AN IMPROVED ALGORITHM FOR HIDING SENSITIVE ASSOCIATION RULES USING EXACT APPROACH

N V Muthu Lakshmi[1]
[1] Research Scholar: Dept of Computer Science
S.P.M.V.V, Tirupati
Andhra Pradesh, INDIA

Dr. K Sandhya Rani [2]
[2] Professor: Dept of Computer Science
S.P.M.V.V, Tirupati,
Andhra Pradesh, INDIA

ABSTRACT

Many algorithms are proposed by researchers to find association rules which help the users to make strategic decisions to improve the performance of the business or any qualitative organizations. The threat occurs to association rule mining when data or information is required to share to many users to get mutual benefits. Among the existing techniques the exact hiding approaches provides best solutions with minimum side effects. A modified inline algorithm is proposed in this paper to hide sensitive rules by formulating constraint satisfaction problem without any side effects with the concepts of positive and negative border sets. By adopting divide and conquer technique on constraints, the efficiency of the proposed algorithm is increased.

KEYWORDS
Association Rule Hiding, Exact Hiding Approach, Inline Algorithm, Constraints Satisfaction Problem ,Positive and Negative Border sets

I. INTRODUCTION

Privacy preserving data mining is a novel research direction in data mining and statistical databases, which has recently been proposed in response to the concerns of preserving personal or sensible information derived from data mining algorithms. There are two types of privacy concern to data mining. The first type of privacy, called output privacy, is that the data is altered so that the mining result would preserve certain privacy. The second type of privacy, called input privacy, is that the data is manipulated so that the mining result is not affected or minimally affected. The various techniques of data mining such as association rule, clustering, classification, sequence mining and so on also faces privacy threat when the knowledge is required to share to one or more people.
Among many data mining techniques, the discovery of association rules from large databases has proven beneficial for companies since such rules can be very effective in revealing actionable knowledge that leads to strategic decisions. However, there has been growing concern that use of this technology is violating individual privacy. This concern makes, protecting individual's privacy is an important goal in association rule mining technology. In a context, in which there exists the need to share information/knowledge among different parties/users to gain mutual benefits but these persons need not be trusted. So, this rule mining technique also requires great research work in protecting private data before allowing the data to their competitors for gaining knowledge.

This privacy issue allows the users to have access over only non sensitive information and they should not gain knowledge over sensitive information. Unfortunately, the protection of sensitive data may also hides some non sensitive data and also produces some spurious information, because correlation existing between items. The problem of finding association rules from large voluminous of data while preserving sensitive information is called privacy preserving association rule mining. Privacy preserving association rule mining is applied in centralized as well as in distributed systems. In centralized system, data provider is nothing but database owner who is willing to share data/knowledge to their legitimate people for getting mutual benefits. These legitimate people may use this knowledge for making strategic decisions to improve the profits/sales/quality in business or to improve the services in organizations or to strengthen the security and so on. In distributed system, data providers are more than one, who are situated at different sites and association rule mining will be done on collective databases.

II. ASSOCIATION RULE HIDING

Sensitive rule hiding is a subfield of privacy preserving data mining, which can be divided into two categories. One is the preserving of data privacy, which considers all or parts of the data to be sensitive. Its goal is to blur the sensitive data but keep the summary information unchanged. The other is the preserving of information privacy, assuming that only the summary information is sensitive. Its goal is to hide the sensitive information but retain most of the original data. Sensitive rule hiding belongs to the second category. In this paper, information privacy in association rule mining is discussed.

The various approaches proposed by researchers hide sensitive information efficiently and accurately but also face the problem of side effects. The side effects occur due to correlation existsting between items in the database. The side effects may decrease the informational accuracy to the users because of the property of correlation association rules may possess spurious or wrong information, also hides non sensitive rules unnecessarily, and accidentally discloses some sensitive rules. So the challenging task is how to protect sensitive rules from users without effecting informational accuracy to the users that is avoiding side effects as far as possible.

The algorithms which have been developed for hiding association rules can be classified into three. These are heuristic approaches, border based approaches and exact approaches.

