International Journal of Research Studies in Science, Engineering and Technology
Volume 1, Issue 9, December 2014, PP 10-13
ISSN 2349-4751 (Print) & ISSN 2349-476X (Online)

Secure Mining of Association Rules in Horizontally Distributed Databases (Potecting Sensitive Labels in Social Network Data Anonymization)

Bollu Jyothi, K. Venkateswara Rao

P.G.Scholar, Dept. of CSE, Krishnaveni Engineering College for Women, Narasaraopet, Andhra Pradesh, India 1
Professor & Head of the Dept of CSE, Krishnaveni Engineering College For Women, Narasaraopet, Andhra Pradesh, India 2
jyothi.nalajala@gmail.com, venkateswara.gnt@gmail.com

Abstract: *In this paper we propose a protocol for secure mining of association rules in horizontally distributed databases. The current leading protocol is that of Kantarcioglu and Clifton. Our protocol, like theirs, is based on the Fast Distributed Mining (FDM) algorithm of Cheung et al, which is an unsecured distributed version of the Apriori algorithm. The main ingredients in our protocol 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 protocol offers enhanced privacy with respect to the protocol in . In addition, it is simpler and is significantly more efficient in terms of communication rounds, communication cost and computational cost.*

1. INTRODUCTION

Data mining refers to Knowledge Discovery in Databases. The data mining process is the extraction of information from various data sets and transform to an understandable manner. A social network is a social graph made up of actors such as individuals or organizations and connections. A social network service consists of a representation of each users, social links and variety of additional services. Most social network services provide means for users to interact over the Internet, such as e-mail and messaging. Social network sites are varied and they incorporate new information and communication tools. The major drawbacks of social networks opens up the possibility of hackers to commit fraud and increases the risk of people falling prey to outline scams resulting in data or identity theft and potentially results in lost productivity. It refers to confidentiality of employer trade secrets and their private information. In order to provide security to social network users our algorithms issue anonymized views of the graph with significantly smaller information losses and analyze their privacy and communication complexity. Formally, in this social networks always represented as a graph, which we refer to as the social graph. The node of such a graph represents an actor and the edges represent ties between those actors.

2. LITERATURE SURVEY

We study here the problem of secure mining of association rules 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 given minimal support and confidence levels that hold in the unified database, while minimizing the information disclosed about the private databases held by those players. That goal defines a problem of secure multi-party computation. In such problems, there are M players that hold private inputs, x1, . . . , xM, and they wish to securely compute y = f(x1, . . . , xM) for some public function f. If there existed a trusted third party, the players could surrender to him their inputs and he would perform the function evaluation and send to them the resulting output. In the absence of such a trusted third party, it is needed to devise a protocol that the players can run on their own in order to arrive at the required output y. Such a protocol is considered perfectly secure if no player can learn from his view of the protocol more than what he would have learnt in the idealized setting where the computation is carried out by a trusted third party. Yao was the first to propose a generic solution for

this problem in the case of two players. Other generic solutions, for the multi-party case, were later proposed in [2,4,10].2 T. Tassa In our problem, the inputs are the partial databases, and the required out-put is the list of association rules with given support and confidence. As the above mentioned generic solutions rely upon a description of the function f as a Boolean circuit, they can be applied only to small inputs and functions which are realizable by simple circuits. In more complex settings, such as ours, other methods are required for carrying out this computation. In such cases, some relaxations of the notion of perfect security might be inevitable when looking for practical protocols, provided that the excess information is deemed benign (see examples of such protocols in e.g. [12,20,23]). Kantarcioglu and Clifton studied that problem in [12] and devised a protocol for its solution. The main part of the protocol is a sub-protocol for the secure computation of the union of private subsets that are held by the different players. (Those subsets include candidate itemsets, as we explain below.) That is the most costly part of the protocol and its implementation relies upon cryptographic primitives such as commutative encryption, oblivious transfer, and hash functions. This is also the only part in the protocol in which the players may extract from their view of the protocol information on other databases, beyond what is implied by the final output and their own input. While such leakage of information renders the protocol not perfectly secure, the perimeter of the excess information is explicitly bounded in and it is argued that such information leakage is innocuous, whence acceptable from practical point of view. Herein we propose an alternative protocol for the secure computation of the union of private subsets. The proposed protocol improves upon that in terms of simplicity and efficiency as well as privacy. In particular, our protocol does not depend on commutative encryption and oblivious transfer (what simplifies it significantly and contributes towards reduced communication and computational costs).

The protocol that we propose here computes a parameterized family of functions, which we call threshold functions, in which the two extreme cases correspond to the problems of computing the union and intersection of private subsets. Those are in fact general-purpose protocols that can be used in other contexts as well. Another problem of secure multi-party computation that we solve here as part of our discussion is the problem of determining whether an element held by one player is included in a subset held by another.

