www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Volume-7, Issue-4, July-August 2017

International Journal of Engineering and Management Research

Page Number: 155-161

Design of a Secure Computation Protocols for Distributed Association Rule Mining on Partitioned Data

Shivaprasad S[1], Prasad Challa[2], Dr. A Ramaswamy Reddy[3], Dr. M Sadanandam[4]
[1,2]Department of CSE, VFSTR University, Guntur, INDIA
[3]Department of IT, VFSTR University, Guntur, INDIA
[4]Assistant Professor, Department of CSE, Kakatiya University, Warangal, INDIA

ABSTRACT

In several applications, data mining play vital role in a distributed database scheme. In this scenario misuse of data may exist. But sometimes, data owners may be related to such misuse. Therefore, data owners do not require their data to be mined, because which contain sensitive information. Privacy preserving is for providing security of the data in the data mining schema. In this paper, an encryption and decryption protocols for secure distributed association rule mining is proposed. We applied commutative encryption and decryption for privacy preserving distributed mining of association rules. This protocol provides better security and better performance than existing once.

Keywords-- Data Mining, Data Preserving, Encryption, Decryption, Association Rule Mining, Security

I. INTRODUCTION

Nowadays, Data Mining is becoming a lot of improvements in the process of Decision making problem. Data mining process is very much applied, but it's focus to many problems [1]. One of the important problem in data mining is preserving the privacy [2]. Most of the applications, data owners are worried about abuse of their data and they do not want sensitive information to be lost. With this reason, they don't want to provide their data for any data mining-related actions.

By using privacy preserving technique we should maintain the data very secret that is not rising to the other party which is running that method (Algorithm). Privacy preserving data mining has common method.

In this method, the data is separated between two or more different parties: Here the main theme of privacy preserving data mining is, to run the algorithms on different parties but we don't allow each party to observe another party secret data. For example, consider that there are different number of hospitals that are used to join their patient data for that reason of medical research. Here, everyone should have the eligibility to view patient details but only some authentication members have the right to change datais of every patient. Coming to Distributed data, we think privacy preservation only concentrates on useful knowledge which is mined from different sites. But it also concerns about the original data in every site whether it is securely hidden or not.

1.1 Related works

The PPDM job in the central data storage state is addressed by two independent research papers [4, 5]. From the disturbed values of individual records present in the training dataset a decision tree classifier is constructed. The key idea is to exactly estimate the allocation of original data standards through original restoration process.

To regenerate circulation process are using, this classifier to constructs whose efficiency is equal to the efficiency of classifiers to construct the original data.

Meantime, the PPDM function addresses the distributed data storage scheme [5]. The ID3 decision tree learning target to construct the useful cryptographic protocols established secure multi-party computing. PPDM represent two various research directions: Randomized Perturbation Technique (RPT) is one type, and the other one is Cryptography-Based Technique.

In recent years PPDM has expanded in the area of data mining as a main research track. In several data mining areas, are Bayesian network, social network, association rules, stream mining, outlier detection, decision trees, recommender systems PPDM attains convincing results [5][6][7][8][9] [10] [11][12].

From the public databases, the records are prevented from indirect classification as the individual records are exactly identified by combination of attribute records though k-anonymity is supposed to be susceptible. K-anonymity encounters several privacy troubles such as t-closeness and l-diversity then there is some diversity in the receptive attributes the victim addition background data can be known to remember the difficulty these two techniques yet not determined the confidentiality changes the various privacy attacks face insufficient privacy preservation exposes the three closed relative privacy models which are indicated in various research results [5].

The privacy-preserving data mining gives us two issues. The main issue about the customer privacy preserving to prevent the data values. The main intention is

to preventing the data cannot expose the private data, and it is secure for mining. The main result is to be preventing the data and this data sharing the irregular data used to preventing data, and it will use to produce the similar values of real data distribution, the real data values should not be exposed. The vertical distributed data has the issue of secured computing association rules is also placed [10]. The vertical partition issue forms during the one by one transaction is divided into different sites, having various set of attributes in the whole set of transactions. Each site has a set of entire transactions. In the horizontal partition data and in vertical partition data, the similarities of each site have to be union to form the relationship is to be mined.