Heuristic approaches have been getting focus of attention for majority of the researchers due to their efficiency, scalability and quick responses. However in some circumstances these heuristic based approaches suffer from undesirable side effects. So these side effects may degrade the performance of the hiding process of sensitive association rules. This heuristic approaches can be categorized into two which are distortion based schemes and blocking based schemes. In order to hide sensitive item sets, distortion based scheme employs, by altering certain items in selected transactions from 1's to 0's and vice versa where as blocking based scheme, employs by, replacing certain items in selected transactions with unknowns. Based on the criteria of considering support or the confidence of the rule to hide sensitive association rules, distortion and blocking based scheme approaches are further partitioned into support-based and confidence-based methodologies.

The second class of approaches called border revision approach which modifies borders in the lattice of the frequent and infrequent item sets to hide sensitive association rules. This approach tracks the border of the non sensitive frequent item sets and greedily applies data modification that may have minimal impact on the quality to accommodate the hiding sensitive rules. Based on border revision concept, researchers proposed many algorithms such as BBA (Border Based Approach), Max–Min1 and Max-Min2 to hide sensitive association rules. The algorithms utilizes different techniques to hide sensitive item sets by deleting specific items belonging to the sensitive item sets from a set of supporting transactions of these sensitive item sets. Also algorithms attempt to minimize the number of non sensitive item sets that may be lost while sanitization is performed over the original database in order to protect sensitive rules.

Third class of approach is non heuristic algorithm called exact, which conceive hiding process as constraint satisfaction problem. These problems are solved by integer programming. Compared to heuristic and border based, this guarantees quality for hiding sensitive information. This can be achieved by applying good sanitization method which minimally distorts the original database, so no side effects to the process of hiding. This approach can be considered as descendant of border based methodology. It works in the following way as follows. First border revision method is applied to small portion of item sets from the original database whose status is determined (that is frequent versus infrequent) and recorded. Then, exact methodologies incorporate unknowns to the original database and generate inequalities that control the status of selected item sets of the border. These inequalities along with an optimization criterion require minimal modification of the original database to facilitate sensitive knowledge hiding, formulate an optimization problem whose solution is guaranteed to lead to optimal hiding. Menon's algorithm is the first one which adopted integer programming technique to find optimum solution. This solution determines the sanitized items and sanitized transactions for hiding sensitive frequent item sets and then applies heuristic approach to get more efficient optimized solution. Inline algorithm is the other one which adopts integer programming problem solving technique for finding optimum solution for hiding sensitive item sets without hiding any single non sensitive frequent item sets. The drawback with this approach is the algorithms designed for this approach may take several orders of magnitude slower than heuristic ones, especially due to the time that is taken by the implementation of integer programming logic to solve the optimization problem. An algorithm which is a modified inline algorithm to hide sensitive frequent item sets to minimize side effects is proposed in this paper. The related work for association rule hiding in centralized database is discussed as follows:

The authors in [1], were the first to propose an algorithm for hiding sensitive association rules by reducing support of an sensitive item set. In [2], authors presented algorithms for both sensitive frequent item sets and sensitive rules. They also proposed three single rule heuristic hiding approaches based on either support or confidence of antecedent or consequent of the sensitive rule. Approaches were introduced for multiple rule hiding and an interesting novelty feature in these approaches is considering an account of the impact of sanitization on hiding the sensitive patterns, but also the impact related to the hiding of non sensitive knowledge [3]. The authors in [4], proposed two distortion based heuristic approaches Priority based Distortion and Weight based Sorting Distortion algorithm to hide sensitive rules. In paper [5], the authors addressed the problem of preserving privacy in association rule mining using heuristic approaches such as support and confidence based approaches. They also introduced new security metrics which should be considered for general framework of association rule mining.

The border revision process was first introduced by the authors in [6], for hiding sensitive association rules and heuristic approach also used in this new approach which uses the notion of border of the non sensitive frequent item sets to track the impact of altering the transactions. This proposed scheme, first computes positive border and negative border in the lattice of all item sets and focus on preserving the quality of the computed borders during the hiding process.

