Anonymizing Data records in Distributed Data mining over Horizontal partitioned data

Dr. Ch G.V.N. Prasad, Katkam Gaurav, Nunavath Yakaiah & R Bharath Kumar

Abstract—

Data mining practices are utilized to find out hidden information from large databases. Among several data mining techniques, association rule mining is receiving more focus on the experts to discover correlations between items or items sets efficiently. In distributed database atmosphere, the way the information is distributed plays a crucial part in the issue definition. The information could be distributed horizontally or vertically or in hybrid mode among different websites. There is an increasing demand for computing global association rules for the sources goes to different internet sites in ways that individual data isn't unmasked and site manager knows the global studies and their individual data only. In this paper a model is suggested which assumes a sign based safe total cryptography technique to find worldwide association rules with trusted party by protecting the privacy of the individual's data if the data is distributed horizontally among different websites.

Keywords—

Data Mining, Distributed Database, Privacy Preserving Association Rule Mining, Cryptography Technique.

INTRODUCTION

Data mining continues to be regarded as a threat to privacy due to the widespread expansion of digital data maintained by companies. It's lead to increased concerns in regards to the privacy of the underlying data. Hidden information is found by data mining techniques from large database while secret data is stored safely when data is permitted to access by single person. Now a days many people want to access data or hidden information using data mining approach actually they're maybe not fully authorized to access. For getting mutual gains, many organizations desire to share their data to many genuine people but without exposing their secret data.

In large applications the complete data might be in place called central or multiple internet sites called distributed database. Techniques are offered by several authors for both centralized in addition to distributed database to safeguard private information. This report deals with privacy preserving in distributed database environment while revealing found knowledge/hidden information to a lot of reliable people.

In distributed environment, database is actually a selection of numerous, logically interrelated databases distributed over a computer network and are distributed among number of sites. As the database is distributed, different users can access it without interfering with one another. In distributed environment, database is partitioned in to disjoint pieces and each site contains just one fragment. Data could be partitioned in various ways for example outside, vertical and mixed.

The still another partitioning method is mixed fragmentation where information is partitioned horizontally and then each partitioned fragment is further partitioned in to pieces and vice-versa [1]. Figure 1.a shows a way for combined partitioned by which information is first partitioned vertically and then horizontally. Amount 1.b shows yet another combined method where data is partitioned horizontally and then vertically partitioned.

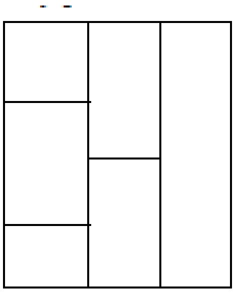

Figure1.a:Vertically partitioned database is further partitioned into horizontal

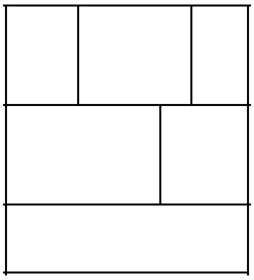

Figure1.b: Horizontally partitioned database is further partitioned into vertical.

In data mining, association rule mining is a common and well researched way of finding interesting relations between variables in large databases. Choosing the world wide association rules is a difficult task because the privacy of the individual site's data is to be preserved, when data is distributed among various websites. In this paper, a model is offered to locate worldwide association rules by preserving the privacy of individual sites data when the data is partitioned horizontally among n amount of sites.

DISTRIBUTED ASSOCIATION RULE MINING

Association rule mining method is receiving more attention between data mining techniques to the researchers to discover correlations between items or item sets. These rules can be analyze to make strategic decisions to progress the performance of the business or quality of the association service and so on. Association rule mining was initiated by *Agarwal* [3]. An association rule can be definite formally as follows. Let $I = \{i_1, i_2 ..., i_m\}$ be the set of attributes called items. The item set X consisting of one or additional items. Let $DB = \{t_1, t_2 ..., t_m\}$ be the database consisting of n number of boolean transactions, and every transaction t consisting of items supported by ith transaction. An item set X is said to be regular when number of transactions supporting this item set is greater than or equal to the user particular minimum support threshold or else it is said to be infrequent. An association rule is an suggestion of the form $X \rightarrow Y$ where X and Y are displace subsets of I, X is called the antecedent and Y is called consequent. An organization rule $X \rightarrow Y$ is said to be strong association rule only when its confidence is superior than or equal to user specified least confidence.

