**Available Online at www.ijcsmc.com**

**International Journal of Computer Science and Mobile Computing**

**A Monthly Journal of Computer Science and Information Technology**

**ISSN 2320–088X**

***IJCSMC, Vol. 3, Issue. 10, October 2014, pg.791 – 798***

**RESEARCH ARTICLE**

# PRIVACY PRESERVING ASSOCIATION RULE MINING

**Ms. V.Uma Rani[1], Dr. M.Srinivasa Rao[2], N.Nikhil Krishna[3]**

[1] Assistant Professor, School of IT, JNTU-Hyderabad, Telangana, India

[2]Professor, School of IT, JNTU-Hyderabad, Telangana, India

[3]PG Scholar, School of IT, JNTU-Hyderabad, Telangana, India

***ABSTRACT: Introduction of high end technologies used by most of the leading software companies needs a keen focus to be kept on Daas (Data Mining as a Service). The data owner transforms data to protect corporate privacy and ships it to the server, server receives mining queries and true patterns are recovered from the extracted patterns which are received from the server. A study is done on the problem of outsourcing the association rule mining task within a corporate privacy-preserving framework. An attack model based on background knowledge and scheme is devised for privacy preserving outsourced mining. The scheme introduced ensures that from at least k-1 other transformed items each transformed item is indistinguishable with respect to the attacker's background knowledge. Results show that the techniques used provide scalability, effectiveness and privacy protection.***

## INTRODUCTION

Communications to all levels of business, social-network transactions and communications are improving. A single entity must process number of transactions that becomes quite prohibitive for a single computer terminal. As such, the offloading and outsourcing of data mining tasks to specialized terminal servers becomes necessary.

 

Association rules or Frequent Item-set mining are the most commonly used mining methods to maintain large datasets. At this offloaded servers, various techniques such as DES algorithm and the building of Frequent Pattern trees etc. The owner get backs the results of all mined rules and their corresponding support values as mined results.

In this process, because the appropriate terminals are semi-honest in nature, this makes a privacy and security leakage which is problem to the data owner. The 3rd party server knows the support values of each item-set and data set in every transaction. Although data encryption is applied, the process is affected to frequency attacks, and leads to a loss of data. The server makes changes to deduct the semantic meanings of each encrypted item-set and data set.

The owner generally encrypts or performs a transformation on the item-sets or data set from a transactional process, and decrypts all the results of frequent items in the post process The general architecture is as in Figure 1 below:

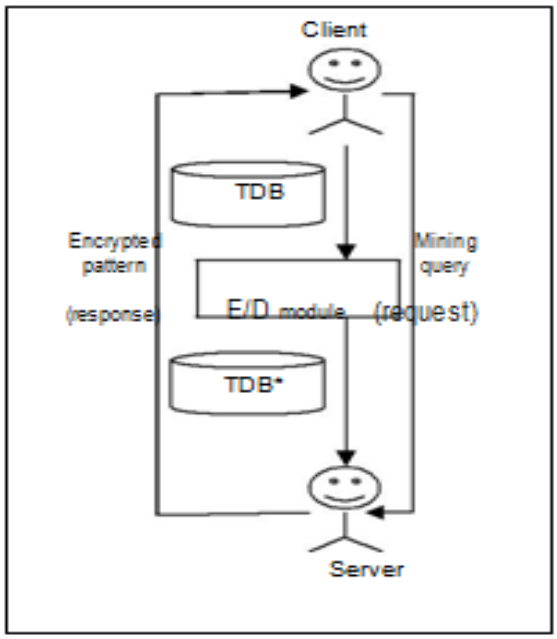


Figure 1: General Architecture of Outsourced 3rd Party Association Rules Mining.

For example, in the mining of data involving faculty members at a university, if an attacker knows the proportions of male or female professors, or technical and non- technical professors, he/she is then able to make deductions on the references and meaning of the data in the mined item-sets and data sets. Clearly, privacy and security systems are required to be in place in such outsourcing activities. This paper attempts to contribute basic body of knowledge by making a case for the distribution of outsourced association rules mining.

 

## RELATED WORK

Privacy preserving data mining is important in data mining research. The main reasons are needs of private data, improved technology, allowing ease of storage, transfer, access, manipulation of distributed data. To save it from attacks and unauthorized access to get the knowledge many different techniques have been used by researchers. The attacker may have information related to the dataset. The important difference between our scenario and other scenarios is that, the results as well as sensitive attributes are not intended to be open to others. There are many techniques that are prevalent for privacy preserving data mining.