3. EXISTING METHODOLOGY

That goal defines a problem of secure multi-party computation. In such problems, there are M players that hold private inputs, $x1, \ldots, xM$, and they wish to securely compute $y = f(x1, \ldots, xM)$ for some public function f. If there existed a trusted third party, the players could surrender to him their inputs and he would perform the function evaluation and send to them the resulting output. In the absence of such a trusted third party, it is needed to devise a protocol that the players can run on their own in order to arrive at the required output y. Such a protocol is considered perfectly secure if no player can learn from his view of the protocol more than what he would have learnt in the idealized setting where the computation is carried out by a trusted third party. Yao was the first to propose a generic solution for this problem in the case of two players. Other generic solutions, for the multi-party case, were later proposed in

4. PROPOSED SYSTEM

Herein we propose an alternative protocol for the secure computation of the union of private subsets. The proposed protocol improves upon that in terms of simplicity and efficiency as well as privacy. In particular, our protocol does not depend on commutative encryption and oblivious transfer (what simplifies it significantly and contributes towards much reduced communication and computational costs). 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. In addition, we claim that the excess information that our protocol may leak is less sensitive than the excess information leaked by the protocol of. The protocol that we propose here computes a parameterized family of functions, which we call threshold functions, in which the two extreme cases correspond to the problems of computing the union and intersection of private subsets. Those are in fact general-purpose protocols that can be used in other contexts as well. Another problem of secure multiparty computation that we solve here as part of our discussion is the set inclusion problem; namely, the problem where Alice holds a private subset of some ground set, and

Bob holds an element in the ground set, and they wish to determine whether Bob's element is within Alice's subset, without revealing to either of them information about the other party's input beyond the above described inclusion.

5. IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

6. COST WITH FINISH TIME-BASED ALGORITHM

The CwFT algorithm is a workflow scheduling algorithm extended from the HEFT algorithm for distributed environments with multiple heterogeneous processing nodes. Instead of optimizing only the workflow makespan as usual, CwFT algorithm also considers reducing the monetary cost that CCs need to pay in a computing framework with the combination between numerous Cloud node and a local system. Similar to HEFFT, the CwFT algorithm is comprised of two phases: *Task Prioritizing* to mark the priority level for all tasks and *Node Selection* to select tasks in a descending order by the priority level and then schedule each selected task on an appropriate processing node to optimize the value of the utility function.

7. CONCLUSION

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. 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. The latter protocol exploits 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 in Section 3 namely, to devise an efficient protocol for set inclusion verification that uses the existence of a semi-honest third party. Such a protocol might enable to further improve upon the communication and computational costs of the second and third stages of the protocol of , as described in Sections 3 and 4. Another research problem that this study suggests is the extension of those techniques to the problem of mining generalized association rules.

REFERENCES

[1] R. Agrawal and R. Srikant. Privacy preserving data mining. In SIGMOD Conference , pages 439–450, 2000.

[2] D. Beaver, S. Micali, and P. Rogaway. The round complexity of secure protocols.In STOC, pages 503–513, 1990.

[3] M. Bellare, R. Canetti, and H. Krawczyk. Keying hash functions for message authentication. In Crypto, pages 1–15. Springer-Verlag, 1996.

[4] A. Ben-David, N. Nisan, and B. Pinkas. FairplayMP - A system for secure multiparty computation. In CCS, pages 257–266, 2008.

[5] J.C. Benaloh. Secret sharing homomorphisms: Keeping shares of a secret secret. In Crypto, pages 251–260, 1986.

[6] D. W.-L. Cheung, J. Han, V. Ng, A. W.-C. Fu, and Y. Fu. A fast distributed algorithm for mining association rules. In PDIS, page 3142, 1996.

[7] T. ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Transactions on Information Theory, 31:469472, 1985.

[8] A.V. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke. Privacy preserving mining of association rules. In KDD, pages 217–228, 2002.

[9] R. Fagin, M. Naor, and P. Winkler. Comparing Information without Leaking It. Communications of the ACM, 39:77–85, 1996.

[10] Theorem for protocols with honest majority. In STOC, pages 218–229, 1987.

[11] W. Jiang and C. Clifton. A secure distributed framework for achieving k- anonymity. The VLDB Journal, 15:316–333, 2006.

[12] M. Kantarcioglu and C. Clifton. Privacy-preserving distributed mining of association rules on horizontally partitioned data. IEEE Transactions on Knowledge and Data Engineering,16(9):10261037, september 2004.

[13] X. Lin, C. Clifton, and M.Y. Zhu. Privacy-preserving clustering with distributed em mixture modeling. Knowl. Inf. Syst., 8:68–81, 2005.

[14] Y. Lindell and B. Pinkas. Privacy preserving data mining. In CRYPTO, pages 36–54, 2000.

[15] B. Pinkas M. Freedman, Y. Ishai and O. Reingold. Keyword search and oblivious pseudorandom functions. In TCC, pages 303–324, 2005.

AUTHORS' BIOGRAPHY

Bollu Jyothi received the B.Tech degree in Computer Science and Information Technology in the year 2006 and pursuing M.Tech degree in Computer Science and Engineering from Krishnaveni Engineering College for Women.

K.Venkateswara Rao received his M.Tech degree in Computer Science from JNTUK, M.Sc degree in Computer Science from ANU. He is currently working as Professor and head of the Dept in CSE in Krishnaveni Engineering College for Women.