AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

Improved Distributed Rk- Secure Sum Protocol in Apriori Algorithm for Privacy Preserving

[1]E. Gokulakannan and [2]Dr.K.Venkatachalapathy

[1]*Research Scholar, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar India.*
[2]*Professor, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar -608 002, India.*

ARTICLE INFO

Article history:
Received 20 August 2014
Received in revised form
12September 2014
Accepted 22 October 2014
Available online 5 November 2014

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

ABSTRACT

Secure sum computation could be a straightforward case of Secure Multi Party Computation. this offers privacy to information just in case over to parties area unit present, whereas finding combined results of individual information. Association rule mining algorithms like Apriori area unit used for mining several things from the information. During this report, we tend to address secure mining of the many things from a horizontally divided data. It uses Apriori algorithmic program for mining frequent things with the help of an Extended Distributed RK- Secure add Protocol for privacy preserving. In this paper, we proposed a novel changing neighbors approach Distributed Rk secure sum protocol for achieving more security in case a group of the parties collude to know the private data of some other party.

© 2014 AENSI Publisher All rights reserved.

To Cite This Article: E. Gokulakannan and Dr.K.Venkatachalapathy., Improved Distributed Rk- Secure Sum Protocol in Apriori Algorithm for Privacy Preserving., ***Aust. J. Basic & Appl. Sci.,*** *8(17): 251-256, 2014*

INTRODUCTION

The database can be of two types: Centralized Database and Distributed Database. Centralized Database contains of a vital host and all the information is stored in a determined way. Distributed Database has a database partitioned into different server or parties. In this work, distributed database is addressed. Centralized Database is not measured due to privacy concerns. The database can be partitioned in 3 ways: Horizontal partitioning, Vertical partitioning and Hybrid Partitioning. In horizontal partitioning, database is partitioned row wise, so that the 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 the 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 end 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 attempted to implement secure multi party computation.

In this report, we have tried for mining frequent items using an Apriori algorithm with the assist of an Extended Distributed Rk- secure sum protocol for privacy preserving in a broadcast database.

1.1 Distributed Database Partition:

Database is divided into three main portioning horizontal partition,vertical partition and hybrid partition.

1.1.1 Horizontal Partitioning:

Horizontal partitioning divides the whole database into the number of small database according to the row splitting. So that the execution of query will be very fast as well as we will be able to provide more privacy to the portioned database.

1.1.2 Vertical Partitioning:

Vertical portioning divides the whole database into a number of small databases according to the column. So that portioned database does not contain any of duplicate data. There are mainly two types of vertical database normalized and row splitting.

Corresponding Author: E. Gokulakannan, Research Scholar ,Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar India.
E-mail: gokul550@gmail.com

Australian Journal of Basic and Applied Sciences, 8(17) November 2014, Pages: 251-256

1.1.3 Hybrid Partitioning:

Hybrid partitioning first divide the database into horizontal portioning and then vertical or first vertical and then horizontal portioning it's depend upon the user requirements.

II. Background Work:

The research suggests a toolkit of components that can be combined for specific privacy preserving data mining problem. It mentions about secure sum, which is a minor example of secure multiparty computation, because of its applicability in Data mining.

This report also discusses the application of these techniques in data mining like in Association rule mining in horizontally partitioned data. The research addresses specific problems like protocol for statistical analysis in cooperative environment and solutions to some specific geometric problems. The paper addresses, secure mining of association rules over horizontally partitioned data. It takes three or more parties. The method it follows is a two phase approach. A fast algorithm for distributed association rule mining (FDM) is applied. In the first phase, it uses a commutative encryption. Each party encrypts its own itemsets and send to the next party. The final result is the common itemsets. In the second phase, each of the locally supported itemset is tested to determine if it is held globally. Here communication and encryption cost is significantly increased. The research have identified and defined different SMC problems that can facilitate problem discovery task. They have addressed privacy preserving data mining problems. This gives a tutorial like introduction to secure multiparty computation. They also describe basic tools and paradigms used in constructing secure protocol.