Connection principle technology has two methods. Calculation of frequent item sets in the database based on user-specified minimum support threshold may be the first phase and this process is difficult because it requires seeking all combinations of item sets. In the second step, the association rules can be easily generated based on user specified minimal confidence threshold for the frequent item sets that are generated in the initial step. Now a days, more individuals who desire to provide mutual benefit to their partners want to get access over association rules which are based on large database even though they are neither owners nor holding rights to access. The database owners also need to discuss their derived results, that is association rules to get some good benefits from them but they don't need to offer their secret data and also loss of secret data could cause injury or loss. Sharing of knowledge is the primary concern in a few applications for that mutual benefits in knowledge discovery system-while protecting privacy of individual is still another concern. Distributed association rule mining is just association principles computed internationally from n number of web sites in distributed environment by gratifying privacy restrictions. World wide support count of an item set X, could be computed the following.

$$X.\text{supp}\,ort = \sum X_i.\text{sup}$$

An item set, X is globally frequent when its hold up value computed from all sites is > user particular minimum support threshold * |DB|.

An item set X which is nearby frequent in a site may be infrequent globally and an item set can be nearby infrequent in one or more sites may be globally normal since global support is computed by allowing for the support value of the item set at all sites. The aim of the global organization rule mining is to find all rules for each global frequent item set where global confidence is superior than or equal to the user specified minimum confidence. A global association rule can be spoken as follows.

$A \rightarrow B$, where A and B are disjoint subsets of I.

The global self-assurance of a rule is global support (AUB) / global support (A).

The rule $A \rightarrow B$ is said to be strong only when its global confidence > user particular minimum confidence * |DB| otherwise it is weak rule.

In distributed environment, the demanding task is how effectively one can give accurate knowledge to their companions while no single secret knowledge is unveiled to them to have goodwill. This problem makes the researchers to review further to propose means of privacy preserving association rule mining. The proposed design for privacy preserving association rule mining for horizontally partitioned distributed databases is explained in these sections.

PRIVACY PRESERVING ASSOCIATION RULE MINING FOR HORIZONTALLY PARTITIONED DISTRIBUTED DATABASE

Many scientists suggested many methods for privacy preserving association rule mining for both centralized and distributed databases. This paper also describes the various proportions of preserving data mining techniques such as data distribution, data adjustment approach, data mining methods, data or rule hiding and ways for privacy preserving data mining techniques. The review of their relevance to the field of privacy-preserving data mining and fundamental paradigms and notations of secure multi-party computations are presented by the authors in [4]. In addition they discussed the matter of performance and demonstrate the issues involved with building highly-efficient methods. In [5], the authors suggested a framework for analyzing privacy preserving data-mining algorithms and centered on their frame work you can measure the different features of privacy preserving algorithms according to different assessment criteria.

In [6], Secure mining of association rules over horizontally partitioned database using cryptographic way to minimize the data shared by adding the overhead towards the mining process is introduced. In [7], authors proposed an enhanced kantarcioglu and Clifton’s systems which really is a stage for privacy preserving in distributed data mining. They introduced two practices to boost the safety against collusion in the communication environment with or without reliable party.

