Secure Mining of the Outsourced Transaction Databases

Ms. Trupti A. Kumbhare [#1], Prof. Santosh V. Chobe [*2]

[#]*Research Scholar, DYPIET, Pimpri, Pune, India*
[1]trupti.kumbhare@gmail.com

[*]*Associate Professor, DYPIET, Pimpri, Pune, India*
[2]sanchobe@yahoo.com

Abstract— **Data mining is an analytical tool which allows users to analyze data, categories it and summaries the relationships among the data. It discovers the useful information from large amount of relational databases. Security is an important issue while performing mining task at the end of third party service provider. When data owner outsources its mining needs to third party service provider, there should be security. Third party service provider has access to the data and can learn sensitive information from it. The result is security may break. As data is an important private property of the owner, it should be secure. This paper presents the algorithms used for mining purpose as well as for securing the data while transacting between Data owner and Third party service provider. It gives secure, efficient and faster results.**

***Keywords*— Data Security, Data mining, Association rule mining, FP-Growth algorithm, AES algorithm.**

I. Introduction

As Data mining is a need of future, it is not necessary that every organization or data owner have sufficient resources, infrastructure to fulfil their customer's needs. Keeping these things in mind, outsourcing of data mining is required, so that customer's requirements can be fulfilled in time. Though it is advantageous to achieve sophisticated analysis on large volumes of data in a cost effective way, there are various serious security issues of the data-mining as- a-service that is data mining performed at the server side and client has to request for the data mining. One of the issues is that when data is outsourced for mining purpose, server has access to the private valuable data of the owner and can read sensitive information. In this way security may break.

II. Proposed System

We proposed a method to solve the problem of security while performing data mining by using k-privacy, i.e. each item in the outsourced dataset should be indistinguishable from at least k1 items regarding their support. , Goal of the proposed system is to devise an encryption scheme which enables formal privacy guarantees to be proved, and to validate this model over real-life transaction database.

The proposed system is to devise encryption schemes such that formal privacy guarantees can be proven against attacks conducted by the server using background knowledge, while keeping their source requirements under control. From system perspective, secure mining is used to provide security for the data which is outsourced to the server machine. The purpose of outsourcing data is to perform pattern mining task at server side. While performing it, it should maintain security. Proposed system generates dataset synthetically. It also accepts real transaction datasets.

Figure 1 shows the proposed system architecture. The client encrypts its data using Encrypt / Decrypt (ED) module. ED module transforms the input data into an encrypted database. The server conducts data mining and sends the patterns to the client.

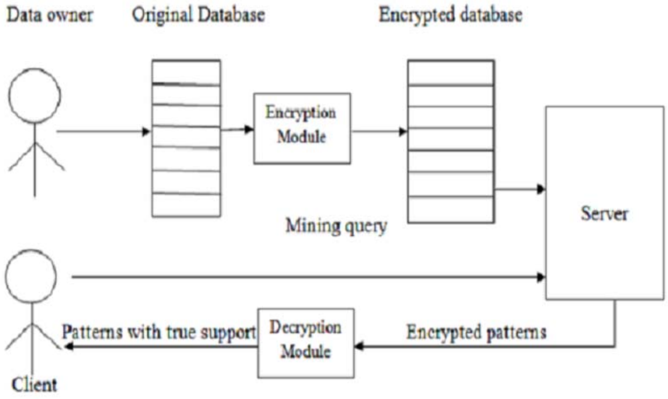

Fig 1. System Architecture

III. Algorithm

The system uses AES algorithm for Encryption / Decryption and FP-growth algorithm for pattern mining task. It is very efficient and requires less time for large dataset.

A. *Encryption / Decryption module*

It is responsible for encrypting the original database which is sent by the data owner and store into Encrypted database and also responsible for receiving the encrypted patterns from server and extracting the true patterns from it.

Encryption / Decryption process make use of three different steps:

- Clustering of items into group of k-items.
- Addition of fake transactions to the original dataset so that k-privacy can be achieved.
- Encryption of every original item of dataset using AES algorithm.

Dataset is uploaded to the server after encryption. Decryption is simply the reverse process of encryption. It is done by the client to get the result.

B. Pattern mining task

Client requests for pattern mining task to the server. Server performs data mining at its end and sends result back to the client. It will send the extracted patterns to the client. For the pattern mining task, FP-Growth algorithm is used. This algorithm is very efficient and faster. It works faster for large database.