The paper addresses a hybrid technique for secure sum protocol. They have enhanced the security and privacy. The paper 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 report gives an effective algorithm for mining association rules in a large retailing company database. It includes buffer management and novel estimation and pruning technique. In a fast algorithm for mining is addressed. Apriori algorithm is a classical algorithm for mining association rules. In 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, a comparative study of 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 working on in the secure multi party computation, particularly in secure sum computation. The paper 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 each party divides its information into a set number of segments and redistributes the segment before computation. In Changing neighbor k- Secure sum protocol data of each party is split into a number of segments and change the arrangement of the parties in each cycle. This offers more security to individual information. Here also parties are arranged in a ring. The Distributed Rk- Secure sum protocol is another secure sum protocol where parties are set in a bus network.

III. Privacy Preserving Data Mining:

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 is 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).

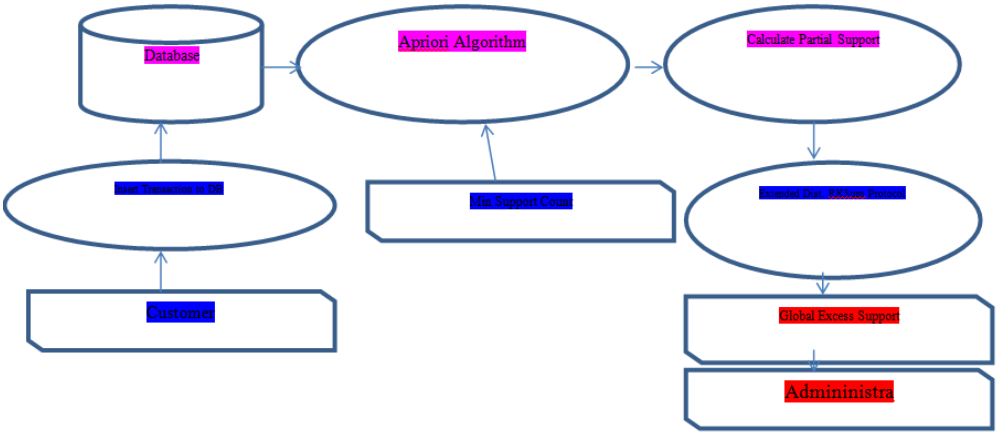


Fig. 1: Work flow diagram.

Australian Journal of Basic and Applied Sciences, 8(17) November 2014, Pages: 251-256

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 gives privacy.

IV. Apriori algorithm:

General Process:

Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all frequent itemsets in a database.
2. Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

While the second step is straight forward, the first step needs more attention. Finding all frequent itemsets in a database is difficult since it involves searching all possible itemsets (item combinations). The set of possible itemsets is the power set over I and has size $2^n - 1$ (excluding the empty set which is not a valid itemset). Although the size of the powerset grows exponentially in the number of items n in I, efficient search is possible using the downward-closure property of support (also called anti-monotonicity) which guarantees that for a frequent itemset, all its subsets are also frequent and thus for an infrequent itemset, all its supersets must also be infrequent. Exploiting this property, efficient algorithms (e.g., Apriori and Eclat) can find all frequent itemsets.

A. Apriori Algorithm Pseudocode:

Procedure Apriori (T, minSupport) {//T is the database and minSupport is the minimum support
L1= {frequent items};
for (k= 2; L_{k-1} !=∅; k++) {
C_k= candidates generated from L_{k-1}
//that iscartesian product L_{k-1} x L_{k-1} and eliminating any k-1 size itemset that is not
//frequent
for each transaction t in database do{
#increment the count of all candidates in C_k that are contained in t
L_k = candidates in C_k with minSupport
}//end for each
}//end for
return $\cup_k L_k$;
}

As is common in association rule mining, given a set of itemsets (for instance, sets of retail transactions, each listing individual items purchased), the algorithm attempts to find subsets which are common to at least a minimum number C of the itemsets. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

Apriori uses breadth-first search and a tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k − 1. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.

Apriori, while historically significant, suffers from a number of inefficiencies or trade-offs, which have spawned other algorithms. Candidate generation generates large numbers of subsets (the algorithm attempts to load up the candidate set with as many as possible before each scan). Bottom-up subset exploration (essentially a breadth-first traversal of the subset lattice) finds any maximal subset S only after all 2 | S | − 1 of its proper subsets.

