© 2015 IJSRSET | Volume 1 | Issue 5 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Privacy Preserving Secure Data Mining of Association Rules in Horizontally in Distributed Computation Databases

[1]Ch. Narsimha Chary, [2]K. Aparna, [3]K. Nagasindhu

[1]Associate Professor, Department of Computer Science & Engineering in Sri Indu College of Engineering& Technology, Hyderabad, India
[2]pursuing M.Tech (CS), Sri Indu College of Engineering& Technology, Affiliated to JNTU-Hyderabad, India
[3]pursuing M.Tech (CS), Sri Indu College of Engineering& Technology, Affiliated to JNTU-Hyderabad, India

ABSTRACT

We propose a protocol for secure mining of association rules in horizontally distributed databases. Our protocol, like theirs, is predicated on the quick Distributed Mining FDM rule that is Associate in unsecured distributed version of the Apriori rule. the most ingredients in our protocol area unit Two novel secure multi-party algorithms ,one that computes the union of personal subsets that every of the interacting players hold, and another that tests the inclusion of part command by one player in an exceedingly set command by another. Our protocol offers increased privacy with regard to the protocol. Additionally, it's easier and is considerably additional economical in terms of communication rounds, communication value and procedure value.

Keywords: Secure Mining, Fast Distributed Databases, Distributed Mining, Multi Party Rule

I. INTRODUCTION

We study here the matter of secure mining of association rules in horizontally partitioned off databases. In this setting, there square measure many sites (or players) that hold undiversified databases, i.e., databases that share constant schema however hold data on totally different entities. The goal is to seek out all association rules with support a minimum of s and confidence a minimum of c, for a few given lowest support size s and confidence level c, that hold within the unified information, whereas minimizing the knowledge disclosed concerning the personal databases command by those players. The information that we'd wish to defend during this context isn't solely individual transactions within the totally different databases, however conjointly additional world data like what association rules are supported regionally in every of these databases. That goal defines a haul of secure multi-party computation. In such issues, there square measure M players that hold personal inputs, x1, . . . , xM, and that they would like to firmly figure y = f(x1, . . . , xM) for a few public perform f. If there existed a trust worthy third party, the players might surrender to him their inputs and he would perform the perform analysis and send to them the ensuing output. Within the absence of such a trustworthy third party, it's required to plot a protocol that the players will run on their own so as to make the desired output y. Such a protocol is taken into account utterly secure if no player will learn from his read of the protocol over what he would have learnt within the idealized setting wherever the computation is dole out by a trustworthy third party. The primary to propose a generic resolution for this downside within the case of two players. Different generic solutions, for the multi-party case, were later projected in. In our downside, the inputs square measure the partial databases, and also the needed output is that the list of association rules that hold within the unified information with support and confidence no smaller than the given thresholds s and c, severally. Because the on top of mentioned generic solutions depend on an outline of the perform f as a mathematician circuit, they'll be applied solely to little inputs and functions that square measure realizable by easy circuits. In additional

IJSRSET151547 | Received: 06 October 2015 | Accepted: 14 October 2015 | September-October 2015 [(1)5: 276-279]

