International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 3 1142 – 1144

Survey on Secure Mining of Association Rules in Vertically Distributed Databases

P. Kalaivani
PG Scholar Department of CSE
Bharath University, Chennai, India

D. Kerana Hanirex
Asst.Professor/CSE
Bharath University, Chennai, India

Dr. K. P. Kaliyamurthie,
HOD/CSE
Bharath University, Chennai, India

***Abstract*—**A distributed database system is a collection of sites connected on a common high bandwidth network. Logically, data belongs to the same system but physically it is spread over the sites of the network, making the distribution invisible to the user. The advantage of this distribution resides in achieving availability, performance, modularity and reliability. In this paper, I have done a survey of papers related to Mining of Association Rules over distributed databases. From this survey, we have come up with a proposed solution to address the problem of secure mining of association rules where transactions are distributed in vertically distributed databases. Each site holds some attributes of each transaction and the sites wish to participate in the identification of globally valid association rules However, the sites should not reveal individual transaction data. The Protocol is based on Apriori Algorithm [2] and MultiParty Algorithm [3] for efficiently discovering frequent item sets with minimum support levels, without either site communicating individual transaction values.

Keywords - *Frequent Item sets, Apriori Algorithm, Multi Party Algorithm, Association Rules*

*****

I. INTRODUCTION

Here we study the problem of secure mining of association rules in vertically partitioned databases. In such a setting, there are many sites (or players) that hold separate databases, i.e., each site maintain separate databases by itself where the transactional attributes are distributed across the databases. The goal is to find all association rules globally with support at least s and confidence at least c for some given minimal support size s and confidence level c that hold in vertical databases while minimizing the information disclosed about the private databases held by those players.

We present Apriori algorithm to mine association rules from vertically partitioned databases. By vertically partitioned, we mean that each site contains some elements of a transaction. Using the traditional "market-basket" example, one site may contain food and other commodity purchases (grocery), while another may have clothing purchases. Using a key such as transaction id and date, we can join these to identify relationships between purchases of clothing and groceries. For Secure Mining, we present Multi Party Algorithm. In an MPC, a given number of participant's p1, p2... pN each have private data, respectively d1, d2... dN. Participants want to compute the value of a public function F on N variables at the point (d1, d2... dN). An MPC protocol is secure if no participant can study more from the description of the public function and the result of the global calculation than what he/she can learn from his/her own entry under particular conditions depending on the model used.

The layout of the paper is as follows. In section 2, brief summaries of the papers referred to are given. In Section 3, presents the Proposed Solution. Lastly, we concluded in section 4 and section 5 contains references.

II. RELATED WORK

Tamir Tassa, has presented a protocol for secure mining of association rules in horizontally distributed databases. The protocol is designed based on Fast Distributed Mining (FDM) Algorithm and Secure MultiParty Algorithm. The Protocol offers enhanced privacy with respect to the protocol in [4]. In addition to that, it is simpler and is significantly more efficient in terms of communication rounds, computational cost and communication cost. All experiments were implemented in C# (.net 4) and were executed on an Intel(R) Core(TM)i7-2620M personal computer with a 2.7 GHz CPU, 64-bit operating system Windows 7 Professional SP1 and 8 GB of RAM[1].

RakeshAgrawl, Ramakrishnan Srikant, have presented a system for discovering association rules between items in Large Databases. The speed for processing the data using AIS [7] and SETM [8] in the databases was not fast and also they tend to find a lot of item sets which were small and thus end up wasting time. They Proposed Apriori Hybrid Algorithm for fast computation. Apriori and Apriori-Tid combined into a Hybrid Algorithm called Apriori Hybrid Algorithm. Experiments have demonstrated that Apriori Hybrid Algorithm is faster than AIS [7] and SETM [8] Algorithm [2].

A.Ben-David, N. Nisan and B. Pinkas, have proposed a system for Multi Party Computation. Secure computation is one of the big achievements of modern cryptography, enabling a set of untrusted parties to compute any function of their private inputs while disclosing nothing but the result of the function. They have presented FairplayMP, a generic system for Secure

Multiparty Computation. This is an extension of the Fairplay system which supported secure computation by two parties. The extension to the multi-party case is needed since cryptographic protocols for the multi-party scenario are completely different from protocols for the two-party case [3].

A.Ben-David, N. Nisan and B. Pinkas, have proposed a system for Multi Party Computation. Secure computation is one of the big achievements of modern cryptography, enabling a set of untrusted parties to compute any function of their private inputs while disclosing nothing but the result of the function. They have presented FairplayMP, a generic system for Secure Multiparty Computation. This is an extension of the Fairplay system which supported secure computation by two parties. The extension to the multi-party case is needed since cryptographic protocols for the multi-party scenario are completely different from protocols for the two-party case [3].

M. Kantarcioglu and C. Clifton, have presented a protocol for Privacy-Preserving distributed mining of Association Rules on Horizontally Partioned data. The paper addresses the problem of computing association rules where the data may be distributed among various custodians, none of which are allowed to transfer their data to another site. Databases are homogeneous where all sites have the same schema but each site has information on different entities. Association Rules have been computed based on Fast Distributed Algorithm (FDM) and Secure Multiparty computation [4].

