ISSN:0975-9646

N V Muthu Lakshmi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 340-349

Privacy Preserving Association Rule Mining of Mixed Partitioned Model in Distributed Database Environment

Dr. N V Muthu Lakshmi[1], K. Sandhya Rani[2]

[1]Dept of Computer Science, MITS College, Madanapalle, AP, INDIA
[2]Dept of Computer Science, SPMVV, Tirupati, AP, INDIA

***Abstract:*---- A big challenge to researchers is, how efficiently, one can extract knowledge in an application from a large database using appropriate data mining techniques. A well known research data mining technique is association rule mining whose outcomes are being successfully used in various real life applications to take strategic decisions. But, in recent years, people wish to share their knowledge which is predicted from association rule mining to their collaborators to get mutual benefits. But the problem is, they do not want to give their sensitive data to the collaborators because disclosure of sensitive data may cause harm to themselves. Here, the main issue is, how to achieve the two goals that is sharing accurate non sensitive information while protecting sensitive information. To address this issue, researchers have found many efficient methods for finding privacy preserving association rule mining. To enhance the availability and reliability of the customer services, distributed database applications are well being used which can also find association rules at global level, called global association rules. In this paper, privacy preserving association rule mining at global level is addressed. A global database usually partitioned in different ways but basically in three common ways such as horizontal, vertical and mixed model is a combination of both. In mixed partitioned model, a global database can be partitioned first in horizontal then each horizontally partitioned database is further partitioned into vertical or vice versa or in any order with different levels. This paper presented two efficient methods for determining privacy preserving association rule mining for two common mixed models by adopting cryptographic techniques, Sign based secure sum concept and scalar product technique.**

Keywords: Association Rules, Privacy Preserving Association Rule Mining, Distributed Database Applications, Global Association Rules.

I. INTRODUCTION

Association rule mining is one among many data mining techniques which predicts associations between items or item sets from large database. Association rule generation has two steps. Computation of frequent item sets from the database based on user specified minimum support threshold is the first step, where generation of association rules from frequent item sets based on user specified minimum confidence threshold is second step. Data mining has been viewed as a threat to privacy because of the pervasive proliferation of electronic data maintained by corporations. This has lead to increased concerns about the privacy of the underlying data. Data mining techniques find hidden information from large database while secret data is preserved safely when data is allowed to access by single person. Now a day's many people want to access data or hidden information using data mining technique even they are not fully authorized to access. For getting mutual benefits, many organizations wish to share their data to many legitimate people but without revealing their secret data. To address this issue, privacy preserving data mining has been evolved.

The process of preserving privacy in case of association rule mining can be termed as privacy preserving association rule mining. Database may consist of enormous amount of transactions which are extracted from a single source of data or from many sources. Depending on the requirements of applications, database is maintained at single location called centralized database or the database may be distributed at multiple sites called distributed database. The main aim of privacy preserving association rule mining in centralized database is, mining process can be done by hiding sensitive data/information from users other than database owner. In distributed environment, aim is finding the global mining results by preserving the individual sites private data/information from one another. Global results are determined only when the necessary results/information is captured based on all sites' database individually like local frequent item sets and their support values of all sites are required to determine whether an item set is globally frequent or infrequent. As the individual database may possess some private data/information and in case of leakage of private data to anyone causes damage to database owners.

In distributed applications, databases are partitioned basically in two ways such as horizontal and vertical partitioned databases, where each partitioned database is placed in one site or many sites. The site which owns the database has local autonomy over its database and no site can have access to any data/information belongs to any other site. Depending on the hierarchy of the distributed application, any site's partitioned database can be further partitioned into two or more and each partitioning may follow horizontal or vertical and this process of partitioning is called mixed/hybrid. In some distributed applications databases are partitioned into disjoint segments so every database is placed in a single location/site only. In this paper, privacy preserving association rule mining for two commonly used mixed partitioning (disjoint) methods in distributed database environment is considered.

II . PRIVACY PRESERVING ASSOCIATION RULE MINING FOR MIXED PARTITIONED MODEL

Privacy-preserving data mining in a distributed environment is a multidisciplinary field and requires close cooperation between researchers and practitioners from the fields of cryptography, data mining, public policy and law. Now, the question is how to compute the results without pooling the data in a way that reveals nothing but the final results of the data mining computation. This question of privacy-preserving data mining is actually a special case of a long-studied problem in cryptography called secure multiparty computation. This problem deals with a setting where a set of parties with private inputs wishes to jointly compute some function of their inputs. This joint computation should have the property that the parties learn the correct output and nothing else, even if some of the parties maliciously collude to obtain more information. Clearly, a protocol is needed to solve privacy-preserving data mining problems. Earlier work in privacy preserving association rule mining is as follows.

In 1996, Clifton et al. [3] discussed and presented ideas related to the issue of protecting privacy of individuals in the database. The state of the art in the area of privacy preserving data mining techniques is discussed by the authors in [4] [5]. This paper also describes the different dimensions of preserving data mining techniques such as data distribution, data modification technique, data mining algorithms, data or rule hiding and approaches for privacy preserving data mining techniques. In [6], the authors proposed a framework for evaluating privacy preserving data mining algorithms and based on their frame work one can assess the different features of privacy preserving algorithms according to different evaluation criteria.

Evfimievski et al. presented a new framework for preserving privacy association rule mining [7]. In order to find privacy preserving association rule mining in centralized database, a new algorithm is presented in [8] which balancs privacy preserving and knowledge discovery in association rule mining. Gkoulalas Divanis, et al. addressed many issues related to privacy preserving data mining, association rule hiding, classes of association rule hiding methodologies and also rule hiding in classification technique, privacy preserving clustering & sequence hiding [9].

The problem of knowing who is richer without disclosing their wealth is addressed in two milliner's problem and which belongs to secure multi party computation. The authors proposed protocols for two milliner's problem and also proposed for multi party case [10].Clifton proposed a toolkit consisting of Secure sum, Secure set union, Secure size of set intersection and Scalar product are the protocols that can be combined for specific privacy preserving data mining applications [11]. The algorithms for privacy preserving association rules mining over horizontally, vertically and mixed partitioned database are presented in this thesis work [12].Secure mining of association rules over horizontally partitioned database using cryptographic technique to minimize the information shared by adding overhead to the mining process is presented in [13].In [14], authors addressed the problem of association rule mining in vertically partitioned database by using cryptography based approach. In [15], several private scalar product protocols for two party scalar product protocols is proposed with a un trusted third party using algebraic computations.

