*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661,p-ISSN: 2278-8727,*
*PP 49-51*
*www.iosrjournals.org*

# Mining of Association Rules in Horizontally Distributed Databases

D.A.Vidhate1, Abhang Shweta[2], Gadekar Sonali [3],Kadam Priyanka[4], Rahinj Rupali[5]
[1,2,3,4]*(IT Department, P.D.V.V.P.C.O.E Ahmednagar, Pune University, India)*

***Abstract:*** *Data mining is the most fast growing area today which is used to extract important knowledge from large data collections but often these collections are divided among several parties. Privacy liability may prevent the parties from directly sharing the data and some types of information about the data. In this project they propose a protocol for secure mining of association rules in horizontally distributed databases. The current integral protocol is that of Kantarcioglu and Clifton well known as KC protocol. This protocol is based on an unsecured distributed version of the Apriori algorithm named as Fast Distributed Mining (FDM) algorithm of Cheung et al. The main ingredients in our protocol are two novel secure multiparty algorithms one that computes the union of private subsets that each of the interacting players hold and another that tests the whether an element held by one player is included in a subset held by another. This protocol offers enhanced privacy with respect to the earlier protocols. In addition, it is not complicated and is importantly more effectual in terms of communication cost, communication rounds and computational cost .*

***Keywords:*** *mining,kc protocol****.***

## I. Introduction

They study here the problem of secure mining of association rules in horizontally partitioned databases. In that setting, there are several sites (or players) that hold homogeneous databases i.e., databases that share the same schema but hold information different entities. 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, that hold in the unified database, while minimizing the information disclosed about the private databases held by those players.The information that we would like to protect in this context is not only individual transactions in the different databases, but also more global information such as what association rules are supported locally in each of those databases.That goal defines a problem of secure multiparty computation. In such problems, there are M players that hold private inputs, x1, . . . , xM, and they wish to securely compute y = f(x1, . . . , xM) for some public function f. If there existed a trusted third party, the players could surrender to him their inputs and he would perform the function evaluation and send to them the resulting output. In the absence of such a trusted third party, it is needed to devise a protocol that the players can run on their own in order to arrive at the required output y. Such a protocol is considered perfectly secure if no player can learn from his view of the protocol more than what he would have learnt in the idealized setting where the computation is carried out by a trusted third party. Yao was the first to propose a generic solution for this problem in the case of two players. Other generic solutions, for the multiparty case, were later proposed 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 T. Tassa is with the Department of Mathematics and Computer Science, The Open University, Raanana, Israel.than the given thresholds s and c, respectively. As the above mentioned generic solutions rely upon a description of the function f as a Boolean circuit, they can be applied only to small inputs and functions which are realizable by simple circuits. In more complex settings, such as ours, other methods are required for carrying out this computation. In such cases,some relaxations of the notion of perfect security might be inevitable when looking for practical protocols, provided that the excess information is deemed benign Kantarcioglu and Clifton studied that problem in and devised a protocol for its solution.The main part of the protocol is a subprotocol for the secure computation of the union of private subsets that are held by the different players. The private subset of a given player, as we explain below, includes the itemsets that are sfrequent in his partial database 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. While such leakage of information renders the protocol not perfectly secure, the perimeter of the excess information is explicitly bounded in and it is argued there that such information leakage is innocuous, whence acceptable from a practical point of view.

Herein 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. In particular, our protocol does not depend on commutative encryption and oblivious transfer what simplifies it significantly and contributes towards much reduced communication and computational costs. While our solution is still not perfectly secure, it leaks excess information only to a small number three of possible coalitions, unlike the protocol of that discloses information also to some single players. In addition, we claim that the excess information 2 that our protocol may leak is less sensitive than the excess information leaked by the protocol of .The protocol that we propose here computes a parameterized family of functions, which we call threshold functions, in which the two extreme cases correspond to the problems of computing the union and intersection of private subsets. Those are in fact general-purpose protocols that can be used in other contexts as well. Another problem of secure multiparty computation that we solve here as part of our discussion is the set inclusion problem; namely, the problem where Alice holds a private subset of some ground set, and Bob holds an element in the ground set, and they wish to determine whether Bobs element is within Alices subset, without revealing to either of them information about the other partys input beyond the above described inclusion.

