ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 1, Janauary 2013

FP-TREE ALGORITHM USING MODIFIED-CK SECURE SUM METHOD

Jyotirmayee Rautaray[1], Raghvendra Kumar[2]

School of Computer Engineering, KIIT University, Odisha, India[1]

School of Computer Engineering, KIIT University, Odisha, India[2]

Abstract: Privacy consideration is frequently moderation of data mining. This paper addresses the problem of association rule mining using the FP tree algorithm where operation is scattered across multiple parties. Each party holds the number of transaction and the parties wish to work together to recognize globally valid Association rule. We presents the multi party transaction data that discovers frequent item sets with minimum support, without either parties revelling the individual data.
.

Keywords: Horizontal Partitioning, FP tree algorithm, Secure multi party computation, Modified Ck secure sum.

I. INTRODUCTION

Privacy is one of the most significant properties that an information system requires. For this motivation, numerous efforts have been devoted to incorporate privacy preserving techniques [6] with data mining algorithms [1] [2] [3] in order to prevent the disclosure of sensitive information during the knowledge discovery. The existing privacy preserving data mining techniques can be classified according to the subsequent five different dimensions like first one is data distribution (centralized or distributed); second one is the adjustment applied to the data (encryption, perturbation, generalization, and so on), third one is the data mining algorithm in which the privacy preservation technique is designed, fourth one is the data type (single data items or complex data correlations) that needs to be protected from disclosure and last one is the approach adopted for preserving privacy [10] [11] (heuristic or cryptography-based approaches). While heuristic based techniques are mainly conceived for centralized datasets, cryptography-based algorithms are designed for protecting privacy in a distributed state of affairs by using encryption techniques. Heuristic-based algorithms recently proposed aim at hiding sensitive unprocessed data by applying perturbation techniques based on probability distributions. Furthermore, several heuristic-based approaches for hiding both raw and aggregated data from side to side, a hiding technique (k-anonymization, adding noises, data swapping, generalization and sampling) has been developed, first in the context of association rule mining and classification, in addition to more recently for clustering techniques. Given the number of different privacy preserving data mining techniques that have been developed in these years, there is an emerging need of moving toward standardization in this new research area, one step toward this essential process is to provide a quantification approach for privacy preserving data mining algorithms to make it possible to evaluate and compare such algorithms. However, due to the variety of characteristics of privacy preserving data mining algorithms, it is often the case that no privacy preserving algorithm exists that outperforms others on all possible criteria. Relatively, an algorithm may carry out better than another one on specific criteria like privacy level, data quality.

A. *FP Tree Algorithm*

Many algorithms recommended like Apriori algorithms. It is based upon the resistant monotone property. Due to their two main problems i.e. frequent database examine and superior computational cost, there is a need of compacted data structure for mining frequent item sets, which moderates the multi scan trouble and improve the candidate item set generation. Tree shelf is an efficient algorithm based upon the lexicographic tree in which each node represents a frequent pattern [1] [6].
FP-Growth algorithm [6] is a disciplined algorithm for producing the frequent item sets without production of candidate item sets. It is based upon divide and conquers approach. It needs database scan to discover all frequent item sets. This approach compresses the database of frequent item sets into frequent pattern tree recursively in the same order of magnitude as the sequence of frequent patterns. Then in next step divide the compressed database into set of conditional databases.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 1, Janauary 2013

B. *Multi Party Computation*

The problem arises when more than two parties want to share their data and want to find their global result without disclosing their individual data; it is referred to secure multi party computation [6]. Secure multi party computation problem that deals with any type of input or function in scattered database. There are many explanation and difficulty that will resolve using the secure sum multi party computation. But in most of the cases single party is considered as trusted party. Then that party will compute the result and distribute the result to all the party that present in the distributed environment. Principally secure multi party computation has two goals; one is to provide the security to the individual data and another is to correct the result. And secure multi party computation contains two main models one is Real model and Ideal model. Real model uses without trusted party and Ideal model uses for with trusted party (TTP) [12] [13] [14]. According the number of inputs the computation is classified into two models, single input computation model and another is multi input computation. The alteration from input to secure multi party computation contains three ways transformation. First one is convert multi party input to secure multi party computation, after the computation is performed. Second one is convert multi party computation to homogeneous secure multi party computation. Third one is convert multi party computation to heterogeneous secure multi party computation. Figure1 shows the Snapshot of Secure Multi Party Computation.

