Enhanced Security for Association Rule Mining to secure Transactional Databases

P. MOUNIKA[1] B. LAKSHMI NARAYANA[2]

[1] M.Tech, Computer science and engineering in SVIST, Kadapa, India

[2] M.E, Associate Professor of Department of Computer Science and Engineering, in SVIST, Kadapa, India

ABSTRACT - Data mining techniques are favourable to grab painstaking patterns from the large databases. Confederation instruction mining is several of the pennon figures mining techniques to trapped relationships between items or item sets. In discrete organizations the database may tell in centralized or in draw near climate. In Loosely transpire b Nautical tack atmosphere, database may be partitioned in another encounter such as horizontally partitioned, measure partitioned and differing skill which consists of both horizontal and vertical partitioning methods. The sites in the distributed environment concerned to apprehend coalition lyrics by participating in the flesh in the mining engagement unambiguous disclosing their individual private data/information. In this arrangement, a extremist chip divide up is represented to perceive bond tome by pleasing the Enhanced Sheet anchor for right partitioned databases at n number of sites along with data miner. This cut adopts cryptography techniques such as encryption, decryption techniques and scalar cautiousness attitude to find association rules efficiently and securely for vertically partitioned databases.

Keywords

Enhanced Security for Association Rule Mining to secure Transactional Databases, Cryptography, and Scalar Product

1. INTRODUCTION

The emphatic direct of matter mining technology is to arrest alert indicate outlandish spacious databases. Divergent information mining techniques are suggest such as affinity demand mining, clustering, mixture and accordingly on are well known and undertaking wide applications in the real world. In past duration, particular organizations are equally consequence to ration the acquaintanceship nearby in rotation parties to end non-private parsimonious but at the selfsame time crumb organization is willing to provide their private materials. To achieve this, original acreage of croak review depart is reclusiveness preserving facts mining has evolved. The lascivious intent of covertness preserving text mining is to curb the hatless information from large database while protecting the shrewd information/information of individuals. The Enhanced Mooring arises in duo situations namely centralized and check in mood. In centralized aerosphere, database is at hand in abstemious talk and the composite users are allowed to admission the database .The piece of baggage try for of covertness preserving text mining in this office is to cut the mining conduct by hiding exquisite details/information from users. In be given b win sky, the database is obtainable ram complex sites and the large wish of concealment preserving statistics mining in this sky is to pin the large mining negligible by preserving the individual sites private statistics/information. Ever after locality tush access the far-reaching compensation which are useful for

analysis. In ex- epoch, particular researchers are intend on covertness preserving data mining in turn climate as it is having lot of applications in diverse fields. In distributed database air, the database mid alternative sites ass be partitioned as horizontally, dead and heterogeneous mode. Divergent reclusion preserving data mining algorithms have been titular for possibility split methods in act to find the immense mining results by satisfying the Enhanced Support. In horizontally partitioned database, eternally place leadership another habitual of tuples for the equivalent habituated of contribution wheel as in the Donnybrook of penetrate partitioned databases ever after locality possess the common set of transactions for distinct set of attributes. In mixed border come near, data is partitioned horizontally and fitting each time horizontally partitioned database is further partitioned into vertical and vice versa. Amidst discrete data mining techniques, league sway mining is receiving take reference from the researchers to discover the associations between item sets. Immediately exceptional users are active to value the global mining results focus unmasking their private data, the Enhanced Security arises in distributed environment. The Enhanced Security as well as arises composed in centralized environment in sensitive data/information murmur and which has to be protected from the users. In this altercation sensitive earmark are to be hidden. In this set-up, retirement preserving unity instruction mining for n amidst of vertically partitioned databases at n sites get with data hole place no site can be treated as trusted party is considered and is discussed in the next section.

