

# Survey on Privacy-Preserving Association Rule Mining

Madhura Karanjikar<sup>1</sup>, Prof. S. D. Kale<sup>2</sup>

<sup>1,2</sup>Department of Computer Engineering

<sup>1,2</sup>Rajashri Shahu College of Engineering, Savitribai Phule Pune University, Pune, India.

**Abstract**-The primary goal of data mining is to find out the hidden and essential information from large set of data. Such data is collected from various multiple applications or sites, where the privacy and security are the major concern issues. Such data is available in huge amount so it is very difficult to find out the data and relationship among items. For this problem, Association Rule Mining is one of the solutions of data mining techniques, which can efficiently correlate the items. The output of such technique can be used in many real time applications to take the proper decisions. But the data owner, who shares their data for mutual advantages, wants to secure their data in association rule mining process. Because it can reveal the sensitive data, which might be harmful. Therefore it becomes very challenging task to achieve the security of data while mining the knowledge from it. In recent, various methods have been developed, represents the core idea of privacy preserving association rule mining on shared data. Such methods are comparatively examined in this paper, on the basis of technique, advantages and disadvantages.

**Keywords**-Privacy-preserving data mining, association rules mining, frequent itemset mining.

## I. INTRODUCTION

Data mining innovation has developed as methods for identifying patterns and patterns from large quantities of data. Mining incorporates different algorithms, for example, clustering, classification, association rule mining and sequence detection and succession location. Traditionally, every one of these algorithms have been developed within a centralized model, with all information being assembled into a central site, and algorithms being run against that information.

Association rule mining discovers all rules in the databases that fulfill some minimum support and minimum confidence requirements. Numerous algorithms are utilized to improve the privacy and security of information. By vertically partitioned, we imply that every site contains a some elements of a transaction. Utilizing the traditional —market basket illustration, one site may contain basic grocery purchases, while another has clothing purchases. Utilizing a key, for example, credit card number and date, we can join these to recognize relationships between purchases of clothing and groceries. Notwithstanding, this reveals the individual

purchases at every site, possibly violating consumer privacy agreements.

There are more sensible illustrations. In the sub-assembly manufacturing process, distinctive manufacturers give components of the finished product. Cars incorporate several subcomponents; tires, electrical equipment, and so forth; made by independent producers. Once more, authors have proprietary data collected by several parties, with a single key joining all the data sets, where mining would help detect/predict malfunctions. The recent trouble between Ford Motor and Firestone Tire give a real-life illustration. Ford Explorers with Firestone tires from a particular factory had tread separation issues in specific circumstances, resulting in 800 injuries. Since the tires did not have issues on different vehicles, and different tires on Ford Explorers did not represent an issue, neither one of the sides felt responsible. The delay in recognizing the genuine issue prompted to a public relations nightmare and the eventual replacement of 14.1 million tires. Many of these were probably fine – Ford Explorers represented just 6.5 million of the replaced tires. Manufacturers had their own data – early era of association rules based on all of the data may have enabled Ford and Firestone to resolve the safety problem before it became a public relations nightmare.

Casually, the issue is to mine association rules across two databases, where the columns in the table are at various sites, splitting each row. One database is assigned the essential, and is the initiator of the protocol. The other database is the responder. There is a join key present in both databases. The rest of the attributes are present in one database or the other, but not both. The goal is to find association rules involving attributes other than the join key.

In business application associations are especially concerned with privacy issue. Most association gather their data from individual party and their particular need. They find it essential to share data to each other. In such cases every unit need to make sure that privacy of individual party must not be disregarded. If Government wants to survey about some medical disease. Individual hospital don't want to reveal personal information of its patient, it is against law. This information is regarded private and we want to avoid exposing confidential information of patient. So we must design the

system which solve this issue and provide security of data owner.

## II. LITERATURE REVIEW

In paper [1] authors proposed a privacy-preserving outsourced frequent item set mining solution for vertically partitioned databases. This allows the data owners to outsource mining task on their joint data in a privacy-preserving manner. Based on this solution, authors built a privacy-preserving outsourced association rule mining solution for vertically partitioned databases.

