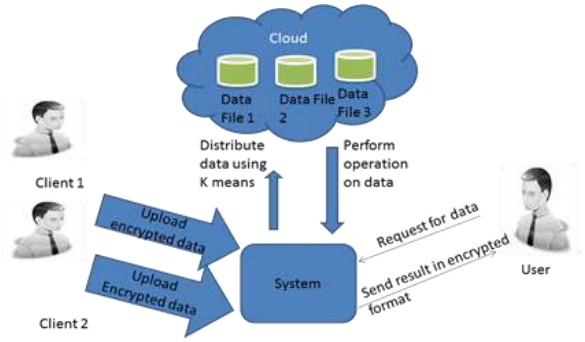

Fig 1. Overview of the Proposed Approach

4. DETAILED APPROACH

The proposed approach uses the public key cryptosystems where M is the message or the plain text which is to be encrypted. The system can be divided into 3 parts (K,E,D):

- A pair of public & private key (l_k,p_k) is generated.
- A ciphertext or encrypted message $c=E_{lk}(m,r)$ is obtained where $m \in M$ & r is a random value.
- Decryption $D_{pk}(c)=m$ is used to obtain plain text again.

A. *Private Data Normalization*

A standard Xml document is used to submit the data so that a data standard is maintained. But as we are dealing with multivariate database, i.e. a multi-attribute database, the value of variable is obtained as a sum of different attributes. Thus, the probability of some variables having large values is high, which can dominate the entire metric. Thus, a normalization method is used to standardize the multi-attribute data, using private mean computation of the data objects.

Let Host A has $d_A=\sum_{i=1}^{A} d_i^A$ with n data entries And Host B has $dB=\sum_{i=1}^{B} d_i^B$ with m data entries

Then mean

$$M=\frac{d^A+d^B}{n+m}$$

This mean is generated using Pallier Homomorphic cryptosystems so it also cannot be intercepted by the adversary. Now, the data is standardized locally using the above mean value as

$x_i = x\text{-}M$ for all data objects x_i

B. *Distance measuring & updation of clusters*

After the standardization of the data a local k-means is performed by all host on their respective datasets & initializes the cluster center for each attribute and assign data objects to the nearest cluster center using Euclidean or Manhattan distance which can be chosen according to the application or database, i.e. $h_1^A,\ldots\ldots h_k^A$, for Host A & $h_1^B,\ldots\ldots h_k^B$, for Host B. As these cluster centers are calculated locally there is no need of any security protocol but in the next step of updating the cluster centers, joint

centers are to be found which needs to be calculated privately.

Cluster Updation: for every data object's values in the j^{th} attribute in i^{th} cluster, calculate sum as $S_j^A = C_{i,j} * n_j$ where, n_j is number of data objects for j^{th} cluster $S_j^B = C_{i,j} * m_j$ where, m_j is number of data objects for j^{th} cluster Now, new i^{th} cluster center for j^{th} attribute is

$$C_{i,j}=S_j^A+ S_j^B/n_i+m_j$$

Pallier Homomorphic cryptosystem [16] is used to do the above computations privately as: Host A, B & Third Party randomly generates a pair of public/private keys(I_R+P_k). Host A & B encrypts their sum value with Third Party's public key and send it to Third Party along with their Public keys. As at the end of an iteration the local cluster centers are combined to get a global cluster center which is used by next local iteration, the correctness of algorithm stands true even in the distributed environment.

C. Iteration Stopping Criteria

As it is know that k-means is iterative in nature, so there must be a criteria which when met stops the iterations. This iteration stopping criteria is reached when output requirement are satisfied. For k-means this criteria is that the Euclidean distance between two consecutive cluster calculations is less then (threshold value).i.e. Dist (C_j,C_{j+1})=Dist$(H_J^{A,i+1}+ H_J^{B,j+1}, H_J^{A,i} + H_J^{B,i}) < \alpha$ or $(H_J^{A,i} + H_J^{B,i})-(H_J^{A,i+1}+ H_J^{B,j+1}) < \alpha$. To check this Host A computes Enc $(H_J^{A,i} - H_J^{A,i+1})$ & host B Enc $(H_J^{B,i} - H_J^{B,i+1})$ locally with third party's public key. Then third party do multiplication of intermediate encrypted values and HOST A & B decrypt with their private key as follows: T = Dec [Enc $(H_J^{A,i} - H_J^{A,i+1})$. Enc $(H_J^{B,i} - H_J^{B,i+1})$] If T<α , then the desired output is reached and the iterations can be stopped.

