IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 2, Ver. II (Mar-Apr. 2016), PP 92-95
www.iosrjournals.org

Two Phase Multi Party Secured Multiplication (2PMSM) In Vertically Distributed Databases

Selva Rathna S[1], Karthikeyan T[2]
[1](*Research Scholar, Manonmaniam Sundaranar University,Tirunelveli,India)*
[2](*Computer Science Department, PSG Arts & Science College,Bharathiyar University, Coimbatore, India)*

Abstract : *Secured multi party computation which is also known as secured computation or multi party computation (SMC) plays a vital role in achieving secured computation in Privacy preserving data mining. Various SMC algorithms are existing for achieving the secured computation. In this paper, Two phase Multi party secured computation for secured multiplication of data in distributed database is presented. The algorithms for vertically distributed database is presented which enables to develop new data mining algorithms in Privacy preserving data mining field.*

Keywords: *Horizontally Distributed Database, multiparty computation, Secured multiparty computation (SMC), Secured multiparty multiplication (SMM), Trusted Third Party (TTP), Vertically Distributed Database*

I. Introduction

In privacy-preserving data mining, the data is divided among two or more different parties with the aim to run a data mining algorithm on the union of the parties' databases without allowing any party to view another individual's private data. Secured Multiparty Sum computation is one of the methods used for handling this type of scenario. Secure Multi party computation is a subfield of cryptography with the goal of creating methods for parties to jointly compute a function over their inputs while keeping those inputs private. 2PMSC protocol has been introduced for secured sum computation. In this paper, 2PMSM is proposed to perform multiplication process in secured way in Horizontal and Vertical distributed database.

II. Background Study

Secure multiparty computation

2.1. Introduction

In an multi party computation (MPC), a given number of participants, P_1, P_2, ..., P_N each have private data, respectively d_1, d_2, ..., d_N. Participants want to compute the value of a public function on that private data: $F(d_1, d_2, ..., d_N)$ while keeping their own inputs secret. For example, suppose we have three parties Alice, Bob and Charlie, with respective inputs x, y and z denoting their salaries. They want to find out which of the three salaries is the highest, without revealing to each other how much each of them makes. The most basic properties that a multi-party computation protocol aims to ensure are:

- **Input privacy:** No information about the private data held by the parties can be inferred from the messages sent during the execution of the protocol. The only information that can be inferred about the private data is whatever could be inferred from seeing the output of the function alone.
- **Correctness:** Any proper subset of adversarial colluding parties willing to share information or deviate from the instructions during the protocol execution should not be able to force honest parties to output an incorrect result. This correctness goal comes in two flavors: either the honest parties are guaranteed to compute the correct output or they abort if they find an error.

2.2. Literature Survey on SMC

A study on various efficient fundamental secure building blocks such as Fast Secure Matrix Multiplication (FSMP), Secure Scalar Product (SSP), and Secure Inverse of Matrix Sum (SIMS) is made to evaluate time/space efficiency on the different protocols in [1]. An algorithm of privacy preserving C4.5 which is applicable to vertically and horizontally partitioned dataset is given in [2]. It gives a detailed computation method of the information gain ratio without revealing privacy. The secure scalar product protocol, the xln(x) protocol and secure sum protocol are used in collaborative computing, which can protect privacy effectively. An excellent review of SMC with a framework for SMC problem discovery and transformation of normal problem to SMC problem is presented. In [3], a novel protocol is discussed to compute the sum of an individual's data given by parties with zero leakage probability. This protocol suggests breaking the data blocks into segments and redistributing the segments among all the parties. Also, neighbor's position is changed to maintain security. Breaking of data into segments and changing location of neighbors is also suggested in [4]. This protocol provides zero probability of data leakage by two colluding parties when they want to attack data

of a middle party. The only drawback of this scheme is that the topology of the computational network changes in each round of the computation. The communication and computation complexity both are $O(n^2)$.

2.3. Two Phase SMC (2PSMC) Protocol

In [5], a new2PSMC protocol with an improved performance on complexity is designed to perform Secured Sum computation on Multiparty environment is proposed. The proposed algorithm 2PSMC runs in two cycles instead of k cycles. Each party breaks the data block into two segments. Also each site generates a random number R_i which will be used for encrypting the sum at each site. Initially, all the sites are arranged randomly and protocol initiator S_1 is also selected randomly. At end of two phases , the protocol initiator will hold the sum of all parties data. This will helps to perform sum operation among data which is distributed horizontally.

2.4. Two Phase SMC Protocol for Multiplication (2PMSM)

As the requirement of secured multiplication is essential for privacy preserved calculation in distributed databases, the Two Phase Secured Sum computation is improved to perform the multiplication process in this paper. This will ensure to perform secured computation of multiplication of data which are distributed horizontally as well as vertically. The algorithm for vertical distributed database is explained in Section 3.

