www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 4 April 2015, Page No. 11453-11456

Two Phase Secured Multiparty Sum Computation Protocol (2PSMC) for Privacy preserving data mining

Selva Rathna[1], Dr.T. Karthikeyan[2]

[1]Manonmaniam Sundaranar University,
Tirunelveli, Tamil Nadu, India
sselvarathna@gmail.com

[2]P.S.G Arts and Science College, Bharathiyar University,
Coimbatore, Tamil Nadu, India
t.karthikeyan.gasc@gmail.com

Abstract: ***Secured Multiparty Sum Computation is an important algorithm designed in Privacy preserving Data mining to perform aggregated computation on data distributed between multi parties. In this paper, a new protocol with an improved performance on complexity is designed to perform Secured Sum computation on Multiparty environment which will enable to develop better algorithms for Privacy preserving data mining process such as Classification, Clustering etc.***

Keywords: Privacy preserving data mining (PPDM), Secured Multiparty Sum Computation (SMC), Trusted Third Party (TTP), Two phase Secured multiparty sum computation (2PSMC)

1. Introduction

Privacy-preserving data mining considers the problem of running data mining algorithms on confidential data that is not supposed to be revealed even to the party running the algorithm. There are two classic settings for privacy-preserving data mining. In the first, the data is divided among two or more different parties; the aim being 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. In the second, some statistical data that is to be released may contain confidential data; hence, it is first modified so that (a) the data does not compromise anyone's privacy, and (b) it is still possible to obtain meaningful results by running data mining algorithms on the modified data set. In this paper, we will mainly refer to scenarios of the first type using Secured Multiparty Sum Computation (SMC) methods. A new algorithm for SMC is proposed in this paper to have better performance in the aspect of complexity.

2. Background Study

A special case of a long-studied problem in cryptography called secure multiparty computation. This problem deals with a setting where a set of parties with private inputs wishes to jointly compute some function of their inputs. Loosely speaking, this joint computation should have the property that the parties learn the correct output and nothing else, even if some of the parties maliciously collude to obtain more information. Clearly, a protocol that provides this guarantee can be used to solve this problem.

In [1], a study 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.

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 is provided in [12] where they developed a framework for SMC problem discovery and transformation of normal problem to SMC problem.

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 $O(n^2)$.

In this protocol, each party partitions its data into k segments where k = n-1 which is the number of parties involved in computation. Let P_1 to be the protocol initiator. The position of the protocol initiator is kept fixed in each round of computation. For the first round of the computation parties are arranged in a serial fashion as $P_1, P_2, \ldots, P_n$. The protocol

initiator starts computation using k-secure sum protocol to get the sum of first segment of each party. Before second round of computation starts P_2 exchanges its position with P_3. In next round of the computation P_2 exchanges its position with P_4 and so on until P_2 exchanges its position with P_n. Generalizing the method we can say that in i_{th} round of the computation P_2 exchanges its position with P_{i+1} until P_n is reached. In each round of computation segments are added using k-secure sum protocol and the partial sum is passed to the next party until all the segments are added and the sum is announced by the protocol initiator party. The description of ck-Secure sum Protocol is given in Figure 1 and Snapshots for a four-party case are shown in Figure 2.