C. Security

To improve the security of the outsourced database, random key is generated by the system. After encryption when the data is uploaded to the server, a random key is generated. It is needed when client want to get the result i.e. at the time of decryption.

IV. RESULTS AND DISCUSSION

The following result shows the performance of the proposed system. As stated earlier, the proposed system provides security while performing pattern mining task. To check the performance of the system, we assess the encryption / decryption overhead and support overhead. We have used Supermarket dataset for evaluation as well as datasets are generated by the system and tested. The proposed system shows the result for 7000 records. Datasets are tested for both the algorithms i.e. Apriori and FP-Growth algorithm. In our system we have used FP-Growth algorithm for data mining task. Figure shows the good results of FP-Growth algorithm than Apriori algorithm.

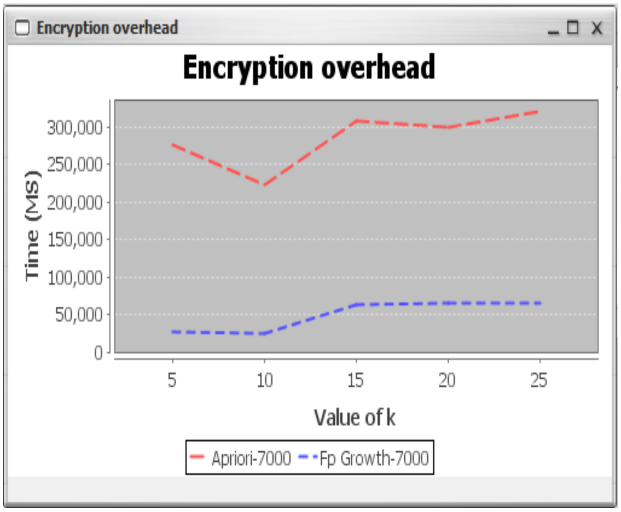

Fig 2. Encryption overhead

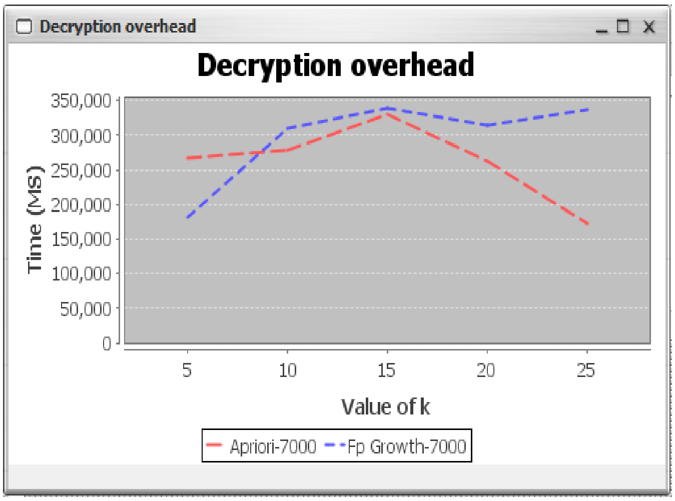

Fig 3. Decryption overhead

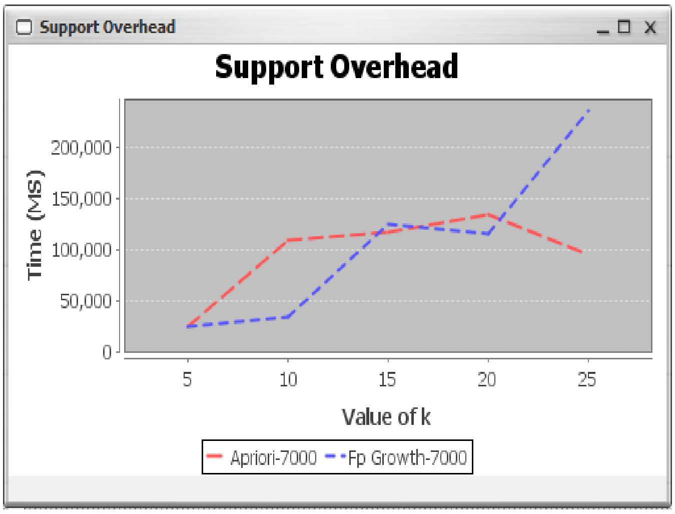

Fig 4. Support overhead

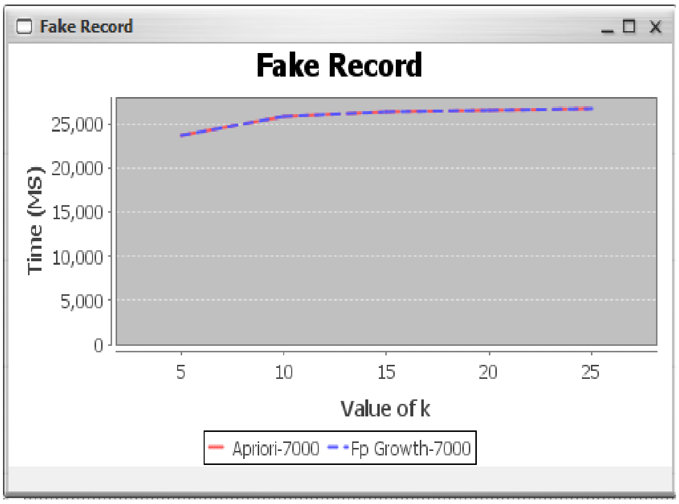

Fig 5. Fake record

V. Conclusion

Data mining is one of the primitive branches of computer science. Due to modernization of handling of data through large flow of influx nodes in case of online and warehouse data, security is at stake. All the existing security algorithms follow their own rigid approach towards conquering security requirements. For the project's sake, secure mining approach for outsourced transaction databases using association rules is proposed. Such rules provide the user with the goodness measure in the transactional databases based on the deep rooted association mining theory. For data mining FP-growth algorithm is used which is very efficient than apriori algorithm without any candidate generation process. On the other hand, the safety of the system tangibly depends on the security algorithm used plus the inner black box description of the algorithm. For instance AES block cipher algorithm is used for both encryption and decryption process. AES provides novice level advantages over all the exiting algorithms which promise similar functionalities on the transactional databases. Strong data protection is achieved by developing an attack model. By using some techniques with the encryption algorithm, privacy is maintained against cryptographic attacks mentioned above.