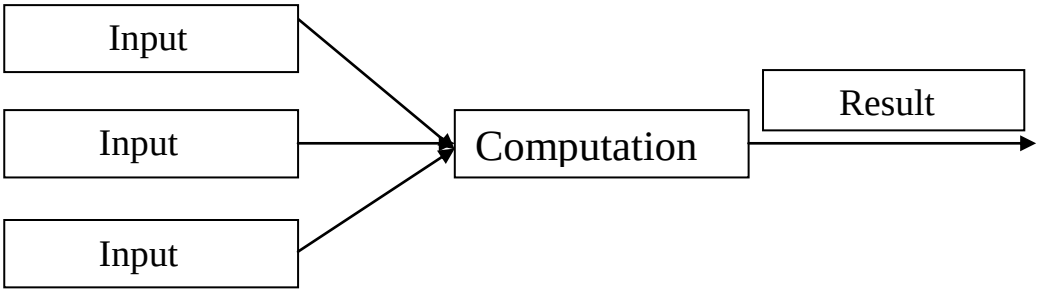


Figure1: - Snapshot of Secure Multi Party Computation

II. PROPOSED WORK

In secure sum protocol basically divide the whole database into number of horizontal partition database and each party have their particular sequence number (p1, p2,……p n) so that if p2 want the value of data of p1 and p3 then party p2 will never be able to know the data of other parties. So in modified Ck secure sum [9] [10] [11] [12] very useful technique to provide a security to database with data leakage is zero percent. Basically secure multi party computation problem uses two computation models, ideal model and real model, in ideal model their exit a trusted third party which will compute the result and broadcast the result to all other parties present in database. But in real model the parties agree on some protocol which allows all the parties to evaluate the function and find the global result. But in modified Ck Secure sum protocol here we change the position of every party at every time after computation of the first. If there are four parties then number of round is four (if there are n number of parties then numbers of round is n), but the global result will always be same in all round. So that in protocol we will able to provide high security to database and percentage of data leakage is zero. Figure 2 shows the movements of parties.
The algorithm can be described as following:
Step 1:- Separate the entire database into number of parties P1, P2, P3, P4, P5… Pn (n≥2).
Step2:- Every party will produce their own random number R1, R2, R3…… Rn.
Step3:- Join the number of parties into the circular ring (P1, P2……Pn).
Step4:-Give permission the party P1 as a protocol initiator.
Step5:- Party P1 will determine their partial support by using the following formula
Psi=X. support-minimum support*|DB|+RN1-RNn (RN is random number).
Step6:- Party 2 will calculate their partial support by using the following formula
Psj=Ps(j-1)+X. support-minimum support*|DB|+RN1-RN(i-1)
Step7:-In this way party P1 will adjust the position to the next party nearby in the environment till Pn-1.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 1, Janauary 2013

Step8:-Subsequent to that party P2 will calculate their global support and transmit the global support to all party presents in the surroundings.
Party P1 will decide the mask value by using the following formula
Mask value is determine by using two different hash function because here double hash based function is used
Key1=Hash (Key) =Key (Mod N)
And after that calculated another hash function
Mask Key=Hash2 (Key1) = Key+M^{key1}