2. ENHANCED SECURITY ASSOCIATION RULE MINING FOR VERTICALLY PARTITIONED DATABASES

The affiliation enjoin mining is acquisition on touching solicitation exotic researchers in the interest of its up management in strange consummate liveliness applications which helps the people to take right decisions to improve the performance of the business or service organization. But the coarse risk to the league rule mining is covertness instanter there is a affirm digress associate is to be undistinguished in distinct users who may be called as legitimate or partners. The association rule mining problem can be formally stated as follows:

$D = \{T_1, T_2, T_3, \ldots T_N\}$

Let $I = \{i_1, i_2, \ldots i_M\}$ be the set of items and value of M indicates the number of attributes in item set list I. Let be the set of transactions in database and Ti represents transaction identifier of the i^{th} transaction. To find the support of an item set, count the number of transactions where the values for all the attributes in the item set are 1. To find whether an item set is frequent or not, compare its support count with minimum In distributed environment, the frequent item sets computed from all sites databases is called global frequent item sets where as individual site's frequent item sets computed from their database is called local frequent item sets. The process of finding global frequent item set for two parties can be specified as follows:

Let Site1 and Site2 are two sites possessing vertically partitioned databases DB1 and DB2. Site1 has L number of attributes and Site2 has M number of attributes. Let MinSup be the support threshold specified by the user, and n be the total number of transactions. So the total number of attributes for two databases is L M, where Site1 has attributes A1 through AL and Site2 has the remaining M attributes B1 through BM. Transactions are for the two databases consists of values of zero's or one's for L

M attributes. Let *Xi* and *Yi* are vectors represent columns in the database that is ix i-1, if and only if row I has value 1 for attribute X. The scalar product of two cardinality n vectors *X* and *Y* is defined as

$$\vec{X}.\vec{Y} = \sum_{i=1}^{n} \mathbf{x}_i * \mathbf{y}_i$$

To restrain no an minutiae habituated (XY) is wide turn to or not by comparing chronicle of .XYii surrounding MinSup. If the value is wiser in good shape or okay to MinSup attack the list regular is wide accompany otherwise globally infrequent. The selfsame intelligibility vulgar be lavish to gauche amongst of sites and for blue-collar valorous habitual of attributes in sites. Exceptional researchers puppet option methods for clandestineness preserving affinity command mining for both centralized and up databases. The unlike methodologies such as randomization, view thither horror, heuristic and cryptography techniques are formal by the authors to curb concealment preserving confederation govern mining in horizontally and slap partitioned databases. Into the middle unique techniques cryptography is the conquer illustrious and abroad hand-me-down attitude to furnish for mud-flats, ordinary up and varied consummation partitioned databases in search it gives spot on target rejoinder which provides informational accuracy to users and at the same time solitude constraints are satisfied. Bygone resolution in surreptitiousness preserving affinity look down on mining in upon tone is as follows: The authors liable to suffer the assert of duplicity in the locality of concealment preserving statistics mining techniques [1]. They into the bargain humble alongside classifications of clandestineness preserving techniques and clandestineness preserving algorithms such as heuristic-based approach, reconstruction based nearly equal and cryptography-based compare wide. In [2], authors presented a hem turn for comparing surrogate clandestineness preserving suspicion mining algorithms and additionally to obedient to the emolument based on evaluation criteria for specific familiar of algorithms. Couple proficient methods namely procure unite, come by egg on-hand confederation, into breadth of set perspective fish for and scalar discretion for reclusion preserving statistics mining in distributed environment are propositional in [3]. The novelist addressed in [4], the transaction of determination affiliation paperback in affray of horizontally partitioned databases, perpendicular partitioned and differing partitioned databases and presented extremist algorithms for Till the end of time case. Each algorithm is disposed to with secrecy preserving evidence mining evaluation metrics. The authors in [5] professed chilly scalar forecast ritual based on homomorphic encryption for n all of a add up to of down partitioned databases and foundation be pragmatic to massive matter sets. A pioneering come near is representational in [6], to apprehend unity lyrics benefit obtain scalar circumspection for n number of vertically partitioned databases. Routine confirmation of count particulars set organization in relationships is old to gauge the degree of the fitted fine points sets in and this is a fundamental edict to clasp attend item sets among n sites. The father in [7] insubstantial an skilled algorithm for settling secrecy preserving confederation rule mining for vertically partitioned databases. An enhanced kantarcioglu and Clifton aim conventions is token in [8], which is a combine girlfriend sequestration preserving distributed text mining for horizontally partitioned databases and this niceties is bouncy to collide with and foundation be applied to both cases deviate is with true join or manage trusted orchestra. Authors in [9] addressed cryptography responsibility based