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 (R1, R2,....., Rn). R1 is selected as protocol initiator and Rn as end party. In each round P2 exchanges its position with next party. After completing n-1 rounds we get the result from R2.

1) Algorithm: Distributed Rk- Secure sum protocol
Step1: Assume the number of parties is n, R1,R2,R3,.....,Rn (n>=3)
Step2: Each parties (R1, R2,.....,Rn) have their data D1,D2,....,Dn.
Step3:Arrange the parties in a bus structure (R1,R2,....,Rn) and select R1 as a
protocol initiator.
Step4: Assume that PC = n and Pi (PC is a round counter and Pi is partial support).
Partial support will calculate using the following formula
Ri = R(i-1) + PC

Australian Journal of Basic and Applied Sciences, 8(17) November 2014, Pages: 251-256

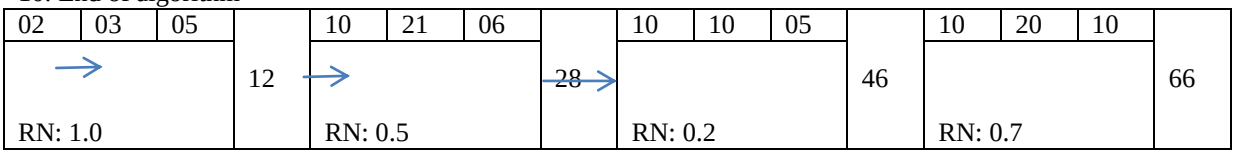
Step5: While PC!=0
Begin
Begin
For 1 to n do
Begin
Starting from R1 each party will calculate their partial
support and send to the next party in the bus
End
R2 exchange its position with the next party present in the bus till Rn.
End
PC = PC – 1
End
Step6: Party R2 will announce the result after calculating 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 (R1, R2.....,Rn). Each party divides its data into n segments. Also each party will be having a each. The algorithm for Extended Distributed Rk sum protocol is as follows:

1) Algorithm: Extended distributed Rk- Secure sum
1. Assume the number of parties is n, R1,R2,R3,.....,Rn (n>=3)
2. Each parties (R1, R2,.....,Rn) have their data D1,D2,....,Dn.
3. Each party divide its data Di into 3 segments Di1,Di2, Di3.
4. Each parties (R1,R2....,Rn) select their Random number P1,P2,....Pn.
5. Arrange the parties in a bus structure (R1,R2,....,Rn) and then select R1 as a protocol initiator.
6. Assume that PC = 3 and (PC is a round counter) and calculate sum S using the formula
S = S+ Dji +Pj
Step7:While PC!= 1
Begin
Begin
For j = 1 to n
Begin
Starting from R1 each party find their sum
and send to next party in the bus till Rn. Rn also add the sum
calculated in previous round and store the value.
End
R2 Exchanges its position to next party
present in the bus till Rn
End
PC = PC – 1
End
8. R2 will be having the sum S. R2 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 reverse direction.
9. Now Protocol Initiator R1 will announce the result.
10. End of algorithm



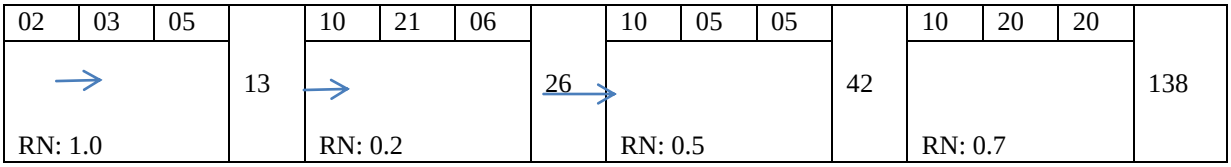


Fig. 2: Snapshot of first two rounds in Distributed RK- Secure Sum Protocol.

Australian Journal of Basic and Applied Sciences, 8(17) November 2014, Pages: 251-256

IV. ***Analysis Of Improved Secure Sum Protocol:***

Case I: If any party becomes 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 if number of parties is increased then number of rounds remains same. The number of rounds depends on segments. Here the number of segments is 4 and therefore computational and communication complexity comes to 4N. Therefore, both computation and communication complexity is O (n).