The authors in [7], proposed two different heuristic approaches called max-min1 and max-min2 which uses positive and negative border concept. The proposed algorithms try to remove all the sensitive item sets that belong to the negative border from the database while maintaining all the positive border item sets to be frequent. A novel methodology was first proposed by the authors in[8], consisting of both exact part and heuristic part to facilitate the hiding of the sensitive knowledge. This methodology adopted Integer Programming Problem Solving technique for exact part to find minimum number of transactions which are to be sanitized to hide sensitive frequent item sets and heuristic algorithm helps to sanitize the database based on the solution obtained from first part. The first algorithm called Inline which was proposed in [9], and which do not rely on any heuristics to protect the sensitive item sets while finding privacy preserving association rule mining. This inline algorithm formulates hiding problem as constraints satisfaction problem in which distance is the optimization criterion and this method is guaranteed to identify hiding solutions of superior quality than heuristic and border based approaches. In [10], authors discussed exact approaches using constraints satisfaction problem solving technique without using any heuristic methods to hide all sensitive item sets and to preserve all non sensitive item sets.

III. PROPOSED METHODOLOGY

One of the important algorithm which belongs to the category of exact hiding approaches is inline algorithm and does not depend on any heuristics to secure the sensitive knowledge obtained by association rule mining. In inline algorithm, hiding process is formulated as an optimization problem in order to optimize (minimize) the distance between the original database and the sanitize database. When compared to heuristics and border based approaches, the inline algorithm provides high quality hiding solutions. However the inline algorithm mainly concentrates on hiding sensitive rules and maintains the frequent items as it is in the process of obtaining sanitized database. But it does not concentrate on preventing the generation of new frequent item sets and consequently there is a possibility of generating new rules. To overcome this problem, a modification of inline algorithm. is proposed in this paper which considers both positive and negative border sets and also to speed of the process a divide and conquer procedure is applied to the constraints. The terminology used in the proposed algorithm is specified in the table 1.

Table 1. Terminology in the proposed algorithm

SN	Abbreviation	Explanation
1	DB = $\{t_1,t_2,..t_N\}$	A original database consisting of N number of transactions
2	I =$\{i_1,i_2,\ldots i_M\}$	An item set of length M
3	L_k	An item set of length k
4	T_{nm}	The n^{th} transaction of m^{th} item
5	S =$\{ s_1, s_2, \ldots s_r\}$	Set of sensitive item sets
6	Sup(J)	Number of transactions supporting item set J
7	Min-Support	Minimum support
8	Min-Confidence	Minimum confidence
9	N	Size of original database, DB
10	F_{DB} =$\{ L_1, L_2, L_3, \ldots L_k\}$	A set consists of all frequentitem sets

SN	Terminology	Explanation
11	A → B	Association rule between item sets A and B
12	AR	Set of association rules
13	S_{min}	Minimal number of sensitive item sets
14	S_{max}	Maximal number of sensitive item sets
15	F'_{DB}	A set of frequent item sets in expected distorted database
16	DB′	Distorted database
17	P_{BR} for F'_{DB}	Set consisting of maximally frequent item sets in F'_{DB}, also called revised positive border
18	N_{BR} for F'_{DB}	Set consisting of minimally infrequent item sets in F'_{DB}, also called revised positive border
19	(BIP) Binary integer programming	Binary integer programming is the problem of finding a binary vector x that minimizes a linear function $f^T x$ subject to linear constraints: $\min_x f^T x$ such that $A \cdot x \le b$, $Aeq \cdot x = beq$, x binary.
20	Border set	It is the union of P_{BR} and N_{BR}, identifies key item sets which separates all frequent patterns from their infrequent patterns

This algorithm utilizes various procedures such as generation of frequent itemsets using Apriori, generation of Positive Border set and negative border set, divide and conquer approach to devide the constraints based on the independent attributes and BIP program to solve the each sub optimization problem

