International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Enhancing Security of Database Stored in Cloud Environment

Rajeshkannan .R[1], T. Saranya[2]

[1]Assistant Professor (Senior) School of Computing Science and Engineering, VIT University

[2] Student School of Computing Science and Engineering, VIT University

Abstract: ***Cloud computing, which refers to an emerging computing model can be defined as set of resources and services provided through the internet with more advantage features such as virtualization, data storage, high expansibility, large amount computation and low cost service. The rapid development in cloud computing has brought data mining in the eyes of researchers as a promising service. The data is outsourced to the third party server by the data owner because of inadequate expertise and computational resources. But the data as well as the association rules of the outsourced database becomes most important property of the data owner. In this paper, we implemented Jenkins hash function encryption algorithm in order to achieve encryption. To provide security, the data owner performs Jenkins hash function encryption scheme and sends data to the server. To recover true pattern of the encrypted data, the data owner will send data mining queries to the server and it will return required pattern. The efficiency of the proposed algorithm is analyzed and compared with the existing algorithm.***

Keywords: Privacy preserving data mining, Jenkins hash function encryption algorithm, Cloud computing

1. Introduction

To ensure corporate privacy, the data owner will perform Jenkins encryption algorithm and sends it to the server. Data owner will send the mining queries to the server and retrieves original patterns from the extracted patterns obtained from the server. Jenkins encryption scheme guarantees that every transformed item is identical with respect to the attacker's knowledge. At the point when a query is submitted to the server, the related data are grouped together by association rule. The correlation and categorical analysis is performed on each and every set to get the irrelevant patterns with regard to eliminate from the main pattern. Security and privacy, especially keeping up confidentiality of data, have turned into a challenging issue because of the rapid growth in information and communication technology.

Data mining is defined as the process of extracting knowledge and patterns from large amounts of data which is stored in database or data warehouses. The extracted pattern and useful information can be used to decision making, information management, process control and query processing. Data mining has become one of the most important data analysing tool in the information industry. It brings conflict between data mining and privacy. Privacy preserving has started as an essential thing because of the success of data mining.

Privacy preserving data mining ensuring the security of individual data or precise knowledge without relinquishing the utility of the data. Individuals have ended up very much aware of the privacy interruptions on their own data and are extremely hesitate to import their sensitive data. This may prompt the unintentional results of data mining. Within the requirements of security, a few methods have been proposed yet at the same time this branch of examination is in its outset.

Privacy preserving data mining techniques can be classified into five categories depends on data distribution. The techniques are

- Anonymization based PPDM
- Condensation approach based PPDM
- Randomized response based PPDM
- Perturbation based PPDM
- Cryptography based PPDM

Anonymization is a technique which is used to hide the individual's sensitive data from attackers by using generalization and suppression method. Here k anonymity is used for generalization and suppression for data hiding. Generalization of data refers to replacing the value of data with specific value and suppression refers to blocking the values. The main disadvantages of this method are the attackers can easily get the data information if they come to know about background knowledge or any attributes of the data.

Condensation approach was introduced by aggarwal which creates clusters from the given data set and then creates pseudo data. The fundamental idea of the method is to consolidate the data into different groups of predefined size and for every group certain measurements are maintained. This approach is utilized as a part of dynamic data upgrade, for example, stream

Randomized response based technique is first check whether the data obtained from client is true or false and after that all the data got mixed up together if the quantity of data is huge. Then the data provider transfers the randomized data to the data receiver then the data receiver will retrieve the original data by using reconstruction algorithm. The real weakness over this technique is it is not required for different databases and it leads to large amount of data loss.

Data perturbation is a method which is used to adjust data using random process. This method obviously misshapes

sensitive data values by transforming them by adding, subtracting or whatever other mathematical operations. This method can deal with different kind of data types such as integer type, character type, Boolean type and classification type. The data set is collected from various sources then pre-processing of data is required to proceed further.

Data perturbation is called as data distortion or data noise. Data perturbation plays a vital role in preserving the sensitive data because it is very crucial to the data owner. Data distortion is performed by applying distinctive methods, for example, adding noise, Data transpose matrix and so forth. The main drawback is the dispersion of every data is restored individually. This implies that any distributed based data mining algorithm lives up to expectations under an understood suspicion to treat each dimension individually.