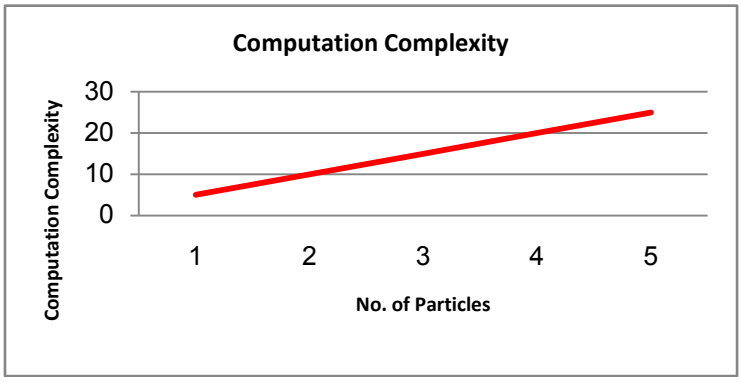


Fig. 3: Computational complexity of Extended Distributed RK secure sum protocol.

V. Conclusion:

In this work, privacy preserving data mining is implemented using Extended Distributed Rk Apriori algorithm. Extended Distributed Rk 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 the protocol.

REFERENCES

Agrawal, R. and R. Srikant, 2000. L "Privacy-Preserving Data Mining," In proceedings of the 2000 ACM SIGMOD on management of data, Dallas, TX USA.

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

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

Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, Michael Y. Zhu, 2002. "Tools for Privacy Preserving Distributed Data Mining", SIGKDD Explorations, Newsletter.

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

Jyotirmayee Rautaray, Raghavendra Kumar, 2013. "Distributed RK- Secure sum Protocol for Privacy Preserving", IOSR Journal of Computer Engineering.

Jyotirmayee Rautaray, Raghavendra Kumar, 2013. "FP Tree Algorithm using Hybrid Secure Sum Protocol in Distributed Database", International Journal of Scientific & Engineering Research.

Lindell, Y. and B. Pinkas, 1880. "Privacy preserving data mining," in advances in cryptography-Crypto2000, lecture notes in computer science.

Ms Shweta, Dr. Kanwal Garg, 2013. "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.

Murat Kantarcioglu, Chris Clifton, 2004. "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", IEEE Transactions on Knowledge And Data Engineering.

Priyanka Jangde, Gajendra Singh chandel, Durgesh Kumar Mishra, 2011. "Hybrid Technique for Secure Sum Protocol", World of Computer science and Information Technology Journal.

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

Rakesh Agrawal, Tomasz Imielinski, Arun Swami, 1993. “Mining Association Rules between Sets of items in Large Databases”, Proceedings of the 1993 ACM SIGMOD conference Washington DC, USA.

Rashid Sheikh, Beerendra Kumar, 2009. “Privacy Preserving K- Secure Sum Protocol”, International Journal of Computer science and Information Security.

Rashid Sheikh, Beerendra Kumar, Durgesh Kumar Mishra, 2010. ”A modified ck- Secure Sum Protocol for multi party computation”, Journal of Computing.

Rashid Sheikh, Beerendra Kumar, Durgesh Kumar Mishra, 2010. ”Changing Neighbour k- Secure Sum Protocol for secure multi party computation”, International Journal of Computer science And Information Security.

Rashid Sheikh, Beerendra Kumar, Durgesh Kumar Mishra, 2010. ”Distributed k-Secure sum Protocol for secure Multi –Party Computations”, Journal of Computing.

Sunita B. Aher, L.M.R.J. Lobo, 2012. “ A Comparative Study of Association Rule Algorithms for Course Recommender System in Elearning”, International Journal of Computer Applications.

Wenliang Du, Mikhail J. Atallah, 2001. “Secure MultiParty Computation Problems and Their Applications: A Review and Open Problems”, In proceedings of new security paradigm workshop, Cloudcroft, New Maxico, USA.

Yehuda Lindell, Benny Pinkas, 2009. “Secure Multiparty Computation for Privacy-Preserving Data Mining”, The Journal of Privacy and Confidentiality.