Our survey is related to beyond papers, expect it is exactly relevant to [13]. It can identify a following task of [13]. In similarity, it can be addressed a various problem with a various technique, and that it is a encrypted one on the other hand it is not a different confidential [36].

a. Motivations

In this method to perform the privacy preserving distributed association rule the mining protocols are proposed by taking more than three parties in horizontal scenario. When the participating parties number is equal to two, the Security is not guaranteed because it build on a secure multi-party compact and union calculation.

In all over the usual itemsets are revealing the all parities participating into two party case it can be now that itemset is promoted to one's site to this leakage can be avoided [13]. After all, we observed in several cases that knowing whether a locally usual itemset at one site is over all accepted but knowledge is not distributed to the other site by analysis. For example, Walmart, may want to know the rule diaper beer is too strong and kept it under the transitions joint of Walmart and target is to known the frequent of a given pattern to be strong and to generate the database is still frequent under the combination of the database with another usual pattern theme to motivation does exit in a hospital, where its president to known together patient databases but with some privacy. In the two party of security has been computation begin the secure the party estimate to security information [20] [21]. The general complexity are not applicable to the datamining tasks as seen to complement to the concrete as well as the application is general secure to the computation protocol [22].

b. Improvements

To expertise the address to our improvement no reviews has been organized earlier. From commutative encryption and decryption, a secure analysis computing protocols to build. The core technique is to build the protocol which perform the design of a secure computation protocols for distributed association rule mining on partitioned data to establish the major improvement to this paper proposed. Here, we considering at one site at which they are frequent in the two-party case which is not supported at one's own site and which is supported globally at others site itemset by corresponding itemset globally are expected locally works only to check the problem by exhibiting the universal itemset. in this paper previously only encryption is used now we are proposing decryption in addition to encryption.

c. Design

The remaining paper is classified as follows, category B defines the outlining, key foundations, and security framework. Category C defines Mining Association Rule Protocols for security and Two party protocol. Category D defines a Integrity Text and category E defines a Preservation Text. category F is results and Discussions is conclusion. At last defines category G conclusion.

II. LITERATURE SURVEY

Here, we will present a few key outlining of Association mining rules, Distributed Association Rule Mining, Secure Distributed Association Rule Mining and some mining problems of Commutative Encryption and Decryption System, security infrastructures.

a. Mining Association Rules

A list of items K={a1, a2,. . .,an}.let DB be a transactions of database, where each transaction I consists of a list of items such that I ⊆ K.

The list of items can be illustrated as K= {a1, a2,. . .,an} association mining rules[14]. Let DB be a transaction of database, where each transaction I consists of a list of items such that I ⊆Given an itemset A ⊆ K, a transaction I contains A if and only if A ⊆I. An association rule is an implication of the form A ⇒ B, where A ⊆ K, B ⊆ K and A ∩ B = ∅. The rule A ⇒ B has support sin the database transaction D if the probability of a transaction in DB containing both A and B is s. The association rule holds in the transaction database DB with confidence c if the probability of a transaction in DB containing A and thus B is c. An item set A with I items is called a I-item set. Here whose confidence is above a minimum confidence threshold and whose support is greater than minimum support threshold which is the problem of mining association rules that one must mind.

Here, an proportion of transactions in a DB its support depends on, A itemsets, and its support count Asup, this indicates the count of transactions in DB. an itemset A frequent A(or also absolutely frequently develop)minimum support threshold is less than support. Two sub problems that defined to show the problem of association rule mining (a) frequent itemsets are found by given minimum threshold value found (b).from the frequent itemsets found, association rules are generated. A focus on the development of efficient methods by the association rule mining research to solve the first sub problem which plays vital rule in cost of mining association rule.