Cryptography is a method which is used to encrypt and decrypt the sensitive data to provide security. It is a very simple and effective method to preserve sensitive data. There are lot of different cryptography algorithms available which is used to prevent privacy leakage of computation of data but it fails to prevent the output of computation.

2. System Analysis

a. Existing System

The service provider who can be an attacker might generate an attack model based on some background details about data which is stored in server even though he may not aware of encryption techniques used by data owner. Rob Frugal algorithm serves to give security to the database on server, with the help of association rule mining and original support of mined patterns can be regained.

Rob Frugal Algorithm:

Rob Frugal Algorithm is used to transform a plain Transaction database into encrypted database with the help of three steps.
Step 1: Uses 1-1 substitution encryption method.
Step 2: Uses K-grouping methods.
Step 3: Add fake transaction to achieve privacy.

Disadvantage of existing system:

- Rob Frugal algorithm focuses only cipher text attacks.
- The encryption algorithm can easily hacked by attacker, if they come to know about any of background details about data.
- To get plain text of encrypted database, the server can send database to the third party server which leads to leakage of information.

b. Proposed System

Jenkins Encryption Algorithm:

- Take any one of the following data owner information like Username, password mobile number, uploaded document name, document size.
- Convert bit to byte conversion.
- Byte conversion answer placed in 64*64 cells.
- Insertion order is not unique.
- Re orders the element and placed in outside the cell.
- Finally add the mobile number or document size to reorder element.

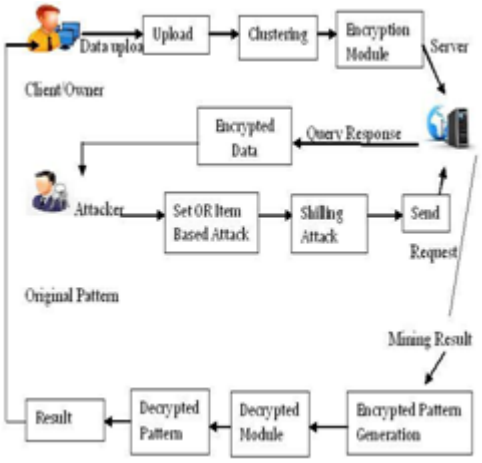


Figure 1: Architecture of proposed system

3. Result & Discussion

The proposed system consists of five modules namely admin, data owner, service provider, Encryption and clustering module.
Admin:

This module focuses on both data owner and service provider. Each member in a data owner and service provider is given a user name and password which is used to identify him uniquely.

Data Owner:

Data owner has to choose which file should be send to server. After that he should encrypt the file using Jenkins algorithm and send to the server.

Service Provider:

The main purpose of service provider is to maintain the server with proper manner. He can see who all are upload their database to the server and how many files got downloaded from server.

Encryption:

Encryption is done with Jenkins encryption algorithm which provides efficient security to the outsourced data. After encryption the encrypted key and file id is sent to the data owner registered mail id.Using that key the data owner can download their data.