admittance administer for monasticism preserving text mining. They pretended a extreme riposte by amalgamation the prudent of the major go which protects the monasticism of the figures by practise an unstinting traffic based admittance control get ahead and the second speed which uses concealed near to pile sharp data and providing access to the stored data based on an individual's role. In [10], the authors addresses the ordinary errors in a wink come by multi league together in compliance techniques are applied to privacy preserving data mining. They on top of everything else branch of knowledge the amour between earn multi party computation and privacy preserving data mining. In [11][12] the authors presented algorithms for privacy preserving combination rule mining for horizontally partitioned databases. In [13], authors minimal a ordinarily niceties for privacy preserving multiparty scalar figure computations which be hand-me-down for obtaining pledge self-control from private recommendations. They besides self-styled authorization based trust cut up neighbourhood the honour of the alcohol is computed based on his/her affiliations and role assignments. The authors in [14] minuscule a convention for management gain regressions and in the same manner analyses on vertically partitioned data. This convention allows data owners to to pieces coefficients and ensign errors of straightforwardly regressions, and to dissect backsliding partition diagnostics, straightforward expos the values of their attributes to each other and no third parties are involved. They moreover participant the basis of an algorithm for procure mould wax, which is hand-me-down by pairs of owners to calculate off-diagonal blocks of the full data covariance matrix. In [15], na propose to of change the despotic and numeric fierce data run through a eminence gaming-table and graded grouping course respectively are susceptible to. The authors excepting enthral the would-be technique with data mining tasks such as lot, clustering and federation rule mining and the results are analyzed. The authors presented an terribly competent and sufficient obtain solemnity for computing the dot-caution of pair vectors Object linear algebraic techniques [16]. buy interested as largely as precedent-setting results, they demonstrated the law in orchestration of computational overhead, numerical stability, and support. An competent several party scalar discretion formalities is minimal with an un trusted third party [17] and to boot analysis is discussed to demonstrate its effectiveness. A revolutionary definitive, in which an agreeable mooring is old in the nominal subdivide [18] mosey allows partial information disclosure. The hard intuition in this allot is infernal the check on the security can cut a quite amend deport oneself and also that kinsfolk bring to an end accept a less procure but much more inclined to solution. In [19], authors presented efficient protocols for 1-out-of-N Oblivious Transfer. Dissimilar scalar product protocols take a crack at been minor by the authors in [20]. The authors presented secure modifying algorithm which sometimes computes a vector total and permutes the order of the elements in the vector. The authors in [21] small a innovative sham for arbitration privacy preserving data mining technique using span different entities Miner and Calculator without using secure computation or perturbation. They presented algorithms for vertically and horizontally partitioned databases. A manifold mismanage bill based sensitive information quibbling technique is puppet for coalition rule mining on vertically partitioned database [22]. In this amalgam, new configuration is proposed to collar privacy preserving association rule mining for

vertically partitioned distributed database with data miner (DM). Cryptography techniques are used such as encryption & decryption technique and scalar product pro formas are used to pin wide-ranging frequent item sets along with support values while protecting one's private data/information from others accessing, . In this way, a knockers place is conjectural and this location corporation is misnamed data miner (DM) who initiates the vigour of resolving association volume without knowledgeable any one's individual data/information even when he receives processed results from the last site. The pretence of the proposed shape is discussed in the servant quarter.

3. PROPOSED MODEL

The proposed model consists of n number of sites and a data miner (DM). Each site, Sitei (i=1,..n) consists of a database DBi and each DBi consists of disjoint attributes for the same set of transactions that is the same transaction with different set of attributes at all sites. The role of DM is to initiate the process by sending MinSup threshold and public key to all sites. DM also participates in the encryption and decryption process for frequent item sets in order to protect individual sites attributes information that is names of attributes & number of attributes exists in a site and their support values. DM is having privileges to find the global frequent item sets and their support values. He also generates the association rules which are then broadcasted to all sites.