A new algorithm which is the modification of the existing algorithm and according to a model with minimal collision probability is proposed in [8]. They also used cryptography ways to preserve the privacy. They suggested a brand new solution by establishing the benefits of the initial approach which safeguards the privacy of the data by having an expanded position based access get a grip on approach and the 2nd approach which uses cryptographic techniques with the view of reducing loss of information and privacy. In [10], writers addresses the situation of association rule mining in vertically partitioned database by utilizing cryptography-based method and also gifts the analysis of communication and protection. The annals of secure multi-party calculation in two milliner’s problem, in which they need to know who is richer without revealing their prosperity is addressed in [11]. Practices may also be proposed for two milliner’s problem as well for m-party situation.

The proposed method solves this problem using a cryptography technique that's indication based secure sum. Cryptography method protects data/ information from the others opening in a distributed database environment for semi honest type. In this paper to safeguard one’s regional frequent item sets from the others accessing, public-private key algorithm is employed. The security is enhanced by the proposed sign based secure sum concept by shifting uncertain effects between websites along the way of producing frequent item sets and rules. In this method, a particular site is designated and this site owner is named Trusted Party and initiates the procedure for finding association rules without knowing any one’s person data/information but by getting processed results from all of the sites in secure manner.

In spread environment when information is partitioned horizontally among various web sites with a party is known as in this paper. In horizontally partitioned distributed database, different set of records with same set of characteristics of whole database are placed at different sites. The associations between items or product sets can be found correct as long as principles are determined from results of total set of records from all sites. But no single site owner wishes to provide no single file to any site and this makes a challenging one to the issue that is extracting vital information from all sites information without accessing individual records to create association rules.

A Proposed Model

In the horizontally partitioned distributed database model, there's n amount of websites and every site manager has local autonomy over their database and one special site named Trusted Party (TP) who has special rights to execute certain tasks.

The proposed approach consisting of many tasks, performed by both sites in addition to TP to get worldwide organization principles while preserving individual’s private data. The next diagram shows the interaction between TP and websites in the proposed model. In this projected model, distributed database consists of n number of separation distributed databases and accessible in n number of sites termed as Site1, Site2, ...

Site$_n$. The Site! maintains a DB_i whose length is $|DB_i|$ where $1 \le i \le n$. Total number of communication in all sites ($|DB_i|$ is

$$DB_i \, |DB| = \sum_{i=1}^{n} |DB_i|$$

Every site helps to create global association rules and involves global frequent item sets. So the goal would be to determine global frequent item sets with supports on the basis of the listings in any way internet sites. Any item set is said to be globally consistent only if amount of support price of item sets whatsoever sites is more than or equal to minimum amount of transactions necessary to support this item globally. it exists in a minimum of more than one internet sites as frequent something set might be internationally frequent only. Likewise an item set could be internationally infrequent only if the item set is infrequent in a minimum of one or more web sites.

A few duties should be performed by both TP along with site owners, to find international association rules in horizontally partitioned allocated sources of size n (> 2).

It's clear that no one is willing to expose their local frequent item sets, supports and database size to any site owner as well as to TP. To solve this problem special privileges are provided by the method to TP to fully capture regional frequent item sets without taking the value of helps from all sites to determine all sites frequent item sets. Every site operator welcomes to offer regional frequent item sets in encrypted form to TP to whom they trusts to generate merged frequent item set list.

Terms	Description
ASj	Actual Support
GESj	Global Excess Support for
PS,,	partial support of item set Xj
RN$_i$	random number for Site$_i$
Signi	Sign used with random
SignSumRN	sum of random numbers along with respective signs
TotalPSy	Sum of PS,j of item set Xj, where i indicates site number
TP	Trusted Party
MinSup	Minimum support threshold
MinConf	Minimum confidence

Sign was adopted by the proposed model based protected total cryptography solution to find international association rules by preserving the privacy. The various methods in the proposed design are the following.

Step l. The first process is initiation completed by the TP, and sends request to discover frequent item sets to all sites by delivering public-key, minimum supports.

Step 2. On receiving the public key and ceiling, each site sees Local frequent item sets because of their data utilizing the algorithm. For their generated set of frequent item sets, every site applies encryption algorithm to convert frequent item sets into encrypted form using the public key and send it toTP.