Jaideep Vaidya, Chris Clifton, have proposed a protocol for privacy preserving Association Rule Mining in Vertically Partitioned Data. The protocol is implemented through a two-party algorithm for efficiently discovering frequent item sets with minimum support levels, without either site communicating individual transaction values. They have demonstrated that it is possible to achieve good individual security with communication cost comparable to that required to build a centralized data warehouse [5].

Hassan. I., F. Marir, have presented vertical partitioning problem during the design of distributed databases by conducting a comparative study for different vertical partitioning algorithms to reach the most efficient vertical fragmentation scheme that leads to proper data allocation and replication. They have proposed Bond Energy Algorithm Binary Vertical Partitioning Algorithm shown how Graph-based vertical partitioning algorithm has contributed towards the optimization of data fragmentation problem by providing an efficient way of improving performance of applications [6].

R. Agrawal, T. Imielinski, and A. Swami, have presented a protocol for mining association rules between sets of items in a large database system. They presented an efficient algorithm for extracting association rules between items in the databases. The Algorithm incorporates novel estimation method, buffer management and pruning techniques. They have presented results of applying this algorithm to sales data obtained from a large retailing company which shows the effectiveness of the algorithm [7].

M. Houtsma and A. Swami, have proposed a protocol for set-oriented association rules mining. The protocol presents an algorithm called SETM for mining association rules. The SETM algorithm is simple, fast and stable over the range of parameter values. The paper shows that at least some aspects of data mining can be carried out by using general query Languages such as SQL, rather than by developing specialized algorithms [8].

In the earlier research work [9], [10], [11], [12], [13], [14], [15], [16] the author uses various techniques such as clustering , transaction reduction and algorithms such as apriori, fpgrowth used for mining frequent itemsets using different datasets were analyzed and compared.

III. PROPOSED SYSTEM

We propose a protocol for finding association rules in vertically distributed database. Here we present two algorithms for finding association rules as fast as possible. Apriori Algorithm is for finding the frequent item sets and other one is for securing the private data of the users. Multiparty secure algorithm will secure the data held by the users. Association Rules will be extracted from the frequent item sets found.

Association rule generation is generally split up into two separate steps:

First, minimum support is applied to find all frequent itemsets in a vertically distributed database.

Second, these frequent itemsets and the minimum confidence constraint are used to form the rules

IV. CONCLUSION

In previous [1], this system was implemented in horizontally distributed databases. Now we propose the system in vertically Partitioned databases. Here we use two major algorithms namely Apriori Algorithm and Secure Multiparty Algorithm. We propose to mine the association rules using these algorithms with faster fetching of data and securing the private data of the users

REFERENCES

[1] Tamir Tassa, “Secure Mining of Association Rules in Horizontally Distributed Databases” IEEE

TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 4, APRIL 2014

[2] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," Proc 20th Int'l Conference. Very Large Data Bases (VLDB), pp. 487-499, 1994.

[3] A. Ben-David, N. Nisan and B. Pinkas, "FairplayMP - A System for Secure Multi-Party Computation," Proc. 15th ACM Conference. Computer and Comm. Security (CCS), pp. 257-266, 2008.

[4] 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,September 2004

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

[6] Vertical Partitioning Impact on Performance and Manageability of Distributed Database systems (A Comparative study of some vertical partitioning algorithms) (2006) by Hassan I. Abdalla, F. Marir 18th National computer conference 2006.

[7] R. Agrawal, T. Imielinski, and A. Swami. Mining Association Rules between Sets of Items in Large Databases. In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D.C., May 1993

[8] M. Houtsma and A. Swami. Set-Oriented Mining of Association Rules. Research Report RJ 9567, IBM Almaden Research Center, San Jose, California, October 1993.

[9] D.Kerana Hanirex, Dr.K.P.Kaliyamurthie," Mining Frequent Itemsets Using Genetic Algorithm", Middle-East Journal of Scientific Research, 19 (6): 807-810, 2014.

[10] Kerana Hanirex.D, Dr.K.P.Kaliyamurthie, "Finding the Dominating Amino Acids in Dengue Virus Type1 Study on mining frequent itemsets", 4(3): (B);880 – 89, Int. Journal of Pharma and Bio Sciences, July, 2013.

[11] D.Kerana Hanirex., Dr.K.P.Kaliyamurthie, "Multi-classification Approach for Detecting Thyroid Attacks", IJPBS, 4(3), (B) 1246 – 1251, July (2013).

[12] Kerana Hanirex.D, "An Efficient TDTR Algorithm for Mining Frequent Itemsets", International Journal of Electronics and Computer Science Engineering,2(1):251-256;2012.

[13] D.Kerana Hanirex ,"A Comparative Analysis on Mining Frequent Itemsets", International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), Volume1, Issue 10, P68-71,2012.

[14] D.Kerana Hanirex, Dr.A.Kumaravel, "An Efficient Partition and Two Dimensional Approach for Mining Frequent Itemsets", International Journal of Technological Synthesis and Analysis (IJTSA), Volume 1, Issue 1 ,P14-17, 2012.

[15] Kerana Hanirex.D, Dr.M.A.Dorai Rengaswamy, "Efficient Algorithm for Mining Frequent Itemsets using Clustering techniques",,IJCSE,3(3):1028- 1032;2011

[16] D.Kerana Hanirex ,"Association Rule Mining in Distributed Database System", International Journal of Computer Science and Mobile Computing(IJCSMC), Vol3,Iss 4,pg 727-732,2014.