Algorithm

Input: Database DB, Min-Support, Min-Confidence, S set of sensitive item sets
Output : Distorted database DB′

Step 1. Using Apriori algorithm find frequent item sets F_{DB} based on Min-Support for the database DB

Step 2. For a given S, Find Smin and Smax by using the following formulas.

$$S_{min} = \{ I \in S / \text{ for al } J \subset I, J \notin S\}$$

$$S_{max} = \{ I \in F_{DB} / \exists J \in S_{min}, J \subset I\}$$

Step 3. Find $F'_{DB} = F_{DB} - S_{max}$

Step 4. Find positive border set by using Positive Border procedure

Step 5. Find negative border set by using negative border procedure

Step6. Find minimal number of constraints for positive border elements

Step7. Find minimal number of constraints for negative border elements

Step8. Combine the constraints computed in step 6 & 7.

Step9. Introduce unknown variables U_{ij} in the sensitive item sets supporting transactions in the distorted database. Construct constraints for each positive border element where sum of support value must be greater than or equal to minimum support and also constructs constraints based on elements of negative border whose support value must be less than Min-Support.

Step 10. Apply divide and conquer procedure to divide the constraints based on their dependencies to form sub optimization problems

Step 11. Using Binary Integer Programming solve the each sub CSP to get unknown values.

Step 12. Update the database based on solution obtained from CSP to get distorted database.

Step 13. Generate association rules based on user specified minimum confidence by using the distorted database. The generated rules are related to all the non sensitive frequent item sets and the sensitive rules related sensitive item sets are hidden.

Step 14. Stop the process

Procedure PositiveBorder // Finds revised positive border item sets

Input : F'_{DB}, S
Output: P_{BR} for F'_{DB} // Positive Border item sets

Step1. call sort procedure for F'_{DB} to get F'_{DBsort}
Step2. Create array count [size(F'_{DB}) ,2]
Each item set of F'_{DBsort} is assigned to count and initialized to zero

Step3.
For each k item set $f \in F'_{DBsort}$ do
For each (k-1) item set, $q \in F'_{DBsort}$ do

```
      If  q is a improper subset of f  then
           count[q,1] ++
   End for loop
   For each f  belongs to F´DBsort do
      If count[f,1] = 0
         add f to PBR
  End for loop
```

Step 4. Return P_{BR} for F'_{DB}

Pseudo code for NegativeBorder // Finds revised negative border item sets

Input: F'_{DB}
Output : N_{BR} for (F'_{DB})

Step 1. Initialize k =1

Step 2. while F_{DB} is not empty

Step 3.

```
 If (k = 1 )
     For each item set x belongs to L1 do
         If x ∉ F´D
            add x to NBR
    end for loop
 else
   if ( k =2 ) go to step 4
      else  go step 5
```

```
Step 4. For each item set x ∈ L1 do
            For each item set y ∈ L1 do
                Z=  join( x, y)
                  If ((x < y) and z ∉ F´D
                      add  z to  NBR
              end inner loop
        end outer loop
```

```
Step 5. For each x ∈ Lk-1
            For each y ∈ Lk-1
             If ((x1 < y1) and (x2 < y2) and
                 (xk-1 < yk-1)and z ∉ F´DB
                         z=  x joins y
                         If z ∉  F'DB
  Generate  a set S-set, consists of  proper subsets
   of z
   While   (( S-set ≠Ø) and (flag = False))
     If ( proper subset ∉ F´D)
      flag = True
  End while loop
  If (flag= False) then
    add z to  NBR
  end inner for loop
        end outer for loop
```

Step6. Return N_{BR} set F'_{DB}

Procedure Reverse-Sort(F'_{DB})

// The algorithm first sorts these item sets in decreasing length and then for all item sets of the same length

Input : F'_{DB}
Output: F'_{DBsort} // The set consisting of sorted elements

Step1 : read each item set in F'_{DB}

Step 2 Apply insertion sort procedure to get decreased sorted list based on item set size

Step 3 Apply any sorting procedure to sort based on support and store this in F'_{DBsort}