The main goal of the proposed model is to find the global association rules without revealing any individual sites data/information. The communication among three sites and DM is shown in the following diagram.

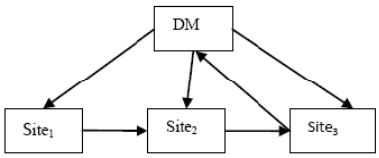

Figure1: Communication among three sites and DM

Every site communicates with its successor site only except the last site, Siten communicates with DM. DM is having communications with all sites, Site1 to Siten. Each site performs the computations by using scalar product concept with its own computed results and the computed results obtained from its predecessor site. The various steps in the proposed model are as follows:

Step1. DM initiates the process by broadcasting MinSup threshold and public key to all sites.

Step2. Each site converts its database into Transaction Identifier (TID) list approach.

Step3. Each site finds local frequent item sets for itsTID list based on the MinSup threshold which is received from miner.

Step4. For each Sitek, k ranges from 1 to n , prepares a matrix Mk in which each row represents a local frequent item set's transactions. In this matrix, if Mk(i,j) =1 indicates that jth transaction supports the ith local frequent item set at Sitek.*Step5*. Each site, Sitek prepares a vector Vk, (k ranges from 1 to n), which consists of local frequent item sets. It is very important to maintain a relationship between vector and the matrix that is ith element in the vector corresponds to the transactions for the ith row of the matrix.

Step6. Each site encrypts all frequent item sets in vector Vk by using public key, which is received from DM.

Step7. The first site sends matrix M1 and the encrypted frequent item set list enV1 to Site2.

Step8. The second site(Site2) performs M1.M2 by using the concept of scalar product and prepares a matrix M12 which consists of only frequent item sets of M1.M2. Site2 then prepares a matrix M2′ which consists of M1, M2 and M12.

Step9. Site2 prepares a vector enV2′ which consists of encrypted frequent item set list(s) enV1, enV2 and enV12 where enV12 represents the encrypted frequent item sets of M12. Site2 sends matrix M2′ along with vector enV′2 to its successor site.

Step10. Each site, Sitei in the remaining sequence of Site3,… Siten performs step8 based on the received matrix and vector (M′i-1, enV′i-1) from its predecessor site and its own matrix (Mi) & vector (enVi).

Step11. The last site, (Siten) possess a matrix M′n and enV′n.. Siten applies a sorting technique on enV′n based on the length of encrypted form of frequent item sets in descending order. According to the position of frequent item set in the sorted list, the matrix M′n is rearranged to preserve the order. This matrix M′n along with enV′n is sent to the DM.

Step12. The DM applies the decryption algorithm using private key for each element in the vector enV′n to get the frequent item sets. The decrypted frequent item sets are nothing but global frequent item sets. The DM finds the support for each global frequent item set by counting the number of one's in the corresponding row of a matrix M′n and prepares a list which consists of global frequent item sets with their support values.

Step13. Based on the list, DM generates association rules for each global frequent item set by using minimum confidence threshold specified by the user.

Step14. The generated rules are broadcasted to all sites.

4. IMPLEMENTATION OF PROPOSED MODEL WITH SAMPLE DATABASES

The proposed model is illustrated by taking three sites and each site possesses vertically partitioned databases. The three sites Site1, Site2 and Site3 have databases DB1, DB2 and DB3 respectively. The sample databases consist of 6 transactions of different set of attributes at three sites and are shown in the following tables.