TABLE I: Signs Table [15]

D	DB is the number of transactions
S	Minsup support threshold
L(n)	n-itemsets are globally frequent
CA(n)	L(n) is design candidate set
X.sup	X is global support count
Dp	DBp is number of transactions
GLp(n)	Sp is n-itemsets of globally frequent

CGp(n)	GLp (n-1) is design candidate sets
LLp(n)	CGp(n) n-itemsets of Locally frequent
X.sup p	Sp of X is local support count

b. Mining Association Rules for Distributed system

The association rule mining searches a distributed environment. Let DB be a database among transactions D.In a distributed system, consider their n number of sites, p1, p2, . . ., pn,and the separation over the database DB with n sites into {DB1,DB2, · · · ,DBn} correspondingly.

Assume that the size of each and every partition of DBi for i=1, . . . , n.In this A. support is the global support count and A.supi is the local support count for A at site Si.For a minimum support threshold s, for that S is said to be globally frequent if A.sup≥s×D; correspondingly, A is said to be locally frequent for site Si,if A.supi ≥ s×Di.In this,L specifies the universal frequent itemsets in DB and L(n) specifies the universal frequent N-itemsets in L.

The main aim of distributed association rule mining algorithm is to find the related globally frequent item sets. A fast algorithm for distributed association rule mining is given in [15]. The method of fast-distributed mining of association rules (FDM) can be encapsulated below:

1.Generation of candidate sets: Generate candidate sets CGi(k) depends on GLi(n−1), itemsets which are supported by Si at the (n− 1)th iteration, using the classic Apriori candidate generation algorithm. Each site produced candidates depends on the common itemsets of globally frequent (n − 1)-itemsets and locally frequent (n − 1)-itemsets.

2.Local Pruning: For each X ∈ Cip(n), scan the database DBi at Sp to compute A.supi. If A is not locally frequent at site Sp,it is eliminated from LLp(n)candidate sets.(At site 1 candidate sets only removes A)but it fixed to other site candidate set.Support Count Exchange: LLp(n) are broadcast the candidate set and each sites computes their global support to find the all frequent globally itemsets in site Sp.

- Broadcast Mining Results: broadcasts Each site the global frequent itemsets to each site. A few main characters on Distributed association rules mining are recorded in Table 1.

c. Distributed Mining Association Rules for security

let m>2 be the number of sites. Various sites has private database DB p transaction. We are given ratios of confidence c and support threshold. Here, the main objective is to design all association rules(itemsets) to satisfying the thresholds, are represented. We are added desire that acknowledgment are limited: If any site cannot read the details of a transaction to other sites, what association rules (itemsets) by any other sites, or the limited value of confidence/support for any association rule at any other site but the ability of message is publish of own message and the final result.

After all, in the case of p=2 we fixed in section 1 that sophisticated an association rule(itemset) is not supported an one's site confess to other sites and it is supported globally, leakage cannot be avoided.In this case two-party defination is little different; this is main purpose to illistrate the secure two-party association rules mining protocols secure the following data privacy:

1.Here, each site size of database and support are secured not reveal to the opposite site.

2.At each site knows at it own's site supported globally and locally frequent association rule but it not revealed to another site.

III. COMMUTATIVE ENCRYPTION SYSTEM

A commutative encryption system is needed to implement the multi-party security computation, it is set as below [16]:

A Commutative Encryption System is F = (M, K, g, h) (where M is the encrypted messages in plain text, K=(E.D) are both encryption and decryption keys correspondingly) if the pair of commutative encryption g : E ×M → M and the commutative decryption function is h : D ×M → M are determinable in polynomial time, a finite determinable functions satisfy all properties as follows. We express ge(x) ≡ g (e, x), hd(x) ≡ h(d, x), and use ∈r to represent evenly at random from A. For all e1,e2..en ∈ E, d1, d2, · · dn ∈ D, each m ∈ M,we have gep1(gep2(· · · (gepn (m)) · ·)) = geq1(geq2(· · (geqn (m)) · ·)) ≡ Tandhdt1(hdt2(· · · (hdsn (N)) · · ·)) = hdu1(hdu2(· · · (hdtn (N)) · · ·)),where (a1, a2, · · , an), (b1, b2, · · · bn),(k1, k2, · · kn) and (m1, m2, · · · mn) are four transformations of (1, 2, · · · , n).B. For all e1, e2,· · · ,en ∈ E, all m1,m2 ∈ M such thatm1 ≠ m2 and big enough k, we have qs(ge1(ge2(· · (gen(m1)) · ·)) = ge1 (ge2 (· · (gen(m2)) · ·))) <1/2n C. For p, q, z ∈r M, e ∈r.

