IJCSN International Journal of Computer Science and Network, Volume 5, Issue 5, October 2016
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 1.02

Secure Extraction of Association Rules in Parallel Disseminated Data

[1] Vinay Bamane, [2] Vipul Bag

[1] Department of Computer Science,
N.K.Orchid College of Engineering Solapur, University Solapur,
Maharashtra, India

[2] Associate Professor Department of Computer Science,
N.K.Orchid College of Engineering Solapur, Solapur University
Solapur, Maharashtra, India

Abstract - In horizontally distributed databases for secure mining of association rules. unsecured distributed version of the Apriori algorithm overcome the problem of FDP Using fast Distributed Mining (FDM) algorithm. There are two rules, one that computes the union of private subsets that each of the interacting group of actors hold, and another that tests the inclusion of an element held by one player in a subset held by another. This paper proposed some rule that offers improved privacy with respect to the proposed rule. In addition, it is simpler and is significantly more efficient in term of communication rounds, communication cost.

Keywords - *Data Mining, FDM Distributed Computation, Frequent Itemsets, Association Rule, Privacy Preserving.*

1. Introduction

The problem of mining safety regulations in split horizontal database. In that setting, there are many sites (or players) that homogeneous database, ie, database schema that is part of the same, but different institutions are holding hold information. The goal at least minimal support and confidence , all with minimal support given to association rules is to find the size and confidence level, integrated database that will hold information held by those private databases revealed about the players at least. This target defines a problem of secure multiparty computation. Yao first two players in Clifton Kantarcioglu and to propose solutions to this problem for a normal [2] and has prepared a proposed rule for the solution of that problem was studied.

The main part of the proposed rule for a given player's personal subset, as we explain below, that are organized by different players, private safe subset of the Union for the calculation is a sub-proposed rule. (Item sets that are s included in the partial database -frequent.) Is the most expensive part of the proposed rule and its implementation, such exchangeable encryption, oblivious transfer, and depends on cryptographic primitives such as hash functions. The proposed rule also other players out of my database of information on the proposed rule can input their own by what lies beyond the final production and is the only part. Such leakage of information renders the proposed rule is not completely safe, 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 proposed rule for the secure computation of the union of private subsets. The proposed proposed rule improves upon that in [1] in terms of simplicity and efficiency as well as privacy. In particular, our proposed rule does not depend on commutative encryption and oblivious transfer (what simplifies it significantly and contributes towards much Communication and reduce the computational cost). While our solution is still not completely safe, for more information, see http://www.ieee.org/publications_standards/publications/rights/index.html leaks.

The only possible to combine a small number (three) for additional information contrary to proposed rule [3] for a few players to disclose information.

Furthermore, we claim that the additional information that may leak our proposed rule additional information leaked by the project is less sensitive than excess information leaked by the project.

2. Literature Review

2.1 Secure Mining of Association Rules in Horizontally Distribute Databases

We propose a proposed rule for secure mining of rules in horizontally distributed databases. The current leading proposed rule is that of Kantarcioglu and Clifton. Our proposed rule, like theirs, is based on the Fast Distributed Mining (FDM)[2] algorithm of Cheung et al. which is an unsecured distributed version of the Apriori algorithm. The main ingredients in our proposed rule are two novel secure multi-party algorithms — one that computes the union of private subsets that each of the interacting players hold, and another that tests the inclusion of an element held by one player in a subset held by another. Our proposed rule offers enhanced privacy with respect to the proposed rule. In addition, it is simpler and is significantly more efficient in terms of communication rounds, communication cost and computational cost. We study here the problem of secure mining of association rules[5] 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 on 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.

2.2 Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data

Data mining can This article addresses the rules of safe mining data associated horizontal partition. The method includes encryption technology to minimize the sharing of information, while adding little overhead to the mining task. Data mining technology has been determined from large data patterns and trends as a means. Data mining and data warehousing to go hand in hand: most tools operate, all collected data to a central site, and then run the algorithms of these data.extract important knowledge from large data collections – but sometimes these collections are split among various parties. Privacy concerns may prevent the parties from directly sharing the data[4], and some types of information about the data. However, privacy concerns can prevent building a centralized warehouse – data may be distributed among several custodians, none of which are allowed to transfer their data to another site. This paper addresses the problem of computing association rules within such a scenario. We assume homogeneous databases[9]: All sites have the same schema, but each site has information on different entities. The goal is to produce association rules that hold globally, while limiting the information shared about each site.