## II. Problem Statement

**1. EXISTING SYSTEMS:**

Kantarcioglu and Clifton studied that problems 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. (The private subset of a given player, as we explain below, includes the item sets that are s-frequent in his partial database. 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. While such leakage of information renders the protocol not perfectly secure, the perimeter of the excess information is explicitly bounded and it is argued there that such information leakage is innocuous, whence acceptable from a practical point of view.

**2. PROPOSED SYSTEM:**

The protocol that we propose here computes a parameterized family of functions, which we call threshold functions, in which the two extreme cases correspond to the problems of computing the union and intersection of private subsets. Those are in fact general-purpose protocols that can be used in other contexts as well. Another problem of secure multiparty computation that we solve here as part of our discussion is the set inclusion problem; namely, the problem where Alice holds a private subset of some ground set, and Bob holds an element in the ground set, and they wish to determine whether Bobs element is within Alices subset, without revealing to either of them information about the other partys input beyond the above described inclusion.

**Advantages of Proposed System:**

We proposed a protocol for secure mining of association rules in horizontally distributed databases that improves significantly upon the current leading protocol in terms of privacy and efficiency. The main ingredient in our proposed protocol is a novel secure multi-party protocol for computing the union (or intersection) of private subsets that each of the interacting players holds..

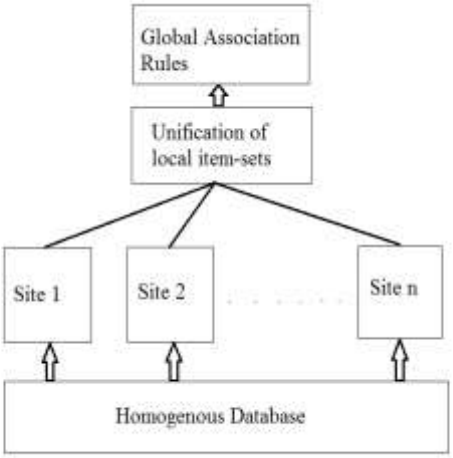


Fig1.homogenous database

## III. CONCLUSION

They proposed a protocol for secure mining of association rules in horizontally distributed databases that improves significantly upon the current leading protocol in terms of privacy and efficiency. One of the main ingredients in their proposed protocol is a novel secure multi-party protocol for computing the union (or intersection) of private subsets that each of the interacting players hold. Another ingredient is a protocol that tests the inclusion of an element held by one player in a subset held by another. Those protocols exploit the fact that the underlying problem is of interest only when the number of players is greater than two. One research problem that this study suggests was described namely, to devise an efficient protocol for inequality verifications that uses the existence of a semihonest third party. Such a protocol might enable to further improve upon the communication and computational costs of the second and third stages of the protocol . Other research problems that this study suggests is the implementation of the techniques presented here to the problem of distributed association rule mining in the vertical setting the problem of mining generalized association rules and the problem of subgroup discovery in horizontally partitioned data..

## Acknowledgements

Inspiration and guidance are invaluable in every aspect of life, especially in the field of academics, which I have received from our beloved and respected guide **Prof.D.A.Vidhate** and Head of Department **Prof.D.A.Vidhate** who has put his careful guidance through which I can complete my project work. Also I want to express my gratitude to his untiring devotion. He undoubtedly is the member of artistic gallery who is masters in all respect. I wish to express my sincere thanks to the departmental staff members for their support. I would also like thank our Principal **Dr. H.N.KUDAL** for his kind support I would also like to thank to my friends for listening to my ideas, asking questions and providing feedback, suggestions and helping for improving my ideas. At last, I would like to express my gratitude towards my parents and my family members who are inspiration of my life.

### REFERENCES

[1] C.N. Modi, U.P. Rao and D.R. Patel [1]To Reduce Data Leakage in Horizontally Distributed Database Using Association Rules 2004.
[2] b) V K S K Sai Vadapalli and G Loshma Privacy Preserving Association Rule Mining in Retail Industries[2] 2009.
[3] c) Ahmed HajYasienPRESERVING PRIVACY IN ASSOCIATION RULE MINING
[4] d) R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In VLDB, pages 487499, 1994. [2]
[5] e) R. Agrawal and R. Srikant. Privacy-preserving data mining. In SIGMOD Conference, pages 439450, 2000.