IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 2, Issue 1, Feb-Mar, 2014
ISSN: 2320 - 8791
www.ijreat.org

Extended Distributed Rk- Secure Sum Protocol in Apriori Algorithm for privacy preserving

Meera Treesa Mathews[1], Manju E.V[2]
Department of Computer Science and Engineering
Sathyabama University, Chennai

Abstract- Secure sum computation is a simple example of Secure Multi party Computation. This provide privacy to data in case more than two parties are present, while finding combined results of individual data. Association rule mining algorithms like Apriori are used for mining frequent items from database. In this paper we address secure mining of frequent items from a horizontally partitioned data. It uses Apriori algorithm for mining frequent items with the help of a Extended Distributed Rk- Secure Sum Protocol for privacy preserving.

Keywords- Secure sum computation, secure multi party computation, Apriori, Extended Distributed Rk- secure sum protocol, frequent items, Global support

I. INTRODUCTION

Database can be of two types: Centralized Database and Distributed Database. Centralized Database consist of a central server and all the data is stored in a centralized manner. Distributed Database have database partitioned into different server or parties. In this paper, distributed database is addressed. Centralized Database is not considered due to privacy concerns. Database can be partitioned in 3 ways: Horizontal partitioning, vertical partitioning and hybrid Partitioning. In horizontal partitioning, database is partitioned row wise so that number of attributes remains same in all the partitions but number of rows may vary. In vertical partitioning, database is partitioned column wise so that number of rows remains same but number of attributes varies. Hybrid partitioning is a combination of horizontal and vertical partitioning. Here we consider only a horizontally partitioned distributed database.

The goal is to produce frequent items that hold globally without breaching privacy. In order to provide privacy while finding global result we use protocols like secure sum protocol and secure multi party computation. Secure sum protocol is used in case of two parties whereas Secure Multi party computation is used in case more than two parties are present. Here we have tried to implement secure multi party computation.

In this paper we have tried for mining frequent items using Apriori algorithm with the help of a Extended Distributed Rk- secure sum protocol for privacy preserving in a distributed database.

II. BACKGROUND WORK

The research [12] suggests a toolkit of components that can be combined for specific privacy preserving data mining problem. It mention about secure sum, which is a small example of secure multiparty computation, because of its applicability in Data mining. This paper also discusses application of these techniques in data mining like in Association rule mining in horizontally partitioned data. The research [13] [14] addresses specific problems like protocol for statistical analysis in cooperative environment and solutions to some specific geometric problems.

The paper [11] addresses secure mining of association rules over horizontally partitioned data. It assumes three or more parties. The method it follows is a two phase approach. A fast algorithm for distributed association rule mining (FDM) is used. In the first phase, it uses a commutative encryption. Each party encrypts its own itemsets and send to next party. The final result is the common itemsets. In the second phase each of the locally supported itemset is tested to see if it is supported globally. Here communication and encryption cost is significantly increased. The research [15] have identified and defined different SMC problems that can facilitate problem discovery task. They have addressed privacy preserving data mining problems. The [9] gives a tutorial like introduction to secure multiparty computation. They also describe basic tools and paradigms used in constructing secure protocol.

The paper [3] addresses a hybrid technique for secure sum protocol. They have enhanced the security and privacy. The paper [1] tried to implement privacy preserving data mining on horizontally partitioned distributed database. They have used FP tree algorithm for data mining and implemented Hybrid secure sum protocol for preserving privacy.

The paper [17] presents an efficient algorithm for mining association rules in large retailing company database. It includes buffer management and novel estimation and pruning technique. In [16] a fast algorithm for mining is addressed. Apriori algorithm [18] [19] is a classical algorithm for mining association rules. In [4] author uses apriori algorithm for mining association rule. Author considers bank data and tries to obtain result using Weka, a data mining tool. In paper [5], a comparative study of 4 association rule mining algorithms. It compares Apriori association rule, Predictive Apriori association rule, Tertius Association rule and Filtered associator and they found Apriori association rule algorithm performance is better.

