IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 12, December 2016

Survey on Mining Association Rules Securely in Horizontally Distributed Databases

Sayali J. Bhole[1], Pankaj Vanwari[2]

Dept of Computer Engineering, Vidyalankar Institute of Technology, Wadala, Mumbai, India[1, 2]

Abstract: Database has become a very important part in today's world. Database is a system for storing and taking care of data. It can store large data items from various areas and therefore sometimes become difficult to find right information. Thus, Data mining is used to retrieve and view necessary information with less effort. Sometimes, distributed databases are shared among various servers where insecurity occurs due to data leakage. The main reason behind this is trusted third party and to avoid this, we have to propose a protocol which will securely perform its work without involvement of any other party. Our protocol also depends on Fast Distributed Mining (FDM) algorithm.

Keywords: distributed databases, association rules, item-sets, FDM.

I. INTRODUCTION

Secure Mining means provide security to the data obtained during data mining. No other party should come to know private data or information of some other party. The only thing they should know is the required output and not the entire information. For example, if some employees need to know who has highest salary in their department, then everyone will share their salary to each other, but in the way that the only output they will get is the highest salary amount and not the employees name or details. Their privacy should be maintained.

Nowadays, each and every user surf through different websites, where they can enter their private details. It is the responsibility of that site to maintain their privacy. Thus, they can use any of the cryptographic techniques to encrypt and decrypt data, so even if anyone tries to steal information will not get the original data. There is a trusted third party in many of the organization which is intermediary between user and admin. This increases chances of data leakage. To reduce this, it is needed to devise a protocol that the players can directly interact with each other in order to arrive at the required output.

In section 2, we have done literature survey on various cryptographic techniques used for secure mining of horizontally distributed databases based on association rules. In section 3, the problem of secure mining is explained. In section 4, the problem and how to overcome it is defined, in section 5 we have discussed our proposed system and concluded the topic in section 6.

II. LITERATURE REVIEW

The Database partitioning can be done in three ways:

A. Horizontal partitioning:
The database can be partitioned horizontally where each fragment consists of a subset of the records of relation R. Horizontal partitioning divides a table into several small tables. Here, the table is partitioned based on rows. SELECT is a horizontal partition of the relation into two set of tuples.

B. Vertical partitioning:
The database can be partitioned vertically where each fragment consists of a subset of relation R. Vertical partitioning divides a table into several small tables. Here, the table is partitioned based on columns. PROJECT is a vertical partition of the relation into two relations.

C. Mixed partitioning:
The data is first partitioned horizontally and each partitioned fragment is further partitioned into vertical fragments and vice versa.
The idea is to build up a well-organized method that enables a secure computation along with minimizing the amount of private data that each party discloses to other. Privacy preserving association rule mining may be used to solve these problems for horizontally partitioned database.

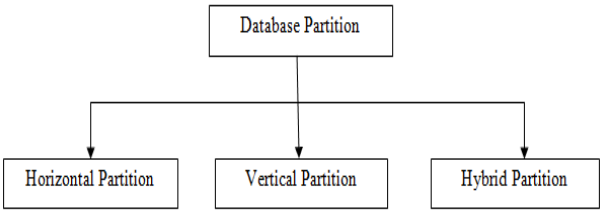

Fig. 1 Types of Database partitioning

There is a survey of the literature to understand the domain of problem and possible approaches to solve them and improvements suggested by different authors. In this section, the summary of different algorithm, problems associated with them and the different approaches used by different authors to solve those problems are discussed. This is required to understand the problem and propose a new model based on their suggestions.

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 12, December 2016

Kantarcioglu and Clifton studied the problem in [2] and devised a protocol for its solution. The main part of the protocol is a sub-protocol for the secure computation of the union of private subsets that are held by the different players. That is the most costly part of the protocol and its implementation relies upon cryptographic primitives such as commutative encryption, oblivious transfer, and hash functions. This is also the only part in the protocol in which the players may extract from their view of the protocol information on other databases, beyond what is implied by the final output and their own input.

Pattnaik Dr. Prasant Kumar [3], proposed Privacy Preservation in Distributed Database. To provide the security to the distributed database, they proposed hash based secure sum cryptography technique without trusted party and another concept proposed in that paper was trusted party that controls all the details of all the party present in the distributed database environment. After comparing the result, it was proved that data leakage with trusted party is more as compared to without trusted party. Privacy is also more in without trusted party as compared to with trusted parties.

As suggested by J. Danasana, R. Kumar and D. Dey [6], they used Double hash function to find mask value enhance the privacy making the partial support (association rule) in more disguised form. When sites try to find the global frequent item sets from its local frequent item sets then the sites also includes some of the infrequent item sets to its local frequent items and send to the next sites so that the current site is unable to know the frequent item set of previous site. The only required site can find the frequent item-sets and thus maintain privacy.

G.K. Chaturvedi, R.M. Gawande [8] proposed Secure Mining of Association Rules in Horizontally Distributed Databases Using FDM and K&C Algorithm. In their work they used FDM algorithm and Unifi-KC algorithm. They compared its efficiency, computational cost and communication cost and clearly explained Privacy in data mining. In their work they concluded that FDM algorithm is sufficient compared to Unifi-KC algorithm.