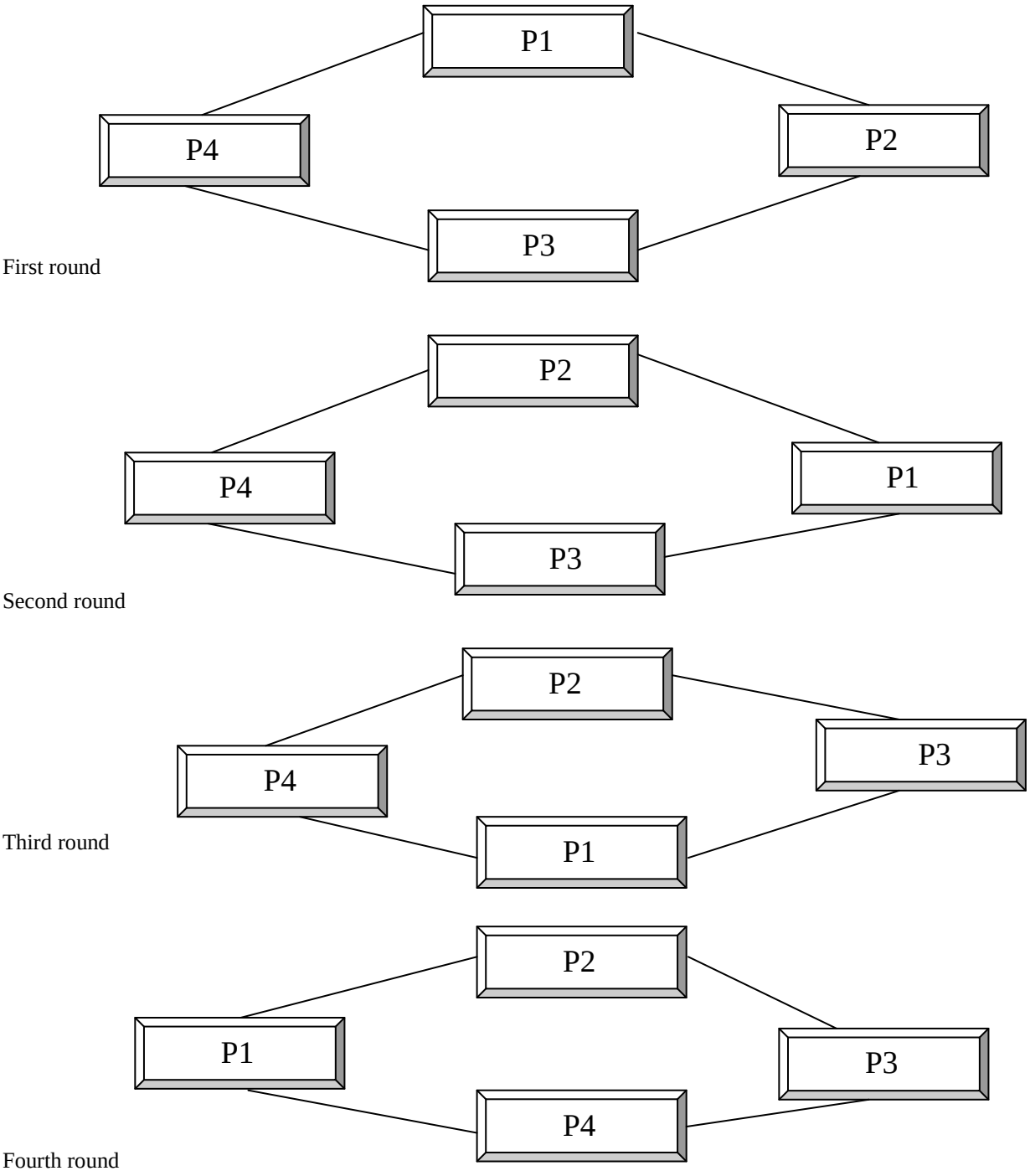


Figure 2: Movement of Party to Provide the Security to the Database

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 1, Janauary 2013

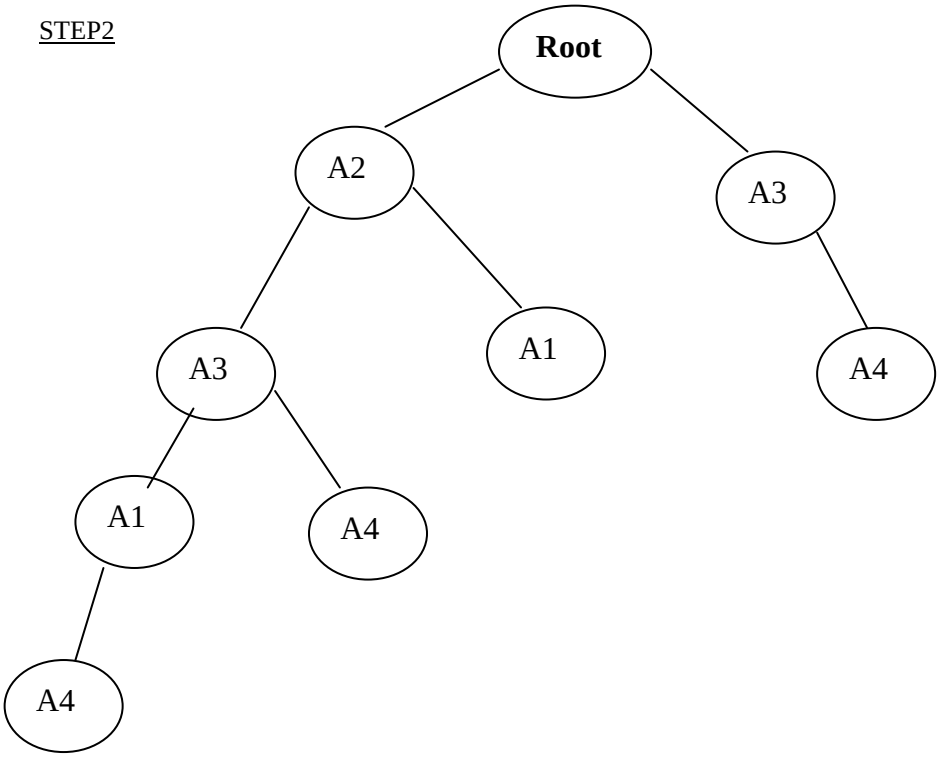
TABLE I
Set of data for Party1 MIN SUPPORT=40%

TID	A1	A2	A3	A4
T1	1	1	1	0
T2	0	1	1	1
T3	1	1	1	1
T4	1	1	0	0
T5	0	0	1	1

TID	LIST
T1	A1, A2, A3
T2	A2, A3, A4
T3	A1, A2, A3, A4
T4	A1, A2
T5	A3, A4

STEP1
A1:3, A2:4, A3:4, A4:3
SORT IN DESENDING ORDER