The authors in [16] proposed architecture for privacy preservation in classification technique for mixed partitioned distributed database model which is a combination of vertical and horizontal for Breast cancer dataset. In [17], algorithm is presented for finding privacy preserving association rule mining in mixed partitioned database model.

In most of the real life applications, mixed partitioning models are used and partitioning follows the organization structure.

Consider the two mixed models which are commonly used and its partitioning are shown in the Figure 1 and Figure 2. The first type partitioning method is one in which all or some horizontally partitioned databases are further partitioned into two or more vertically partitioned databases and second type is one in which all or some vertically partitioned database are further partitioned into two or more horizontally partitioned databases. Mixed Model-1 is shown in following figure.

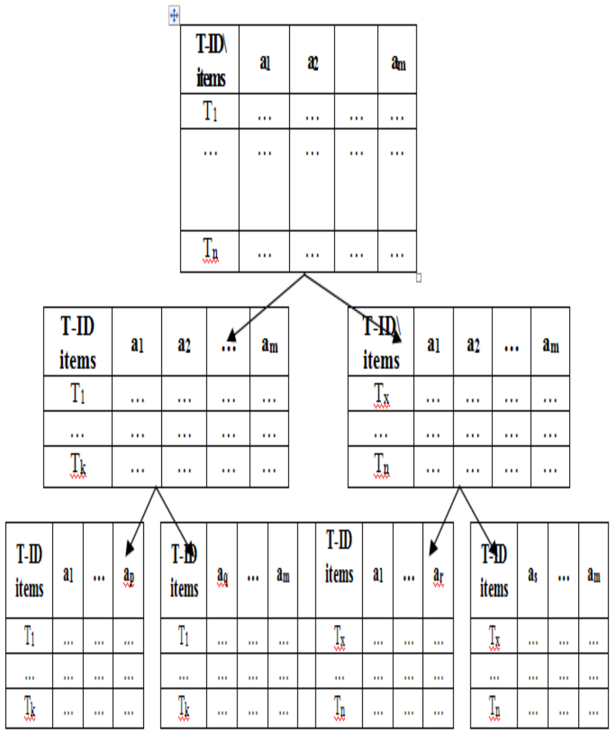

Fig.1. Two Horizontally Partitioned Databases are Further Partitioned into Two Vertical (Model-1)

Initially the database is partitioned into two horizontal partitioned databases. As partitioning is based on horizontal, all two partitioned sites possess same set of attributes but possessing different set of disjoint transactions. Each horizontally partitioned database is further partitioned into two vertical databases. The other Mixed Model-II is as shown below.

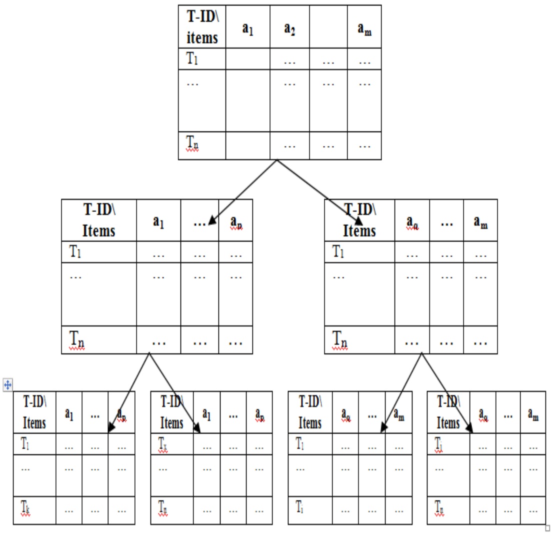

Fig. 2. Two Vertically Partitioned Databases are Further Partitioned into Two Horizontal (Model - 2)

In the above diagram, the database is partitioned into two vertical partitioned databases. As partitioning is based on vertical, all two partitioned sites possess disjoint subset set of attributes but possessing same set of transactions. In addition to the above common types of mixed partitioned databases, many other models exist like combinations of horizontal, horizontal and then vertical or vice versa. In this paper, the above two types of mixed model partitioning strategies are considered to find the global association rules.

Two different methodologies are proposed in this paper to deal with the two mixed partitioning methods and are considered as two different cases Case I and Case II. These methods incorporates algorithms presented in [18] to find privacy preserving association rule mining for n number of horizontally partitioned databases with Trusted Party (TP) using sign based secure sum technique and can be referred in this paper as Horiz-TP-Algorithm. Similarly the algorithm [19] is adopted for n vertically partitioned databases with Data Miner (DM) by using scalar dot product technique and can be referred as Vert-DM-Algorithm. Methodology for each case is discussed in the next section.

III. Proposed Methods

Case I: The mixed Model-1 specified in Fig. 1. is considered as Case I.

In this case, Trusted Party (TP) situated at $Level_0$ and the database is horizontally partitioned into two or more disjoint fragments situated at $Level_1$ and then each horizontally partitioned database is further partitioned into two or more vertically partitioning databases which will be at $Level_2$ in the hierarchy.

Case II: The mixed Model-2 specified in Fig. 2. is considered as Case II.

In this model, data miner(DM) exist at $Level_0$ and the database is vertically partitioned into two or more disjoint fragments situated at $Level_1$ and then each vertically partitioned database is further partitioned into two or more horizontal partitioning databases which will be at $Level_2$ in the hierarchy.

The algorithm for Case I is specified in the next subsection.

