

A Novel Secure Multiparty Algorithms in Horizontally Distributed Database for Fast Distributed Database

V.Bangaru Lakshmi¹ & R.Murugadoss²

¹PG Scholar, Dept of MCA, St. Ann's College of Engineering & Technology, Chirala, AP-India

²Associate Professor, Dept of MCA, St. Ann's College of Engineering & Technology, Chirala, AP-India

Abstract

The proposed a protocol for protected mining of association rules in horizontally distributed database. The current leading set of rules is that of Kantarcioglu and Clifton. Our procedure, like theirs, is based on the Fast Distributed Mining (FDM) algorithm of Cheung et al. which is an unsecured spread version of the Apriori algorithm. The main ingredient in our procedure are two novel secure multi-party algorithms. 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 actor in a subset held by another. Our protocol offers improved separation with respect to the protocol. In addition, it is simpler and is extensively more efficient in terms of announcement rounds, announcement cost and computational cost.

Keywords: Privacy Preserving Data Mining; Distributed Computation; Frequent Itemsets; Association Rules.

I. Introduction

The most precious part of the protocol and its implementation relies upon cryptographic primitives such as commutative encryption, oblivious remove, and confusion functions. The main part of the procedure is a sub-protocol for the secure computation of the union of private subsets that are held by the different group of actors. This is also the only part in the procedure in which the players may extract from their view of the protocol information on other

database past what is implied by the final output and their own input. Leakage of information renders the protocol not perfectly protected, the perimeter of the excess information is explicitly bounded. It is argued there that such information leakage is safe whence acceptable from a practical point of view.

Insufficient security, simplicity and efficiency are not well in the databases, not sure in isolation in an existing system. While our solution is still not perfectly protected, it leaks excess information only to a little number (three) of probable coalitions, unlike the protocol of that discloses information also to some single group of actors. Our protocol may leak is less sensitive than the excess information leaked by the protocol.

II. Problem Statement

The proposed protocol improves upon that in terms of simplicity and efficiency as well as privacy. The protocol does not depend on commutative encryption and oblivious transfer. The proposed here computes a parameterized family of functions, which we call doorstep functions, in which the two great cases communicate to the problems of computing the union and intersection of private subsets.

The excess information that our protocol may leak is less sensitive than the excess information leaked by the protocol. This project proposed 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. The problem of secure multiparty computation that we solve here is the set inclusion problem.

Objective

The 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 multiparty protocol for compute the union (or intersection) of private subsets that each of the interacting players holds.

III. Proposed System

II. The proposed protocol improves upon that in terms of simplicity and efficiency as well as Privacy. Our protocol does not depend on commutative encryption and oblivious transfer. We suggest here computes a parameterized family of function, which we call threshold functions, in which the two extreme cases correspond to the troubles of computing the union and intersection of private subsets. The overload information that our procedure may leak is less sensitive than the excess information leaked by the protocol. The problem of secure multiparty computation that we solve here is the set inclusion problem.

The advantages are as following.

The proposed a procedure for protected 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 projected procedure is a novel secure multi-party protocol for computing the union (or intersection) of private subsets that each of the interacting players holds.

IV. Algorithm as Proposal

Algorithm Analyzed:

The first pass of the algorithm simply counts item occurrence determine the big 1-itemsets. A successive pass, say pass k , consists of two phase. First, the large itemsets L_{k-1} set up in the $(k-1)$ th pass are used to generate the candidate itemsets C_k , using the apriori- production function. Next, the database is scanned and the support of candidates in C_k is counted. For fast excluding we need to proficiently determine the candidates in C_k that are contained in a given transaction t . The subset function used for this purpose.

Table 1:

Notation associated with the candidates.