A2:4, A3:4, A1:3, A4:3

STEP2



STEP3
A4= {(A3:1), (A1:1, A3:1, A2:1), (A3:1, A2:1)}
A3= {(A2:3), (A3:1)}
A2= {A2:4}
A1= {(A3:2), (A2:2), (A1:1)}

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 1, Janauary 2013

STEP4
A4= {(A3:3), (A2:2)}
STEP5
A4A3:2, A4A2:2
SUPPORT (A4, A3) = COUNT (A4, A3)/T=3/5=O.6>40%
STEP6
SUPPORT (A4, A2) = COUNT (A4, A2)/T=2/5=O.4>40%
CANDIDATE SET = {A2, A3, A4}

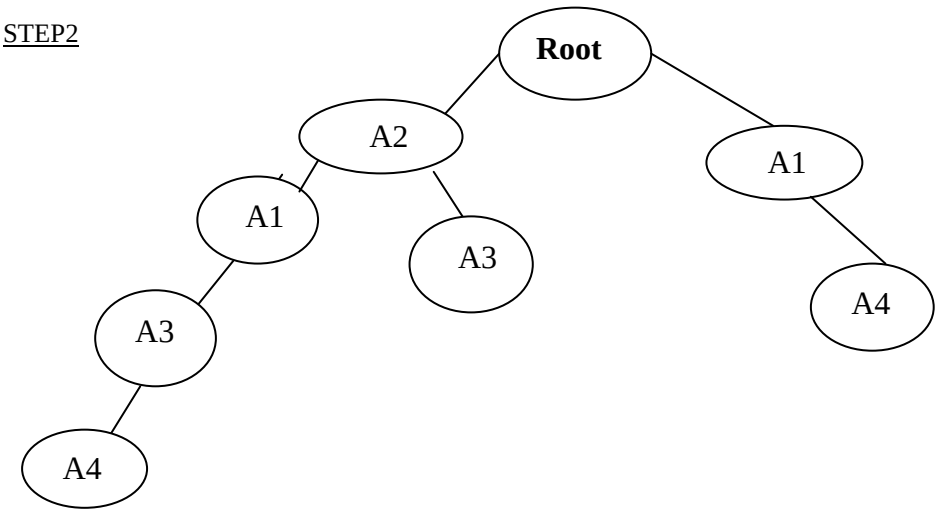
TABLE II
Set of data for Party2 MIN SUPPORT=40%