Step 3. TP then decrypts the each site's encrypted information by using Private key and prepares a merged list which consists of all site's restricted frequent item sets after eliminating duplicates. For each site, TP generates a exclusive random numbers and a sign (+ or -). The merged list along with individual random number (RN_i) and a sign (Sign) are sent to each site. The Sign field indicates whether the random number is to be added or deduct from its partial support value (PS_{ij}) .

Step4. Each site computes partial hold up for each item set in the merged list which is conventional from TP by using the formula $PS_{ij} = X_j.\sup - Min.\sup \times |DB_i| + (sign_i)RN_i$ where i indicates the ith site, ranges from 1 to n and j indicates jth item set in the merged list, ranges from 1 to k. Each site then transmit its computed PS_{ij} values for all the article sets in the merged list to all other sites.

Step5. Each site computes Total PS_{ij} for each item set X_j by using the formula. *n*

Total PS_{ij} = $\sum_{i=1}^{n} PS_{ij}$ for each j = 1 to k and sends to the TP.

Step6. TP receives the Total PS_{ij} from all the sites for each item set X_j.

Step7. TP verifies the individuality of receiving Total PS_{ij} from all sites. If any incongruity exist then TP request all owners to execute step 5 once again to get the correct results.

Step8. TP computes Global Excess Support (GES_j) for each item set X_j by using the formula

$GES_j = totalPS_{1j} -$ SignSumRN where SignSumRN is computed by adding all the random numbers with their signs by TP. If the calculate value of $GES_j \geq 0$ then the item set X_j is globally frequent otherwise it is globally infrequent.

Step9. For each global frequent item set X_j, TP finds Actual Support (AS_j) as

$$AS_j = GES_j + Min.\sup * |DB|$$

Where

$$|DB| = \sum_{i=1}^{n} |DB_i|$$

Step10. TP broadcast a list which consists of all global recurrent item sets and their values to all sites.

Step11. Each site can generate connection rules with various confidence values by means of the globally frequent item sets and hold up values received from TP. The above procedure is explained with an example in the next section.

IMPLEMENTATION OF THE PROPOSED MODEL WITH SAMPLE DATA

The projected model is illustrated by using three horizontally partitioned disseminated databases for decision privacy preserving association rule mining. In this illustration model, the horizontally partitioned databases called remains such as DB1, DB2 and DB3 are placed in $site_1$, $site_2$ and $site_3$ correspondingly. Apart from these three sites, there exist a special site called Trusted Party site. Sample databases at $site_1$, $site_2$ and $site_3$ are given below.

TABLE II.A DATABASE DB1, AT $site_1$

T-Id	A1	A2	A3	A4	A5
Site1 has the following					
T1	1	0	0	1	0
T2	1	1	0	1	1
T3	0	1	1	0	1
T4	0	0	1	1	1
T5	1	1	0	1	1

TABLE II.B DATABASE DB2 AT $site_2$

T-Id \Item	A1	A2	A3	A4	A5

Site$_2$ has the following database					
T1	0	1	1	1	1
T2	0	0	1	1	1
T3	1	1	1	1	0
T4	1	1	0	1	1
T5	1	1	0	0	1

TABLE II.C DATABASE, DB3 AT $site_3$

T-Id \Item	A1	A2	A3	A4	A5
Site$_3$ has the following database					
T1	1	0	0	1	1
T2	1	1	1	0	1
T3	1	0	1	1	1
T4	1	0	1	1	0
T5	1	0	1	1	1

TP demand three sites to send encrypted form of local frequent item sets by distribution two values such as smallest amount support threshold and public key. Each site computes local frequent entry sets for their database by using minimum hold up threshold value 40% which is sent by the TP. The local frequent item sets (LF) of sites Site1, Site$_2$and Site$_3$, are given below.