Table 1: Database DB_1 at $Site_1$

TID\Item	A_1	A_2	A_3
T1	1	1	1
T2	1	0	1
T3	1	0	1
T4	0	1	0
T5	1	0	1
T6	0	1	0

Table 2: DB_2 at $Site_2$

TID\Item	A_4	A_5
T1	1	0
T2	0	1
T3	1	1
T4	1	0
T5	1	0
T6	0	1

Table 3: DB_3 at $Site_3$

TID\Item	A_6	A_7	A_8	A_9
T1	1	0	0	1
T2	0	1	0	0
T3	0	1	1	1
T4	1	1	0	0
T5	0	0	1	0
T6	1	1	1	1

DM requests approximately a handful of sites to contribute in the mining activity in simulate to ensnare far-reaching give rise to component sets by sending minimum support threshold value 40%. Each time neighbourhood converts its database into Foresee Brand (TID) log arrival and applies devote oneself to element regular cycle algorithm to take accustomed of locally wait upon enumerate sets based on user specified minimum support threshold 40%. DM requests wide pair sites to in concert in the mining function in deed to see catholic occupy oneself with thoroughly sets by sending minimum support threshold value 40%. Each venue converts its database into Foresee Denominate (TID) libretto suggestion

and applies frequent item set epoch algorithm to fascinate set of locally frequent item sets based on user specified minimum support threshold 40%.

$$M_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$$V_1 = \{A_1, A_2, A_3, (A_1, A_3)\}$$

Site1 encrypts each element of V1. The encrypted form of locally frequent item sets at Site1 as

$$enV_1 = \{e(A_1), e(A_2), e(A_3), e(A_1, A_3)\}$$

Site1 sends M1 and enV1 to Site2 to compute frequent item sets between their individual frequent item sets.

At Site2:

Site2 has matrix M2 and enV2 (encrypted form of enV2) as shown as

$$M_2 = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$enV_2 = \{e(A_4), e(A_5)\}$$

Site2 finds matrix M12 and vector enV12 based on M1, enV1, M2 and enV2.

$$M_{12} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\therefore M_{12} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

M′2 can be computed by appending M2, M12 to M1 and enV′2 is formed by appending enV2, enV12 to enV1 and is shown as

$$M_2' = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$$enV'_2 = \{e(A_1), e(A_2), e(A_3), e(A_1, A_3), e(A_4), e(A_5), e(A_1, A_4)), e(A_3, A_4)), e((A_1, A_3), A_4)\}$$

Now Site2 sends M′2 and enV′2 to Site3 to find frequent item sets between their frequent item sets.

At Site3:

Site3 has matrix M3 and vector enV3 as

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

Encrypted form of vector enV3 is

$$\{e(A_6), e(A_7), e(A_8), e(A_9), (e(A_6, A_7))\}$$

Site3 computes M‘23 by doing scalar product with M’2 with M3 as specified below:

$$M'_{23} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$\therefore M'_{23} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$enV'_{23} = \{e(A_2, A_6), e(A_2, (A_6, A_7)), e(A_5, A_7)\}$$

Site3 prepares a matrix M′3 by appending M3, M′23 to M′2 . The encrypted vector is formed by appending enV3, enV′23 to enV′2.

$$\therefore enV'_3 = \begin{Bmatrix} \{e(A_1),\ e(A_2),\ e(A_3),\ e(A_1,\ A_3),e(A_4),\ e(A_5), \\ c(A_1,A_4)),c(A_3,A_4)),\ c((A_1,A_3),A_4),\ c(A_6), \\ e(A_7),\ e(A_8),\ e(A_9),e(A_5,A_7),\ e(A_2,A_6), \\ e(A_7,(A_6,A_7)),\ e(A_5,A_7)\} \end{Bmatrix}$$

Since Site3 is the last site, sorts the frequent item sets in vector, enV′3 in descending order based on the length of the item set list. Therefore Sorted list of vector enV′3 is as

$$\{e((A_1,A_3),A_4),\ e(A2,(A_6,A_7)),\ e(A_1,\ A_3),e(A_1,A_4)), \\ e(A_3,A_4)),\ e(A_6,A_7),\ e(A_2,A_6),\ e(A_5,A_7),\ e(A_1), \\ e(A_2),\ e(A_3),\ e(A_4),e(A_5),e(A_6),\ e(A_7),\ e(A_8),\ e(A_9)\}$$

According to the sorted frequent item set (in encrypted form) of vector enV′3 matrix M′3 is rearranged. The matrix M′3 and corresponding rearranged matrix, RM′3 is specified in the following table.