IV. MINING ASSOCIATION RULE PROTOCOLS FOR SECURITY

a. Current Protocols

The FDM algorithm follows a general path,with limited protocols replacing the support count of information LLn and its broadcasts the LLi(n) [13][15].The resultant data is a union of internally supported itemsets without revealing the special itemset of designer, and then the method gives for secure of threshold exceeds of support count.

The main approach as follows the two-phase,but generally combines the support counts and locally develop rules are passed by the encrypted values inserted into two sites.In the two-phase observes candidate itemsets and resolve of these appropriate global support confidence thresholds.

In figure 1 explains commutative encryption. Each site encrypts its usual itemsets. The itemsets are encrypted and then extended to all other sites until all sites have itemsets are encrypted and then itemsets are passed to the common site are eliminated the duplicate itemsets and then start decryption. For example, suppose we have two websites like Amazon and flipkart.If someone purchase an item through the amazon website then amazon only knows

the details of the purchased item.

As well as the same case with flipkart.But if two sites want to know the frequent items which are purchased by end users then the data controller is maintains data of two sites. Where both sites send their data which is encrypted based on that data,the data controller calculates frequent itemsets. If any itemsets are exceeding the minimum threshold then the value is passed to that sites or to eliminate the duplicate itemsets [40].

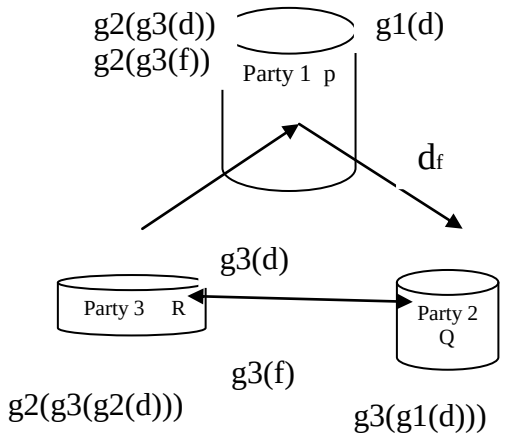

Fig. 1 Concluding the itemsets of global candidates.

In figure 2 shows each itemset is tested locally and see it is supported globally. Assume this method presented three or more sites. In the two-party case does not support at individual site confess to other sites and its supported globally. Here try to preserve the perfectly secure method for count the global results. In the two-party case reduce the collision.

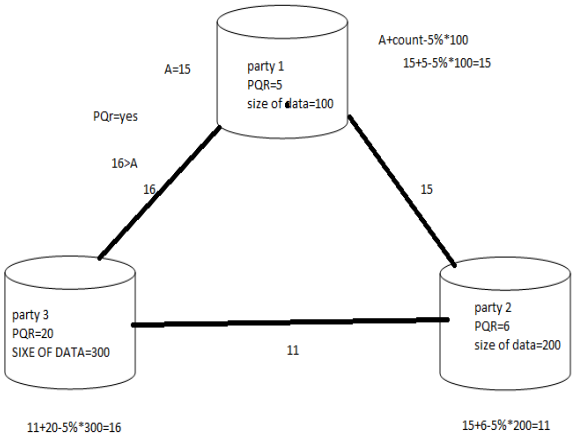

Fig. 2: Determining if item set support exceeds the 5 percent threshold [13]

In this diagram shows each party calculates its local support count ,and an itemset PQR are supported at one or more parties. Here, first party selects a random value A(15) and the random value A(15) added to the amount by which support count and database size and then value exceeds the minimum threshold and result is (15).After that value(15) is passed to the party 2 same procedure will be calculated and then value is exceeds the minimum threshold and the result is (11).Afterthe value(11) is passed to the third party and value is again added a support count and the result is (16).The resulting value A(16) has exceeds the random value (15)[40].

2. Two-Party Case

To preserve the data privacy of participating parties is to discuss the basics of how to found globally frequent I-itemsets.