2.3 Anonymization of Centralized and Distributed Social Networks by Sequential Clustering

Efficient Computation anonymizations problem partitioned database. Given shared across multiple sites, horizontal or vertical, we have designed a secure distributed algorithms, so that different locations get k Check database - Anonymous ℓ - various data bases of their union not to divulge sensitive information. Our algorithm is based on [7] sequential algorithms, and provide anonymizations Gongyongshiye than other anonymization algorithms, especially those so far implemented in a distributed environment significantly better. Our algorithm can be applied to a wide variety of technical and practical measures and a number of sites. While previous cryptographic algorithms distributed algorithms rely on expensive, password assume that our solution is surprisingly minimal.

3. Existing System

The problem of secure mining of association rules in horizontally partitioned databases. In that setting, there are several sites (or links) that hold homogeneous databases, I.e. databases that share the same schema but hold information on different entities. The inputs are the partial databases[6], and the required output is the list of association rules that hold in the unified database with support and confidence no smaller.

Disadvantages of Existing System:

The inadequate security, privacy and current system disadvantages: confidence in the existing system is not in the database simplicity and functionality. The solution is still not perfectly safe, additional information that lead to possible leaks of the proposed rule, Only a small number (three) of unlike some of the same players also released information. is less sensitive than the additional proposed rule could leak information leaked by our proposed rule.

4. Proposed System

- **Proposed rule** works that we propose here, we work threshold that has two serious problems of computing the intersection of Union and private subset of analog phone calculates a parameterized family. The fact that those general purpose proposed rule used in other contexts as well, that

can be. [7] As part of our discussion here is that we solve the problem, another set of secure multi-party computation problem of inclusion; Namely, the problem of where to set some ground Alice holds a private subset, and Bob ground holds an element in the set, and to determine whether they are within the subset Alice Bob elements want any of them to another party investments without revealing the information described above for inclusion beyond. The module holder applying for the loan. He / she account number, bank details like the name of the branch are and we also have car loans, home loans, business loans, etc. are providing loan categories.; Details will be sent to bank loans. Apriori [1] is designed to operate on databases containing transactions. Apriori algorithm purpose is to find associations between different sets of data, it is sometimes as "market basket analysis" is called. Each set of data and a number of items is called a transaction. Apriori's production rules that tell us how often the items are contained in the data set is set.

The proposed rule works we propose here, we work threshold that has two serious problems of computing the intersection of Union and private subset of analog phone calculates a Prmatarizaid family. In other contexts the fact that general-purpose proposed rule that can be used as well in. [Vii] as part of our discussion we have to solve the problem is another set of secure multi-party computation problems to be included; That is, the problem of where to set some ground Alice holds a private subset, and Bob ground holds an element in the set, and to determine what would Bob elements within Alice subset of them to another party investments without revealing the information described above for inclusion beyond. The module runs the application for the account holder. He / she account number, bank details like the name of the branch are and we run the car, Planet Run, run, etc. run business categories are provided.; Details will be sent to the concerned bank run.

Apriori algorithm Apriori [one] is designed to operate on databases containing transactions. Apriori algorithm purpose is to find associations between different sets of returns, market basket analysis, as it is sometimes called ". Deta ke pratyek set madon kee ek sankhya hai aur ek saude kaha jaata hai. Apriori ke utpaadan the rules tell us how often the items are contained in the set is the set of returns.

❖ Google Translate for Business: Translator Tulkitvebsite Tronsletrglobl Market Finder

❖ Definition Of Apriori Algorithm

- The Apriori Algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules.
- Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation, and groups of candidates are tested against the data.
- Apriori is designed to operate on database containing transactions (for example, collections of items bought by customers, or details of a website frequentation).