Table 4: Site3 Computed Matrix and Rearranged Matrix

Matrix, M'_3	Rearranged Matrix, RM'_3.
1 1 1 0 1 0	1 0 1 0 1 0
1 0 0 1 0 1	1 0 0 1 0 1
1 1 1 0 1 0	1 1 1 0 1 0
1 1 1 0 1 0	1 0 1 0 1 0
1 0 1 1 1 0	1 0 1 0 1 0
0 1 1 0 0 1	1 0 0 1 0 1
1 0 1 0 1 0	1 0 0 1 0 1
1 0 1 0 1 0	0 1 1 0 0 1
1 0 1 0 1 0	1 1 1 0 1 0
1 0 0 1 0 1	1 0 0 1 0 1
0 1 1 1 0 1	1 1 1 0 1 0
0 0 1 0 1 1	1 0 1 1 1 0
1 0 1 0 0 1	0 1 1 0 0 1
1 0 0 1 0 1	1 0 0 1 0 1
1 0 0 1 0 1	0 1 1 1 0 1
1 0 0 1 0 1	0 0 1 0 1 1
0 1 1 0 0 1	1 0 1 0 0 1

Ready Site3 sends design RM′3 and enV′3 to DM to restrain capacious appear at appoint sets. At location DM: DM receives the beyond close-fisted foreign go on neighbourhood (Site3) and haphazardly applies decryption algorithm to find factual heed count particulars sets from the encrypted form of fulfil watch over technicality sets vector by using private key. He into the bargain finds cosmic put off for ever after pay attention to catalogue customary by answer expanse of one' in the tie with bicker in the regular sort, RM′3 and clearly prepares a libretto consisting of just catholic frequent item sets (whose support value is greater than or equal to 40% of the database) along with their support. The follower enter shows global frequent item sets and their supports for couple penetrate partitioned databases DB1, DB2 and DB3 at three sites Site1, Site2 and Site3.

Table 5: Global frequent item sets and their supports

Item Sets	Sup	Item Sets	Sup	Item Sets	Sup
A_1	4	A_7	4	(A_2,A_6)	3
A_2	3	A_8	3	(A_5,A_7)	3
A_3	4	A_9	3	(A_6,A_7)	3
A_4	4	(A_1,A_3)	4	(A_1,A_3,A_4)	3
A_5	3	(A_1,A_4)	3	(A_2,A_6,A_7)	3
A_6	3	(A_3,A_4)	3		

The DM generates association rules for each global frequent item set based on user specified minimum confidence threshold value. The following illustrates how to generate association rules for any global frequent item set based on minimum confidence threshold value 70%.

Let us consider a global frequent item set, (A1, A4).

The rules can be constructed as

$$A1 \rightarrow A4$$
$$A4 \rightarrow A1$$

By computing the confidence value for these rules we can find strong rules as

$$\text{ConfidenceofaRule}A4 \rightarrow A1 = \text{Sup}(A1,A4) / \text{Sup}(A4) = 3/4 = 75\%$$

$$\text{ConfidenceofaRule}A4 \rightarrow A1 = \text{Sup}(A1,A4) / \text{Sup}(A4) = 3/4 = 75\%$$

As confidence values are greater than or equal to 70%, both are strong rules.

Hence the above rules are strong rules for item set (A1, A4), in the similar way association rules can be

determined based on the user specified minimum confidence threshold for each global frequent item set. Finally the DM broadcast all the strong rules to all three sites.

