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

Secure Mining of Association Rules from Homogeneous Database

Mahale Mohini V.[1], Shaikh I.R.[2]

[1]Department of Computer Engineering, SND COE & RC, Yeola, Nashik-423401

[2]Professor, Department of Computer Engineering, SND COE & RC, Yeola, Nashik-423401

Abstract: ***Data mining extracts interesting information and patterns from large database. One of the most widely used technic in data mining is association rule mining. This extracts correlation of attributes from the large database. These extracted rules are used for define some predicates in various application. In the case of distributed database the data is get shared from number of users. The data shared by those users should not disclose to other users while the efficient rules are get calculated from this share database. This paper provides the protocol which wills securely extract the association rules from the shared database. Here we are considering the example of online shopping store. Where owner of the shop will extract the rules from the transactional database.***

Keywords: Privacy preserving data mining, distributed computation, frequent item sets, association rules, secure multiparty computation.

1. Introduction

Data mining technology has developed to identify patterns and trends from large database. In data mining and data warehousing has a tools which operate by gathering all data into a central site, then the algorithm will run on that central data. But if the data is distributed among several sites then data should be share, but none of them ready to share their data due to the security issues. This paper addresses the problem of computing association rule in such a scenario where data is distributed among several sites or clients. Here we assume homogeneous database where all databases from different client have same schema but entities are different. Goal of this system is to produce association rules from distributed database while limiting the information shared about each site.

Recently distributed data mining becomes more popular in various areas like medical, government and in business application. Where more than user mine their data for generate useful information. In such environment more than one user are connected to each other wanted calculate some common function without disclosing the sensitive information to each other, this kind of computation is called as secure multiparty computation.

This paper provides the solution to the problem of secure multiparty computation. The protocol proposed here provides a secure algorithm to extract association rules from the distributed environment. The proposed algorithm here is Fast Distributed Mining (FDM) algorithm. This uses the Apriori algorithm to get the frequent itemset from the database. And to provide security to sensitive data MD5 algorithm is get used.

2. Related Work

Previous work in privacy preserving data mining has reflects two correlated settings. One in which data miner and data owner are two different objects, and another in which data is distributed between several parties who wanted to mine data on the combined quantity of data that they hold. In the first setting the goal is to protect individual data from the data miner, here the data owner uses perturbed data to protect his data from data miner [2].In the second setting the goal is to perform data mining while securing the data records from the other data owner in the mining process. The common method here is the cryptographic

In the second setting, the goal is to perform data mining while protecting the data records of each of the data owners from the other data owners. This is a problem of secure multi-party computation. The usual approach here is cryptographic, Lindell and Pinkas proposed a protocol where they build secure ID3 decision tree [3]. Lin et al. [5] has proposed an algorithm called EM algorithm for secure clustering over horizontally distributed data. Problem of association rule mining in the distributed environment also studied in [1] in which to provide the security commutative cryptography is get used along with FDM .The solution to association rule mining in vertical distributed environment has given in[4]and[6] where each database holds different attributes. In case of large scale horizontally distributed environment protocol given in [7], which considers top manager computers that, assist the resources to decrypt the messages.

L. Kissner and D.X. Song gives priacy preservation using set intersection operation [8].While in paper [9] public key cryptographic system is get provided.

3. Proposed Methodology

The proposed system gives the alternative protocol which will overcome from the problem which are occurred in Farst distributed Mining (FDM) proposed by Kantarcioglu and Clifton(1).The proposed system is more efficient than the existing system in terms of privacy, communication rounds, communication cost and computational cost. The existing and proposed system both are based on FDM [1], which is an unsecured version of the Apriori algorithm. The proposed

system computes a parameterized family of function which is called as threshold function In which two cases correspond to the problems of computing the union and intersection of private subsets. The protocol used for this function can be used in other cases as well. The major problem of extraction of association rule is set inclusion problem; the problem where Bob holds a private subset of some ground set, and Alice holds an element in the ground set, and they wish to determine whether Alice's element is within Bob's subset, without revealing to either of them information about the other party's input beyond the above described inclusion.