The literature gives a gist of the methods that could be used for privacy preserving data mining and extensive work has been done out of that major emphasis is only on anonymization perturbation and many others. Y-H Wu et al. proposed method to reduce the side effects in sanitized database. They present a novel approach that strategically modifies a few transactions in the transaction database to decrease the supports or confidences of sensitive rules without producing the side effects.

**Issue:** The most important issue is of maintaining privacy of not only the individual but also important association rules that a corporate or a company has through his transaction data base or warehouse which can help them to transform their business and help in maintaining competitive edge on their competitors.

**Space Complexity:** The space complexity is high because of following reasons: They are using fake transactions that are increasing the data base size which is not useful as the data set is stored in server and needs to be used through internet. Maintaining table which stores data about the perturbation done on both the sides (from where the dataset is uploaded and where the data set is downloaded).

**Time Complexity:** Time complexity is also high

## ASSOCIATION RULE

The support is a measure of the frequency of a rule and the confidence is a measure of the strength of the relation between set based and item based. Support function "s" of an association rule is defined as the percentage/fraction of records that contain (A ∪ B) to the total number of records in the database. Apriori is a breadth-first, level-wise algorithm is used to implement the association rule. This algorithm have a main steps follow : Exploits monotonicity as much as possible, Search Space is pass across bottom-up, level by level, Support of an item set is only counted in the database if all its subsets were frequent.

 

Apriori algorithm approach is A rule X=>Y satisfies min sup an sup (X∩Y) / sup(X) >= minconf. Hence, first find all item set I so that. sup (I) >= min sup. Then for every frequent I: Split I in all possible ways X∩Y and Test if sup (X∩Y) / sup(X) >= min conf. Data mining, association rules. In privacy preserving are useful for analyzing and predicting customer behavior and pattern of purchase. They play an important part in market analysis, data of basket shopping, product clustering, classification, and catalog design and store layout. Similarly in this work Association rules are generated from the pre-processed dataset. These rules are generated by the Aprior Algorithm. Now, those rules whose support value is above the minimum support value are to be hidden. Here for hiding these rules, manipulation is done in transaction where other item is inserted into the transaction.

SUB SYSTEMS:

1. The Pattern Mining Task

2. Privacy Model

3. Attack Model

i. Item-based attack

ii. Set –based attack

4. Encryption/Decryption Scheme

THE PATTERN MINING TASK

The reader is assumed to be familiar with the basics of association rule mining. We let I = i1... in be the set of items and D = t1... tm a transaction database (TDB) of transactions, each of which is a set of items. We denote the support of an item set S ⊆ I as supp D(S) and the frequency by freq D(S). Recall, freq D(S) =supp D(S)/|D|. For each item i, supp D (I) and freq D (I) denote respectively the individual support and frequency of I. The function supp D (.), focused over items, is also called the item support table. The popular frequent pattern mining problem: given a TDB D and a support threshold σ, search all item sets whose support in D is at least σ. In this paper, we limit ourselves to the study of a (corporate) privacy preserving outsourcing framework for frequent pattern mining.

 

PRIVACY MODEL

We let D denote the original TDB that the owner has to protect the identity of individual items, the owner put in an encryption function to D and transforms it to D∗, the encrypted database. We refer to items in D as plain items and items in D∗ as cipher items. The idiom item shall mean plain item by default. The idea of plain item sets, bare transactions, bare patterns, and their cipher counterparts are well defined in the understandable way. We use I to denote the set of plain items and E to refer to the set of cipher items.

ATTACK MODEL

The server or an intruder who gains access to it may possess some background knowledge using which they can on the encrypted database D∗. We generically refer to of this agent as a attacker. We adopt a conservative model and assume that the attacker knows exactly the set of (plain) items I in the original transaction database D and their true supports.

We assume the service provider (who can be an attacker) is semi-honest in the sense that although he does not know the details of our encryption algorithm, he can be exited and thus can use his skills and knowledge to make inferences on the encrypted transactions. We also predict that the attacker always returns (encrypted) item sets together with their exact support. The data owner (i.e., the corporate) considers the true identity of:

(1) Every cipher item,

(2) Every cipher transaction, and

(3) Every cipher frequent pattern as the intellectual property which should be protected. We consider the following attack model:

• Item-based attack:

The semi honest service provider can attack the owners data depend upon the single item identity.

