ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Securing Data of Horizontally Distributed Databases Using RSA

Sayali J. Bhole, Pankaj Vanwari

ME Student, Dept. of Computer Engineering, Vidyalankar Institute of Technology, Wadala, Mumbai, India

Assistant Professor, Dept. of Computer Engineering, Vidyalankar Institute of Technology, Wadala, Mumbai, India

ABSTRACT: Database is becoming most important part of today's world .It stores large amount of data taken from different areas. Thus, it becomes difficult to find required information quickly. Data mining solves this problem as it helps to retrieve and view necessary information within few minutes. Distributed databases are shared among servers which increases data leakage. The main reason behind this is inclusion of trusted third party. Thus, we propose a protocol which insures security without involvement of any other party. Our protocol also depends on FDM algorithm.

KEYWORDS: distributed databases, association rules, item-sets, FDM

I. INTRODUCTION

Secure Mining means provide security to the data obtained during data mining. The data transferred between any two parties should be secured, so that no other party gets to know about your data. The party other than communicating parties should know only desired output and not the full 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.

Internet surfing is most common thing nowadays. People surf through internet for finding information, shopping, and online payment and so on. For this purpose, it is necessary to secure their private details from masqueraders. Thus, they have to use cryptographic techniques to encrypt and decrypt data using any of the techniques to maintain its secrecy. 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 required to design a protocol so that players can interact with each other directly without any third party to arrive at the required output.

In section 2, we have discussed the previous methods proposed by various authors for securing data. In section 3, we explained our proposed system, modules included in it. In section 4, experimental results are discussed along with comparison charts between existing and proposed system. In section 5, we have concluded our project topic results.

II. LITERATURE SURVEY

The Database partitioning can be done in three ways:

A. *HORIZONTAL PARTITIONING:*

This partitioning divides the table into several small tables by row. SELECT is a horizontal partition of the relation into two set of tuples.

B. *VERTICAL PARTITIONING:*

This partitioning divides the table into several small tables by columns. PROJECT is a vertical partition of the relation into two relations.

C. *MIXED PARTITIONING:*

This partitioning first divides the table horizontally and then divides that fragmented part of table vertically or vice-versa.

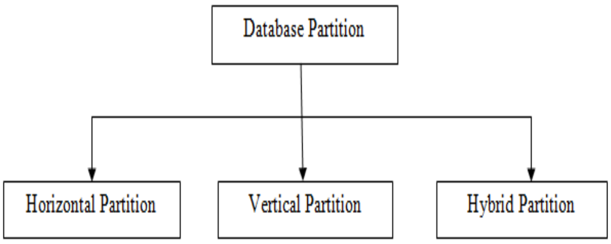


Fig. 1 Types of Database Partitioning

The survey about a particular topic of author is done to understand the problem of a domain and how they solved that problem in different ways, so that we get an idea of the topic. We also get some suggestions from the authors to propose our own new model without much effort.

In the paper proposed by Kantarcioglu and Clifton [2] it devised a protocol for the solution of its problems. 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. His implementation relies upon commutative encryption, oblivious transfer and hash functions. This protocol helps the players to extract information on other databases from their point of view.

Pattnaik Dr. Prasant Kumar [6], 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 trusted party that handles 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.

The paper secure mining of association rules in horizontally distributed databases using FDM and K & C algorithm[8] proposed by G.K. Chaturvedi, R.M. Gawande. They compared efficiency, computational cost and communication cost of both the algorithms and clearly explained privacy in data mining. They concluded that FDM is sufficient compared to UNIFI-KC algorithm.

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.