The main notion of FDM is that any frequent item set must be also locally frequent in at least one of the sites. Hence, in order to find all globally frequent item sets, each player discloses his locally s-frequent item sets. Then the players check each of them to see if they are s-frequent also globally.
The FDM algorithm proceeds as follows:

1) Initialization: All the players should calculate all k-item sets that are s-frequent that is calculate $F^k_{s.}$
2) Generation of candidate set: The set of all local and global frequent item sets are get calculated by each player $P_{m.}$ Specifically Pm computes $F^{k-1,m}_s \cap F^{k-1}_s$. Then the Apriori algorithm is get performed to generate the set $B^{k,m}_s$.
3) Local Pruning: Each player computes $supp_m(X)$. He then maintains only locally frequent item which is denoted by $Cs^{k,m}$.
4) Unifying the candidate item sets: Each player broadcasts his own set of items $Cs^{k,m}$ which is calculated in above step. Then all players computes $C^k s$
5) Computing local supports: Local supports of all item sets that is $C^k s$ is get calculated.
6) Broadcast mining results: Each player broadcasts his own local support. So that everyone can compute the global support of every item set. Finally the set of all globally frequent item sets F^k_s whish is subset of C^k_s is get produced.

4. Mathematical Modeling

The proposed system access a transactional database and extracts association rules from it.
Mathematical model for the system: -

The proposed system S is define as
$S = \{D, P, L, F^k_s, C^k_s, Km, EC^k_s\}$
Where
$D = \{D1,D2,D3,\ldots\ldots Dm\}$ Set of databases.
$P = \{P1,P2,P3,\ldots\ldots Pm\}$ Number of players.
$F^k_s = \{F^1_s, F^2_s, F^3_s, \ldots\ldots F^k_s\}$ Globally frequent item sets.
C^k_s = Frequent item sets.
Km = Private randome key.
$E C^k_s$ = Encrypted frequent item sets.
F = {f1,f2,f3,f4,f5}
Main functions in system design are given below.

1. Function f1 takes Dm and calculate s-frequent item sets for each transaction.
   $F1(Dm) \rightarrow .f^1s, f^2s \ldots f^k s$
2. Function f2 will find item sets which are locally frequent as well globally frequent.
   $f2(Dm) \rightarrow F^{k-1}_s \cap F^{k-1}$
3. Function f3 computes frequent item sets C^k_s
4. Function f4 will encrypt the generated C^k_s using key km.
5. $f4(C^{k,m}_s) \rightarrow EC^k_s$

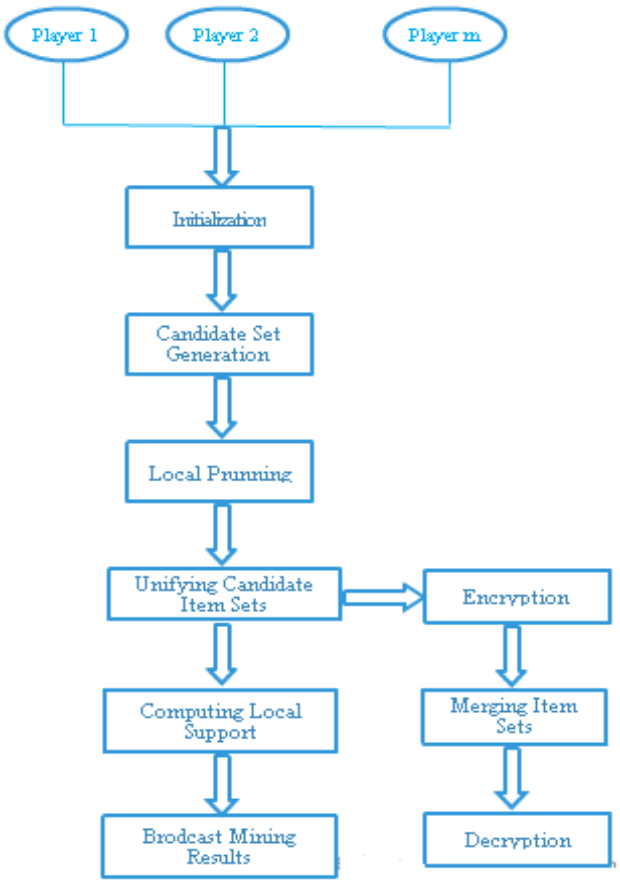

Figure 1: Architecture Diagram

4.1 Data Flow Diagrams

4.1.1 Level 0- Data Flow Diagram

Level 0 DFD for extraction of association rule is as shown in the figure given bellow. The strew database given as the input to the system. And this system is responsible to generate strong rules.