A. Case I

Database is partitioned into two or more horizontal, $\{DBH_1, DBH_2, \ldots, DBH_K\}$ and then each horizontally partitioned database, DBH_i is vertically partitioned into two or more $\{DBHiV_1, DBHiV_2, DBHiV_3, \ldots, DBHiV_L\}$ or zero. The major tasks in the proposed method are as follows:

Algorithm

Input: Hierarchy of Partitioned databases, Databases of all sites, MinSupport, MinConfidence

Outpt: Global association rules

Step 1 For each site $Site_i$ possessing DBH_i , $i = 1$ to k

If (DBH_i is vertically partitioned into $DBHiV_1, \ldots, DBHiV_L$ and $L \geq 2$)

{ Apply Vert-DM-Algorithm for vertically partitioned databases of L number of sites with a DM (at $Site_i$)

Now $Site_i$ consisting of Global Frequent Item sets (GFI) along with supports for its vertically partitioned databases }

Else

$Site_i$ finds frequent item sets for its database DBH_i

Step 2 Apply Hort-TP-Algorithm for k- horizontally partitioned databases DBH_i , $i = 1$ to k with TP (at $Site_0$)

Step 3 The TP declares globally frequent item sets based on support values

Step 4 The TP broadcast the GFI & supports to the sites possessing DBH_i , $i = 1$ to k

Step 5 Every site $Site_i$, $i = 1$ to k

Generates global association rules based on user specified MinConfidence

Broadcast global association rules to its partitioned sites //holding $\{DBHiV_1, .., DBHiV_L\}$

Step 6 Stop the process.

The algorithm for Case II is specified as follows:

B. Case II

Database is partitioned into two or more vertical $\{DBV_1, \ldots, DBV_p\}$ where $p \geq 2$ and then each vertically partitioned database DBV_j (j ranges from 1 to p) is further partitioned into horizontal partitioned databases $\{DBV_jH_1, \ldots, DBV_jH_q\}$ or zero. The major tasks in the proposed method are as follows:

Algorithm

Input: Hierarchy of Partitioned databases, Databases of all sites, MinSupport, MinConfidence

Output: Global association rules

Step 1 For each site $Site_j$ possessing DBV_j , j =1 to p

If (DBV_j is horizontally partitioned into{DBV_jH_1,.., DBV_jH_q} and q ≥ 2}

{Apply Hort-TP-Algorithm for q-horizontally partitioned databases with a TP (at $Site_j$)

Now $Site_j$ consisting of GFI along with supports for its horizontally partitioned databases }

Else

$Site_j$ finds frequent item sets for its database DBV_j

Step 2 Apply Vert-DM-Algorithm for p-vertically partitioned databases, p≥2 with DM (at $Site_0$).

Step 3 DM finds GFI & supports based on user specified MinSupport.

Step 4 DM generates global association rules using GFI and support values based on user specified MinConfidence threshold value.

Step 5 Every site $Site_j$, j= 1 to p

Broadcast global association rules to its partitioned sites // holding {$DBVjH_1$, …, $DBVjH_q$}

Step 6 Stop the process.

The implementation of two mixed models which are considered in Case I and Case II are illustrated with sample database in the next section.

IV. Implementation Of The Model-I With Sample Data

Consider the global/whole database which is initially partitioned horizontally into two databases DB_1 and DB_2 and are further partitioned into DB_3&DB_4, DB_5&DB_6 respectively. The hierarchy of this model can be treated as a 3 level hierarchy model that is at level zero ($Level_0$) TP exist and at level one ($Level_1$) two horizontally partitioned sites exist and at level two ($Level_2$), two vertical partitioned databases exist. The TP will be situated at $Site_0$, horizontally partitioned database are DB_1 and DB_2 situated at $Site_1$ and $Site_2$ and at $Level_2$ vertically partitioned database DB_3, DB_4, DB_5, DB_6, situated at $Site_3$, $Site_4$, $Site_5$ and $Site_6$ respectively. The sites and their databases the considered model as shown in the following diagram.

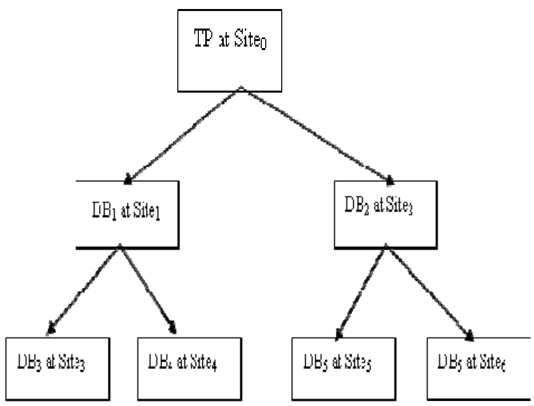

Fig. 3. Mixed Model Consisting of Databases and its Sites in Different Levels

The sample database, DB consists of 10 transactions (T=, T_2, …,T_n} and five attributes (A_1,A_2,..A_5} is considered for implementation purpose. A special site exist called $Site_0$ where TP (TP) resides who has special privileges to find global frequent items from its partitioned sites without violating individual sites' privacy constraints. The sample database is shown below:

TABLE I.
Sample Database (DB)

TID\ItemSet	A_1	A_2	A_3	A_4	A_5	A_6
T_1	1	1	1	0	1	0
T_2	0	0	0	1	1	1
T_3	1	0	1	0	0	1
T_4	0	1	0	1	0	1
T_5	1	1	1	1	0	1
T_6	0	1	1	0	1	0
T_7	0	1	0	1	0	1
T_8	1	0	1	1	0	1
T_9	1	1	1	0	1	0
T_{10}	0	1	0	1	1	0

Initially, the database DB is partitioned horizontally into two disjoint databases DB_1 and DB_2. DB_1 is at $Site_1$ and DB_2 is at $Site_2$. The following two TABLES II & III shows horizontally partitioned databases DB_1 and DB_2 each consisting of 5 transactions (T_1, T_2, T_3, T_4, T_5) for same set of attributes (A_1,A_2,..A_5).