In paper [2] authors we presented an efficient result integrity verification approach that can provide deterministic guarantee for outsourced frequent item set mining. The key idea of the approach is to construct cryptographic proofs of all (in) frequent item sets. They discussed how to optimize the number of proofs to improve the performance.

In paper [3] authors tackled issues of privacy preserving association rule mining are addressed here. In particular, privacy preserving algorithms over horizontal and vertical partitioned databases are discussed and results are compared.

In paper [4] authors have proposed CRYPPAR, a full-fledged framework for privacy preserving association rule mining based on cryptographic approach over vertically partitioned data. We also conducted empirical evaluation on CRYPPAR. The results indicated that the method of building it is efficient and may become a general way to do PPDM in real life.

In this paper [5], authors concentrate on the situation when the database is distributed vertically, and propose an effective multi-party protocol for evaluating itemsets that preserves the privacy of the individual parties. The proposed protocol is algebraic and recursive in nature, and depends on a recently proposed two-party protocol for the same issue. It is not only appeared to be much faster than similar protocols, additionally more secure. They likewise show a variation of the protocol that is resistant to collusion among parties.

Privacy considerations [6] often constrain data mining projects. This paper addresses the issue of association rule mining where transactions are distributed across sources. Every site holds some attributes of each transaction, and the sites wish to collaborate to identify globally valid association rules. Notwithstanding, the destinations must not reveal individual transaction information. Authors show a two-party algorithm for efficiently discovering frequent itemsets with minimum support levels, without either site revealing individual transaction values.

This paper [7] proposes a protocol for secure mining of association rules in horizontally distributed databases. The present leading protocol is that of Kantarcioglu and Clifton. Their protocol, like theirs, depends on the Fast Distributed Mining (FDM) algorithm of Cheung et al., which is an unsecured distributed variant of the Apriori algorithm. The fundamental ingredients in their protocol are two novel secure multi-party algorithms — one that computes the union of private subsets that each of the associating players hold, and another that tests the inclusion of an element held by one player in a subset held by another.

Table 1. Survey Table

Sr. No.	Title	Paper Details	Method Used	Advantages	Disadvantages
1.	Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases	Propose a cloud-aided frequent itemset mining solution, which is used to build an association rule mining solution.	privacy-preserving association rule mining and frequent itemset mining	These solutions are very efficient and suitable to be used by data owners wishing to outsource their databases to the cloud but require a high level of privacy without compromising on performance.	Not focus on demonstrating the utility of the proposed homomorphic encryption scheme and outsourced comparison scheme in other settings.
2.	Integrity Verification of Outsourced Frequent Itemset Mining with Deterministic Guarantee	Focus on the integrity and verification issue in UFI mining problem during outsourcing process, i.e., how the data owner verifies the	UFI Mining Methods	Reduce the verification to the expected support verification and provide good probabilistic guarantees about the correctness of verification.	Time consuming process.

		mining results.			
3.	Privacy preserving mining of Association Rules on horizontally and vertically partitioned data: A review paper	Addresses the problem of association rule mining where transactions are distributed across sources.	Privacy preserving association rule mining	Achieve good individual security with communication cost comparable to that required to build a centralized data warehouse.	This work is limited to boolean association rule mining. Non-categorical attributes and quantitative association rule mining are significantly more complex problems
4.	CRYP PAR: An efficient framework for privacy preserving association rule mining over vertically partitioned data	Propose CRYP PAR, a novel, full-fledged framework for privacy preserving association rule mining based on a cryptographic approach.	CRYP PAR framework	It is efficient, secure method to perform association rule mining over vertically partitioned data.	Need to improve accuracy.
5.	Towards a Collusion-Resistant Algebraic Multi-Party Protocol for Privacy-Preserving Association Rule Mining in Vertically Partitioned Data	Focus on the case when the database is distributed vertically, and propose an efficient multi-party protocol for evaluating itemsets that preserves the privacy of the individual parties.	Multiparty Protocol	Preserves the privacy of the individual parties.	Not design and test parallel variants of the proposed multi-party protocol.
6.	Privacy preserving association rule mining in vertically partitioned data	Addresses the problem of association rule mining where transactions are distributed across sources.	Two-party algorithm	Given a privacy preserving scalar product protocol, and an efficient protocol for computing scalar product while preserving privacy of the individual values.	Privacy issues face other types of data mining, such as Clustering, Classification, and Sequence Detection.
7.	Secure mining of association rules in horizontally distributed databases	Propose a protocol for secure mining of association rules in horizontally distributed databases.	secure mining of association rules	It is simpler and is significantly more efficient in terms of communication rounds, communication cost and computational cost.	The problem of mining generalized association rules.