Algorithm - Fast Distributed Mining (FDM)

The FDM Algorithm proceeds as follows:

- Initialization
- Candidate Sets Generation
- Local Pruning
- Unifying the candidate item sets
- Computing local supports
- Broadcast Mining Results

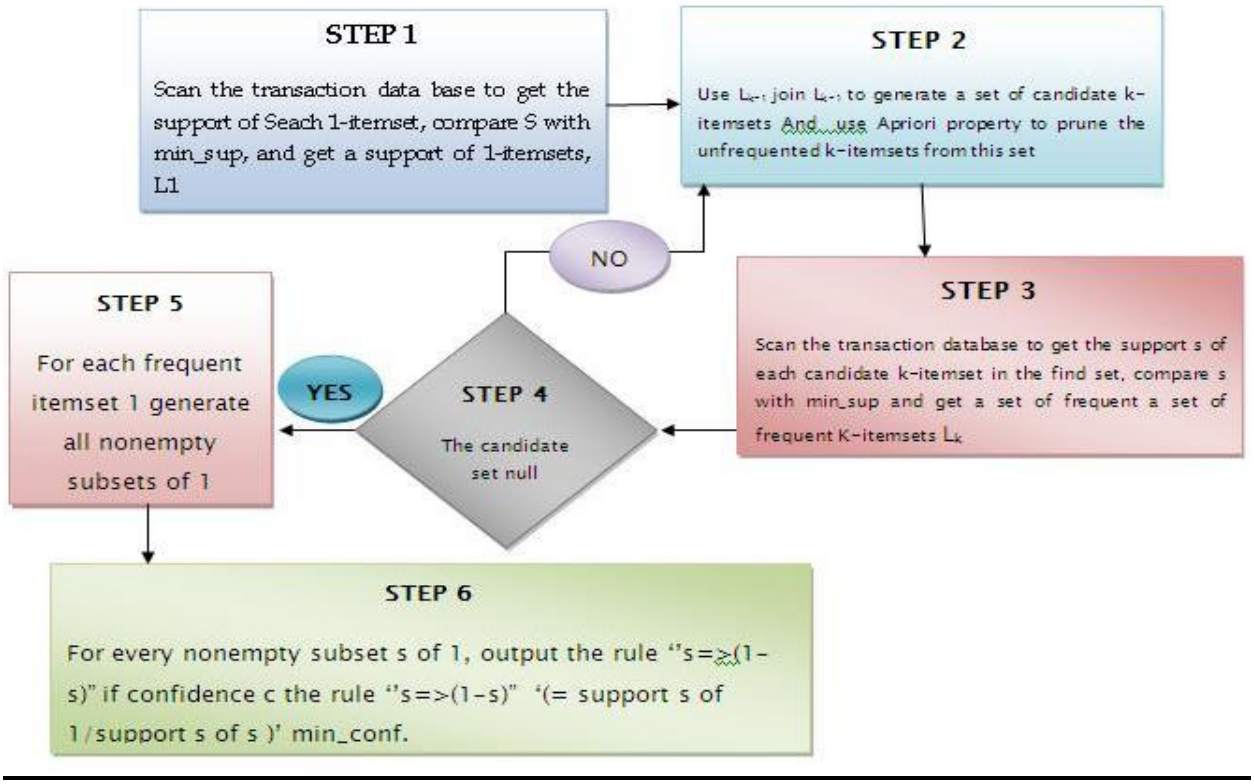

Figure 1. Architecture of advanced FDP

4.1 Modules

1. Privacy Preserving Data Mining
2. Distributed Computation
3. Frequent Item sets
4. Association Rules

4.2 Modules Description

4.2.1.Privacy Preserving Data Mining

Wherein the data owners and data miners are two different entities , and another , in which the data is distributed between several parties [11] aims to unify the corpus of data on who they are holding joint implementation data mining . In the first set , the goal is to protect the data record from the data miners. Therefore, the data of the owner intended .data prior to its release. The main approach in this context is to apply data perturbation. The idea is that. Computation and communication costs versus the number of transactions N the perturbed data can be used to infer general trends in the data, without revealing original record information. In the second setting, the goal is to perform data mining[10] while protecting the data records of each of the data owners from the other data owners. This is a problem of secure multiparty computation. The usual approach here is cryptographic rather than probabilistic.