TABLE II.
Database (DB_1) at $Site_1$

TID\ Item	A_1	A_2	A_3	A_4	A_5	A_6
T_1	1	1	1	0	1	0
T_2	0	0	0	1	1	1
T_3	1	0	1	0	0	1
T_4	0	1	0	1	0	1
T_5	1	1	1	1	0	1

TABLE III.
Database (DB_2) at $Site_2$

TID\ Item	A_1	A_2	A_3	A_4	A_5	A_6
T_6	0	1	1	0	1	0
T_7	0	1	0	1	0	1
T_8	1	0	1	1	0	1
T_9	1	1	1	0	1	0
T_{10}	0	1	0	1	1	0

The Database DB_1 at $Site_1$ is further partitioned into two disjoint vertically partitioned databases DB_3 and DB_4 and which are in $Site_3$ and $Site_4$ respectively. The database DB_3 has 5 transactions and 4 attributes such as {A_1, A_2, A_3, A_4}. The database DB_4 has 5 transactions and

2 attributes such as $\{A_5, A_6\}$ and these databases are shown in the below tables:

TABLE IV
Database (DB_3) at $Site_3$

TID\Item	A_1	A_2	A_3	A_4
T_1	1	1	1	0
T_2	0	0	0	1
T_3	1	0	1	0
T_4	0	1	0	1
T_5	1	1	1	1

TABLE V
Database (DB_4) at $Site_4$

TID\Item	A_5	A_6
T_1	1	0
T_2	1	1
T_3	0	1
T_4	0	1
T_5	0	1

Apply Verti-TP-Algorithm for finding privacy preserving association rule mining for vertically partitioned databases is applied. $Site_3$ and $Site_4$ finds frequent item sets for their databases DB_3 and DB_4 based on user specified minimum support threshold value by using apriori algorithm. The determined frequent item sets at two sites are shown below:

At $Site_3$, for Database DB_3:
The set of frequent item sets and its support is $\{A_1 \rightarrow 3, A_2 \rightarrow 3, A_3 \rightarrow 3, A_4 \rightarrow 3, <A_1,A_2> \rightarrow 2, <A_1,A_3> \rightarrow 3, A_2,A_3> \rightarrow 2, <A_2,A_4> \rightarrow 2, <A_1.A_2.A_3> \rightarrow 2\}$

At $Site_4$, for Database DB_4:
The set of frequent item sets and its support is $\{A_5 \rightarrow 2, A_6 \rightarrow 4\}$

$Site_3$ prepares a matrix M_3 consisting of supporting transactions for frequent item sets and vector V_3 possessing frequent item sets. Similarly $Site_4$ prepares a matrix M_4 for representing supporting transactions for frequent item sets and also constructs vector V_4 to store frequent item sets. These matrices and vectors are shown below:

$$M_3 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad M_4 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$V_3 = \{A_1,A_2,A_3,A_4,<A_1.A_2>, <A_1,A_3>, <A_2,A_3>, <A_2,A_4>, <A_1,A_2,A_3>\}$ and $V_4 = \{A_5,A_6\}$

$Site_3$ sends matrix M_3 and vector V_3 to $Site_4$. $Site_4$ finds scalar dot product $M_{3.4}$ by using the above two matrices M_3 and M_4.

$$M_{3.4} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Based on the matrix $M_{3.4}$, frequent item sets and its support are computed.

$\{<A_1,A_5> \rightarrow 2, <A_1,A_6> \rightarrow 2, <A_2,A_6> \rightarrow 2, <A_3,A_6> \rightarrow 2, <A_4,A_6> \rightarrow 3, <A_1,A_3.A_6> \rightarrow 2\}$The matrix M'_4 is formed by $Site_4$ by augmenting M_4, and the computed $M_{3.4}$ to the eceived matrix M_3 from $Site_3$.

$$M'_4 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The computed M'_4 is sent to $Site_1$. After receiving the matrix from $Site_4$, $Site_1$ declares GFI based on MinSupport threshold value and are shown in the following TABLE VI.

TABLE VI
GFI and Supports at Site$_1$ (DB$_1$)

S No.	Item set	Sup	S. No	Item set	Sup	S.No	Item set	Sup
1	A_1	3	7	$\langle A_2,A_3\rangle$	2	13	$\langle A_1,A_6\rangle$	2
2	A_2	3	8	$\langle A_2,A_4\rangle$	2	14	$\langle A_2,A_6\rangle$	2
3	A_3	3	9	$\langle A_1,A_2,A_3\rangle$	2	15	$\langle A_3,A_6\rangle$	2
4	A_4	3	10	A_5	2	16	$\langle A_4,A_6\rangle$	3
5	$\langle A_1,A_2\rangle$	2	11	A_6	4	17	$\langle A_1,A_3,A_6\rangle$	2
6	$\langle A_1,A_3\rangle$	3	12	$\langle A_1,A_5\rangle$	2	18	$\langle A_2,A_4,A_6\rangle$	2

Now the next horizontally partitioned database DB_2 at $Site_2$ is considered to find GFI for its vertically partitioned databases DB_5 and DB_6. The methodology proposed in [2] for finding privacy preserving association rule mining for vertically partitioned databases is applied for these databases DB_5 and DB_6 where DM is at $Site_2$. The Database DB_2 and its vertically partitioned databases DB_5 and DB_6 and are at $Site_2$, $Site_5$ and $Site_6$ respectively. The database DB_5 has 5 transactions and 3 attributes such as $\{A_1, A_2, A_4\}$. The database DB_6 has 5 transactions and 3 attributes such as $\{A_3, A_5, A_6\}$ and these databases are shown below:

TABLE VII
Database DB$_5$

TID\Item	A_1	A_2	A_4
T_1	0	1	0
T_2	0	1	1
T_3	1	0	1
T_4	1	1	0
T_5	0	1	1

TABLE VIII
Database DB$_6$

TID\Item	A_3	A_5	A_6
T_1	1	1	0
T_2	0	0	1
T_3	1	0	1
T_4	1	1	0
T_5	0	1	0

$Site_5$ and $Site_6$ finds frequent item sets for their databases DB_5 and DB_6 based on user specified minimum support threshold value by using apriori algorithm. The determined frequent item sets at two sites are shown below:

At Site$_5$, for database DB$_5$:
The set of frequent item sets and its support is $\{A_1 \rightarrow 2, A_2 \rightarrow 4, A_4 \rightarrow 3, <A_2, A_4> \rightarrow 2\}$

At Site$_6$, for database DB$_6$:
The set of frequent item sets and its support is $\{A_3 \rightarrow 3, A_5 \rightarrow 3, A_6 \rightarrow 2, <A_3, A_5> \rightarrow 2\}$

$Site_5$ prepares a matrix M_5 consisting of supporting transactions for frequent item sets and vector V_5 possessing frequent item sets. Similarly $Site_6$ prepares a matrix M_6 to represent supporting transactions for each frequent item sets and also constructs vector V_6 to store all frequent item sets. The matrices $M_{5.}$,$M_{6,}$ and vectors V_5,V_6 are shown below:

$$M_5 = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \quad V_5 = \{A_1, A_2, A_4, <A_2, A_4>\}$$

$$M_6 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \quad V_6 = \{A_3, A_5, A_6, <A_3, A_5>\}$$

$Site_5$ sends M_5 and V_5 to $Site_6$.
$Site_6$ finds scalar dot product $M_{5.6}$ using the two matrices M_5 and $M_{6.}$

$$M'_{5.6} = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \quad V_{5.6} = \{<A_1,A_3>, <A_2,A_3>, <A_2,A_5>, <A_2,A_3,A_5>\}$$

From matrix $M_{5.6}$, based on MinSupport threshold value, frequent item sets and its support are computed.