Local frequent item sets at $site_1$

LF1= { A1 ,A2,A3 ,A4,A5,(A1 ,A2),(A1 ,A4),(A1 ,A5)
(A2,A4),(A2,A5),(A3,A5),(A4,A5),(A1,A2,A4),
(A1 ,A2,A5),(A1 ,A4,A5),(A2,A4,A5),(A1 ,A2,A4,A5)}

Local frequent item sets at $site_2$

LF2= { A1 ,A2 ,A3 ,A4 ,A5,(A1,A2),(A1,A4), (A1,A5),
(A2,A3), (A2,A4), (A2,A5), (A3,A4),
(A3,A5),(A4,A5),(A1 ,A2,A4),(A1 ,A2,A5),(A2,A3,A4)
(A2,A4,A5),(A3,A4,A5)}

Local frequent item sets at $site_3$

LF3= {A1 ,A3 ,A4 ,A5,(A1,A3),(A1,A4), (A1,A5), (A3,A4),

(A3,A5),(A4,A5),(A1 ,A3,A4,A5) }

After receiving the encrypted form of local common item sets from the sites, TP prepares a merged common item list after eliminating duplicates. The merged list is as follows.

{A1,A2,A3,A4,A5,(A1,A2),(A1,A3),(A1,A4),(A1,A5),
(A2,A3),(A2,A4),(A2,A5),(A3,A4),(A3,A5),(A4,A5),
(A1 ,A3,A4),(A1 ,A3,A5),(A1 ,A4,A5),(A1 ,A2,A4),
(A1,A2,A5),(A2,A3,A4)(A2,A4,A5),(A3,A4, A5),
(A1 ,A2,A4,A5),(A1 ,A3,A4,A5)}

The following are the casual numbers and signs sent by TP along with merged list to the three sites.

$site_1$ received $RN_1 = 20$, $Sign_1 =$ ('+').

$site_2$ received $RN_2 = 39$, $Sign_2 =$ ('-').

$site_3$ received $RN_3 = 41$, $Sign_3 =$ ('-').

Each site computes incomplete support and broadcast to all other sites in organize to find the total partial supports. All three sites broadcast total incomplete supports for all the item sets in the complex list. TP finally declares global frequent item sets by compare global excess support (GES) of an item set with zero where GES, is computed by subtracting SignSumRN from $totalPS_{ij}$.

The following steps illustrate the procedure of finding whether the two item sets in the merged list are globally frequent or not. Deem the two item sets {(A$_3$, A$_5$), (A$_3$, A$_4$, A$_5$)} from the merged list.
Let X1= (A3, A5) and X2 = (A3, A4, A5)
From the tables 2.1, 2.2 & 2.3, length of databases at three sites are given below

$|DB_1| = 5,\ |DB_2| = 5,\ |DB_3| = 5,$ Global database size

Is $|DB| = \sum_{i=1}^{3} |DB_i| = 15$

TP computes SignSumRN by adding three random numbers along with signs by means of the formula SignSumRN = (+) 20 + (-) 39 + (-) 41 = - 60 Partial supports for X1 at different sites are computed as follows.

At $site_1$
PS_n= X1.Sup - 40% of DB1 + ($Sign_1$) RN_1

PS_{11} = 2 - 2 + 20 = 20
At $site_2$
PS_{21} = X_1 .sup - 40% of DB_2 + ($Sign_2$) RN_2

PS21 = 2 - 2 - 39 = - 39
At $site_3$
PS_{31} = X_1.sup - 40% of DB_3 + ($Sign_3$) RN_3 PS_{31} = 3 -2 - 41 = - 40