III. PROBLEM STATEMENT

The problem is of secure mining of association rules in horizontally partitioned databases. In this, there are several sites (or players) that hold homogeneous databases, i.e., databases that share the same schema but may have different information, where all sites use same DBMS products.

The goal is to find all association rules with support at least s and confidence at least c, for some given minimal support size s and confidence level c, from data held in a unified database. The information that we would like to protect in this context is not only individual transactions in the different databases, but also that are supported locally in each of those databases.

IV. PROBLEM DEFINITION

The problem depends on secure multi-party computation. In such problems, there are several players that hold private inputs and they wish to securely compute it for some public function. In previous systems, there existed a trusted third party where the players could send their inputs which perform the function evaluation and send them the resulting output. Thus, the trusted-party would come to know the private information of users. In order to avoid this, it is needed to devise a protocol so that the players can directly interact with each other in order to arrive at the required output. In our problem, the inputs are the partial databases, and the required output is the list of association rules that hold in the unified database with support and confidence no smaller than the given minimal support s and confidence c. The algorithm used is Fast Distributed Mining (FDM) which is an unsecured distributed version of the Apriori algorithm.

V. PROPOSED WORK

In Proposed System, we propose an alternative protocol for the secure computation of the union of private subsets. The proposed protocol improves upon that in terms of simplicity and efficiency as well as privacy. The data stored at each site is encrypted and then it is send so as to protect data from malicious party. The required party will decrypt using particular key and get original data. The modules included in our project are as follows:

A. User module:
User has to login and feed inputs into any required sites, perform transactions. This data should be encrypted using any cryptographic technique. Earlier, they used AES technique, now we are going to use RSA and perform encryption-decryption.

B. Admin module:
Admin can view the user details. It can also view item-sets based on the user processing details using association rule with apriori algorithm.

C. Association Rule:
Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true. The final outcome is to find association rules for the frequent item-sets.

D. Apriori algorithm:
The purpose of the Apriori Algorithm is to find associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items and is called a transaction. The output of Apriori is sets of rules that tell us how often items are contained in sets of data. We have used Fast

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 12, December 2016

Distributed Mining (FDM) algorithm in our system. FDM algorithm consists of following steps:

1. Initialization
2. Candidate sets generation
3. Local pruning
4. Unifying the candidate item-sets
5. Computing local supports
6. Broadcast mining results

VI. CONCLUSION

There are various methods to secure private data of users from any third party which results into some output. The proposed system will try to obtain the best results in terms of time efficiency and computation cost. The purpose is to secure mined databases in its temporary storage, so that the server need not be hit every time when the data is required. The main aim is to highly focus on the security of the databases, maintain privacy. We will use RSA as a cryptographic technique and then compare it with AES results.

ACKNOWLEDGEMENT

I would like to thank my guide, **Prof. Pankaj Vanwari** for his expert guidance, encouragement and valuable suggestions at every step and also to the Head of the Department **Prof. Sachin Deshpande,** Computer Engineering for encouraging and inspiring me to carry out the project in the department lab.

I am extremely happy to acknowledge and express my sincere gratitude to my parents for their constant support and encouragement and last but not the least, friends and well wishers for their help, encouragement and cooperation during the course of the project.

REFERENCES

[1] Shikha Sharma, "An extended method for Privacy Preserving Association Rule Mining", International journey of advanced research in computer science and software engineering, vol 2, October 2012.
[2] M. Kantarcioglu; C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", IEEE Trans. Knowl. Data Eng. 16(9): 10261037, 2004.
[3] Dr.P.K. Pattnaik, K Raghvendra, Dr. Y Sharma, "Privacy preservation in distributed database", European Journal of Academic Essays 1(2): 35-39, 2014, ISSN: 21831904.
[4] Y. Lindell; B. Pinkas, "Secure Multiparty Computation for Privacy-Preserving Data Mining", Journal of Privacy and Confidentiality, 2008.
[5] A.A., Veloso; Jr.W. Meira; S. Parthasarathy; M.B. de Carvalho, "Efficient, accurate and privacy preserving data mining for frequent itemsets in distributed databases," Proceedings of the Brazilian Symposium on Databases, Manaus, Amazonas, Brazil, pp.281-292, 2003.
[6] Jayanti Danasana, Raghvendra Kumar and DebaduttaDey, "Mining Association Rule For Horizontally Partitioned Databases Using Ck Secure Sum Technique", International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.6, November 2012.
[7] R.L. Rivest, A. Shamir, and L.M. Adleman,"A method for obtaining digital signatures and public-key cryptosystems" Commun. ACM, 21(2):120–126, 1978.
[8] G.K. Chaturvedi, R.M. Gawande, "Secure Mining of Association Rules in Horizontally Distributed Databases Using FDM and K&C algorithm" International Journal of Emerging Trends & Technology in Computer Science, Volume 3, Issue 3, May-June 2014.