$\{<A_1,A_3> \rightarrow 2, <A_2,A_3> \rightarrow 2, <A_2, A_5> \rightarrow 3, <A_2, A_3, A_5> \rightarrow 2\}$

The matrix M'_6 is formed by $Site_6$ by augmenting M_5, and the computed $M_{5.6}$ to the received matrix M_5 from $Site_{5.}$

$$M'_6 = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

The computed M'_6 is sent to the $Site_2$. After receiving the matrix M'_6 from $Site_{6,}$ $Site_2$ declares GFI based on MinSupport threshold value and is shown in the following TABLE VIII.

TABLE IX
GFI and Supports at Site$_2$

S. No	Item set	Sup	S. No	Item set	Sup	S. No	Item set	Sup
1	A_1	2	5	A_3	3	9	$\langle A_1,A_3\rangle$	2
2	A_2	4	6	A_5	3	10	$\langle A_2,A_3\rangle$	2
3	A_4	3	7	A_6	2	11	$\langle A_2,A_5\rangle$	3
4	$\langle A_2,A_4\rangle$	2	8	$\langle A_3,A_5\rangle$	2	12	$\langle A_2,A_3,A_5\rangle$	2

After computing GFI for its vertically partitioned databases, $Site_1$ stores these item sets in a set denoted as FI_1. Similarly $Site_2$ also prepares a set FI_2 which consists of GFI of its vertically partitioned databases. Each site applies the procedure specified in [1] which uses a sign based secure sum cryptography technique to preserve the individual's privacy in the process of finding GFI for the database DB_1 and DB_2.

$FI_1=\{A_1,A_2,A_4,<A_2,A_4>,A_3,A_5,A_6,<A_3,A_5>,<A_1,A_3>,<A_2,A_3>,<A_2,A_5>,<A_2,A_3,A_5>\}$

$FI_2=\{A_1,A_2,A_3,A_4,<A_1,A_2>,<A_1,A_3>,<A_2,A_3>,<A_2,A_4>,<A_1,A_2,A_3>,<A_5,A_6>,<A_1,A_5>,<A_1,A_6>,<A_2,A_6>,<A_3,A_6>,<A_4,A_6>,<A_1,A_3,A_6> <A_2,A_4,A_6>\}$

$Site_1$ and $Site_2$ sends FI_1 & FI_2 to TP in order to find GFI. Trusted authority prepares a merged list and stores in set MList.

$MList=\{A_1,A_2,A_3,A_4,,A_5,A_6,<A_1,A_2>,<A_1,A_3>,<A_1,A_5>,<A_1,A_6>,<A_2,A_3>,<A_2,A_4>,<A_2,A_5><A_2,A_6>,<A_3,A_5>,<A_3,A_6>,<A_4,A_6>,<A_1,A_2,A_3>,<A_1,A_3,A_6><A_2,A_4,A_6>,<A_2,A_3,A_5>\}$

After preparing the merged list, $Site_0$ (Trusted Party) broadcast MList to $Site_1$ and $Site_2$ to initiate to do further processes. Using Sign based secure sum, $Site_1$ and $Site_2$ provides support in disguised form for each item set in the merged list MList. $Site_0$ also sends unique random number and unique signs to Site 1 and Site2 respectively along with MList.

The following are the random numbers and signs sent by $Site_0$ (Trusted Party) to $Site_1$ and $Site_2$.

$Site_1$ has $RN_1 = 26$, $Sign_1$ = ('-') and $Site_2$ has $RN_2 = 370$ $Sign_2$ = ('-').

TP computes SignSumRN by adding random numbers with signs as

$$SignSumRN = (-)\ 26 + (-)\ 370 = -396$$

Computations for some of the item sets in merged list are illustrated below:

Consider the three item sets from MList which are $\{A_1,<A_2,A_5>,<A_2,A_4,A_6>\}$

Let us assume user specified minimum support threshold value as 40%.

The computations of partial supports and total supports for the above three item sets at $Site_1$ are shown below:

At $Site_1$:

For Item Set $<A_1>$:

Partial support = $A_1 \cdot sup - 40\% |DB_1| + (Sign_1)*RN_1 = 2 - 2 + (-)*26 = -26$

For Item Set $<A_2, A_5>$:

$Site_1$ do not know possess $<A_2, A_5>$ since this item set is infrequent from its partitioned sites. But there is a chance to become global frequent when local supports are taken into consideration even if it is infrequent at one site. So $Site_1$ itself finds support of $<A_2, A_5>$ by doing scalar product between $<A_2>$ and $<A_5>$.

$<A_2> = < 1, 0, 0, 1, 1>$ and $<A_5>= < 1, 1, 0, 0, 0>$

$<A_2, A_5>= <1, 0, 0, 0, 0>$

$Site_1$ computes partial support value for $<A_2, A_5>$ in order to compute global support.

Partial support = $<A_2, A_5>.sup - 40\% |DB_1| + (Sign_1)*RN_1 = 1 - 2 + (-)*26 = -27$

For Item set $<A_2, A_4, A_6>$:

Partial support = $<A_2,A_4,A_6> \cdot sup - 40\% |DB_1| + (Sign_1)*RN_1 = 2 - 2 + (-)*26 = -26$

The computations of partial supports for the above three item sets at $Site_2$ are shown below:

At $Site_2$:

For Item set $<A_1>$:

Partial support = $A_1 \cdot sup - 40\% |DB_2| + (Sign_2)*RN_2 = 3 - 2 + (-)*370 = -369$

For Item set $<A_2, A_5>$:

Partial support = $<A_2, A_5> \cdot sup - 40\% |DB_2| + (Sign_2)*RN_2 = 3 - 2 + (-)*370 = -369$

For Item set $<A_2, A_4, A_6>$:

Partial support = $<A_2,A_4,A_6> \cdot sup - 40\% |DB_2| + (Sign_2)*RN_2$ But $Site_2$ do not know possess the support value of $<A_2, A_4, A_6>$ since this item set is infrequent in its partitioned sites. But there is a chance to become global frequent item set when local supports are taken into consideration. So $Site_2$ itself finds support of $<A_2, A_4, A_6>$ by doing scalar dot product between $<A_6>$ and $<A_2, A_4>$.