Acknowledgment

The authors would like to express thanks to the reviewers for helpful comments.

References

[1] M. Arunadevi, R. Anuradha, "Privacy Preserving Outsourcing for Frequent Itemset Mining", International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801, Vol.2, Special Issue 1, March 2014.

[2] 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 Systems Journal, Vol. 7, No. 3, September 2013.

[3] Gagandeep Kaur, Shruti Aggarwal, "Performance Analysis of Association Rule Mining Algorithms", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, August 2013, ISSN: 2277 128X.

[4] Rahul Mishra, Abha Choubey, "Comparative Analysis of Apriori Algorithm and Frequent Pattern Algorithm for Frequent Pattern Mining in Web Log Data", International Journal of Computer Science and Information Technologies (IJCSIT), Volume. 3 (4), 2012, 4662-4665.

[5] F. Giannotti, L. V. Lakshmanan, A. Monreale, D. Pedreschi, and H.Wang, "Privacy-preserving data mining from outsourced databases", in Proc. SPCC2010 Conjunction with CPDP, 2010, pp. 411-426.

[6] Hamdan.O.Alanazi, B.B.Zaidan, A.A.Zaidan, Hamid A.Jalab, M.Shabbir and Y. Al-Nabhani "New Comparative Study between DES, 3DES and AES within Nine Factors", Journal of Computig, Volume 2, Issue 3, March 2010, ISSN 2151-9617.

[7] Ian Molloy, N. Li, and T. Li, "On the (in) security and (im) practicality of outsourcing precise association rule mining", in Proc. IEEE International Conference Data Mining, Dec. 2009, pp. 872-877.

[8] W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis, "Security in outsourcing of association rule mining", in Proc. International Conference Very Large Data Bases, 2007, pp. 111-122.

[9] Gosta Grahne, Jianfei Zhu, "Fast Algorithms for Frequent Itemset Mining Using FP-Trees", IEEE Transactions on Knowledge and Data Engineering, Vol. 17, NO. 10, October 2005.

[10] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data", IEEE Transaction Knowledge Data Engineering, vol. 16, no. 9, pp. 1026-1037, September 2004.

[11] R. Agrawal and R. Srikant, "Privacy-preserving data mining", in Proc. ACM SIGMOD International Conference Management Data, 2000, pp. 439-450.

[12] Y. Lindell and B. Pinkas, "Privacy

[13] preserving data mining", in CRYPTO, 2000, pp. 36-53.