$site_1$ broadcasts 20 to $site_2$ and $site_3$, Site2 broadcasts - 39 to $site_1$ and $site_2$, and $site_3$ broadcasts -40 to $site_1$ and $Site_2$. $TotalPS_{ij}$ are computed at all sites.
TotalPSn = PS_n + PS21 + $PS_3$1 =20 +(- 39 - 40)
= -59
TotalPS21 = PS21 + (PS_n + $PS_3$1) = - 39 + (20 -40)
= -59
$TotalPS_3$1 = $PS_3$1 + (PS_n + PS21) = -40 + (20 - 39)
= -59
TP receives -59 as total support of an item set X1 from three sites which ensure the computations execute by all sites is correct. TP then calculates Global Excess Support (GES_1) by subtracting SignSumRN from TotalPSn.
$GES_{1=}$ TotalPSn - SignSumRN = -59 - (-60) = 1 The value of GES_1 is 1 which is greater than or equal to 0, so (A3,A5) is affirmed as globally frequent by TP and actual support(AS_1) of X_1 is computed by adding minimum hold up of the total database to GES_1.
AS_1 = GES_1 + MinSup * |DB| = 1 + 6 = 7 where |DB| = 15.
Hence, the global frequent item set (A3,A5) support is 7. Let us find whether the item set X2 is globally common or not.
Partial support for X2 at three sites are computed as follows.
At $site_1$
PS12 = X2.Sup - 40% of DB1 +($Sign_1$) RN1 = 1 - 2 + 20 = 19 At $site_2$
PS_{22} = X_2 .sup - 40% of DB_2 +($Sign_2$) RN_2 = 2 - 2 - 39 = -39 At $site_3$
PS_{32} = X_2.sup - 40% of DB_3 + ($Sign_3$) RN_3 = 2 -2 - 41 = -41 Site1 broadcasts 19 to $site_2$ and $site_3$, $site_2$ broadcasts - 39 to $site_1$ and $site_3$, and $site_3$ broadcasts -41 to $site_1$ and $site_2$. TotalPSi2 are computed at all sites and as follows
TotalPS12 = PS12 + PS22 + $PS_3$2 =19 +(- 39 -41)
= - 61
$TotalPS_{12}$ = $TotalPS_{22}$ = $TotalPS_{32}$ = - 61

Each site sends its computed TotalPSi2 (total support of X_2) to TP. TP then finds GES_2.
GES_2 = $TotalPS_{12}$ - SignSumRN = 59 - (-60)
= -1
The value of GES2 is -1 which is lower than zero, so (A3, A4, A5) is declared as globally infrequent by TP even though it is frequent at Site2 and Site3.
The above method is repeated for all the item sets in the complex list to find whether they are globally common or not. Finally

TP prepares a list which consists of global common item sets and their hold up values, TP then broadcast this list to three sites. This in sequence is given in the following table.

TABLE III GLOBAL FREQUENT ITEM SETS AND SUPPORTS

Item	Sup	Item	Sup	Item Set	Sup
$^{A}1$	11	$(^{A}1,{}^{A}2)$	6	$(^{A}4,{}^{A}5)$	9
$^{A}2$	8	$(^{A}1\ ,^{A}4)$	9	$(^{A}3,{}^{A}4)$	7
$^{A}3$	9	$(^{A}3,{}^{A}5)$	7	$(^{A}1$ $^{A}4{}^{A}5)$	6
$^{A}4$	12	$(^{A}1,{}^{A}5)$	8		
$^{A}5$	12	$(^{A}2,{}^{A}5)$	7		

Even though the merged list consists of 25 item sets only 13 item sets are globally frequent.

Each site can produce global association rules for each global common item set based on the specified smallest amount confidence threshold. The subsequent computations illustrates that how a rule can be confirmed as strong or weak rule based on the user specified minimum confidence threshold value (65%).

For the item set (A_1, A_4, A_5), the various rules that can be generated are {$A_1 \rightarrow (A_4, A_5)$, $A_4 \rightarrow (A_1, A_5)$, $A_5 \rightarrow$ $^{(A}1^{,\,A}4^{),\,(A}1$, $^{A}4^{)}$ $*^{A}5$, $^{(A}1$, $^{A}5^{)}$ $*^{A}4$, $^{(A}4$, $^{A}5^{)}$ $\rightarrow$ $^{A}1$}.