Figure 2: Admin Module

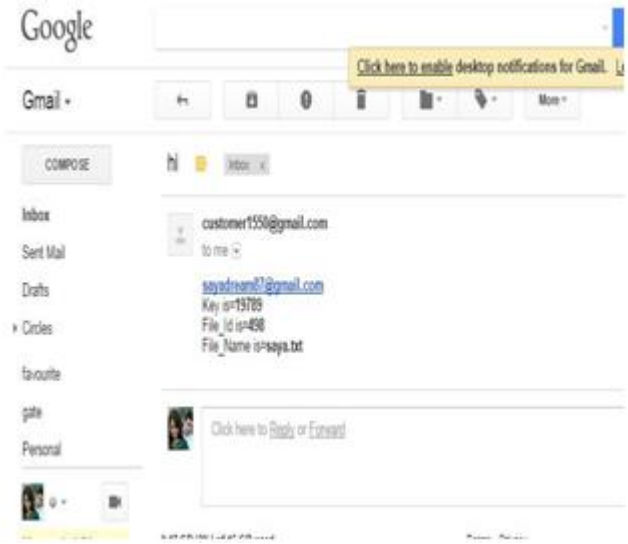
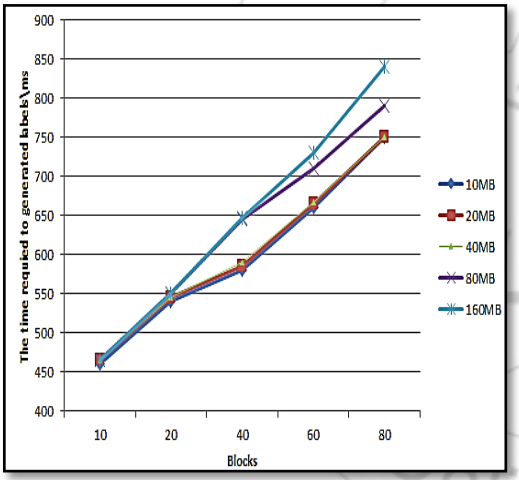
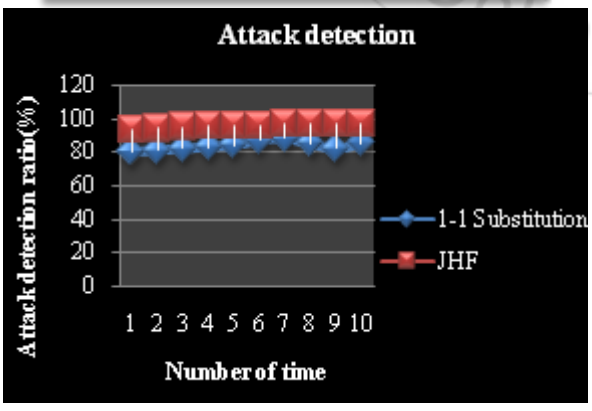


Figure 3: Encryption Module

Performance Graph

The following graph shows that the required time to generate the blocks according to data owner file size. It clearly says that the time will increase if the file size increase.





4. Conclusion

Now days, data mining as a service become very popular because of the rapid growth of cloud computing model. Here privacy preserving data mining techniques are discussed and compared with other. From our analysis and observation it was concluded that the performance of Jenkins encryption algorithm is better than the other algorithms. There are large numbers of cryptography algorithm present and many other security issues also present while outsourcing database to the server. So the future work will be based on other techniques that can be used to encrypt the data in efficient manner such that the best encryption technique can be identified to provide security to the database.

References

[1] Fosca giannotti, laks v.s. lakshmanan, anna monreale, dino pedreschi, and hui (wendy) wang," Privacy-preserving mining of association rules from outsourced transaction databases", *IEEE transactions on knowledge and data engineering vol:7 no:3 year 2013.*

[2] Ambika vishal pawar, ajay dani, "Enhancing privacy-preserving cloud database querying by preventing brute force attacks", *International Journal of Computer, Information, Systems and Control Engineering Vol:8 No:1, 2014.*

[3] Dwipen laskar ,geetachri lachit, "A review on privacy preservation data mining (ppdm)",*International Journal of Computer Applications Technology and ResearchVolume 3-Issue.[4]Murat kantarcioglu and chris clifton," Privacy-preserving distributed mining of association rules on horizontally partitioned data", IEEE transactions on knowledge and data engineering, vol. 16, no. 9, september 2004.*

[4] Vineet richhariya & prateek chourey, "A robust technique for privacy preservation of outsourced-transaction database", *International Journal of Research in Engineering & Technology IMPACT: IJRET) ISSN(E): 2321-8843; ISSN(P): 2347-4599 Vol. 2, Issue 6, Jun 2014, 51*

[5] Neha Jain and Gurpreet Kaur "Implementing DES Algorithm in Cloud for Data Security" *VSRD International Journal of CS & IT Vol.2 Issue 4, 2012, pp. 316-321.*

[6] Simarjeet Kaur "Cryptography and Encryption In Cloud Computing", *VSRD International Journal of CS & IT Vol. 2 Issue 3, 2012,pp. 242-249.*

[7] D.H. Patil "Data Security over Cloud" *International Journal of Computer Applications 2012.*

[8] G. Jai Arul Jose "Implementation of Data Security in cloud Computing" *International Journal of P2P Network Trends and Technology – vol 1, Issue1-2011.*

[9] Sriram Ramanujam "Data Security in Cloud Computing" *J.Comp. & Math. Sci –vol 2(1),2011.*