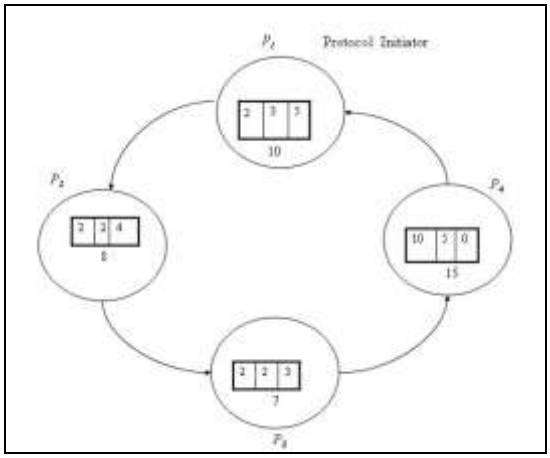

Figure 1: Initial Architecture of ck-Secure Sum

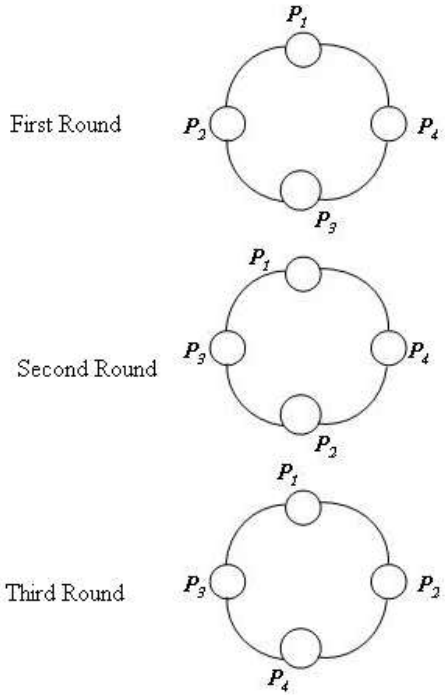

Figure 2: Snapshot of ck-Secure Sum for four party case

Eventhough ck-Secure sum Protocol provides zero probability of data leakage by two colluding parties when they want to attack data of a middle party and an appreciable improvement over previous protocols available in the literature.

Considering the requirement to reduce the computation and the communication complexity along with preserving the security, Two Phase secured multiparty sum computation protocol (2PSMC) is proposed in this paper.

3. Proposed Architecture and Protocol Description of 2PSMC

3.1 Description of Protocol

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 *Ri* 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. The two phases of the protocol is explained in Section 3.2 and Section 3.3.

3.2 Phase One of 2PSMC

Consider there are N number of sites where N >=3. Each site S_i where i = 1 to N has value D_i which is partitioned into two segments randomly as D_{i1} and D_{i2} such that $D_i = D_{i1} + D_{i2}$. Protocol initiator S_1 is also selected randomly and Site S_1 generate a random number R_1. At Site S1, V1 is generated using Equation 1.

$$V_i = R_1 + D_{11} \quad (1)$$

At each Site Si, partial sum V_i is calculated using Equation 2 where $2<=i<=n$

$$V_i = Vi\text{-}1 + Ri + D_{i1} \quad (2)$$

While the cycle reaches Site *Sn*, the partial sum with first segment data of all sites along with random number of each site will be available with Site *Sn*.

3.3 Phase Two of 2PSMC

In the 2nd phase, again the sites are arranged randomly. In the second phase, each site will subtract its random number from the partial sum received from the previous site and adds its second segment of data. The calculation of *Vi* will be done using Equation 3 at each site.

$$Vi = V_{i-1} - R_i + D_{i2} \quad (3)$$

3.4 Sample Demonstration of Protocol

For example consider a scenario with five parties. Let S1, S2, S3, S4 and S5 are the parties involved in the computation and each party hold the values 25, 23, 15, 9 and 11 respectively. Each party breaks their data block into two segments. The arrangement of parties is made randomly in the both phases. The sample segments of the parties in two phases and its arrangement are shown in Table 1. S_3 and S_2 are chosen as protocol initiators in first and second phase respectively. The computation in each cycle is shown in Figure 3 and Figure 4.

Table 1 : Sample organization of 2PSMC

Site	*Value of the site*	*Position (First Phase)*	*Segment Value (First Phase)*	*Position (Second Phase)*	*Segment Value (Second Phase)*	*Random Number*
S_1	25	3	17	5	8	3
S_2	23	5	6	1	17	56
S_3	15	1	4	4	11	54
S_4	9	4	0	3	9	87
S_5	11	2	6	2	5	4

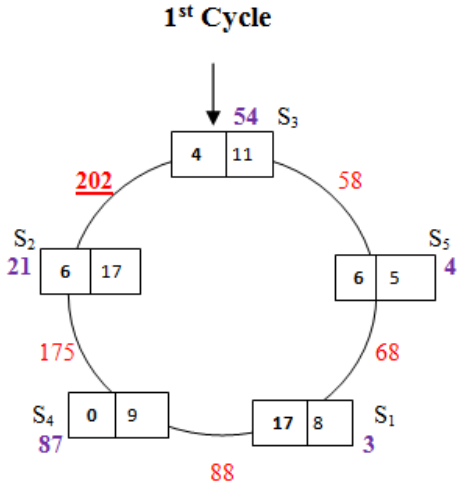

Figure 3: 2PSMC for five party case (1st Phase)

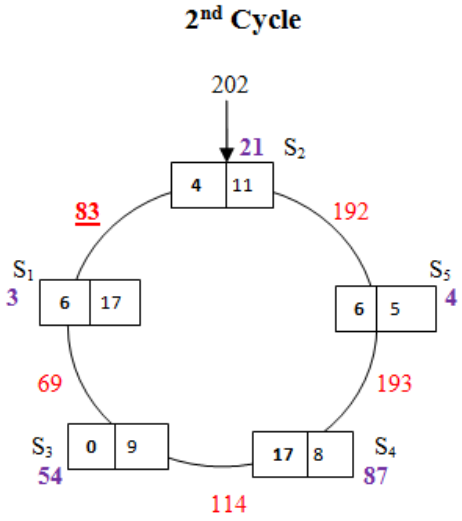

Figure 4: 2PSMC for five party case (2nd Phase)

The algorithm of 2PSMC is given in section 3.3.

