ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

Inviolable Data mining in Cloud using AES and Paillier Cryptosystem

Nahan Rahman M.K.

Student, Dept. of Computer Science and Engg., Cochin College of Engg., and Technology, Valanchery, India

Abstract: An enormous amount of composite and permeate digital data, frequently figured as big data has been multiplied at a exponential rate with the promotion in technology. To stem the big data with the typical storage systems are not possible and canvassing the big data by using the traditional study tools has become a dispute to the analyzers The apportion of the big data to the cloud-lets or hosts is carried out by cloud computing .cloud computing dissolves the trouble of handling, storing and inspecting big data. Even though cloud computing is the logical solution for the problem of big data interpretation and storage, there shows a immense liability to the security of big data storage in cloud computing, demands more consideration. Major concern in storing big data in cloud environment is data privacy. Intrusion based on data-mining is the major obstacle which is needed to be encountered very seriously. This paper recommends a inviolable data mining way such as k means algorithm and also describes a secure and efficient encryption technique AES. Upon these encrypted data homomorphic computations are performed by following paillier cryptosystem. This looks forward for an inviolable system as there is a chance for the intruder's attempt for becoming failed.

Keywords*:* Cloud computing, K-means, AES, homomorphic encryption.

INTRODUCTION

Readily available resources are the main concern for the people nowadays. These include computing, storage and applications etc. Probable result for this kind of problem is the cloud computing environment. Services like Software as a Service, Security as a service, Platform as a Service, framework as a Service and even one can think of everything as Service (XaaS) are provided by the cloud computing..Cloud computing primarily point out to the web-oriented computing, providing devices or users with shared pool of resources, information or software on demand and pay per-use basis. Microsoft, Amazon, Google 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 storage and processing capabilities for large datasets [1]. Mutual easily constructible enumerable resources on need measured access accelerated planning, autonomous location and self-service are some of the important peculiarities of a cloud environment.

Even though these are the particularities offered by cloud computing, nowadays people are losing their passion in cloud computing. Confidentiality is the critical problem for cloud computing which we needed to be encountered, as we are outsourcing the data and application into the service providers premises, the place where the customers are losing physical control over their data and application. Reason being the cloud issues which includes: availability or business continuity, Data confidentiality, performance unpredictability, data transfer bottleneck, software licensing, bugs in large distributed systems, scalability, and reputation fate sharing and scalable storage [2]. Out Of the above security issues which are listed data privacy issue or confidentiality of data is one of the important.

The main security problem arises when the service provider misuses with the data and application. Also any intruder or adversary having an unauthorized access to the storage on cloud can get the data and retrieve large amount of data. There are many data analysis tools for getting the large datasets especially by analyzing the statistical value.

This thesis implements an effective and an inviolable method to securely mine the data preserving its confidentiality. We are mostly awaking that the process of dealing with large dataset seems to be more critical. Upon these large datasets namely the big data, we are performing the K means clustering. As the data is vulnerable to attacks, it demands a inviolable data retrieval. There we performs a secure and effective most popular AES encryption, upon these encrypted content homomorphic encryption is done, specifically the additive homomorphic technique namely paillier cryptosystem.. Homomorphic assessment of AES decryption also has captivating applications: When data is encrypted under AES and we need to compute on the data, then homomorphic AES decryption would alter this AES encrypted data into an FHE encrypted data, and then we could perform whatever computation we desired [5].

RELATED WORK

Data has endured generally stalled to the cloud for high admittance due to the enhanced cloud platform. But we generally avoid the importance of security of the data in the cloud and entrusted on cloud service provider. [8] Describes the confidentiality problems on social and business applications and the methodologies to make the accessible data inviolable. Also, critically studying homomorphic encryption and Crypt DB schemes which are admissible to protect data from awful third party service environments (cloud computing) and also from a bettor for

DOI 10.17148/IJARCCE

these applications. Evaluating ongoing encryption and key authority proficiency in [6] gives a better idea of the system described. And also describing the effective and popular methods for increasing the performance of the system. Calculations of arbitrary functions [8] by fully homomorphic encryption are described. It also gives the definitions and properties of homomorphic encryption. Not confront the gentility and observation of the blue-print, schemes established along these lines influenced from a number of deficiencies.