4.2.2. Distributed Computation

The part of the safety performance achieved in the first embodiment 13] we use to perform protocol UNIFI -KC, wherein the switching password is 1024 RSA In the second implementation (note FDM) unified step, we use our agreement UNIFI, wherein the hash function is keyed HMAC.

In both embodiments, the step we have achieved in a safe manner described later FDM algorithm 5. We tested two implementations with respect to the three measures:

1) over the entire agreement of all players (FDMKC and FDM) total computation time. The measures include a priori calculation time, and to recognize the global S-frequent item sets, as described later in.

2) Total computation time of the unification proposed rules only (UNIFI-KC and UNIFI) over all players.

3) Total message size. We ran three experiment sets, where each set tested the dependence of the above measures on a different parameter: • *N* — the number of transactions in the unified database.

4.2.3. Frequent Itemsets

They believe that two possible settings. If the desired output includes all global S- frequent item sets , as well as their support of the size , the value of Δ (x) can be displayed for all .Here we describe the proposed Kantarcioglu and Clifton solutions. SNM - in this case , these values can be summed security protocol , which Pm private addend is suppm (x) is calculated . Even more interesting is set , however, is not the size of a part of the support in the desired output . We continue to discuss

4.2.4. Association Rules

The main ingredient in the proposed rules we proposed in a new security rules for multi-party proposed to calculate private subset union (or intersection) , each player holds interactive .Once set Fs all S- frequent item sets are found , we will continue to look for all the (S, C) -association rule (rule , at least SN support and confidence , at least C). In order to derive all from fs (S, C) , we rely on simple lemma effective way -association rules.

5. Experimental Setup and Result

The databases that we used in our experimental evaluation are synthetic databases that were generated using the same techniques that were introduced in [1] and then used also in subsequent studies such as [8]. gives the parameter values that were used in generating the synthetic database. The reader is referred to for a description of the synthetic generation method and the meaning of each of those parameters. Examples of Bank Transaction.