$<A_6> =<0, 1, 1, 0, 0>$ and $<A_2, A_4>= <0, 1, 0, 0, 1>$

$\therefore \{<A_6> \cdot <A_2, A_4>\} = <0, 1, 0, 0, 0>$

So the item set $<A_2, A_4, A_6>$ support value is found and it is 1 which can be substituted in the formula for finding partial support value.

Partial support = $<A_2,A_4,A_6>.sup - 40\% |DB_2| + (Sign_2)*RN_2 = 1 - 2 + (-)*370 = -371$

$Site_1$ and $Site_2$ find the total support by exchanging its computed partial support values. The total support computations for the considered three item sets in the MList are as follows:

$Site_1$ and $Site_2$ finds Total support of an item set by adding it's computed partial support of an item sets with its received partial support of the same item set.

At $Site_1$ & At $Site_2$:

For Item set $<A_1>$:

Total support = Partial support of $<A_1>$ at $Site_1$ + Partial support of $<A_1>$at $Site_2$ = -26-369=-395

For Item set $<A_2, A_5>$:

Total support = Partial support of $<A_2, A_5>$ at $Site_1$ + Partial support of $<A_2, A_5>$at $Site_2$ = -27-369= -396

For Item set $<A_2, A_4, A_6>$:

Total support = Partial support of $<A_2, A_4, A_6>$ at $Site_1$ + Partial support of $<A_2, A_4, A_6>$at $Site_2$ = -26 -371 = -397

$Site_1$ and $Site_2$ sends it's computed total support values to TP in order to find GFI.

At $Site_0$ (Trusted Party):

After receiving total partial support values from $Site_1$ and $Site_2$ thetrusted party computes Global excess support and then computes actual support values for these item sets to determine GFI.

For Item set $<A_1>$:

Global excess support = Total support of $<A_1>$ - (SignSumRN) = -395– (-396)=1

Actual support = Global excess support of <A_1> + 40% of |DB|)= 1+4=5

Hence the item set A_1 is globally frequent as its global support is 5 which is greater than 4

For Item set <A_2, A_5>:

Global excess support = Total Support of <A_2, A_5> - (SignSumRN) = -396 –(-396)=0

Actual support = Global excess support of <A_2, A_5> + 40% of |DB|) = 0+4=4

Hence the item set <A_2, A_5> is globally frequent as its global support is 4 even though this item set infrequent at $Site_1$.

For Item set <A_2, A_4, A_6>:

Global excess support = Total support of <A_2, A_4, A_6> - (SignSumRN) = -397 –(-396)= -1

Actual support = Global excess support of <A_2, A_4, A_6> + 40% of |DB|) = -1 +4=3

Hence the item set <A_2, A_4, A_6> is globally infrequent as its global support is 3 which is less than 4.

The above steps will be repeated for the remaining item sets in merged list MList to find its global support. From the computed global support values, TP decides which are globally frequent and which globally infrequent item sets are. The eleven GFI and their support values are specified as {A_1->5, A_2->7, A_3->6, A_4->6, A_5->5, A_6->6, <A_1, A_3> -> 5, <A_2, A_3> ->4, <A_2, A_4> -> 4,<A_2, A_5> -> 4, <A_4, A_6> -> 5}

By this way no site can predict any information about local support values of any site since no site can predict random numbers or signs assigned to other sites by TP. Even TP cannot predict individual site's information based on the received total support values which are in disguised form.

$Site_1$ sends the GFI with support values to its partitioned database sites $Site_3$ and $Site_4$. Similarly $Site_2$ sends the same results to its partitioned database sites $Site_5$ and $Site_6$. So any site can generate global association rules based on the received GFI along with their support values and user specified minimum confidence threshold value.

Case II:

In Case II, second mixed model (Model-2) is considered. The database is initially vertically partitioned into two databases DB_1 and DB_2 and which are further horizontally partitioned into DB_3 & DB_4, DB_5 & DB_6 respectively.

Consider the sample database given in the TABLE I which has 10 transactions and 5 attributes.

The hierarchy of this model can be treated as a three level hierarchy model that is at level zero ($Level_0$) DM exist and at level one ($Level_1$) two vertically partitioned sites exist and at level two ($Level_2$), two horizontally partitioned databases exist. The DM is at $Site_0$, Vertically partitioned databases of DB_1 and DB_2 are in $Site_1$, $Site_2$ respectively and which are at $Level_1$. The horizontally partitioned data bases that is DB_3, DB_4, DB_5 and DB_6 are at $Site_3$,$Site_4$,$Site_5$ and $Site_6$ respectively and which are at $level_2$. The DM who has special privileges to find global association rules from the databases without violating any individual's privacy constraints. The databases and its sites in the considered model are shown in the following diagram.

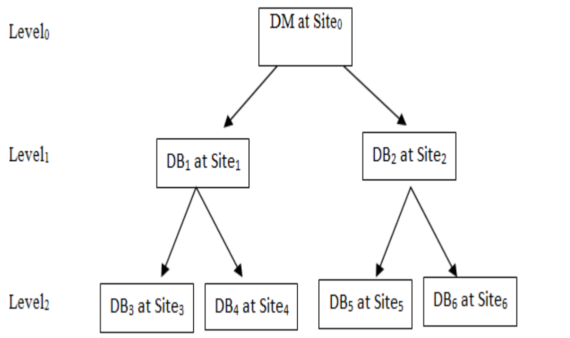

Fig. 4 A Mixed Model Consisting of Horizontal and Vertical Databases

Initially, the database DB given in TABLE I partitioned into two disjoint vertical databases DB_1 and DB_2 and are shown in TABLE X and XI The database DB_1 consisting of 10 transactions (T_1, T_2, .. T_{10}) with 3 attributes (A_1, A_2, A_4) where as DB_2 consisting of 10 transactions (T_1, T_2, .. T_{10}) with 3 attributes (A_3, A_5, A_6).