A functioning implementation of the (leveled) HE scheme of BGV in [4], use this to examine the AES circuit homomorphically by discharging a standard against which other implementations can be allotted. Homomorphic encryption can be applied in any environment on the light of various public key algorithms [6] for preserving confidentiality. [7] Discusses the security problems in the virtual machines and network and then propose a remedy in the form of an infrastructure to control these security problems.

As the popular clustering algorithm is K-means sometimes it takes time for the clustering processes. [10] Gives the better way for finding the initial cluster center and it needs less time complexity but still needs to compute the k value.

PROPOSED APPROACH

Consider a multivariate database, which holds the user's data. Upon these data k-means clustering is performed .In order to protect the data from the antagonist we need to encrypt it by using the most secure encryption technique AES. We want to perform homomorphic encryption on these encrypted data to preserve it from attacks.

A) k-means

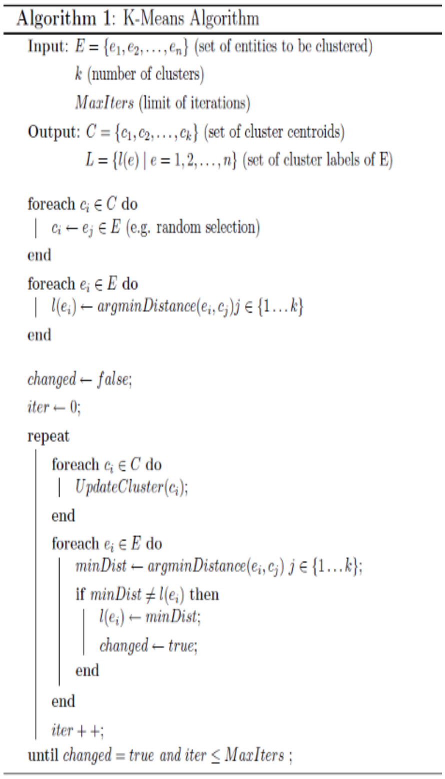
The main intention of the k-means clustering algorithm is to aggregate the objects into distinct group according to the attributes or features of the objects.

The grouping is completed or done by simply calculating the sum of squares of the distance of the objects to their agnate cluster centroid.

Thus, the motive of K-mean clustering is to classify the data. Figure 1 shows the clustering algorithm.

Here's how the algorithm works:

1. The algorithm approximately chooses k points as the initial cluster centers ("means").
2. Each point in the dataset is accredited to the closed cluster, based upon the Euclidean distance between each point and each cluster center.
3. Each cluster center is recomputed as the average of the points in that cluster.
4. Steps 2 and 3 repeat until the clusters converge. Convergence may be defined



```
Algorithm 1: K-Means Algorithm
Input: $E = \{e_1, e_2, \ldots, e_n\}$ (set of entities to be clustered)
       $k$ (number of clusters)
       $MaxIters$ (limit of iterations)
Output: $C = \{c_1, c_2, \ldots, c_k\}$ (set of cluster centroids)
        $L = \{l(e) \mid e = 1, 2, \ldots, n\}$ (set of cluster labels of E)

foreach $c_i \in C$ do
    $c_i \leftarrow e_j \in E$ (e.g. random selection)
end
foreach $e_i \in E$ do
    $l(e_i) \leftarrow argminDistance(e_i, c_j) j \in \{1 \ldots k\}$
end

$changed \leftarrow false$;
$iter \leftarrow 0$;
repeat
    foreach $c_i \in C$ do
        $UpdateCluster(c_i)$;
    end
    foreach $e_i \in E$ do
        $minDist \leftarrow argminDistance(e_i, c_j)\ j \in \{1 \ldots k\}$;
        if $minDist \neq l(e_i)$ then
            $l(e_i) \leftarrow minDist$;
            $changed \leftarrow true$;
        end
    end
    $iter{+}{+}$;
until $changed = true$ and $iter \leq MaxIters$ ;
```

Figure 1: K-means clustering algorithm

B) Advanced Encryption Standard

After the clustering is accomplished we need to perform the AES cryptographic technique. As we are very familiar with AES, it is need to reveal that AES has security but it is not obvious. It is just a connection between time and cost. Clearly saying how long a attacker take to get the key and how much it costs too. As we all know the various key lengths is the striking feature of the AES encryption technique. That key length is decided only by knowing how long the security is needed to be required and about the cost. There are many controversies regarding the weakness of AES encryption technique. Till now we can't prove .i.e., simply we can say there is no other way for the intruder to get the key except exhaustive searching. Even the AES -128 provides large number of possible keys; it is impractical for the intruder as it takes centuries for searching the key [5]. Key lengths for AES encryption can be 128,192,256 and it can be referred as accordingly. Large key size gives for security. However large key computation requires more computation resources and it affects the performance adversely. It works on 128 bit block size in 10, 12 and 14 rounds according to the key size. Encryption consists of following operations:

1) Initial Round

Add Round Key: Round key is associated with each byte state and that is the initial round.