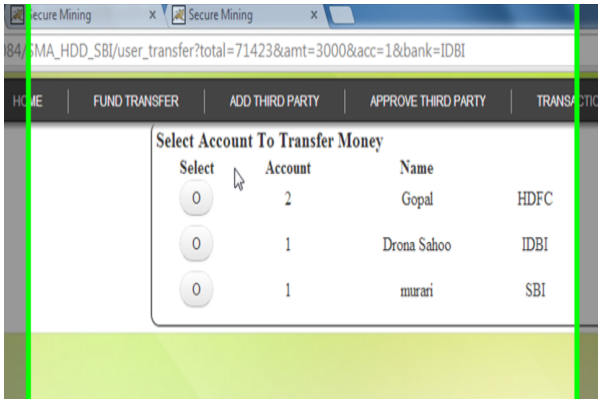

Figure 2. Select Account to Transfer Money

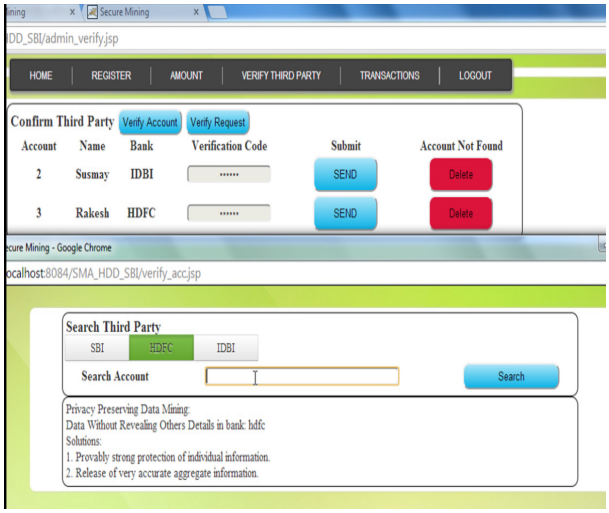

Figure 3. Search Third Party

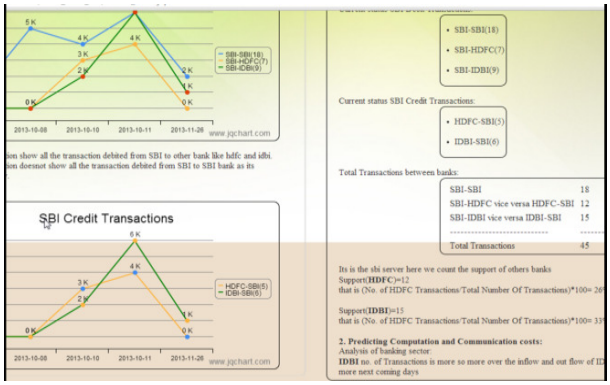

Figure 4. Debit and Credit Transaction

6. Conclusion

This paper proposed a new rule for secure mining in horizontally distributed databases that improves drastically upon the current leading procedure in terms of retreat and competence.

References

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

[2] 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,Sept. 2004.

[3] D. Beaver, S. Micali, and P. Rogaway, "The Round Complexity of Secure Proposed rules," Proc. 22nd Ann. ACM Symp. Theory of Computing (STOC), pp. 503-513, 1990.

[4] M. Bellare, R. Canetti, and H. Krawczyk, "Keying Hash Functions for Message Authentication," Proc. 16th Ann. Int'l Cryptology Conf. Advances in Cryptology (Crypto), pp. 1-15, 1996.

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

[6] J.C. Benaloh, "Secret Sharing Homomorphisms: Keeping Shares of a Secret Secret," Proc. Advances in Cryptology (Crypto), pp. 251-260, 1986.

[7] J. Brickell and V. Shmatikov, "Privacy-Preserving Graph Algorithms in the Semi-Honest Model," Proc. 11th Int'l Conf. Theory and Application of Cryptology and Information Security (ASIACRYPT), pp. 236-252, 2005.

[8] D.W.L. Cheung, J. Han, V.T.Y. Ng, A.W.C. Fu, and Y. Fu, "A Fast Distributed Algorithm for Mining Association Rules," Proc. Fourth Int'l Conf. Parallel and Distributed Information Systems (PDIS), pp. 31-42, 1996.

[9] D.W.L Cheung, V.T.Y. Ng, A.W.C. Fu, and Y. Fu, "Efficient Mining of Association Rules in Distributed Databases," IEEE Trans.Knowledge and Data Eng., vol. 8, no. 6, Dec. 1996.

[10] T. ElGamal, "A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms," IEEE Trans. Information Theory,vol. IT-31, no. 4, July 1985.

[11] A.V. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy Preserving Mining of Association Rules," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD),pp. 217-228, 2002.

[12] R. Fagin, M. Naor, and P. Winkler, "Comparing Information without Leaking It," Comm. ACM, vol. 39, pp. 77-85, 1996.

[13] M. Freedman, Y. Ishai, B. Pinkas, and O. Reingold, "Keyword Search and Oblivious Pseudorandom Functions," Proc. Second Int'l Conf. Theory of Cryptography (TCC), pp. 303-324, 2005.

Author:

Mr. Vinay Bamane received B.E degree in Computer Science and Engineering from Dr.Babasaheb Ambedkar, Maratwada University of Aurangabad, Post Graduation diploma in CDAC (Head Office) from Barti Pune, Master of Business Administration in Human Resource from Y.C.M.O.U Nashik and pursuing the M.E. degree in Computer Science and Engineering in Nagesh Karajagi Orchid College of Engineering & Technology, Solapur, India. He is doing her dissertation work under the guidance of Mr. Vipul Bag Associate Professorat Nagesh Karajagi Orchid College of Engg. & Technology, Solapur, Maharashtra, India.

Mr. Vipul Bag is working as Assosiate Professor in Department of Computer Science and Engineering in NK Orchid College of Engineering and Technology, Solapur, Maharashtra, India. He has 17 years of teaching experience. He has co-authored over 20 International Journal Publications. He is pursuing PhD from SGGSIET, Nanded, Maharashtra, India. The current research interests are Recommendation systems, Data Mining and Machine Learning.