Many researches are going on in the secure multi party computation, especially in secure sum computation. The paper [10] presents a secure sum computation where the parties are arranged in a ring and each party is divided into a fixed number of segments. They have used a randomization technique for privacy. In the paper [6] each party divides it data into fixed number of segments and redistributes the segment before computation. In Changing neighbour k- Secure sum protocol [7][8] data of each party is broken into number of segments and change the arrangement of the parties in each round. This provides more security to private data. Here also parties are arranged in a ring. The Distributed Rk- Secure sum protocol [2] is another secure sum protocol where parties are arranged in a bus network. They also present a changing neighbour approach for more security.

III. PRIVACY PRESERVING DATAMINING

Privacy preserving data mining is done by first mining frequent items from individual parties using Apriori algorithm, then applying Extended Distributed Rk- secure sum protocol to get a global result.

Fig. 1 shows the work flow. After mining the frequent items from each party, the partial support of these frequent items are calculated in every Party.

Partial Support is calculated using the following formula:

Partial support (PS) = X.Support – Minimum support * size of Database. (where X is the frequent item)

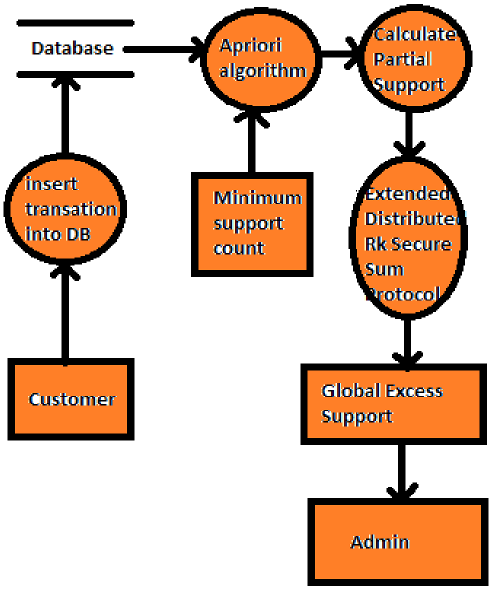
Database
Apriori algorithm
Calculate Partial Support
insert transation into DB
Minimum support count
Extended Distributed Rk Secure Sum Protocol
Customer
Global Excess Support
Admin

Fig.1 Work flow diagram

Now Global excess support of an item can be found by calculating the sum of partial support of that item in all the parties. This is done using Extended Distributed Rk- Secure sum Protocol. This provides privacy.

A. Apriori algorithm

Apriori algorithm is one of the most famous algorithms for finding frequent item sets. It follows a bottom up strategy. Frequent item sets are sets of items having minimum support. Support is an indication of how frequently an item occurs in a database. These frequent item set can be used to build association rules.

Apriori algorithm starts by scanning the database and finding all the frequent 1-itemsets (L_1) with their counts. Now candidate itemsets are generated by performing join operation on L_1 and then scan database to find the count of each candidate generated. Now we remove all the candidates with count less than minimum support. This creates L2 (all frequent 2-itemset). By performing join operation on L2 next candidate itemset is generated and the process continues till table L_k is empty. Now the frequent items are generated. Fig. 2 shows an example of generating frequent items using Apriori algorithm.