- $L_1 = f$ large 1-itemsets g ;
- for $(k = 2; L_{k-1} = 6; k++)$ do begin
- $C_k = \text{apriori-gen}(L_{k-1});$ // New candidates
- for all transactions $t \in D$ do begin
- $C_t = \text{subset}(C_k, t);$ // Candidates contained in t
- for all candidates $c \in C_t$ do
- $c: \text{count}++$
- end
- $L_k = \{c \in C_k \mid c: \text{count} \geq \text{minsup } g\}$
- end
- Answer = $\bigcup_k L_k$;

V. System Architecture

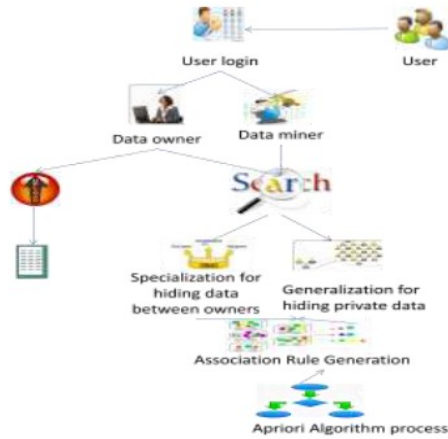


Fig 1: System Architecture

The above system architecture explains the user module which enlists the privacy preserving data mining has considered two connected settings. One, in which the data owner and the data miner

k-itemset	An itemset having k items.
LK	Set of large k-itemsets (those with minimum support). Each member of this set has two <u>elds</u> itemset and ii) support count.
Ck	Set of candidate k-itemsets (potentially large itemsets). Each member of this set has tw: i) itemset and ii) support count.
Ck	Set of candidate k-itemsets when the TI of the generating transactions arekept

are two different individual, and another, in which the data is scattered among several parties who aim to jointly perform data mining on the unified corpus of data that they grip. In the first location, the goal is to

protect the data records from the data miner. Hence, the information holder aims at anonymizing the data prior to its release. The main approach in this framework is to apply data perturbation. He disconcerted data can be used to infer general trends in the data, without informative original documents information. In the second setting, the goal is to perform data mining while protecting the data records of each of the data owners from the other data owners. The work of the administrator is to view user details. Direction to view the item set based on the user processing details using association role with Apriori algorithm connection rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information depository. An example of an association rule would be "If a customer buys a dozen seed, he is 80% likely to also purchase develop Association set of laws are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important associations support is an indication of how regularly the items appear in the database. Self confidence indicate the number of period the if/then statement have been found to be true.

VI. EXPERIMENTAL RESULTS

Table1. Parameter For Generating the synthetic Database

Parameter	Genration Method	Value
N	Number of transactions in the whol database	500000
L	Number of items	70000
At	Transaction Average Size	10
Af	Average size of maximum potentially large itemsets	4
Nf	Number of maximum potentially large itemstes	20000
CS	Clustering size	5
PS	Pool size	60
cor	Correction level	0.5
MF	Multiplying factor	1800

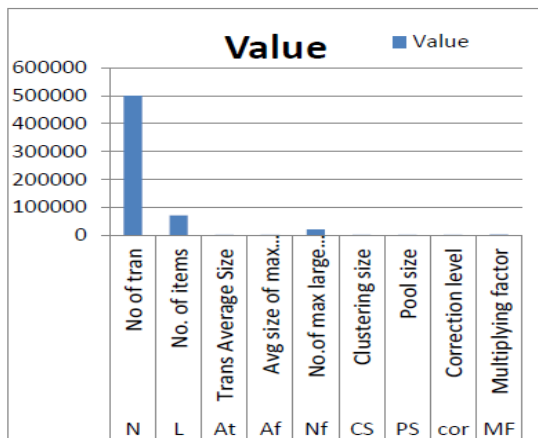


Fig 2. Parameter Value

We compared the performance of two secure implementations of the FDM algorithm. In the first implementation (denoted FDM-KC), we executed the unification step (Step 4 in FDM) using procedure UNIFI-KC, where the commutative nobody was 1024-bit RSA in the second implementation (denoted FDM) we used our procedure UNIFI, where the keyed-hash purpose was HMAC. In both implementations, we implemented in FDM algorithm in the secure manner that was described. We tested the two implementations with respect to three measures:

- 1) Total working out time of the complete protocols (FDMKC and FDM) over all players. That measure include the Apriori computation time, and the time to identify the globally s-frequent itemsets.
- 2) Total working out time of the unification protocols only
3. (The latter two procedures are implement in the same way in both Protocols FDM-KC and FDM.)
(UNIFI-KC and UNIFI) over all players.
- 4) Total message size.