3.5 Algorithm of 2PSMC Protocol

The algorithm: *2PSMC Protocol*

1. *Split data of each site into 2 segments*
2. *Arrange all sites randomly. Select a site as Protocol Initiator.*
3. *Protocol initiator will initialize* $V_1 = R_1 + x_1$ *where x1 is the first segment value of protocol initiator and R1 is it's random value*
4. *for i =2..n*
5. *Calculate* $V_1 = Vi\text{-}1 + R_i + x_{i1}$
6. *Send* V_i *to next random site*
7. *Arrange all sites randomly. Start 2nd cycle from Protocol initiator site.*
8. *for i =1..n*
9. *Calculate* $V_1 = Vi\text{-}1 - R_i + x_{i1}$
10. *Send* V_i *to next random site*
11. *At the end Vi will hold the sum of all sites*
12. *End of Algorithm*

3.6 Performance Analysis of the Protocol

In this protocol, each party breaks its data into two segments secretly on its own. If two neighbor parties collude they can know only their own data segments in the computation. The 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.

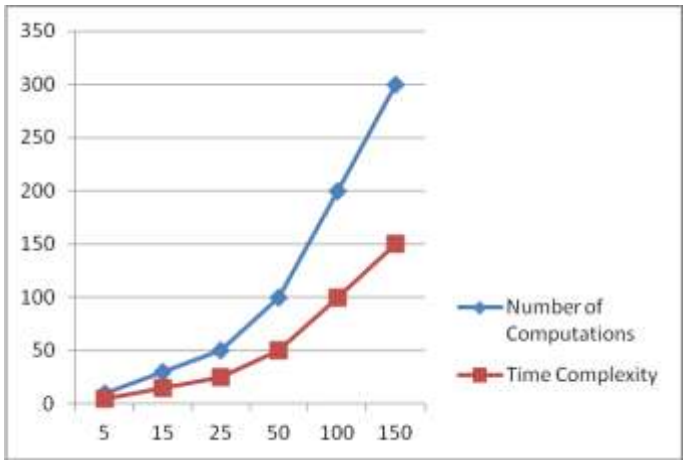

Figure 5: Performance of number of sites versus number of computations and time complexity

4. Conclusion

In this paper, a new protocol 2PSMC is proposed to compute secured sum 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 improve this algorithm with fuzzy logic and neural network learning.

References

[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. (journal style)
[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. (journal style)
[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. (journal style)
[5] 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)
[6] 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)
[7] 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)
[8] Charu. C. Agarwall., Philip.S.Yu, "Privacy Preserving Data Mining, Models and Algorithms", ISBN 978-0-387-70991-8,2008 (book style)
[9] Jaiwan.H., Michaline J., Jain P., "Data Mining Concepts and Techniques", Third Edition, 2012, ISBN 978-0-12-381479-1, (book style)
[10] Jaideep.V., Chris C., Michael Z., "Privacy preserving Data Mining", 2006, ISBN-13: 978-0-387-25886-8 (book style)

Author Profile

S. Selva Rathna received M.C.A degree in 2000 and M.Tech in Information Technology in 2010 in Manonmaniam sundaranar university, Tirunelveli, Tamil Nadu, India. She is presently doing Ph.D in Computer Science in Manonmaniam sundaranar university, Tirunelveli, Tamil Nadu, India.. She is highly interested in topics like data mining, data ware housing, privacy preservation, image processing etc. Her 14 years of experience in Oracle data base has supported her a lot in successful completion of this paper.

Dr. T.Karthikeyan has completed his Ph.D degree in Computer Science and presently working as Associate Professor in P.S.G. Arts and Science College, Coimbatore, Tamil Nadu, India. His extensive knowledge in Data mining, Image processing, Image mining, Security and privacy preserving data mining has supported a lot in conceptualizing this research.