Here, from the global results and its own input one party can easily reduce the others party's information, this is the possibility that need to adjust the above leakage that cannot be avoided.

We all know the possibilities of association rules (itemsets) are locally supported or not and its own site supports globally without revealing to the other site. The possibility of applications concrete on the following. For example walmart may want to known the rule diaper->beer is too strong and kept it under the transtions joint of walmart and target is to known the frequent of a given pattern to be strong and to generate the database is still frequent under the combination of the database with another usual patterns theme to motivation does exit in a hospital where its president to known together patient databases but with some privacy. In the first step of protocol 1,it encrypts CGp(n) through both the sites. In protocol 1 step 2 at Commutative Encryption encrypts CGp(n) at each site and then it passes to the step3 chipertext and it is send to the other site, and again step 6 the chipertext encrypts and received its individual Commutative Encryption System and sends back to step7.For this purpose in Protocol 2 to plan the identity testing for security, there is no importance to order of encryption to play any role in the encryption results of Commutative Encryption System.

In the code 2 conducted in the second step is to find GLp(n). In this protocol 1 (step 2 and step3), one site chipertexts sends its two-time encryption one-by-one to other sites, and then step 6 and step 7 demonstrate support count at opposite site. To generate random numbers step 11,14,15 and then conducts scalar product computation. The chipertext global support is securely computed by sub protocol of secure scalar product in the two sites (in parallel plaintext, itemset); if the threshold is greater than the support then the globally n-item set frequent will be added to GLp(n).

V. INTEGRITY TEXT

Code 1: Here encryption can be done two times providing better security. First one site encrypts $CG_{p(n)}$ and is sent to another site for the second-time encryption. The main theme of protocol 2 is to define and determine $GL_{p(n)}$. It first go through the steps from 1 to 9 between the two sites and make a match of equal cyphertexts. Next it moves from step 10 to 27 and calculates the global support for the cipher texts using a secure sub protocol called two-party division.

Assuming the support count for site 1 as $x_{1(p)}$,sup1 and transaction database as D_1 . As well as support count of site 2 is sup2 and size of transaction database is D2 . The

global support is $x_{1(n)+}sup_2/D_1+D_2$ then,
$W'=a_1*a_2*W_1/W_2$
$=a_1*a_2*(a^1_1*a^2_1*(x_{1(n)}.sup_1+sup_2)/a^2_1*a^2_2*(M_1+M_2))$
$=a^2_1/a^1_1*a^2_2/a^1_2*(a^1_1*a^2_1*(x_{1(n)}.sup_1+sup_2)/a^2_1*a^2_2*(M_1+M_2)$
$= x_{1(n)}.sup_1+sup_2/M_1+M_2.$