5. EXPERIMENTAL SETUP

1. Tools:

• **Hadoop** [27] - Hadoop is a Java framework that runs applications on large clusters of commodity hardware & comprise of features like Google File System (GFS) & the Map Reduce computing prototype. Hadoop's HDFS is distributed file system that is extremely faulttolerant &, is designed to be mounted on lowcost hardware. It is most suitable for applications with large datasets & has a high throughput access to the data being used by application.

• **Mahout** [26] - Mahout is a machine learning library provided by Mahout & is open source. Currently it primarily implements recommender System, clustering, & classification algorithms. It's also provides scalability across machines. It can be the machine learning tool for the processing of collection of large data, which may be too large for a single machine. Mahout should be run on top of Hadoop when a large amount of data is to be processed.

2. Parameters Used:

• k –no of the clusters. Default is 6 clusters.

• x – No. of iterations which is taken to be 10.

• dm – distance measure used is Cosine Distance.

• cd – convergence delta or threshold value which is taken as 0.5.

3 Testbed

Dataset used: Synthetic_control data, which is control charts exhibiting the time series comprising of 6 different classes, from UCI Machine learning repository is used [28].

• 600 records are there with 60 attributes per record.

Technology used:

• Linux 12.04, 64-bit – 40GB hardisk,1.5 GB RAM

• Jdk 1.7.0_60

• Hadoop-1.2.1

• Apache maven-3.2.1

• Mahout-0.9

4. RESULTS AND ANALYSIS

A. *Evaluation Parameters*

1. Correctness

Correctness refers to the validity of the final results obtained or the outcome of the experiments performed using the proposed approach, on the same hardware & software platform as compared to the original or base approach. The correctness is checked by comparing the deviation of the results from the anticipated results.

2. Security

This parameter evaluates the proposed algorithm in terms of security i.e. the capability of the algorithm to prevent the attackers, with malicious intent, to gain access to the confidential user data & valuable information inferred from the raw data.

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 kmeans on the partitioned data with same parameters & computation environment as the original single party k-means, produced the same end results & same inference, thus, validating the correctness of the proposed approach.

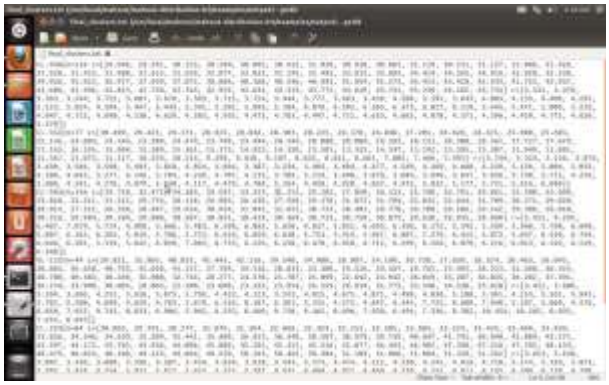

Fig 2: Final Clusters on Decentralized data

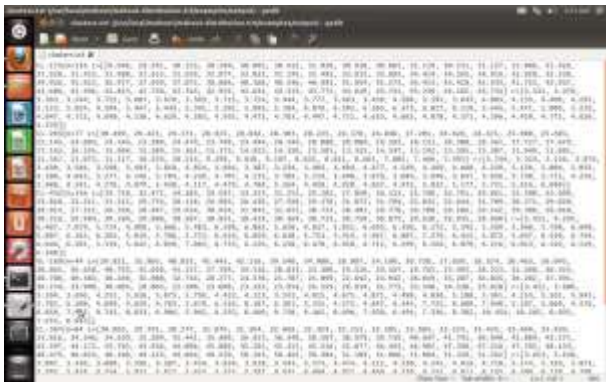

Fig 3: Final Clusters on Centralized data

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. The correctness can be further seen as both the algorithms were run on same environment & platforms with same hardware and software configuration.

Thus, it is proved that the algorithm maintains the correctness & 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 & 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 & simple data mining techniques to extract valuable information from the data. But, as the data is fragmented & 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, & that too in an encrypted form, & it performs the computation on the encrypted data & returns the encrypted results to each party. Thus, each party only knows their intermediate values & 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 & to simulate the approach with those values. This prevents Sniffing attack on the data-in transit.

6. CONCLISION

Security & privacy is the major issue concerning the clients as well as the providers of cloud services as a lot of confidential & sensitive data is stored in cloud which can provide valuable information to an attacker. This paper proposes a method to solve the privacy issues of the cloud. It assumes that the user data is distributed on two hosts & performs a combined k-means clustering using the Pallier Homomorphic encryption system for security purpose so as to prevent any interpretation of intermediate results by an attacker. The proposed approach can further be extended by adding a digital signature or hashing technique to authenticate the third party so as to prevent an adversary from posing as the third party to host's. Also it can be generalized or extended to more number of hosts if required.