Step 4 Return F'_{DBsort}

Procedure (GenAssRuls) // Generates association rules based on frequent item set

Input : F_{DB}, Min-Confidence
Output: Association rules (AR)

```
Step1.    Initialize j with 2;  create a file, Rule-set
consisting of three fields rule-id, antdnt,
            consqnt ;
Step 2.   For each item set Lj in FDB
  Generate ropoersubsets and saved in set SSUB for
    Lj,   // SSUB = {S1,S2,…,Sm }
     Initialize i with one
      While (  i <= m ) do
         read  Si from SSUB
     antdnt = Si   and  consqnt = Lj - Si
    read  sup of antdnt and consqnt
   confd=  sup (antdnt U consqnt)
                  /sup(antdnt
 if  (con >= Min-Confidence)
    add (rule-id, antdnt, consqnt,confd) to
       AR
    increment i by one
    End while loop
 End for loop
```

Step 3. Return a set consisting of all association rules AR

IV. A IMPLEMENTATION

By taking the following sample database consisting of 5 attributes, named A_1, A_2, A_3, A_4, A_5 and 8 transactions. Each transactions has id where T_i denotes i^{th} transaction. If a transaction supports a particular item then the corresponding value will be one otherwise it is zero. The sample database is given in table 2.

Table 2 Sample database

Items/Transactions	A_1	A_2	A_3	A_4	A_5
T_1	1	0	1	1	0
T_2	0	1	0	1	1
T_3	0	0	0	0	1
T_4	1	1	0	0	0
T_5	0	0	1	1	0

T_6	0	0	1	1	1
T_7	1	1	0	1	1
T_8	0	0	1	0	0
SUP	3	3	4	5	4

Let the sensitive items in S = {A3,$<$A2,A4$>$}

The first step of this algorithm is generating frequent item sets based on Min-Support,25% and then saving these results in F_{DB} which is specified below.

F_{DB}={ $A_1, A_2, A_3, A_4, A_5, <A_1,A_2>, <A_1,A_4>,<A_2,A_4>,<A_2,A_5>,<A_3,A_4>,<A_4,A_5>,<A_2,A_4,A_5>$ }

The frequent item sets along with their support specified in the following table.

Table 3 Frequent Item Sets with its Support

Item set	Sup	Item set	Sup
A_1	3	$<A_1,A_4>$	2
A_2	3	$<A_2,A_4>$	2
A_3	4	$<A_2,A_5>$	2
A_4	5	$<A_3,A_4>$	3
A_5	4	$<A_4,A_5>$	3
$<A_1,A_2>$	2	$A_2,A_4,A_5>$	2

Based on sensitive item sets, minimal sensitive item set (S_{min}) and maximal sensitive item sets (S_{max}) are determined towards finding solution and are given here.

S_{min} = { $<A_3>$, $<A_2,A_4>$ }

F_{DB} = { $A_1, A_2, A_4, A_5, <A_1,A_2>, <A_1,A_4>, <A_2,A_5>, <A_4,A_5>$}

Using these set of values, maximally frequent item sets that is Positive borderset (P_{BR}) and minimally infrequent item set that is negative border set (N_{BR}) are determined by calling corresponding procedures and are shown here.

P_{BR} = { $<A_4,A_5>, <A_1,A_2>, <A_1,A_4>, <A_2,A_5>$}

N_{BR} = { $A_3, <A_1,A_5>, <A_2,A_4>$}

border set is formed by combining positive border (P_{BR}) and negative border (N_{BR}) sets and is given here.

Border set = { $<A_4,A_5>, <A_1,A_2>, <A_1,A_4>, <A_2,A_5>, A_3, <A_1,A_5>, <A_2,A_4>$}

The original database is then copied and stored in a table called distorted database. For each sensitive item set, unknown variables are introduced for sensitive item sets of all transactions and are specified in table 4. The problem is to find values for unknown variables for the distorted database such that no sensitive rule is revealed, no non sensitive rule is hidden and no wrong rule is generated when competitors mine.