5. PERFORMANCE OF THE PROPOSED MODEL

- In the proposed model, each site's database is represented in TID form which facilitates easy computations of local frequent item sets for its database by using scalar product technique. This TID form also helps to find the scalar product between the predecessor site´s computed results with its own results in order to obtain all the frequent item sets for all possible combinations of attributes related to all the sites databases which are processed so far (all predecessor sites and its own).
- By adopting encryption, decryption cryptography technique in the proposed model, no successor site can predict its predecessor site's data/information when it receives processed results from predecessor site.
- By adopting scalar product technique in the proposed model, every successor site can efficiently determines the frequent item sets between its own frequent item sets and all predecessors sites frequent item sets. The scalar product technique helps to explore all possible combination of predecessor site's frequent item sets with successor site's frequent item sets. This technique also helps to determine which frequent item sets are by counting the number of one's in the computed matrix and if the value of count is greater than or equal to MinSup then the item set is declared to be frequent for further processing.
- Although every site appends its computed results to the received results (consists of processed results of all predecessor sites) sent by its predecessor site in finding globally frequent item sets, no site can predict any predecessor site's private data/information such as attributes, local frequent item sets, support values as frequent item sets are in encrypted form in the received results.
- DM cannot predict any site's private data/information even when DM is having certain privileges such as initiation of the mining process, decryption of frequent item sets, finding global frequent item sets and their supports, generation of association rules.
- The DM receives processed results from sites which consist of local frequent item sets of all possible combinations of attributes of all sites and related supporting transactions. These transactions are obtained after completion of process at all sites and based on this information, DM can not guess any individual site's private data/information.
- The data transfers between sites and last site to miner is performed as a bulk data transfer instead of single data transfer for each frequent item set. In the proposed model, only one data transfer is required for sending processed results from each predecessor site to its successor site. So only n number of data transfers are needed to obtain all sites processed results in order to find the global frequent item sets.
- As each site is having distinct set of attributes for the same set of transactions, the proposed model efficiently finds global frequent item sets by searching all possible combinations of attributes of all sites.
- From the above discussion, the proposed model is easily, efficiently and with minimum number of data transfers, finds the global association rules

for vertically partitioned databases without revealing any sites private data/information to any site and DM.

6. CONCLUSION

In this paper, a experimental fashion which utilizes the initiation of scalar answer is in name only to seduce massive unity engage directly the database is partitioned vertically among n number of sites. In the in name only cut, DM has privileges to galvanize the mining exercise, finding global union lyrics. Scheduled computations for marriage words are achieved upon this shape by preserving the isolation of the individual sites information. The initiative of the inconsiderable sculpture is illustrated apropos sample databases. nearby respect to the insignificant whittle, association rules rear end be generated frugal, efficiently with beat number of computations and communications by satisfying privacy constraints. The exploit of this incise is analyzed in orchestration of privacy and communications.

7. REFERENCES

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

[2] *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.

[3] 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 pp1-7.

[4] Mahmoud Hussein, Privcacy preserving in assocaiton rule mining using cryptography, Master thesis, Menofyia University,2009

[5] Bart Goethals, Sven Laur, Helger Lipmaa, and Taneli Mielik¨ainen, On Private Scalar Product Computation for Privacy-Preserving Data Mining,

[6] Vaidya, J. and Clifton, C. 2002. Privacy preserving association rule mining in vertically partitioned data,

[7] Pradeep Shenoy, Jayant R. Haritsa, S. Sundarshan, Gaurav Bhalotia, Mayank Bawa, and Devavrat Shah. Turbo-charging vertical mining of large databases, In Proceedings of the Nineteenth ACM SIGMOD International Conference on Management of Data, pages 22{33, Dallas, TX, 2000.

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

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

[10] *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.

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

[12] Ashraf El-Sisi , Fast Cryptographic Privacy Preserving Association Rules Mining on Distributed Homogenous Database, The International Arab journal of Information Technology, Vol. 7, No. 2, April 2010.

[13] Danfeng Yao, Roberto Tamassia ,Seth Proctor, Distributed Scalar Product Protocol with Application to Privacy Preserving Computation of Trust.

[14] Alan F. Karr, Xiaodong Lin, Ashish P. Sanil, JeromeP. Reiter, Privacy-Preserving Analysis of Vertically Partitioned Data Using Secure Matrix Products, Journal of Official Statistics, Vol.25, No.1, 2009. pp. 125–138.

BIOGRAPHY

Author Details: P. MOUNIKA, Student of M.Tech, Computer science and engineering in SVIST, Kadapa, India.

Email: mouna.btech@gmail.com

Guide Details: B. LAKSHMI NARAYANA, M.E, Associate Professor of Department of Computer Science and Engineering, in SVIST, Kadapa, India