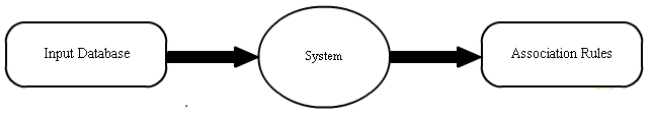

Fig.2: Level 0 Dataflow Diagram

4.1.2. Level - Data Flow Diagram

Level 1 data flow diagram gives a detailed view of the flow of data in the proposed system, in which all the function needed for the system are shown

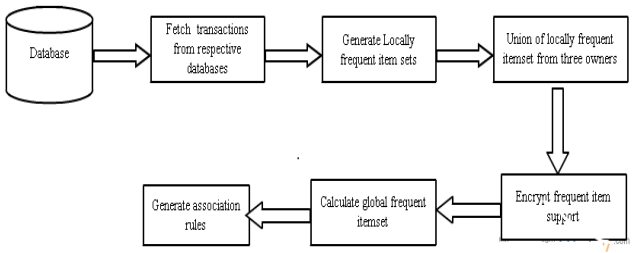

Figure 3: Level1 Dataflow Diagram

5. Result and Dataset

5.1.1 DATASET

Here we use transactional database

5.1.2. Result set

Here we apply Apriori algorithm on the database which calculates the local frequent itemset. And FDM algorithm which will calculate global frequent itemset using which the association rules are get generated.

Figure 4: Local Itemset

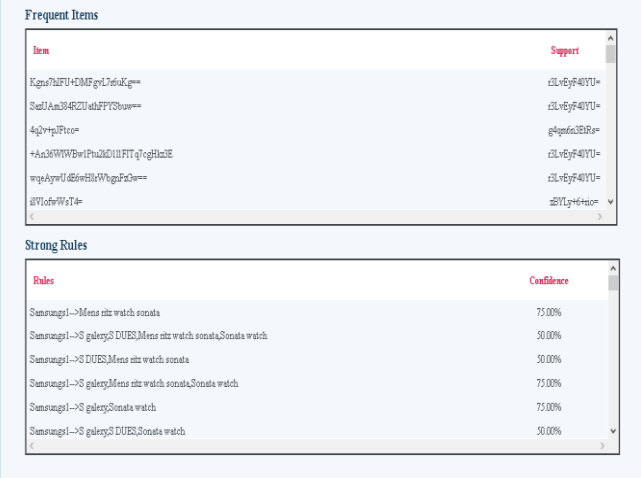

Figure 5: Association Rules

Figure 6: Graph showing Association Rules.

6. Conclusion

Extracting association rules from strew database involves the problem of secure multiparty communication. We proposed a protocol for secure mining of association. Rules from strew database that improves expressively upon the current leading protocol in terms of privacy and efficiency. The main ingredient is a protocol that tests the inclusion of an element held by one player in a subset held by another. n addition, this system Also, it is more simple and significantly more effective in terms of communication rounds, communication cost and computational cost.

References

[1] M. Kantarcioglu and C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 9, pp. 1026-1037, Sept. 2004.

[2] R. Agrawal and R. Srikant, "Privacy-Preserving Data Mining,"Proc. ACM SIGMOD Conf., pp. 439-450, 2000.

[3] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," Proc. Crypto, pp. 36-54, 2000.

[4] J. Vaidya and C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 639- 644, 2002.

[5] X. Lin, C. Clifton, and M.Y. Zhu, "Privacy-Preserving Clustering with Distributed EM Mixture Modeling," Knowledge and Information Systems, vol. 8, pp. 68-81, 2005

[6] J. Zhan, S. Matwin, and L. Chang, "Privacy Preserving Collaborative Association Rule Mining," Proc. 19th Ann. IFIP WG 11.3 Working Conf. Data and Applications Security, pp. 153-165, 2005.

[7] A. Schuster, R. Wolff, and B. Gilburd, "Privacy-Preserving Association Rule Mining in Large-Scale Distributed Systems," Proc. IEEE Int'l Symp. Cluster Computing and the Grid (CCGRID), pp. 411-418,y 2004.

[8] L. Kissner and D.X. Song, "Privacy-Preserving Set Operations," Proc. 25th Ann. Int'l Cryptology Conf. (CRYPTO), pp. 241-257, 2005.

[9] T. ElGamal, "A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms," IEEE Trans. Information Theory, vol. IT-31, no. 4, July 1985

Author Profile

Mahale M. V., Post Graduate Student, was with Pune University, Maharashtra, India. She is now with the Department of Computer Engineering, SND COE, Pune University, Maharashtra. India.