TABLE XI
Database (DB_1)

TID\ Item	A_1	A_2	A_4	TID\ Item	A_1	A_2	A_4
T_1	1	1	0	T_6	0	1	0
T_2	0	0	1	T_7	0	1	1
T_3	1	0	0	T_8	1	0	1
T_4	0	1	1	T_9	1	1	0
T_5	1	1	1	T_{10}	0	1	1

TABLE XII
Database (DB_2)

TID\ Item	A_3	A_5	A_6	TID\ Item	A_3	A_5	A_6
T_1	1	1	0	T_6	1	1	0
T_2	0	1	1	T_7	0	0	1
T_3	1	0	1	T_8	1	0	1
T_4	0	0	1	T_9	1	1	0
T_5	1	0	1	T_{10}	0	1	0

The Database DB_1 at $Site_1$ is further partitioned into two disjoint horizontally partitioned databases DB_3 and DB_4 and which are in $Site_3$ and $Site_4$ respectively. The database DB_3 has 5 transactions and 3 attributes such as {A_1, A_2, A_4}. The database DB_4 has 5 transactions and 3 attributes such as {A_4, A_5, A_6}.

$Site_3$ and $Site_4$ finds frequent item set list FI_3 &FI_4 for their databases DB_3 and DB_4 respectively based on user specified MinSupport threshold by using apriori algorithm and are as follows.

FI_3 = { A_1→3, A_2→3,A_4→3, <A_1,A_2>→2, <A_2,A_4>→2}

FI_4 = { A_1→2, A_4→3}

$Site_5$ and $Site_6$ finds frequent item set list FI_4 &FI_5 for their databases DB_5 and DB_6 respectively based on user specified MinSupport threshold by using apriori algorithm and are as follows.

FI_5 = { A_3→3, A_5→2, A_6→4, <A_3,A_6>→2}

FI_6 = { A_3→3, A_5→3,A_6→2,<A_3,A_5>→2}

$Site_1$ applies the methodology specified in [1] to find the GFI for its horizontally partitioned databases DB_3 and DB_4. The same procedure is also followed by $Site_2$ to find GFI for its partitioned databases DB_5 and DB_6. Here the $Site_1$ and $Site_2$ acts as a TP to its partitioned databases. After completion of this process, $Site_1$ and $Site_2$ consists of GFI of its partitioned databases.

$Site_1$'s global frequent item set is $\{A_1 \to 5, A_2 \to 7, A_4 \to 6, \langle A_2, A_4\rangle \to 4\}$

$Site_2$'s global frequent item set is $\{A_3 \to 6, A_5 \to 5, A_6 \to 6\}$

Now the methodology specified in [1] is applied for finding global association rules for vertically partitioned databases DB_1 and DB_2 of DB with DM at $Site_0$.

$Site_1$ and $Site_2$ obtain the transactions related to global frequent items sets from its partitioned databases.

$Site_1$ prepares a matrix M_1 & vector V_1 consisting of supporting transactions related to frequent item sets & frequent item sets respectively. Similarly $Site_2$ prepares a matrix M_2 & vector V_2 consisting of supporting transactions related to frequent item sets & frequent item sets respectively. $Site_1$ sends matrix M_1 and Vector V_1 to $Site_2$ to in order to find GFI. On receiving matrix M_1 and V_1 from $Site_1$, $Site_2$ computes scalar dot product over M_1 and M_2 to get $M_{1.2}$. The matrices M_1, M_2 & $M_{1.2}$ and vectors V_1,V_2 & $V_{1.2}$ are shown below:

$$M_1 = \begin{bmatrix} 1010100110 \\ 1001111011 \\ 0101101101 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 1010110110 \\ 1100010011 \\ 0111101100 \end{bmatrix} \quad M_{1.2} = \begin{bmatrix} 0001101001 \\ 1010100110 \\ 1000110010 \\ 0001101001 \\ 1000010011 \\ 0101101100 \end{bmatrix}$$

Vectors $V_1 = \{\{A_1, A_2, A_4, \langle A_2, A_4\rangle\}$, $V_2 = \{A_3, A_5, A_6\}$, $V_{1\cdot 2} = \langle A_2,A_4\rangle, \langle A_1,A_3\rangle, \langle A_2,A_3\rangle, \langle A_2,A_4\rangle, \langle A_2,A_5\rangle, \langle A_4,A_6\rangle$. Finally, matrix M'_2 is formed by augmenting M_2 and the computed matrix $M_{1\cdot 2}$ to the received matrix M_1 from $Site_1$. The matrix M'_2 is shown below:

$$M'_2 = \begin{bmatrix} 1&0&1&0&1&0&0&1&1&0 \\ 1&0&0&1&1&1&1&0&1&1 \\ 0&1&0&1&1&0&1&1&0&1 \\ 0&0&0&1&1&0&1&0&0&1 \\ 1&0&1&0&1&1&0&1&1&0 \\ 1&1&0&0&0&1&0&0&1&1 \\ 0&1&1&1&1&0&1&1&0&0 \\ 1&0&1&0&1&0&0&1&1&0 \\ 1&0&0&0&1&1&0&0&1&0 \\ 0&0&0&1&1&0&1&0&0&1 \\ 1&0&0&0&0&1&0&0&1&1 \\ 0&1&0&1&1&0&1&1&0&0 \end{bmatrix}$$

The DM declares GFI using the values in the received matrix M'2 based on user specified MinSupport threshold value. The eleven GFI are $\{A_1\text{->}5, A_2\text{->}7, A_3\text{->}6, A_4\text{->}6, A_5\text{->}5, A_6\text{->}6, \langle A_1, A_3\rangle \text{->} 5, \langle A_2, A_3\rangle \text{->}4, \langle A_2, A_4\rangle \text{->} 4, \langle A_2, A_5\rangle \text{->} 4, \langle A_4, A_6\rangle \text{->} 5\}$

V. Privacy Preservation For The Proposed Models

- The individual site's privacy is achieved even when the database is partitioned into many disjoint segments in different ways and in different levels. Other than parent site no site can predict any site's private data or information.
- For solving horizontally partitioned databases at a level, privacy is achieved by using Horiz-TP-Algorithm which incorporates Encryption, Decryption and sign based secure sum techniques. The sites which are at same levels partitioned from the parent, exchange information in disguised form between them in order to find the GFI of the data bases. So they cannot predict other site's database information.
- In case of solving vertically partitioned databases, privacy preserving association rule mining can be determined efficiently with Verti-DM-Algorithm which adopted cryptography and Scalar product techniques.
- When the site's databases are not partitioned from the same parent but are at same level. Privacy is still achieved as there is no single communication is allowed between them. In a case even different parent nodes are at same level, privacy is achieved since a node which acts as a parent of its child nodes have communication only with own parent node.
- The data transfer between sites is done as a bulk data transfer instead of single data transfer. So minimum number of communications is required to find global frequents item sets while achieving privacy at all levels for each site's database.