2) Rounds

DOI 10.17148/IJARCCE

Sub Bytes: This is the first transformation as we are substituting with the byte .we illustrate it with hexadecimal digits.

Shift Rows: This is the transforming step in which the first row is kept unchanged .Last three rows is shifted cyclically in a convinced number of steps.

MixColumns: In this step mixing the columns is performed by merging four bytes of every column.

Add Round Key: Final step in the rounds is the adding of the round key

3) Final Round

Sub Bytes.
Shift Rows.
AddRoundKey.

Computations are performed on the encrypted data which are encrypted by using AES, after this step homomorphic decryption of the encrypted data is carried out and we will be able to perform the whatever computations we need to.

C) Homomorphic Encryption

To the encrypted data homomorphic encryption is carried out. This is the technique that can be performed directly on the encrypted data. Applying public key algorithms homomorphic encryption can be adapted to any system. There are many encryption algorithms to secure the operations and the storage of the data, when data is brimmed into the accessible area. To perpetuate confidentiality and to rule data discovered on remote server, homomorphic encryption is advantageous that concedes the operations on the encrypted text, which can give the same outcomes after computations when working directly on the plain text. This technology has the abeyant to change the way that we defend arbitrary computation on the cloud platform [4]. The aim of homomorphic encryption experimentation is to chart an encryption scheme that is purposefully amenable in a distinct way to make computation on encrypted data. For this principle we follow the Paillier cryptosystem which delights the demands of this approach. We practice E (a).E (b) =E (a+b) and E (a) b = E (a*c) in this path, where E is the desired encryption pattern.

The mathematical significance of homomorphic encryption is shown below: Where N is message to be encrypted, x= (N), Encryption of message N with key k, N= (x), decryption of cipher text x into plain message N with key k. Now, let us consider the homomorphic property: X1= (N), X2= (N1). Now, adequate to bring out approximate calculations on X1 and X2, if X1+X2 then it is known as additive homomorphic encryption, i.e. X3= X1+X2. If the computation is like X1*X2 then we call it as multiplicative homomorphic encryption, i.e. X4=X1*X2. The decryption result of X3, X4 is the sum of N1, N2 and product of N1, N2 respectively. As we all know the main idea of the fully homomorphic encryption (FHE) is the computations or arbitrary operations upon these data.

D) Paillier Cryptosystem

Additive Homomorphic Encryption: A Homomorphic encryption is additive, if Ek (X1⊕X2) = Ek (X1) ⊕ Ek (X2) As the encryption function is additively homomorphic, the following circumstances can be explicated. The product of two cipher texts will decrypt to the sum of their analogous plaintexts, D (E (m1, r1) · E (m2, r2) mod n2) = m1 + m2 mod n.

The product of a cipher text with a plaintext construct g will decrypt to the sum of the corresponding plaintexts, D (E (m1, r1) · gm2 mod n2) = m1 + m2 mod n.

For key generation consider two prime numbers p and q, such that gcd (pq, (p-1)(q-1))=1,this property is needed to be satisfied.

1. **Key generation**:
Step 1: n = pq, the RSA modulus
Step 2: λ = lcm (p − 1, q − 1)
Step 3: g є Z /n2 Z s.t. n|or dn2(g)
Step 4: Public-key: (n, g), secret key: λ, μ

2. Encryption of m :
Step 1: m є {0, 1... n − 1}, a message
Step 2: h єR Z/n Z
Step 3: c = g^m h^n mod n^2, a cipher text

3. Decryption of c :
m = L (c^λ mod n^2) *μ mod n

Suppose there are two ciphers,
CT1 and CT2 such that:
CT1 = gm1x1n mod n2
CT2 = gm2x2n mod n2
CT1·CT2 = gm1x1n·gm2x2n mod n2
Additive Property is: gm1+m2(x1x2) n mod n2

When the user gives the homo query ,that is encrypted query to the cloud will perform the computations asked by the user and returns the homo reply to the user. As the query is encrypted by using AES technique also performed homomorphic encryption will gives more inviolable data retrieval .It gives the intruder a most difficult task to get the data.