TID	A1	A2	A3	A4
T1	1	1	1	1
T2	0	1	1	0
T3	1	0	0	1
T4	0	1	0	0
T5	1	1	1	0

TID	LIST
T1	A1, A2, A3, A4
T2	A2, A3
T3	A1, A4
T4	A2
T5	A1, A2, A3

STEP1
A1:3, A2:4, A3:3, A4:2
SORT IN DESENDING ORDER:-
A2:4, A1:3, A3:3, A4:2

STEP2



STEP3
A4= {(A3:1, A1:1, A2:1), (A1:1)}
A3= {(A1:2, A2:2), (A2:1)}
A2= {(A2:4)}

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 1, Janauary 2013

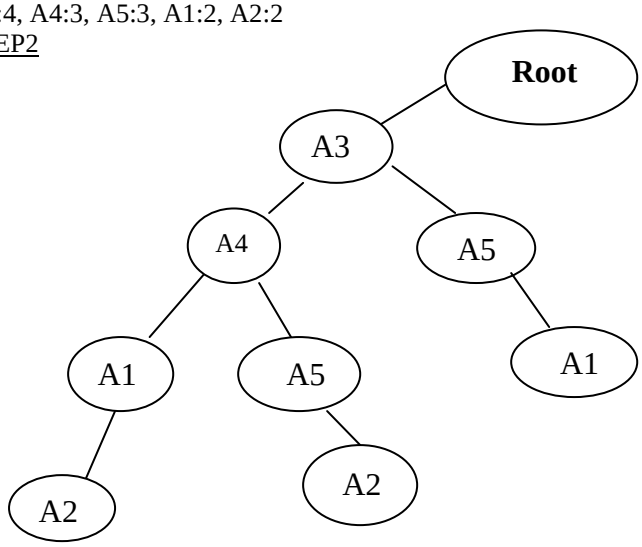
A1= {(A2:2), (A1:1)}
STEP4
A4= {A1:2}
A3= {A2:3}
STEP5
A4A1:2, A3A2:3
STEP6
SUPPORT (A4, A1) = COUNT (A4, A1)/T=2/5=0.4=40%
SUPPORT (A3, A2) =COUNT (A3, A2)/T=3/5=0.6>40%
SUCHTHAT CANDIDATE SET- {A1, A2, A3, A4}