1) Psuedocode: Apriori

C_k: Candidate itemset of size k
L_k: frequent itemset of size k

```
L1={frequent items};
for(k=1; Lk != Ø; k++) do begin
        Ck+1= candidates generated from Lk;
        for each transaction t in database do
                increment the count of all
                candidates in Ck+1 that are
                contained in t
        Lk+1= candidates in Ck+1 with min_sup
End
return UkLk ;
```

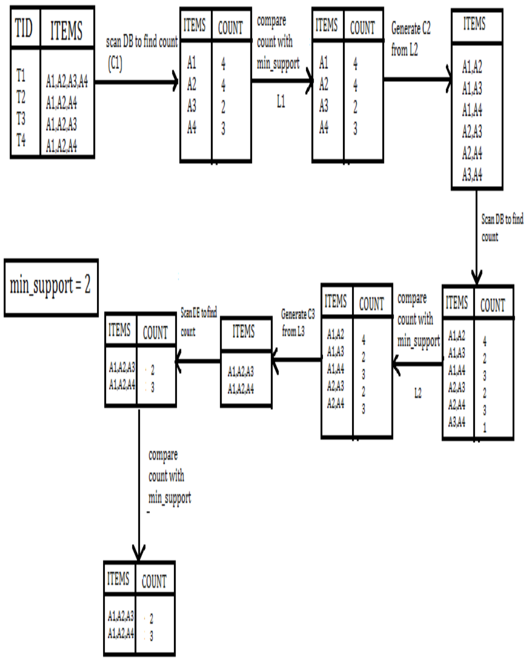

Fig. 2 Generating frequent items

B. Distributed Rk- Secure sum protocol

The Existing distributed Rk- secure sum algorithm is a secure multi party computation protocol. All the parties are arranged in the form of a bus network (P1, P2,......, Pn). P1 is selected as protocol initiator and Pn as end party. In each round P2 exchanges its position with next party. After completing n-1 rounds we get the result from P2.

1) Algorithm: Distributed Rk- Secure sum protocol

```
Step1: Assume the number of parties is n,

        P1,P2,P3,......,Pn (n>=3)
Step2: Each parties (P1, P2,.....,Pn) have their
        data D1,D2,....,Dn.
Step3:Arrange the parties in a bus structure
        (P1,P2,....,Pn) and select P1 as a
        protocol initiator.
Step4: Assume that RC = n and Pi (RC is a
        round counter and Pi is partial
        support).
        Partial support will calculate using the
        following formula
        $P_i = P_{(i-1)} + RC$
Step5: While RC!=0
        Begin
          Begin
            For 1 to n do
               Begin
                  Starting from P1 each party
                  will calculate their partial
            support and send to the next party
            in the bus
        End
         P2 exchange its position with the next
         party present in the bus till Pn.
       End
       RC = RC – 1
     End
Step6: Party P2 will announce  the result after
       calculating the from all the parties.
Step7: End of algorithm
```

C. Extended distributed Rk- Secure sum Protocol

Extended distributed Rk- secure sum protocol is an improvement over the distributed Rk-secure sum Protocol. In Distributed Rk- secure sum protocol if more than two parties join together, they can know the data of some party. The extended Rk- secure sum protocol tries to remove these drawbacks.

Here all parties are arranged in a Bus network (P1, P2.....,Pn). Each party divides its data into n-1 number of segments. Also each party will be having a random number each. The algorithm for Extended Distributed Rk- secure sum protocol is as follows:

1) Algorithm: Extended distributed Rk- Secure sum

```
1.  Assume the number of parties is n,
P1,P2,P3,.....,Pn (n>=3)
2. Each parties (P1, P2,.....,Pn) have their data
D1,D2,....,Dn.
3. Each party divide its data Di into 3 segments
Di1,Di2, Di3.
4. Each parties (P1,P2....,Pn) select their Random
number R1,R2,....Rn.
5. Arrange the parties in a bus structure (
P1,P2,.....,Pn) and then select P1 as a
protocol initiator.
6.  Assume that RC = 3 and  (RC is a round
counter) and calculate sum S using the formula
        $S = S + Dji + Rj$
Step7:While RC!= 1
        Begin
            Begin
               For j = 1 to n
                  Begin
```

```
                    Starting from P1 each party find
their sum and send to next party in the bus till Pn.
Pn also add the sum calculated in previous round
and store the value.
                    End
                    P2 Exchanges its position to next
party present in the bus till Pn
          End
          RC = RC – 1
          End
8. P2 will be having the sum S. P2 subtracts its
Random number multiplied by n-1 and send it to
party present just before in the bus network.
          Each Party does the same and send the
          sum in reverse direction.
9. Now Protocol Initiator P1 will announce the
result.
10. End of algorithm
```