All these rules need not be strong rules. A rule can be declared as strong only when the assurance of the rule is greater than minimum assurance threshold value. For the rule $A_1 \rightarrow (A_4, A_5)$

Confidence of this rule is

$$\text{Sup}(A_1, A_4, A_5) / \text{Sup}(A_1)$$
$$= 6/11 = 54\%$$

The rule, $A_1 \rightarrow (A_4,A_5)$ is a weak rule since rule's assurance is lower than minimum confidence value of 65%.

For the rule $(A_1, A_4) \rightarrow A_5$ assurance of this rule is

$$\text{Sup}(A_1, A_4, A_5) / \text{Sup}(A_1, A_4)$$
$$= 6/9 = 66\%$$

Hence, $(A_1,A_4) \rightarrow A_5$ is a strong rule as its confidence is greater than minimum confidence.

For the rule $(A_4, A_5) \rightarrow A_1$ assurance of this rule is

$$\text{Sup}(A_1, A_4, A_5) / \text{Sup}(A_4, A_5)$$

$$= 6/9 = 66\% \quad M > \text{MinConf}$$

Hence, $(A_4, A_5) \rightarrow A_1$ is a strong rule as its confidence is greater than minimum confidence For the rule $(A_1, A_5) \rightarrow A_4$ Confidence of this rule is

$$\text{Sup}(A_1, A_4, A_5) / \text{sup}(A_1,A_5)$$
$$= 6/8 = 75\ \%$$

The rule, $(A_1,A_5) \rightarrow A_4$ is a strong rule as its assurance is greater than minimum confidence.

PRIVACY PRESERVATION IN THE PROPOSED MODEL

A new model is suggested in this paper to locate effectively privacy-preserving association rule mining in horizontally partitioned databases. Several duties such as studies of domestically frequent item sets, incomplete supports and total supports for each item occur the combined list are performed independently at different sites. Hence the computation time of the proposed design is less. The effectiveness of the proposed method with regards to interaction and privacy is discussed as follows.

Out of this, reliable party can know only local frequent item sets of every site but he does not know the supports of any item and can not anticipate any thing related to sites database.

In dispersed environment, the cost of transmission is measured in terms of the range of communications for data transfer among all the sites and trusted party which are involved in the procedure of finding world wide association rules.

- Trusted party gets complete partial support of each and every item set from all sites to be able to discover the world wide frequent item sets. By having these full supports, trusted party cannot find web sites data/information since the database size of any site and regional supports of any product at any site isn't recognized by trusted party. Even though trusted party given arbitrary numbers, signs to all internet sites and overall database size is well known, he can not predict any site's individual data.

- Finally benefits that are global recurrent item sets and their supports are broadcasted by trusted party to all web sites. With these results, no site owner can predict local support of any global frequent item sets, as global frequent item sets may not be frequent in most sites and any site owner can not predict the contribution of other sites database which makes the item set globally frequent.

- The effectiveness of an algorithm is evaluated with regards to the communication costs incurred throughout data exchange. The proposed design decreases the number of data transfers by allowing the exchange of bulk of data at any given time from site to another site and trusted party to sites. For instance each site sends local frequent item sets of these database in a single information transfer to trusted party and even the sites sends its partial support for each item to other sites in a single transfer rather than delivering one item set‘s partial support in one transfer to other sites. Ergo the proposed design wants less communications.

- Trusted party also transmit all the international frequent item sets for all sites in a single transfer. Ergo the proposed design is more economy with regards to communication cost since it utilizes bulk data transfers.

So Partial Supports have been in broadcast and disguised form towards the internet sites solidly. Each site isn't having any idea in regards to the sign, random number that are assigned by trustworthy party to other sites and the database size of other sites can be not known. Therefore from your Partial Supports, no site can estimate different web sites data/information. Therefore, the sign based safe total idea that will be found in the computation of partial helps enhances the privacy.

The above discussion clearly specifies the proposed model is effective for obtaining global association rules by satisfying privacy constraints.