Code 2: Generate $GL_{1(n)}$, $GL_{2(n)}$ (1)
input: M(= 2) sites number 1, 2; Minimum support, s
output: $GL_{p(n)}$ at each p
1. At site 1 begin
2. Locate the first
$<x_{1(n)}, x_{1(n)}.sup1, g_{f2}g_{f1}(x_{1(n)})>$in its ordered triples;
3. Send $g_{f2}g_{f1}(x_{1(n)})$ to site 2;
4. end
5. At site 2 begin
6. Search the $g_{f1}g_{f2}(x_{2(n)})$ that equals $g_{f2}g_{f1}(x_{1(n)})$;
7. if *Found* then Set $sup2 = x_{2(n)}.sup2$;
8. else Set $sup2 = 0$;
9. end
10. At site 1 begin
11. produced two new non-zero random numbers
a^1_1, a^2_1 and set $a1 = a^2_1/a^1_1$;
12. end
13. At site 2 begin
14. yield two new non-zero random numbers
a^1_2, a^2_2;
15. Send $a2 = a^2_2/a^1_2$ to site 1.
16. end
17. Site 1 and Site 2 begin
18. Collaborate to compute the Scalar Product securely on $(a^1_1\times x_{1(n)}.sup_1, a^1_1)$ and $(a^1_2, a^1_2\times sup_2)$; Only Site 1 obtains:
$W_1 = a^1_1\times a^1_2\times x_{1(n)}.sup_1 + a^1_1\times a^1_2\times sup_2 = a^1_1\times a^1_2\times x_{1(n)}.sup_1 + sup_2)$;
19. End
20. begin site 1 and 2.
21. Collaborate to securely compute the Scalar Product on $(a^2_1\times M_1, a^2_1)$ and $(a^2_2, a^2_2\times M_2)$; Only Site 1 obtains:
$W2 = a^2_1\times a^2_2\times M_1+a^2_1\times a^2_2\times M_2 = a^2_1\times a^2_2\times (M_1+M_2)$;
22. End
23. At the site 1
24. Compute $z\_ = a_2\times a_1\times W_1/W_2 = a_2\times a_1\times(a^1_1\times a^1_2\times(x_{1(n)}.sup_1+sup_2))/(a^2_1\times a^2_2\times(M_1+M_2)) = (x_{1(n)}.sup1+sup2)/(M_1+M_2)$;
25. if$W\geq s$ then Insert $x_{1(n)}$ into $GL_{1(n)}$;
26. delete the triple
$<x_{1(n)}, x_{1(n)}.sup1, f_{e2}f_{e1}(x_{1(n)})>$;
27. End
28. Repeat step begin
29). Site 2 takes the first $< x_{2(n)}, x_{2(n)}.sup_2, f_{e1}f_{e2}(x_{2(n)})>$in its respective triples which are ordered.Also repeats the process form step 1 to step 27 in the same way. Now ,Site 1 takes the next triple and excutes the same process;
30. The above procedure repeats for all the triples at both Site 1 and at Site 2.By the end of the protocol ,both Site 1 and Site 2 get $GL_{1(n)}$ and $GL_{2(n)}$ respectively.
31. In Decryption calculates the global frequent item sets, if any message is not supported at any site, that message will be deleted.
32.end

5.1. PRESERVATION TEXT

1. Code 1: Here we apply simulation technology which is described in definition 2 indicate the security of this code. And communication done at step 3 and step 7 where the security measured is commutative encryption. By the end of step 3, we get cipher texts that are obtained applying commutative encryption. By the definition1, the simulation views could be obtained by selecting the numbers, uniformly from the set of cypher texts in a random form. As the selected numbers, cipher texts are bounded in a uniform way in the same set they are not distinguishable. Because of the constraints of space, the security of step 7 is ignored which is like step 3.
2. Code 2: In this code, the data transfer can be done at the step 3,15,18 and 21. As the cipher text $g_{f2}g_{f1}(x_{1(n))}$ was obtained from step 2 in the step 3, indicates not a private data. And, at step 15, randomly produced number a1 is not a private data. To get the scalar product securely, steps 16 and 21 will be combined. Here, he security depends on the security of scalar product protocol. The security of protocol 2 can be decided from the composition theorem by observing the analysis described.

VI. RESULTS AND DISCUSSION

By using code1 and code 2 some of the strings are encrypted and decrypted information is shown in the given tables.

In the table1 shows that, at site1 the message is “Hello”, and the support count is 3.the resulting encryption is shown table1.At site 2 the message is “Hello” ,and the support count is 2.the resulting encryption is shown table1.And site1 the message is “Mango”, and the support count is 3.the resulting encryption is shown table1.At site 2 the message is “Apple”, and the support count is 3.the resulting encryption is shown table1. And site1 the message is “Grape”, and the support count is 2.the resulting encryption is shown table1.At site 2 the message is “Grape”, and the support count is 2.the resulting encryption is shown table1.

Site1			Site2		
message	encryption	Support count	message	encryption	Support count
Hello	125n116n108n108n99n	3	Hello	125n116n108n108n99n	2
Mango	124n97n122n115n99n	2	Apple	97n127n127n108n116n	3
Grape	115n126n97n127n116n	2	Grape	115n126n97n127n116n	2

In table 2 shows that, at site 1 message is “Hello’,

and generate Random value 15,16 and also generate global support count with minimum support count 0.37. At site 2 message is “Hello’, and generate Random value 16,17 and also generate global support count with minimum support count. And at site 1 message is “Mango’, and generate Random value 15,16 and also generate global support count with minimum support count 0.37. At site 2 message is “Apple’, and generate Random value 16,17 and also generate global support count with minimum support count .And, at site 1 message is “Grape’, and generate Random value 15,16 and also generate global support count with minimum support count 0.37.At site 2 message is “Grape’, and generate Random value 16,17 and also generate global support count with minimum support count.