advanced settings, like ours, different strategies square measure needed for finishing up this computation. In such cases, some relaxations of the notion of good security could be inevitable once craving for sensible protocols, given that the surplus data is deemed benign see samples of such protocols that downside in and devised a protocol for its resolution. The most a part of the protocol may be a sub-protocol for the secure computation of the union of personal subsets that square measure command by the various players. The personal set of a given player, as we have a tendency to justify below, includes the item sets that square measure s-frequent in his partial information. That the foremost expensive a part of the protocol and its implementation depends upon crypto logical primitives like independent cryptography, oblivious transfer, and hash functions. This can be conjointly the sole half within the protocol within which the players might extract from their read of the protocol data on different databases, on the far side what's implicit by the ultimate output and their own input. Whereas such outflow of knowledge renders the protocol not utterly secure, the perimeter of the surplus data is data outflow is innocuous, wherefrom acceptable from a sensible purpose of read. Herein we have a tendency to propose an alternate protocol for the secure computation of the union of personal subsets. The projected protocol improves upon that in [18] in terms of simplicity and potency also as privacy. Specifically, our protocol doesn't rely on independent cryptography and oblivious transfer (what simplifies it considerably and contributes towards abundant reduced communication and process costs). Whereas our resolution remains not utterly secure, it leaks excess data solely to a tiny low variety (three) of potential coalitions, in contrast to the protocol of that discloses data conjointly to some single players. Additionally, we have a tendency to claim that the surplus data Digital Object Identifier .This article has been accepted for publication in an exceedingly future issue of this journal however has not been absolutely emended. Content might amendment before final publication. That our protocol might leak is a smaller amount sensitive than the surplus data leaked by the protocol of. The protocol that we have a tendency to propose here computes a parameterized family of functions, that we have a tendency to decision threshold functions, within which the 2 extreme cases correspond to the issues of computing the union and intersection of personal subsets. That square measure if truth be told all-purpose protocols that may be employed in different contexts also. Another downside of secure multiparty computation that we have a tendency to solve here as a part of our discussion is that the set inclusion problem; particularly, the matter wherever Alice holds a personal set of some ground set, and Bob holds part within the ground set, and that they would like to work out whether or not Bob's component is among Alice's set, while not revealing to either of them data concerning the opposite party's input on the far side the on top of delineated inclusion. In this paper, we have a tendency to study the matter of. The system architecture is given fig1.

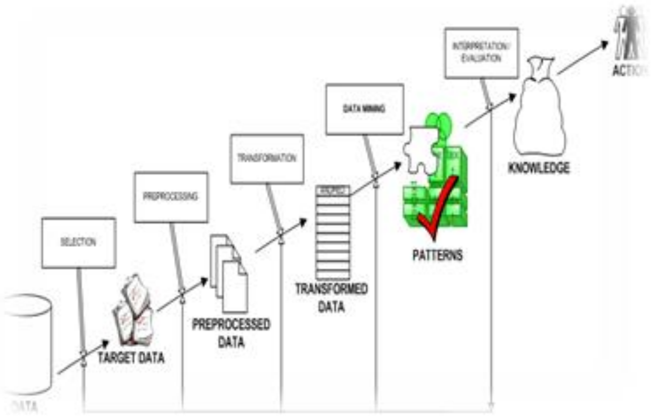

Figure 1: System Architecture

II. EXISTING SYSTEM

Some analysis folks studied that issues and devised a protocol for its resolution. The most a part of the protocol could be a sub-protocol for the secure computation of the union of personal subsets that area unit command by the various players. The non-public set of a given player, as we tend to justify below, includes the item sets that area unit s-frequent in his partial information. That's the foremost pricey a part of the protocol and its implementation depends upon cryptologic primitives like independent secret writing, oblivious transfer, and hash functions. This is often additionally the sole half within the protocol within which the players might extract from their read of the protocol data on alternative databases, on the far side what's understood by the ultimate output and their own input. Whereas such escape of data renders the protocol not absolutely secure, the perimeter of the surplus data is expressly finite and it's argued there that such data escape is innocuous, wherefrom acceptable from a sensible purpose of read.

DRAWBACKS OF EXISTING SYSTEM:

1. Insufficient security, simplicity and efficiency are not well in the databases, not sure in privacy in an existing system.
2. 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.
3. Our protocol may leak is less sensitive than the excess information leaked by the protocol.

III. PROPOSED SYSTEM

The protocol that we have a tendency to propose here computes a parameterized family of functions, that we have a tendency to decision threshold functions, during which the 2 extreme cases correspond to the issues of computing the union and intersection of personal subsets. Those square measure indeed all-purpose protocols which will be employed in alternative contexts additionally. Another downside of secure multiparty computation that we have a tendency to solve here as a part of our discussion is that the set inclusion problem; particularly, the matter wherever Alice holds a non-public set of some ground set, and Bob holds a component within the ground set, and that they would like to work out whether or not Bob's component is at intervals Alice's set, while not revealing to either of them info regarding the opposite party's input on the far side the on top of delineated inclusion.