• Set-based attack:

The service provider attacks the owners data depend upon the many item identities. By this method the attacker can easily attacks the data correctly but they can't use that data because that data's are in cipher text form. Data owners are using the separate E/D Module.

 

ENCRYPTION/DECRYPTION SCHEME:

Encryption:

An encryption scheme is introduced which transforms a TDB D into its encrypted version D∗. Our scheme is parametric w.r.t. k > 0 and consists of three main steps: (1) using 1-1 substitution ciphers for each plain item; (2) using a specific item k-grouping method; (3) using a method for adding new fake transactions for achieving k-privacy. The constructed fake transactions are added to D (once items are replaced by cipher items) to form D and transmitted to the server.

Decryption:

When the client requests the execution of a pattern mining query to the server, showing a minimum support threshold σ, the server send back the computed frequent patterns from D∗. Clearly, for every item set S and its corresponding cipher item set E, we have that supp D(S) ≤ supp D_(E).For every cipher pattern E returned by the server together with supp D_(E), the E/D module restores the corresponding plain pattern S. It needs to remake the exact support of S in D and decide on this basis if S is a continuous pattern. To obtain this goal, the E/D module adjusts the support of E by removing the effect of the fake transactions. Supp D(S) = supp D_(E)−supp D_\D(E). This follows from the fact that support of an item set is additive over a disjoint union of transaction sets that is item based or set based. Finally, the "S" pattern with adjusted support is kept in the output if supp D(S) ≥ σ. The calculation of supp D_\D(E) is performed by the E/D module using the synopsis of the fake transactions in D∗ \ D.

**CONCLUSION**

In this paper, the problem of (corporate) privacy preserving mining of frequent patterns (from which association rules can be easily computed) on an encrypted outsourced transaction database is focused or studied. A conservative model is assumed, where the adversary knows the domain of items and their exact frequency and can use this knowledge to identify cipher items and cipher item sets. An encryption scheme called Rob Frugal is proposed which is based on 1-1 substitution ciphers for items and adding fake transactions to make each cipher item share the same frequency as ≥ k − 1 others.

 

**REFERENCES**

[1] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In *VLDB*, pages 487–499, 1994.

[2] Rakesh Agrawal and Ramakrishnan Srikant. Privacy-preserving data mining. In *SIGMOD*, pages 439–450, 2000.

[3] Gilburd B, Schuste A, and Wolff R. k-ttp: A new privacy model for large scale distributed environments. In *VLDB*, pages 563 – 568, 2005.

[4] Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal. Market oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. In *HPCC*, 2008.

[5] Kun-Ta Chuang, Jiun-Long Huang, and Ming-Syan Chen. Power law relationship and self-similarity in the item set support distribution: analysis and applications. *VLDB Journal*, 17(5):1121–1141, 2008.

[6] Valentina Carina, Sabrina De Capitani di Vimercati, Sara Foresti, and Pierangela Samarati.

*K-anonymity*. In *Secure Data Management in Decentralized Systems*, pages 323–353. 2007.

[7] Chris Clifton, Murat Kantarcioglu, and Jaideep Vaidya. Defining privacy for data mining. In *National Science Foundation Workshop on Next Generation Data Mining*, pages 126–133, 2002.

 

**AUTHOR PROFILES:**

| | |
|---|---|
|  | **Ms Uma Rani** is presently working as **Assistant Professor** in School of IT, JNTU Hyderabad. She has more than twelve years of experience. Her area of interest is data mining and information security. |
|  | **Dr. M Srinivasa Rao** is former director of School of IT and he is currently working as professor in it. His articles and publications are published all over the world.<br>His area of interest includes Web Technologies, Artificial Neural Networks, Data mining, Software Testing Tools and IT workshop. |
|  | **N. Nikhil Krishna** is currently pursuing his **M.Tech (Software Engineering)** in School of IT, JNTU-Hyderabad. He did his B.Tech in Computer Science &Engineering from Murthy Institute of Technology and Sciences Engineering and Technology, Hyderabad. His research area interest includes data mining, information security and cloud computing |

 