Table 2

Site 1			Site 2			Minimum support count
Message	Random value	Global support count	Message	Random value	Global support count	
Hello Mango Grape	15,16	0.653302 0.261321 0.522642	Hello Apple Grape	17,18	0.653302 0.391981 0.522642	0.37

In decryption the message is “hello” and “Grape” are supported with minimum support count and global frequent item set.And the message is “Mango” and “Grape” are not supported with minimum support count and global frequent item set so that two message is deleted.

Description

Site1	Site2
hello	hello
Grape	Grape

VII. CONCLUSIONS

In this paper, we proposed design of a secure computation protocols for distributed association rule mining on partitioned data. Here, information will be transformed from one site to another site with more security is achieved. This will be done by means of encryption, global support count values and decryption of information at both the sites. These encryption and decryption protocols use existing security protocols [40].Designing of multiparty data mining protocol; supporting malicious model it is a big challenge in SMC applications. The main issue is how to design a protocol and that can be further reduced the time and communication cost will be deal with future

REFERENCES

[1] Han, Jiawei, and Micheline Kamber. "Data Mining: Concepts and Techniques, 2nd edition Morgan Kaufmann Publishers." *San Francisco, CA, USA* (2006).

[2] Lin, Xiaodong, Chris Clifton, and Michael Zhu. "Privacy-preserving clustering with distributed EM mixture modeling." *Knowledge and information systems*8.1 (2005): 68-81.

[3] O.Goldreich, “Encryption Schemes,” (working draft), Mar. 2003,available: http://www.wisdom.weizmann.ac.il/~oded/PSBook

[4] Agrawal, Rakesh, and Ramakrishnan Srikant. "Privacy-preserving data mining." *ACM Sigmod Record*. Vol. 29. No. 2. ACM, 2000.

[5] Lindell, Yehuda, and Benny Pinkas. "Privacy preserving data mining." *Annual International Cryptology Conference*. Springer Berlin Heidelberg, 2000.

[6] J.Lin,Y.Cheng.Privacypreservingitemsetminingthroughnoisyitems. Expert Systems with Applications. 2009, 36(3): 5711-5717

[7] Sakuma, Jun, and Shigenobu Kobayashi. "Large-scale k-means clustering with user-centric privacy-preservation." *Knowledge and information systems*25.2 (2010): 253-279.

[8] Kantarcioglu, Murat, and Chris Clifton. "Privacy-preserving distributed mining of association rules on horizontally partitioned data." *IEEE transactions on knowledge and data engineering* 16.9 (2004): 1026-1037.

[9] Vaidya, Jaideep, Murat Kantarcıoğlu, and Chris Clifton. "Privacy-preserving naive bayes classification." *The VLDB Journal—The International Journal on Very Large Data Bases* 17.4 (2008): 879-898.

[10] McSherry, Frank, and Ilya Mironov. "Differentially private recommender systems: building privacy into the net." *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009.

[11] Li, Feifei, et al. "Hiding in the crowd: Privacy preservation on evolving streams through correlation tracking." *2007 IEEE 23rd International Conference on Data Engineering*. IEEE, 2007.

[12] Zhou, Bin, and Jian Pei. "Preserving privacy in social networks against neighborhood attacks." *2008 IEEE 24th International Conference on Data Engineering*. IEEE, 2008.

[13] Kantarcioglu, Murat, and Chris Clifton. "Privacy-preserving distributed mining of association rules on horizontally partitioned data." *IEEE transactions on knowledge and data engineering* 16.9 (2004): 1026-1037.

[14] Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." *Proc. 20th int. conf. very large data bases, VLDB*. Vol. 1215. 1994.