BENIFITS OF PROPOSED SYSTEM

1. 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.
2. 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.

IV. METHODOLOGY

1. Privacy Preserving Data Mining:

One, in which the data owner and the data miner are two different entities, and another, in which the data is distributed among several parties who aim to jointly perform data mining on the unified corpus of data that they hold. In the first setting, the goal is to protect the data records from the data miner. Hence, the data owner aims at anonym zing the 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 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.

2. Distributed Computation

We compared the performance of two secure implementations of the FDM algorithm Section In the first implementation (denoted FDM-KC), we executed the unification step using Protocol UNIFI-KC, where the commutative cipher was 1024-bit RSA in the second implementation (denoted FDM) we used our Protocol UNIFI, where the keyed-hash function was HMAC. In both implementations, we implemented Step 5 of the FDM algorithm in the secure manner that was described in later. We tested the two implementations with respect to three measures:

1) Total computation time of the complete protocols (FDMKC and FDM) over all players. That measure includes the Apriori computation time, and the time to identify the globally *s*-frequent item sets, as described in later.
2) Total computation time of the unification protocols 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.

3. Frequent Item Sets

We describe here the solution that was proposed by Kantarcioglu and Clifton. They considered two possible settings. If the required output includes all globally *s*-frequent item sets, as well as the sizes of their supports, then the values of $\Delta(x)$ can be revealed for all. In such a case, those values may be computed using a secure summation protocol, where the private addend of Pm is $suppm(x) - sNm$. The more interesting setting, however, is the one where the support sizes are not part of the required output. We proceed to discuss it.

4. Association Rules

Once the set Fs of all *s*-frequent itemsets is found, we may proceed to look for all (s, c)-association rules (rules with support at least sN and confidence at least c). In order to derive from Fs all (s, c)-association rules in an efficient manner we rely upon the straightforward lemma.

V. CONCLUSION

We proposed a protocol for secure mining of association rules in horizontally distributed databases that improves significantly upon the current leading protocol [18] in terms of privacy and efficiency. One of the main ingredients 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. 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.

VI. REFERENCES

[1] M.D. Atkinson, J.-R. Sack, N. Santoro, and T. Strothotte, "Min-maxHeaps and Generalized Priority Queues,"Comm. ACM,vol. 29,no. 10, pp. 996-1000, 1986.

[2] A. Balmin, V. Hristidis, and Y. Papakonstantinou, "Objectrank:Authority-Based Keyword Search in Databases,"Proc. Int’l Conf.Very Large Data Bases (VLDB),pp. 564-575, 2004.

[3] Z. Bao, T.W. Ling, B. Chen, and J. Lu, "Effective XML KeywordSearch with Relevance Oriented Ranking,"Proc. Int’l Conf. DataEng. (ICDE),2009.

[4] H. Bast and I. Weber, "Type Less, Find More: Fast Autocompletion Search with a Succinct Index," Proc. Ann. Int’l ACM SIGIRConf. Research and Development in Information Retrieval (SIGIR),pp. 364-371, 2006.

[5] H. Bast and I. Weber, "The Completesearch Engine: Interactive,Efficient, and towards Ir&db Integration,"Proc. Biennial Conf.Innovative Data Systems Research (CIDR), pp. 88-95, 2007.

[6] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S.Sudarshan, "Keyword Searching and Browsing in Databases UsingBanks,"Proc. Int’l Conf. Data Eng. (ICDE),pp. 431-440, 2002.

[7] Y. Chen, W. Wang, Z. Liu, and X. Lin, "Keyword Search onStructured and Semi-Structured Data,"Proc. ACM SIGMOD Int’lConf. Management of Data,pp. 1005-1010, 2009.

[8] E. Chu, A. Baid, X. Chai, A. Doan, and J.F. Naughton, "CombiningKeyword Search and Forms for Ad Hoc Querying of Databases,"Proc. ACM SIGMOD Int’l Conf. Management of Data,pp. 349-360,2009