TABLE III
Set of data for Party3 MIN SUPPORT=40%

Tid	A1	A2	A3	A4	A5
T1	1	1	1	1	0
T2	0	1	1	1	1
T3	0	0	1	1	1
T4	1	0	1	0	1

TID	LIST
T1	A1 A2 A3 A4
T2	A2 A3 A4 A5
T3	A3 A4 A5
T4	A1 A3 A5

STEP1
A1:2, A2:2, A3:4, A4:3, A5:3
SECONDARY ORDER:-
A3:4, A4:3, A5:3, A1:2, A2:2
STEP2



STEP3
A1= {(A5:1, A3:1), (A4:1, A3:1)}
A2= {(A1:1, A4:1, A3:1), (A5:1, A4:1, A3:1)}
A3= {(A3:4)}

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 1, Janauary 2013

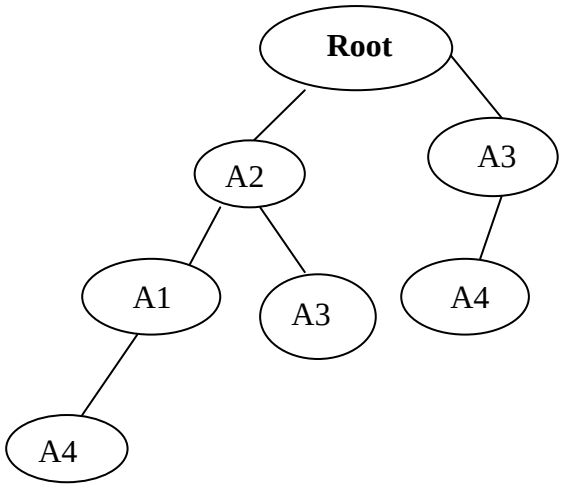
A4= {(A3:3)}
A5= {(A3:1), (A4:2, A3:2)}
STEP4
A1= {A3:2}
A2= {A3:2, A4:2}
A5= {(A3:3)}
STEP5
A1A3:2, A2A3:3, A2A4:2, A5A3:3
STEP6
SUPPORT (A1, A3) = COUNT (A1, A3)/T=2/4=0.5>40%
SUPPORT (A2, A3) = COUNT (A2, A3)/4=2/4=0.5>40%
SUPPORT (A2, A4) = COUNT (A2, A4)/4=2/4=0.5>40%
SUPPORT (A5, A3) = COUNT (A5, A3)/4=3/4=0.75>40%
SUCH THAT CANDIDATE SET-{A1, A2, A3, A4, A5}

TABLE IV
Set of data for Party4 MIN SUPPORT=40%

Tid	A1	A2	A3	A4
T1	1	1	0	0
T2	0	1	1	0
T3	1	1	0	1
T4	0	0	1	1

TID	LIST
T1	A1 A2
T2	A2 A3
T3	A1 A2 A4
T4	A3 A4

STEP1
A1:2, A2:3, A3:2, A4:2
IN DESENDING ORDER:-
A2:3, A1:2, A3:2, A4:2
STEP2



STEP3
A1= {(A2:2)}
A2= {(A2:3)}

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 1, Janauary 2013

A3= {(A2:1), (A3:1)}
A4= {(A1:1), (A2:1), (A3:1)}
AS HERE MIN SUPPORT=40%
SO, HERE NO CANDIDATE SET IS SELECTED.
At step1: After the calculation the list of Candidate set generated in table 1= {A2, A3, A4}
At step2: After the calculation the list of Candidate set in table 2= {A1, A2, A3, A4}
At step3: After the calculation the list of Candidate set in table 3 Candidate set= {A1, A2, A3, A4, A5}
At step4: After the calculation no Candidate set is generated in table 4
Consider the item set {A2}
Each parties select their random number RN1=10, RN2=20, RN3=10, RN4=10
Key =10, M=2
Hash key=key mod M
Mask key=hash key-M^{key}
Hash key=10 mod 2=0
Mask key=10-1=9
Consider the item set I= {A2}
PS=I1Support- Minimum support*DB + (RN I -RN (i-1)) + Mask key
Calculation for Round 1-
PS11= 4-.4*5+ (10-20) +9=1
PS12=4-.4*5+ (20-10) +1=13
PS13=2-.4*4+ (10-20) +13=3.4
PS14=3-.4*4+ (10-10) +3.4=4.8
Global encrypt support (GES) = Partial support-Mask key
GES=4.8-9=-4.2
Calculation for Round 2 –
PS12= 4-.4*5+ (20-10) +9=21
PS11=4-.4*5+ (10-20) +21=13
PS13=2-.4*4+ (10-10) +13=13.4
PS14=3-.4*4+ (10-10) +13.4=14.8
Global encrypt support (GES) = Partial support-Mask key
GES= 14.8-9=5.8
Calculation for Round 3 –
PS12= 4-.4*5+ (20-10) +9=21
PS13=2-.4*4+ (10-20) +21=11.4
PS11=4-.4*5+ (10-10) +11.4=13.4
PS14=3-.4*4+ (10-10) +13.4=14.8
Global excess support (GES) = Partial support-Mask key
GES= 14.8-9=5.8
Calculation for Round 4 –
PS12= 4-.4*5+ (20-10) +9=21
PS13=2-.4*4+ (10-20) +21=11.4
PS14=3-.4*4+ (10-10) +11.4=12.8
PS11=4-.4*5+ (10-10) +12.8=14.8
Global excess support (GES) = Partial support-Mask key
GES=14.8-9=5.8
If global excess support is greater than or equal to Zero than it is frequent or else it is Infrequent. So in this trouble the value of global excess support is positive, it means it is globally frequent. But A2 is frequent at party1, party2 and party3 so it's locally frequent and infrequent in party4. A2 construct it is a frequent after doing some of the computation in various parties. And after calculation the dummy item sets that convert infrequent to frequent or frequent to infrequent that makes that the successor parties will not at all able to know the previous result value of other parties.