Table 4 Distorted Database with Unknown Variables

Items/Transactions	A_1	A_2	A_3	A_4	A_5
T_1	1	0	1	1	0
T_2	0	U_{22}	0	U_{24}	1
T_3	0	0	0	0	1
T_4	1	1	0	0	0
T_5	0	0	U_{53}	1	0
T_6	0	0	U_{63}	1	1
T_7	1	U_{72}	0	U_{74}	1
T_8	0	0	U_{83}	0	0
SUPPORT	3	3	4	5	4

Using this database, constraints are formed for each element of positive and negative border sets.

Constraints formed from positive border item sets are as follows

$<A_4, A_5>$, $U_{24}+1+U_{74} >= 2$

$<A_1, A_2>$, $1 + U_{72} >= 2$

$<A_1, A_4>$ $1+U_{74} >= 2$

$<A_2, A_5>$ $U_{22} +U_{72} >= 2$

With respect to negative border set, constraints are formed as follows.

$<A_3>$ $1+U_{53} + U_{63} + U_{83} < 2$

$< A_1, A_5 >$ $1<2$

$< A_2, A_4 >$ $U_{22} U_{24} + U_{72} U_{74} < 2$

The constraints can be simplified by evaluating constants and are shown here.

$U_{24} +1+U_{74} >= 1$

$U_{72} >= 1$ it implies U_{72} =1 no need to be a constraint.

$U_{74} >= 1$ implies U_{72} =1 no need to be a constraint.

$U_{22} +U_{72} >= 2$

$U_{53} + U_{63} + U_{83} < 1$

$U_{22} U_{24} + U_{72} U_{74} < 2$

Only the constraint related to the A3 is independent to other constraints. So the above CSP can be further subdivided in to sub CSP problems and each problem is solved simultaneously so that overall time and complexity can be reduced.

Sub problem1

Min $Z_1 = U_{22} + U_{24} + U_{72} +U_{74}$

$U_{24} +1+U_{74} >= 1$

$U_{22} +U_{72} >= 2$

$U_{22} U_{24} + U_{72} U_{74} < 2$

$U_{22},U_{24},U_{72},U_{74}$ takes either 0 or 1.

Subproblem 2

Min $Z_2 = U_{53} + U_{63} +U_{83}$

$U_{53} + U_{63} + U_{83} < 1$

U_{53}, U_{63}, U_{83} takes either 0 or 1

The above are simplified constraints however the last constraint in sub problem 1 consists of product of two variables and can not be solved by BIP because it accepts linear variables only. Using constraint degree reduction approach this constraint can be converted into a constraint of single variables but introduces 7 additional constraints and illustrated here.

Let Ψ1, Ψ_2,be the two new temporary variable which replaces product of variables as

$\Psi1= U_{22} U_{24}$ and $\Psi_2 = U_{72} U_{74}$

By using this process this constraint that involves produce of binary variables become linear. The resulting CSP is presented as follows:

Min $U_{22} + U_{24} + U_{72} + U_{74}$

Subject to the constraints

$U_{24} + 1 + U_{74} >= 2$

$U_{22} + U_{72} >= 2$

$\Psi1 <= U22 \quad \Psi1 <= U24$

$\Psi1 >= U22 + U24 + 2 - 1$

$\Psi2 <= U72 \quad \Psi2 <= U74$

$\Psi2 >= U72 + U74 + 2 - 1$

$\Psi1 + \Psi2 < 2$

Where U_{22}, U_{24}, $\Psi1$, $\Psi2$ takes values either 0 or 1.

The solution of the CSP leads to optimal hiding solution for first sub problem is presented as

$U_{22}=1$, $U_{24}=0$, $U_{72}=1$, $U_{74}=1$

The solution of the CSP leads to optimal hiding solution for second sub problem is presented as

$U_{53}=0$, $U_{63}=0$, $U_{83}=0$,

Replace every unknown variable with corresponding value which is obtained after solving the above CSP to get distorted database and is specified in the below table.