III. Two Phase Secured Multi Party Multiplication (2PMSM)

Secure multiparty Multiplication (SMM)

Consider that P_1, P_2, …. P_n are the multi parties where n is number of parties. Each party P_i holds the data P_{ij}. A party P_i, have m number of tuples and n number of attributes. The data for each parties and secured multiplication is as shown in (1) & (2)

$$P_i = \begin{pmatrix} D_{11} & D_{12} & \cdots & D_{1n} \\ D_{21} & D_{22} & \cdots & D_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ D_{m1} & D_{m2} & & D_{mn} \end{pmatrix} \quad \ldots\ldots\ldots (1)$$

$$\pi(P_1, P_2 \ldots P_n) = P_{1_{ij}} \, x \, P_{2_{ij}} \ldots\ldots \quad P_{n_{ij}} \quad \ldots\ldots(2)$$

where $P_{1_{ij}}$ is the value of Party 1,

$P_{2_{ij}}$ is the value of Party 2,

$P_{n_{ij}}$ is the value of Party n and

$\pi(P_1, P_2 \ldots P_n)$ yields the secured multiplication of data of n parties.

The proposed algorithm 2PSMM runs in two cycles. Each party generates a random number *Ri* which will be used for encrypting the product at each site. Initially, all the sites are arranged randomly and protocol initiator S_1 is also selected randomly. The two phases of the protocol is explained in Section 3.1 and Section 3.2.

3.1. Phase One of 2PMSM

Consider there are N number of parties where N >=3. Protocol initiator P_1 is selected randomly and Party P_1 generate a random number R_1. Protocol Initiator P_1 which has value D_1. The protocol initiator Pi, generated value V_1 using Equation 3

$$V_i = R_1 * D_1 \qquad (3)$$

At each Site Si hold value D_i and random value R_i and hence partial product V_i is calculated using Equation 4 where $2<=i<=n$

$$V_i = V_{i-1} * Ri * Di \qquad (4)$$

While the cycle reaches Site *Sn*, the partial product along with random value of all sites will be available with Site *Sn*.

3.3. Phase Two of 2PMSM

In the 2nd phase, again the sites are arranged randomly. In the second phase, each site will divide partial value received from the previous site with its own random number which is used in phase one. The calculation of *Vi* will be done using Equation 3 at each site.

$$Vi = V_{i-1} / R_i \qquad (5)$$

3.4. Algorithm for 2PMSM

1. Arrange all sites randomly. Select a site as Protocol Initiator.

2. Protocol initiator will initialize $V_1 = R_1 + x$ where x is the value of protocol initiator and R1 is it's random value
3. for i =2..n
4. Calculate $V_1 = Vi\text{-}1 * R_i * x_i$
5. Send V_i to next random site
6. Arrange all sites randomly. Start 2nd cycle from Protocol initiator site.
7. for i =1..n
8. Calculate $V_1 = Vi\text{-}1 / R_i$
9. Send V_i to next random site
10. At the end Vi will hold the product of all sites
11. End of Algorithm

IV. Secured Multiplication In Vertically Distributed Database

4.1. Vertically Distributed Database 2PMSM Protocol

In vertically distributed database, the different attributes are present in different sites. Consider the following example with 3 different sites. Site 1, 2 & 3 are holding the attributes respectively shown in Table 1.

Table 1 : Example Vertically distributed database

Site 1(Employee Master)	Site 2 (Working Data)	Site 3 (Allowance Data)
Name	Emp Id	Designation
Emp ID	Hours Worked	Wage Weightage
Wage/Hour	Designation	

To calculate the salary for each employee, it is needed to calculate Wage/Hour * Hours Worked * Wage Weightage. As per the 2PSMM protocol, the protocol initiator will be chosen randomly in both phases and each site will generate its own random share. At the first phase each site will multiply its random share along with its value and at the second phase each site divides its random share from the result. The sample organization for the example as discussed above is shown in Table 2 and the protocol execution is presented in Figure 1.

Table 2 : Sample organization of 2PMSM

Site	Position (First Phase)	Position (Second Phase)
S_1	3	2
S_2	2	1
S_3	1	3

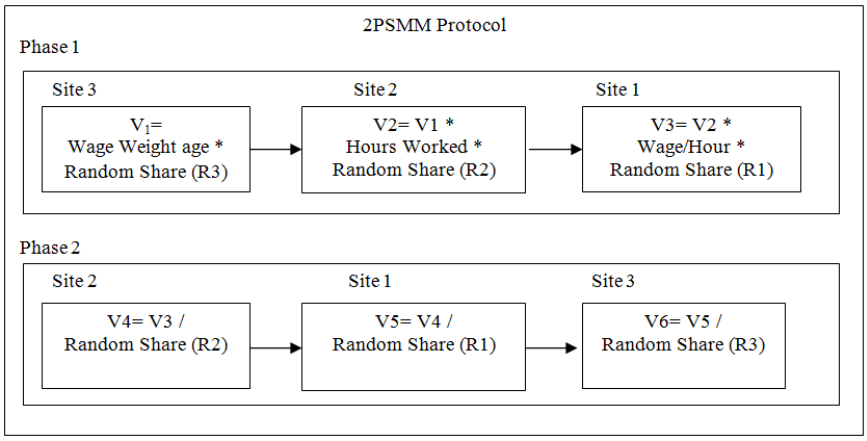

Figure 1: Execution of 2PMSM protocol

4.2. Performance Analysis of the Protocol

This protocol guarantees that a party will not know its position of arrangement since the position arrangement is made by the protocol randomly for each parties. Number of rounds of computation is two and the number of computation in each round is *n*. Hence, the communication and computation complexity both are *O* (*n*) which is better than earlier protocols available for Secured Multiparty computation. Figure 5 shows the performance based on number of sites against number of computations and time complexity.