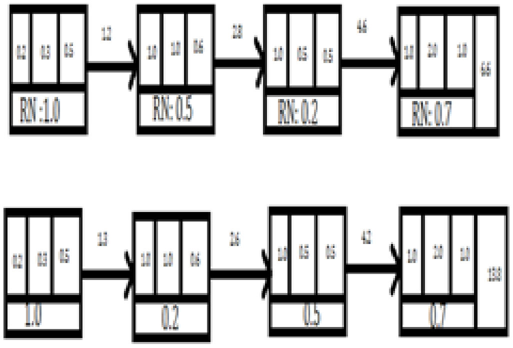

Fig. 3 Snapshot of first two rounds in Extended Distributed Rk- Secure Sum Protocol

IV. ANALYSIS OF EXTENDED DISTRIBUTED RK- SECURE SUM PROTOCOL

Case I: If any party become malicious

If any party become malicious we cannot obtain the correct result.

Case II: If any two parties collude:

Extended Distributed Rk- Secure Sum Protocol provides privacy to two colluding neighbours as the parties change their position in each round.

Case III: If more than two parties collude:

Even if more than two parties join together to know the data of some party it is not possible as the data is divided into segments and random numbers are added with the segment. The colluding parties can know only its own data.

Case IV: If all parties are honest:

If all parties are honest we can obtain the correct result within n-1 rounds. Then the addition of random numbers is not necessary.

In this Extended Distributed Rk- secure sum protocol, even if number of parties is increased then number of rounds remains same. The number of rounds depends on number of segments. Here number of segments is 4 and therefore both computational and communication complexity comes to 4N. Therefore both computation and communication complexity is O(n).

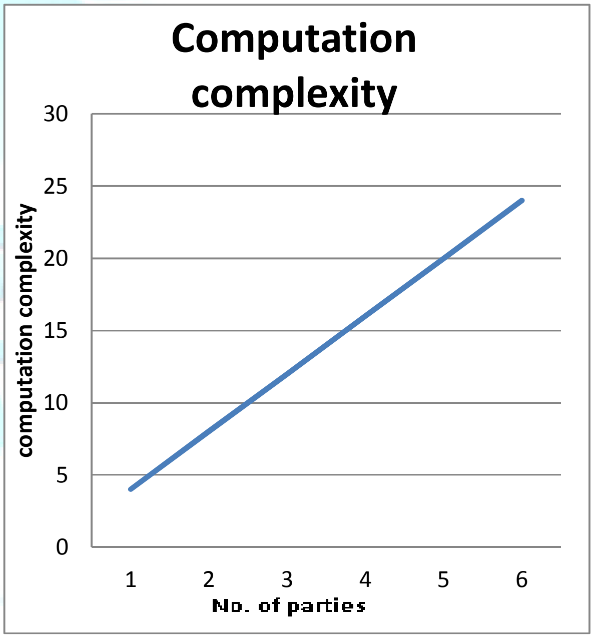

Fig. 4 Computational complexity of Extended Distributed Rk secure sum protocol

V. CONCLUSION

In this paper, privacy preserving data mining is implemented using Extended Distributed Rk-Secure Sum protocol in Apriori algorithm. Extended Distributed Rk- Secure Sum Protocol is a Secure Multi party Computation Protocol. It is an extension of Distributed Rk- Secure Sum Protocol. The proposed algorithm provide more security in computation and more privacy to data. The computation and communication complexity is O(n). In future we can try to reduce the complexity of protocol.

REFERENCES

[1] Jyotirmayee Rautaray, Raghavendra Kumar, "FP Tree Algorithm using Hybrid Secure Sum Protocol in Distributed

PRDG

Database", International Journal of Scientific & Engineering Research, Vol. 4, No. 3,march 2013, pp 1-5.