Hence, the proposed methods are efficient in finding privacy preserving association rule mining for mixed partitioning databases.

VI. Conclusions

Privacy is becoming a great research topic in the process of applying data mining techniques to various real applications. As the necessity makes the people to share the knowledge to the legitimate people in order to gain mutual benefits and this issue made to study privacy preserving data mining. Among many data mining techniques, privacy preserving association rule mining is a popular technique. But finding an efficient solution satisfying both privacy constraints as well as accuracy is a challenging task to researchers. Privacy constraints differ from centralized database environment to distributed database, so methodologies also differ from one environment to other. A database in distributed environment can be partitioned in different ways like horizontal, vertical or mixed mode. In this paper, two new methodologies are presented to find global association rules in distributed environment by satisfying privacy constraints for two common mixed partitioned models. Algorithms are also presented for each mixed model and implementation is discussed with suitable database. The efficiency of the proposed model is discussed in terms of privacy and communication.

REFERENCES

[1] R Agarwal, T Imielinski and A Swamy(1993), "Mining Association Rules between Sets of Items in Large Databases", *Proc.* of the ACM SIGMOD *Intl. Conf.* on Management of Data, pp 207-210, 1993.

[2] R. Agrawal and R. Srikant.(1994.) "Fast, algorithms for mining association rules in large databases", *Proc*. 20th VLDB *Conf.* Santiago, Chile, pp 487-499.

[3] C. Clifton and D. Marks(February 1996), "Security and privacy implications of data mining. In Workshop on Data Mining and Knowledge Discovery", pp 15–19, Montreal, Canada, University of British Columbia, Department of Computer Science.

[4] *Verykios, V.S., Bertino, E., Nai Fovino, I., Parasiliti, L., Saygin,Y., and Theodoridis, Y. 2004*. State-of-the-art in privacy preserving data mining. SIGMOD Record, 33(1):50–57.

[5] Charu C. Aggarwal and Philip S. Yu(July 2008), "A general survey of privacy - preserving data mining models and algorithms", Privacy-preserving data mining: models and algorithms.

[6] *Elisa Bertino , Igor Nai Fovino Loredana Parasiliti Provenza(*2005), A Framework for Evaluating Privacy Preserving Data Mining Algorithms, Data Mining and Knowledge Discovery, 11, pp 121–154.

[7] Alexandre Evfimievski, Ramakrishnan Srikant, Rakesh Agrawal, Johannes Gehrke (, 2002), "Privacy Preserving Mining of Association Rules", SIGKDD 02 Edmonton, Alberta, Canada.

[8] Shaofei Wu1, Hui Wang2(, 2008), "Research On The Privacy Preserving Algorithm Of Association Rule Mining In Centralized Database",Intl. Symposiums on Information Processing ,pp 131-134.

[9] Gkoulalas Divanis, V.S. Verykios(2010), "Association Rule Hiding for Data Mining", Advances in Database Systems, Vol. 41.

[10] A.C. Yao.(1982) "Protocols for secure computations". InProc. of the 23rd Annual IEEE Symposium on Foundations of Computer Science.

[11] Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, and Michael Y. Zhu(2003), Tools for privacy preserving distributed data mining, SIGKDD Explorations, Vol. 4, No. 2 pp 1-7.

[12] Mohamed Hussein(2009), "Privacy Preserving in Association Rule Mining using cryptography", Master's Thesis, Menofia University, Egypt.

[13] M. Kantarcioglu and C. Clifto.(2004), "Privacy-preserving distributed mining of association rules on horizontally partitioned data", In IEEE Transactions on Knowledge and Data Engineering Journal, Vol. 16(9), pp 1026–1037.

[14] Vaidya, J. & C. Clifton, C.(2002), "Privacy preserving association rule mining in vertically partitioned data", in 'The Eighth ACM SIGKDDIntl. Conf. on Knowledge Discovery and Data Mining', SIGKDD, ACM Press, Edmonton, Canada, pp. 639–644.

[15] Yiqun Huang, hengding Lu, HepingHu(2005), "Privacy preserving association rule mining with scalar product", Natural Language Processing and Knowledge Engineering, IEEE NLP-KE '05,*Proc.* IEEE, *Intl*.

[16] Asha Khatri, Swathi Kabra, Shamsher Singh, Durgesh kumar Mishra(2010), "Architecture for Preserving Privacy During Data Mining by Hybridization of Partitioning on Medical Data", Fourth Asia Intl. Conf. on Mathematical/Analytical Modeling and Computer Simulation. pp 93-97.

[17] Hussein, A. El-Sisi, and N. Ismail(2008), "*Mining Association Rules on Mixed Database with Privacy Preserving",* The 4th International Computer Engineering Conf. Information Society Applications in the Next Decade, ICENCO2008, Giza, Egypt.

[18] N V Muthu Lakshmi, Prof. K. Sandhya Rani (2012,), "Privacy Preserving Association Rule Mining in Horizontally Partitioned Databases Using Cryptography Techniques", (IJCSIT) ISSN : 0975-9646,*Intl. J.* of Computer Science and Information Technologies, Vol. 3(1), pp 3176-3182.

[19] N V Muthu Lakshmi, Prof. K. Sandhya Rani(2012), "Privacy reserving Association Rule Mining in Vertically Partitioned Databases", *Intl. J.* of Computer Applications (IJCA), (0975 – 8887) Vol. 39, No.13, pp 29-35.