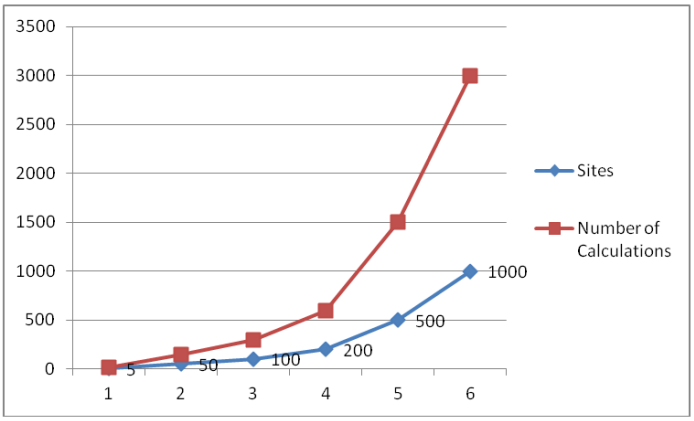

Figure 5: Performance of number of sites versus number of calculation

V. Conclusion

In this paper, a new protocol 2PMSM is proposed to compute secured multiplication for multi party environment. Since the protocol runs in two phases with *n* computations in each phase where *n* is the number of parties, the protocol has a very good performance while comparing with earlier protocols. In future, effort can be made to develop various datamining algorithm for privacy preservation using this protocol.

References

Journal Papers

[1]. Teo, S.G., Lee, V., Shuguo Han, "A Study of Efficiency and Accuracy of Secure Multiparty Protocol in Privacy-Preserving Data Mining", 26th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp: 85-90, 2012 (journal style)

[2]. Yanguang Shen ,. Hui Shao, Jianzhong Huang, "Research on Privacy Preserving Distributed C4. 5 Algorithm", Third International Symposium on Intelligent Information Technology Application Workshops, IITAW '09. pp:216-218, 2009.

[3]. Pathak, F.A.N., Pandey, S.B.S., "Distributed changing neighbors k-secure sum protocol for secure multiparty computation", Nirma University International Conference on Engineering (NUiCONE), pp: 1-3, 2013.

[4]. Sheikh.R, Kumar B., Mishra D.K., Changing Neighbors k-Secure Sum Protocol for Secure Multi-Party Computation, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 1, 2010.

[5]. Selva Rathna S, Karthikeyan T., Two Phase Secured Multiparty Sum Computation Protocol (2PSMC) for Privacy Preserving Data Mining, International Journal of Engineering and Computer Science, ISSN : 2319-7242., Vol.4, Issue 4., April 2015

[6]. W. Du and M.J. Atallah, "Secure Multiparty Computation Problems and Their Applications: A Review and Open Problems," In proceedings of new security paradigm workshop, Cloudcroft, New Maxico, USA, pages 11-20, Sep. 11-13 2001. (journal style)

[7]. Yehuda Lindell, and Benny Pinkas, "Secure Multiparty Computation for Privacy preserving data mining", The Journal of Privacy and Confidentiality, pp : 59-98, 2009. (journal style)

[8]. Jyotirmayee R., Raghvendra K.., "FP Tree Algorithm using Hybrid Secure Sum Protocol in Distributed Database", International Journal of Scientific & Engineering Research Volume 4, Issue3, 2013, pp: 1 – 5, 2013 (journal style)

[9]. Priyanka Jangle, Gajendra Singh, D.Mishra, Hybrid Technique for Secure Sum Protocol, Worldof Computer Science and Information Technology Journal (WCSIT) ISSN : 2221-0741 Vol. 1 No. 5 198-201, 2011.

Books:

[10]. Charu. C. Agarwall., Philip.S.Yu, "Privacy Preserving Data Mining, Models and Algorithms" Springer, 233 Spring Street, New York, NY 10013, USA, ISBN 978-0-387-70991-8,2008

[11]. Jaiwan.H., Michaline J., Jain P., "Data Mining Concepts and Techniques", Morgan Kaugman Publisher, Elsevier Inc, 22, Wyman Street, Waltham, USA., British Third Edition, 2012, ISBN 978-0-12-381479-1,

[12]. Jaideep.V., Chris C., Michael Z., "Privacy preserving Data Mining", Springer, 233 Spring Street, New York, NY 10013, USA , 2006, ISBN-13: 978-0-387-25886-8