We ran three experiment sets, where each set tested the dependence of the above measures on a diverse parameter:

- N — the number of transactions in the unified database,
- M — the number of players, and
- s — the threshold support size.

In our basic configuration, we took $N = 500,000$, $M = 10$ and $s = 0.1$. In the first experiment set, we kept M and s fixed and tested several values of N. In the second experiment set, we kept N and s fixed and varied

M. In the third set, we kept

N and M fixed and varied s. The results in each of those experiment sets are shown above. All experiments were implemented in C# (.net 4) and were executed on an Intel(R) Core(TM)i7-2620M personal computer with a 2.7GHz CPU, 8 GB of RAM, and the 64-bit operating system Windows 7 Professional SP1.

VII. CONCLUSION:

The 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 competence. One of the main ingredient in the proposed protocol is a novel secure multiparty protocol for computing the union (or intersection) of private subsets that each of the interacting group of actors hold. Another ingredient is a procedure 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 in equality verifications that uses the existence of a semihonest third party. Such a protocol might enable to further improve upon the announcement 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 detection in horizontally partitioned data.

VIII. References

[1].A. Ben-David, N. Nisan, and B. Pinkas. FairplayMP - A system for secure multi-party computation.

[2].H. Grosskreutz, B. Lemmen, and S. R^uping. Secure distributed subgroup discovery in horizontally partitioned data. Transactions on Data Privacy.

[3]. M. Kantarcioglu and C. Clifton. Privacy preserving distributed mining of association rules on horizontally partitioned data. IEEE Transactions on Knowledge and Data Engineering.

[4].P. Bogetoft, D.L. Christensen, I. Damgard, M. Geisler, T. Jakobsen, M. Krⁱgaard, J.D.Nielsen, J.B. Nielsen.

[5].P. Bogetoft, I. Damgard, T. Jakobsen, K.Nielsen, J.Pagter, and T. Toft. A practical implementation of secure actions based on multiparty integer computation.

[6]P.JagannadhaVarma,Amruthaseshadri, .M.Priyanka, M.Ajay Kumar, B.L.Bharadwaj Varma, " Association Rule Mining with Security Based on partial Databases.

[7] Prof. Geetika. Narang, Anjum Shaikh, Arti Sonawane, Kanchan Shegar, Madhuri Andhale," Preservation Of Privacy In Mining Using Association Rule Technique".

[8] Zhi Liu,Tianhong Sunand Guoming Sang," An Algorithm of Association Rules Mining in Large Databases Based on Sampling".

[9] Priyanka Asthana, Anju Singh , Diwaka Singh," A Survey on Association Rule Mining Using Apriori Based Algorithm and Hash Based Methods ", IEEE.

Author's Profile:



V.Bangaru Lakshmi received her graduate degree in B.Sc-Computer science from Nagarjuna University,Guntur. The year of 2009-

2012.Pursuing Post graduate degree MCA from St.Ann's college of Engineering and Technology Affiliated to Jawaharlal Nehru Technology University Kakinada in the year of 2012-2015.



Murugadoss Rajan received his graduate degree in B.Sc-Computer Science from Madurai Kamaraj University in 1995-1998. Post graduate degree in M.C.A from

Madurai Kamaraj University in 1998-2001 and M.E (CSE) from Anna University in 2005-2007. Pursuing PhD in Computer Science and Engineering from Sathyabama University Chennai from 2009.He has Ten years of teaching Experience in the field of Computer Science and Engineering .His Research interests are Fuzzy Neural Networks and Soft Computing. He has published many papers in National Conferences, International Conferences and National and International Journals with Scopus Indexed Journa