III. PROPOSED METHODOLOGY

In Proposed System, we propose a protocol to avoid the need of any third party, an intermediary between two communicating parties. The Protocol uses 1024 bit RSA cryptographic technique to encrypt-decrypt data, FDM algorithm to compute frequent itemsets. The flowchart of our system is as follows:

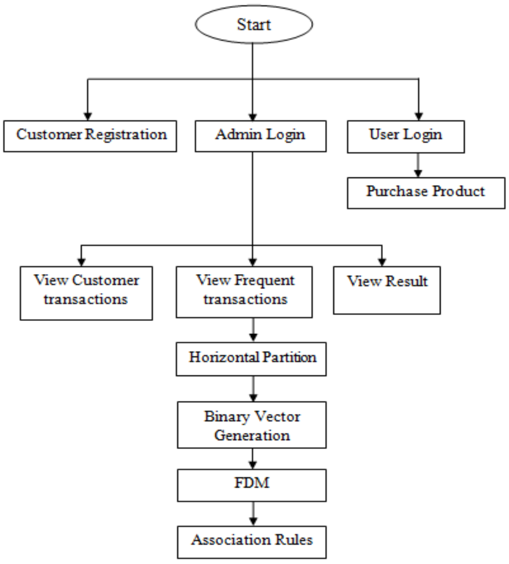


Fig. 2 Architecture of proposed system

The main modules in our system include user module, admin module, association rule and FDM. Let us explain each module in detail:

A. *USER MODULE:*

User has to login and purchase the products by just selecting items and submitting it. If user is not registered, they have to first register themselves and then use that UserId and Password to login. After registering to the site and then logging in only, they can purchase products.

B. *ADMIN MODULE:*

Admin has to login to view customer transactions, frequent transactions and the final output. He then partition data horizontally, performs FDM process over it to get the frequent item-sets and association rules with support and confidence value.

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.

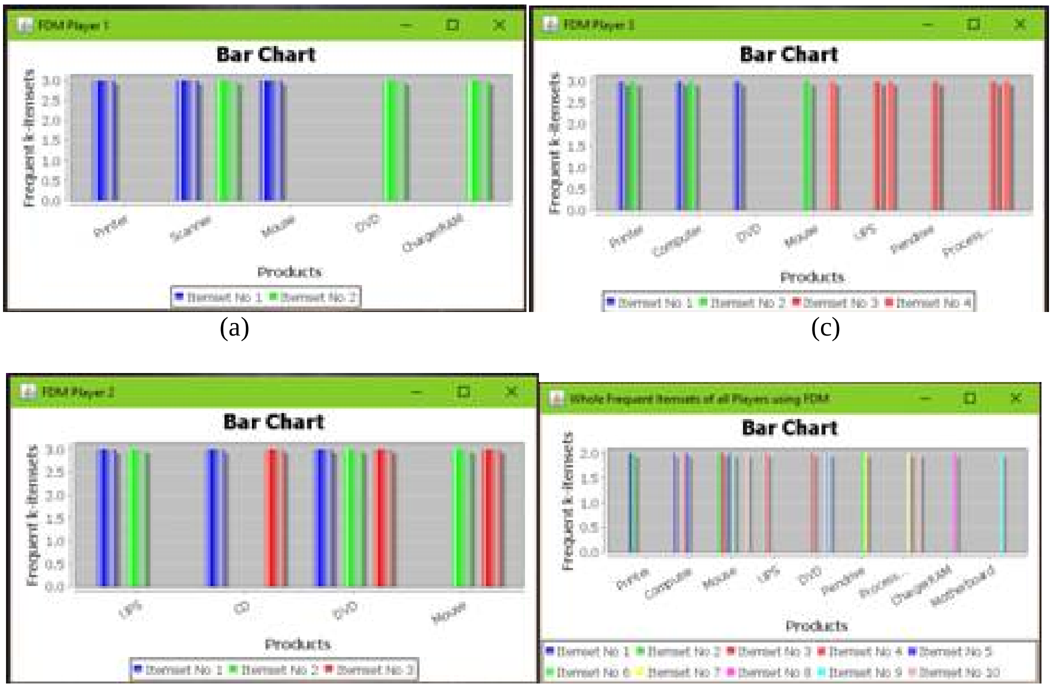
D. *FREQUENT DATA MINING(FDM):*

FDM algorithm consists of following steps:

- Initialization
- Candidate sets generation
- Local pruning
- Unifying the candidate item-sets
- Computing local supports
- Broadcast mining results

IV. EXPERIMENTAL RESULTS

In FDM algorithm, each player first calculates the frequent itemsets present locally in their sets. Then they broadcast the result to each other so that they can find the most frequent item sets present globally between them. The frequent itemsets are thus first calculated locally and then globally using FDM algorithm. In our project, this frequent itemsets are divided into 3-players based on minimum support value i.e. threshold value. The following figures depict the frequent itemsets locally at each player as well as globally.

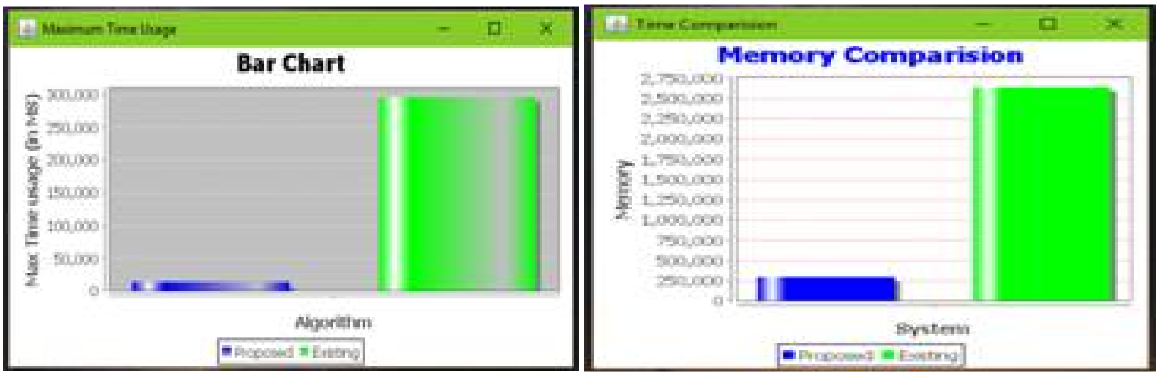


(a) (c)

(b) (d)

Fig. 3 Frequent itemsets present locally as well as globally

The first three charts i.e. Fig 3. (a), (b), (c) shows the frequent items present locally at each player i.e in our case the frequent item sets present at each of the three players and Fig 3. (d) Shows the global frequent items calculated by all the players. Also we calculated the maximum time usage and memory consumption of our system and compared the result with the existing systems shown in Fig 4. The starting time of both the systems are taken, then after finishing the task of our system, we get end time and from this, we calculate the time usage of system. Similarly, we find time usage of existing system.



(a) (b)

Fig. 4 Comparing Maximum time usage (a) and memory consumption (b) of both the systems.

V. CONCLUSION

There are various methods to secure private data of users from any third party which results into some output. Our Proposed system uses RSA cryptographic technique to encrypt-decrypt data. The memory consumption and computation time of proposed system using RSA is calculated and compared with existing system using AES. The results of proposed system are much better than existing system. In future, we can secure data from distributed databases divided vertically or in mixed distributed databases.

VI. 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. 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.
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. 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.
7. 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.
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.
9. W. Du; M. J. Atallah, “Secure multi-party computation problems and their applications: A review and open problems”, In Proceedings of the 2001 New Security Paradigms Workshop, Cloudcroft, New Mexico, 2001.
10. Y. Lindell; B. Pinkas, “Privacy preserving data mining”, Advances in Cryptology, CRYPTO 2000 ,2000.

BIOGRAPHY

Sayali Bhole is pursuing M.E (Comp Engg.) from Vidyalankar Institute of Technology, Mumbai University, Maharashtra. She did her graduation B.E (Comp Engg.) from Mumbai University, Maharashtra.