[2] Jyotirmayee Rautaray, Raghavendra Kumar, "Distributed Rk- Secure sum Protocol for Privacy Preserving", *IOSR Journal of Computer Engineering, Vol. 9, No. 1, January – February 2013, pp. 49 – 52*

[3] Priyanka Jangde, Gajendra Singh chandel, Durgesh Kumar Mishra, "Hybrid Technique for Secure Sum Protocol", World of Computer science and Information Technology Journal, Vol. 1, No. 5, 2011, pp 198 – 201.

[4] . Ms Shweta, Dr. Kanwal Garg, "Mining Efficient Association Rules through Apriori Algorithm Using Attributes and comparative analysis of various Association Rule Algorithms", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, No. 6, June 2013, pp. 306 – 312.

[5] Sunita B. Aher, Lobo L.M.R.J, " A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning", International Journal of Computer Applications (0975 – 8887), vol . 39, No. 1, February 2012, pp. 48 – 52.

[6] Rashid Sheikh, Beerendra Kumar, Durgesh Kumar Mishra," Distributed k-Secure sum Protocol for secure Multi – Party Computations", Journal of Computing, Vol. 2, No. 3, March 2010, pp. 68 – 72.

[7] Rashid Sheikh, Beerendra Kumar, Durgesh Kumar Mishra, "Changing Neighbour k- Secure Sum Protocol for secure multi party computation", International Journal of Computer science And Information Security, Vol. 7, No. 1, 2010, pp. 239 – 243.

[8] Rashid Sheikh, Beerendra Kumar, Durgesh Kumar Mishra," A modified ck-Secure Sum Protocol for multi party computation", Journal of Computing, Vol. 2, No.2, February 2010, pp. 62 – 66.

[9] Yehuda Lindell, Benny Pinkas, "Secure Multiparty Computation for Privacy-Preserving Data Mining", The Journal of Privacy and Confidentiality, vol.1, No.1, 2009, pp. 59-98.

[10] Rashid Sheikh , Beerendra Kumar, "Privacy Preserving K- Secure Sum Protocol", International Journal of Computer science and Information Security, Vol. 6, No. 2, 2009, pp. 184 – 188.

[11] Murat Kantarcioglu, Chris Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", IEEE Transactions on Knowledge And Data Engineering, Vol. 16, No. 9, September 2004

[12] Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, Michael Y. Zhu, "Tools for Privacy Preserving Distributed Data Mining", SIGKDD Explorations, Newsletter, vol.4, no.2, ACM Press, 2002, pp.28-34.

[13] M. J. Atallah and W. Du. "Secure Multiparty Computational Geometry," In proceedings of Seventh International Workshop on Algorithms and Data Structures(WADS2001). Providence, RhodeIsland, USA, 2001, pp. 165-179.

[14] W. Du and M.J.Atallah, "Privacy-Preserving Statistical Analysis," In proceedings of the 17th Annual Computer Security Applications Conference, New Orleans, Louisiana, USA, 2001, pp. 102-110.

[15] Wenliang Du, Mikhail 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, Sep. 11-13 2001, page 11-20.

[16] Rakesh Agrawal, Ramakrishnan Srikant,"Fast algorithm for mining association rules", Proceedings of the 20th VLDB Conference Santiago,Chile, 1994.

[17] Rakesh Agrawal, Tomasz Imielinski, Arun Swami, "Mining Association Rules between Sets of items in Large Databases", Proceedings of the 1993 ACM SIGMOD conference Washington DC, USA, May 1993, pp. 1- 10.

[18] Jiawie Han, Micheline Kamber, Jian Pie, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishing, Third edition, pp 248 – 253.

[19] http://www.cs.sunysb.edu/~cse634/lecture_notes/ 07apriori.pdf accessed on date 02-02-2012.