[15] Cheung, David W., et al. "A fast distributed algorithm for mining association rules." *Parallel and*

Distributed Information Systems, 1996., Fourth International Conference on. IEEE, 1996.
[16] Agrawal, Rakesh, Alexandre Evfimievski, and Ramakrishnan Srikant. "Information sharing across private databases." *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*. ACM, 2003.
[17] Boneh, Dan. "The decision diffie-hellman problem." *International Algorithmic Number Theory Symposium*. Springer Berlin Heidelberg, 1998.
[18] Du, Wenliang. *A study of several specific secure two-party computation problems*. Diss. Purdue University, 2001.
[19] Kantarcioglu, Murat, and Onur Kardes. "Privacy-preserving data mining applications in the malicious model." *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*. IEEE, 2007.
[20] Yao, Andrew Chi-Chih. "How to generate and exchange secrets." *Foundations of Computer Science, 1986., 27th Annual Symposium on*. IEEE, 1986.
[21] Lindell, Yehuda, and Benny Pinkas. "A proof of security of Yao's protocol for two-party computation." *Journal of Cryptology* 22.2 (2009): 161-188.
[22] Goldreich, Oded. *Foundations of cryptography: volume 2, basic applications*. Cambridge university press, 2009.
[23] Samarati, Pierangela. "Protecting respondents identities in microdata release." *IEEE transactions on Knowledge and Data Engineering* 13.6 (2001): 1010-1027.
[24] Machanavajjhala, Ashwin, et al. "l-diversity: Privacy beyond k-anonymity." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1.1 (2007): 3.
[25] Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. "t-closeness: Privacy beyond k-anonymity and l-diversity." *2007 IEEE 23rd International Conference on Data Engineering*. IEEE, 2007.
[26] Ganta, Srivatsava Ranjit, Shiva Prasad Kasiviswanathan, and Adam Smith. "Composition attacks and auxiliary information in data privacy." *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008.
[27] Kifer, Daniel. "Attacks on privacy and deFinetti's theorem." *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. ACM, 2009.
[28] Zhang, Feng, et al. "Privacy-preserving two-party distributed association rules mining on horizontally partitioned data." *Cloud Computing and Big Data (CloudCom-Asia), 2013 International Conference on*. IEEE, 2013.
[29] Dwork, Cynthia, et al. "Calibrating noise to sensitivity in private data analysis." *Theory of Cryptography Conference*. Springer Berlin Heidelberg, 2006.
[30] McSherry, Frank, and Kunal Talwar. "Mechanism design via differential privacy." *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*. IEEE, 2007.
[31] Hay, Michael, et al. "Boosting the accuracy of differentially private histograms through consistency." *Proceedings of the VLDB Endowment* 3.1-2 (2010): 1021-1032.
[32] Mohammed, Noman, et al. "Differentially private data release for data mining." *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011.
[33] Lee, Jaewoo, and Chris Clifton. "Differential identifiability." *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012.
[34] Kasiviswanathan, Shiva Prasad, et al. "Analyzing graphs with node differential privacy." *Theory of Cryptography*. Springer Berlin Heidelberg, 2013. 457-476.
[35] Ji, Zhanglong, and Charles Elkan. "Differential privacy based on importance weighting." *Machine learning* 93.1 (2013): 163-183.
[36] Mohammed, Noman, et al. "Secure two-party differentially private data release for vertically partitioned data." *IEEE Transactions on Dependable and Secure Computing* 11.1 (2014): 59-71.
[37] Goethals, Bart, et al. "On private scalar product computation for privacy-preserving data mining." *International Conference on Information Security and Cryptology*. Springer Berlin Heidelberg, 2004.
[38] Lin, Hsiao-Ying, and Wen-Guey Tzeng. "An efficient solution to the millionaires' problem based on homomorphic encryption." *International Conference on Applied Cryptography and Network Security*. Springer Berlin Heidelberg, 2005.
[39] Blake, Ian F., and Vladimir Kolesnikov. "Strong conditional oblivious transfer and computing on intervals." *International Conference on the Theory and Application of Cryptology and Information Security*. Springer Berlin Heidelberg, 2004.
[40] Zhang, Feng, et al. "Privacy-preserving two-party distributed association rules mining on horizontally partitioned data." *Cloud Computing and Big Data (CloudCom-Asia), 2013 International Conference on*. IEEE, 2013.