REFERENSES

[1] M. Brantner, D. Florescu, D. Graf, D. Kossmann, and T. Kraska, "Building a database on S3." In Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pp. 251-264.ACM, 2008.

[2] 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).

[3] A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, "Above the clouds: A Berkeley view of cloud computing." Dept. Electrical Eng. and Comput. Sciences,University of California, Berkeley, Rep. UCB/EECS 28 (2009): 13.

[4] R. Chow, P. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Masuoka, and J.Molina, "Controlling data in the cloud: outsourcing computation without outsourcing control." In Proceedings of the 2009 ACM workshop on Cloud computing security, pp. 85-90. ACM, 2009.

[5] D. J. Solove, "I've got nothing to hide and other misunderstandings of privacy," San Diego L. Rev. 44 (2007): 745.

[6] P. K. Rexer, “Data miner survey highlights the views of 735 dataminers” 2010.

[7] C. Su, F. Bao, J. Zhou, T. Takagi, and K. Sakurai, "Privacy-preserving two-party k-means clustering via secure approximation." InAdvanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on, vol. 1, pp. 385-391.IEEE, 2007.

[8] Md. Riyazuddin , Dr.V.V.S.S.S.Balaram , Md.Afroze , Md.JaffarSadiq , M.D.Zuber. "An Empirical Study on Privacy Preserving Data Mining".International Journal of Engineering Trends and Technology (IJETT).V3(6):687-693 Nov-Dec 2012. ISSN:2231-5381

[9] K. Che, and L. Liu, "A random rotation perturbation approach to privacy preserving data classification." (2005).

[10] A. Inan, M. Kantarcioglu, and E. Bertino, "Using anonymized data for classification." In Data Engineering, 2009. ICDE'09. IEEE 25th International Conference on, pp. 429-440. IEEE, 2009.

[11] M. V. Dijk, and A. Juels, "On the Impossibility of Cryptography Alone for Privacy-Preserving Cloud Computing." IACR Cryptology ePrint Archive 2010 (2010): 305.

[12] 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.

[13] R.Mishra, S. K. Dash, D. P. Mishra, and A. Tripathy, "A privacy preserving repository for securing data across the cloud." InElectronics Computer Technology (ICECT), 2011 3rd International Conference on, vol. 5, pp. 6-10. IEEE, 2011.

[14] M. D. Singh, P. R. Krishna, and A. Saxena, "A cryptography based privacy preserving solution to mine cloud data." In Proceedings of the Third Annual ACM Bangalore Conference, pp. 14. ACM, 2010.[15] S. Pearson, "Taking account of privacy when designing cloud computing services." In Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing, pp. 44-52. IEEE Computer Society, 2009.

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

[17] K. P. Lin, and M. S. Chen, "Privacy-preserving outsourcing support vector machines with random transformation." In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 363-372. ACM, 2010.

[18] R. Bhadauria, and S. Sanyal. "Survey on security issues in cloud computing and associated mitigation techniques." arXiv preprint arXiv:1204.0764, 2012.

[19] R. Bhadauria, R. Borgohain, A. Biswas and S. Sanyal. "Secure Authentication of Cloud Data Mining API " arXiv preprint arXiv:1204.0764, 2012.

[20] 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.

[21] H. Wu, Y. Ding, C. Winer, and L. Yao. "Network security for virtual machine in cloud computing." 2010 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), IEEE, pp. 18-21, 2010.

[22] I. Agudo, D. Nuñez, G. Giammatteo, P. Rizomiliotis, and C. Lambrinoudakis. "Cryptography goes to the cloud." Data Management,and Applications in Secure and Trust Computing, Springer Berlin Heidelberg, pp. 190-197, 2011.

[23] C. Tai, J. Huang, and M. Chung. "Privacy Preserving Frequent PatternMining on Multi-cloud Environment." 2013 International Symposium on Biometrics and Security Technologies (ISBAST), IEEE, pp. 235- 240, 2013.

[24] ASA. Ansari, and KK. Devadkar. "Secure cloud mining." 2012 IEEE International Conference on Computational Intelligence & Computing Research (ICCIC),IEEE, pp. 1-4, 2012.

[25] Q. Lu, Y. Xiong, X. Gong, and W. Huang. "Secure collaborative outsourced data mining with multi-owner in cloud computing." 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom) , IEEE, pp. 100-108, 2012.

[26] S. Owen, A. Robin, T. Dunning, and E. Friedman. Mahout in Action. Manning Publications, 2012.

[28] http://archive.ics.uci.edu/ml/databases