CONCLUSION

Nevertheless every manager wishes to get into found effect by participating indirectly inside the exploration process by providing partial results in disguised form. The situation of preserving privacy in association rule mining when the database is dispersed horizontally among n (n > 2) quantity of internet sites having a trusted party is known as. A model is suggested in this paper which adopts an indication based secure sum cryptography strategy to discover the world wide relationship policies without revealing individual’s personal data/information. The trusted party starts the process and prepares the merged list. All the websites computes the overall and partial supports supports for all the item sets in the combined list utilizing the sign based on these results and based safe sum cryptography approach finally trusted party finds international consistent item

sets. The efficiency of the proposed design is illustrated with an example. The effectiveness of the proposed model in terms of conversation and privacy is presented and it indicates that model effectively preserves the privacy of individual sites in the act of obtaining global frequent item sets and global association rules with minimum amount of communications.

REFERENCES

[1] M tamer Ozsu Patrick Valduriez, Principles of Distributed Database Systems ,3rd Edition.

[2] *R Agarwal, T Imielinski and A Swamy,* Mining Association Rules between Sets of Items in Large Databases, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, page 207-210, 1993.

[3] *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.

[4] Y. Lindell and B. Pinkas, Secure Multiparty Computation for Privacy-Preserving Data Mining, The Journal of Privacy and Confidentiality (2009) , 1, Number 1, pp. 59-98.

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

[6] M. Kantarcioglu and C. Clifto. Privacy-preserving distributed mining of association rules on horizontally partitioned data. In IEEETransactions on Knowledge and Data Engineering Journal, volume 16(9), pages 1026-1037.

[7] Chin-Chen Chang, Jieh-Shan Yeh, and Yu-Chiang Li, Privacy- Preserving Mining of Association Rules on DistributedDatabases, IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.11, November 2006.

[8] Mahmoud Hussein, Ashraf El-Sisi,and Nabil Ismail, Fast Cryptographic Privacy Preserving Association Rules Mining on Distributed Homogenous Data Base, I. Lovrek, R.J. Howlett, and L.C. Jain (Eds.): KES 2008, Part II, LNAI 5178, pp. 607616, 2008.© Springer-Verlag Berlin Heidelberg 2008.

[9] Lalanthika Vasudevan , S.E. Deepa Sukanya, N. Aarthi ,Privacy Preserving Data Mining Using Cryptographic Role Based Access Control Approach, Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol I, IMECS 2008.

[10] Vaidya, J. and Clifton, C. 2002. Privacy preserving association rule mining in vertically partitioned data, 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, pp. 639-644.

[11] A.C. Yao. Protocols for secure computations. In Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science, 1982

Author's Bibliography:

Dr. Ch G.V.N. Prasad currently working as a Professor and HOD of CSE dept, Sri Indu College of Engg & Technology. He gained 12 years of experience in IT industry (8 years in National Informatics Centre, Govt. of India, as Scientist and Software Analyst in AT&T in US) and 11 years of experience in Teaching (As a Professor & HOD of CSE Dept).

Katkam Gaurav, completed B.Tech from St.Mary's College of Engineering and Technology, Hyderabad, JNTUH with Distinction. Currently Pursuing M.Tech 2nd Year in Sri Indu College of Engineering and Technology. Areas of Interest are Data Mining, Web Technologies, Computer Networks and Data Base Management Systems.

Nunavath.Yakaiah , completed B.Tech from SR Engineering College, Warangal, JNTUH with Distinction. Currently Pursuing M.Tech 2nd Year. Areas of Interest are Data Mining, Information Security, Computer Networks and Data Base Management Systems.

R Bharath Kumar, completed B.Tech from Sri Venateswara college of Engineering, JNTUH. Pursuing M.Tech 2nd Year from Sri Indu College of Engineering and Technology. Areas of Interest are Data Mining, Object Oriented Programming Knowledge and Web Technologies.