**III. PROPOSE SYSTEM**

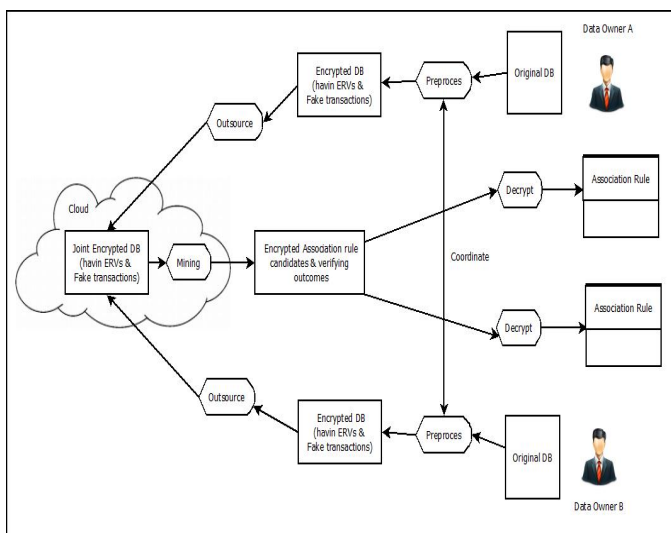


Fig 1. Propose System

The system model is included two or more data owners and a cloud. Every data owners have a private database, and the data owners encrypt their private databases prior to outsourcing the encrypted databases to the cloud. Data owners can also request the cloud to mine association rules or frequent itemsets from the joint database for their behalf. The (honest but curious) cloud is tasked with the compiling and storing of databases got from various data owners, the mining of association rules or frequent itemsets for data owners, and the sending of the mining result to relevant data owners.

**IV. CONCLUSION**

From this survey we conclude that, Privacy becomes great challenge for researchers while applying data mining technique on data in various real time applications. Because data owner share their data for mining the knowledge to obtain the mutual benefits. In this survey we studied various recent

approaches, which are based on association rule mining for knowledge discovery. Such methods also preserving the privacy of distributed data. Some methods perform the association rule mining on horizontal and vertical distributed model of data. Efficiency and accuracy are challenging task along with privacy. We also compare recent approaches along with their limitations, which may be useful for further research study.

## REFERENCES

- [1] L. Li, R. Lu, K. K. R. Choo, A. Datta and J. Shao, "Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 8, pp. 1847-1861, Aug. 2016.
- [2] B. Dong, R. Liu and W. H. Wang, "Integrity Verification of Outsourced Frequent Itemset Mining with Deterministic Guarantee," 2013 IEEE 13th International Conference on Data Mining, Dallas, TX, 2013, pp. 1025-1030.
- [3] M. N. Kumbhar and R. Kharat, "Privacy preserving mining of Association Rules on horizontally and vertically partitioned data: A review paper," Hybrid Intelligent Systems (HIS), 2012 12th International Conference on, Pune, 2012, pp. 231-235.
- [4] D. H. Tran, W. K. Ng and W. Zha, "CRYPPAR: An efficient framework for privacy preserving association rule mining over vertically partitioned data," TENCON 2009 - 2009 IEEE Region 10 Conference, Singapore, 2009, pp. 1-6.
- [5] D. Trinca and S. Rajasekaran, "Towards a Collusion-Resistant Algebraic Multi-Party Protocol for Privacy-Preserving Association Rule Mining in Vertically Partitioned Data," 2007 IEEE International Performance, Computing, and Communications Conference, New Orleans, LA, 2007, pp. 402-409.
- [6] Vaidya, Jaideep, and Chris Clifton. "Privacy preserving association rule mining in vertically partitioned data." Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002.
- [7] Tassa, Tamir. "Secure mining of association rules in horizontally distributed databases." IEEE Transactions on Knowledge and Data Engineering 26.4 (2014): 970-983.