PERFORMANCE EVALUATION

In terms of execution time K-means algorithm is the best clustering algorithm in the scenario of easy implementation and popular partitioning algorithm .In the comparison scenario to representative object based clustering (FCM) k-means algorithm shows more performance in terms of execution time.it takes only less execution time.as FCM produces close results to k-means, it is quite time consuming so, k-means is superior. Taking SOM and k-means which are the two classical methods for clusterin, SOM is very complex than k-means. k-means is simple and easy to implement. query redirection method that improved K-means clustering algorithm in terms of performance and accuracy in distributed environment compared to heirarchichal clustering.K-Means produce tighter clusters than hierarchical clustering, especially if the clusters are globular.

AES encryption technique can use key lengths of 128, 192, or 256 bits, whereas DES uses key length of only 56 bits. Block size for DES encryption technique is 64 bits but for AES it is 128 bits. Based on the text files used and the experimental result it was concluded that AES algorithm consumes least encryption and RSA consume longest encryption time. Decryption of AES algorithm is better than other algorithms. Homomorphic encryption supports arbitrary operations on cipher texts and it enables the construction of programs for any desirable functionality, which can be run on encrypted inputs to produce an encryption of the result. Since such a program need never decrypts its inputs, it can be run by an entrusted party without revealing its inputs and internal state.

CONCLUSION

Cloud Computing is a flexible, cost-effective, and proven delivery platform for providing business or consumer IT services over the Internet. However, cloud Computing presents an added level of risk because essential services are often outsourced to a third party, which makes it harder to maintain data security and privacy, support data and service availability, and demonstrate compliance. This thesis is looking forward to eliminate the security issues and ensuring the confidentiality of the data in the cloud. This paper assumes that upon the large datasets, performs an efficient k-means for clustering the large datasets. As AES is more efficient and secure methodology we are implementing an AES encryption technique to encrypt the data. To this encrypted data additive homomorphic property such as paillier cryptosystem is carried out. Homomorphic encryption system for security purpose so as to prevent any interpretation of intermediate results by an attacker.

We are highly conscious to the demand to protect the individual's personal data and curator and records managers act under a professional agreement to so. Cloud data security is the major issues we are facing today. So this system will provide a better solution for this.

REFERENCES

1. R.Mishra, S. K. Dash, D. P. Mishra, and A. Tripathy, "A privacy preserving repository for securing data across the cloud". In Electronics Computer Technology (ICECT), 2011 3rd International Conference on, vol. 5, pp. 6-10. IEEE, 2011.
2. ASA. Ansari and KK. Devadkar," Secure cloud mining." 2012 IEEE International Conference on Computational Intelligence & Computing Research (ICCIC), IEEE, pp. 1-4, 2012.
3. Mayank Varia1, Sophia Yakoubov1, and Yang Yang2, " HEtest: A Homomorphic Encryption Testing Framework", Work performed while at MIT Lincoln Laboratory.
4. Craig GentrY, Shai Halevi , Shai Halevi ,Nigel P. Smart." Homomorphic Evaluation of the AES Circuit"
5. Chungsik Song, Younghee Park*, Jerry Gao, Sri Kinnera Nanduri, William Zegers." Favored Encryption Techniques for Cloud Storage"
6. Payal V. Parmar, Shafika N. Patel, Niyatee I. Bhatt ," Survey of Various Homomorphic Encryption algorithms and Schemes"
7. 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.
8. Urvi Narang, Dr. D. C. Jinwala ,Sankita Patel," Privacy Preserving Clustering using Fully Homomorphic Encryption"
9. https://sites.google.com/site/journalofcomputing
10. Madhu Yedla, Srinivasa Rao Pathakota, T M Srinivasa," Enhancing K-means Clustering Algorithm with Improved Initial Center" / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 1 (2) , 2010, 121-125
11. Manpreet kaur, Usvir Kaur," Comparison Between K-Mean and Hierarchical Algorithm Using Query Redirection"on International Journal of Advanced Research in Computer Science and Software Engineering 2013.

DOI 10.17148/IJARCCE