III. CONCLUSION

In This paper addresses the problem of computing association rules using FP Tree algorithm within a scenario of homogeneous database. We take for granted that all parties have the identical schema, but every party does not have information on dissimilar Entities. The goal is to produce association rules that hold worldwide while limiting the information shared about each party. Many proposals have been partied to apply secure multi party computation. Secure multi party computation being used in huge scale databases which extends to protect privacy to the private data of dissimilar parties. In this paper our focus is based on horizontal partitioned Distributed data through a accepted association rule mining technique.

REFERENCES

[1]Agrawal, R., et al "Mining association rules between sets of items in large database". In: Proc. of ACMSIGMOD'93, D.C, ACM Press, Washington, pp.207-216, 1993.
[2]. Agarwal, R., Imielinski, T., Swamy, A. "Mining Association Rules between Sets of Items in LargeDatabases", Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data,pp. 207-210,1993.
[3]. Srikant, R., Agrawal, R "Mining generalized association rules", In: VLDB'95, pp.479-488, 1994.
[4]Agrawal, R., Srikant, R, "Privacy-Preserving Data Mining", In: proceedings of the 2000 ACM SIGMOD on management of data, pp. 439-450, 2000.
[5] Lindell, Y., Pinkas, B, "Privacy preserving Data Mining", In: Proceedings of 20th Annual International Cryptology Conference (CRYPTO),2000.
[6][2] J.Han, J.Pei and Y.Yin. "Mining frequent patterns without candidate Generation", in: Proceeding of ACM SIGMOD International Conference Management of Data, pp.1-12, 2000.
[7]Kantarcioglu, M., Clifton, C ,"Privacy-Preserving distributed mining of association rules on horizontally partitioned data", In IEEE Transactions on Knowledge and Data Engineering Journal, IEEE Press, Vol 16(9), pp.1026-1037, 2004.
[8] Han, J. Kamber, M "Data Mining Concepts and Techniques". Morgan Kaufmann, San Francisco, 2006.
[9]Sheikh, R., Kumar, B., Mishra, D, K, "A Distributed k- Secure Sum Protocol for Secure Multi-Party Computations". Journal of Computing, Vol 2, pp.239-243, 2010.
[10]Sugumar, Jayakumar, R., Rengarajan, C "Design a Secure Multi Party Computation System for PrivacyPreserving Data Mining". International Journal of Computer Science and Telecommunications, Vol 3, pp.101-105, 2012.
[11] N V Muthu Lakshmi, Dr. K Sandhya Rani ,"Privacy Preserving Association Rule Mining without Trusted Party for Horizontal Partitioned database", International Journal of Data Mining & KnowledgeManagement Process (IJDKP) Vol.2, pp.17-29, 2012.
[12] N V Muthu lakshmi, Dr. K Sandhya Rani, "Privacy Preserving Association Rule Mining in Horizontally Partitioned Databases Using Cryptography Techniques", International Journal of Computer Science and Information Technologies(IJCSIT), Vol. 3 (1) , PP. 3176 – 3182, 2012.
[13] Goldreich, O., Micali, S. & Wigerson, A. ,"How to play any mental game", In: Proceedings of the 19th Annual ACM Symposium on Theory of Computing, pp.218-229.
[14] Franklin, M., Galil, Z. & Yung, M.,"An overview of Secured Distributed Computing". Technical Report CUCS- 00892, Department of Computer Science, Columbia University.

Biography

Jyotirmayee Rautaray
School of Computer Engineering, KIIT University, Bhubaneswar, Odisha, India

Raghvendra Kumar
School of Computer Engineering, KIIT University, Bhubaneswar, Odisha, India