Table 5. Distorted Database

Items/Transactions	A_1	A_2	A_3	A_4	A_5
T_1	1	0	1	1	0
T_2	0	1	0	0	1
T_3	0	0	0	0	1
T_4	1	1	0	0	0
T_5	0	0	0	1	0
T_6	0	0	0	1	1
T_7	1	1	0	1	1
T_8	0	0	0	0	0
SUPPORT	3	3	1	4	4

The following rules are generated from the distorted database

Table 6 Rules Obtained From Distorted Database

$<A_1,A_2>$	$A_1 \longrightarrow A_2 = 66\%$ $A_2 \longrightarrow A_1 = 66\%$
$<A_1,A_4>$	$A_1 \longrightarrow A_4 = 66\%$ $A_4 \longrightarrow A_1 = 50\%$
$<A_2,A_5>$	$A_2 \longrightarrow A_5 = 66\%$ $A_5 \longrightarrow A_2 = 50\%$
$<A_4,A_5>$	$A_4 \longrightarrow A_5 = 50\%$ $A_5 \longrightarrow A_4 = 50\%$
Total number of rules = 8	

IV.B RESULTS ANALYSIS

This experiment shows that sensitive association rule hiding can be done efficiently and easily with this methodology. By observation optimum solution hides all the sensitive item sets and in turn hides all the specified sensitive association rules. In heuristic and border based approaches, solution is guaranteed but by accepting more side effects where as modified inline algorithm provides solution with least number of side effects. This algorithm finds easily positive border set and negative border set for any given sensitive item sets. The minimum number of constraints based on these two sets are formulated to frame a constraint satisfaction problem. This problem consisting of minimization of objective function and constraints which are ready to solve using BIP. This technique BIP is also guaranteed to provide solution for the represented CSP problem and also provides optimum solution.

It is observed that all the non sensitive item sets are not hidden and also all the non sensitive rules are accessible to any competitor. The algorithm solves the problem when the constraint consisting of product of variables using constraints reduction approach. The proposed Inline algorithm which adopts border revision concept finds suitable constraints to find optimum solution. The positive border is used to form constraints which specifies the item sets that are not to be hidden and the negative border shows constraints for sensitive item sets and some infrequent items which are in border to be hidden when mining is performed.

The rules obtained from distorted database given in table 5 proved that goals are achieved successfully that is all sensitive rules are hidden, all frequent item sets are also frequent in distorted database and no new frequent item is generated. Now the database owner may wish to provide database (distorted) to competitors to share the knowledge for getting mutual benefits. So any competitor can have access over database for performing mining operation to get association rules and these rules helps to make analysis to improve their performance of the system.

The efficiency of this methodology is increased with divide and conquers technique applied on constraints. As the database size increases and number of sensitive items increases, the CSP problem becomes more complex. Dividing the CSP problem in to sub CSP problems based on the constraints will be helpful to overcome this problem and the solution of each sub problems are combined to get the final solution. The distance between original and distorted database will be small and this is represented as minimization of objective function in the CSP. In this way database owner can share accurate knowledge to competitors by hiding sensitive item sets.

The side effect of this methodology (constraint degree reduction approach) is that it increases the number of constraints that participate to the system of inequalities. On the other hand, the resulting inequalities are very simple and allow for fast solutions, thus adhere for an efficient solution of the CSP.

V. CONCLUSION

A modified inline algorithm is proposed in this paper to efficiently perform the hiding process without any side effects by using the concepts such as positive and negative border sets and divide and conquer technique on constraints based on their dependencies. Especially when the number of constraints increases , the constraint satisfaction problem becomes more complex and the proposed algorithm solves this with the application of divide and conquer technique on constraints and it also reduces the overall time by simultaneously solving each sub CSP problem. The proposed algorithm is illustrated with sample database and proved that the sensitive item sets are hidden without any side effects.

REFERENCES

[1] M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim, and V.S. Verkios, Disclosure Limitation of sensitive rules(1999), In proceedings of the 1999 IEEE Knowledge and Data Engineering Exchange Exchange Workshop (KDEX'99), PP 45-52.

[2] E. Dasseni, V.S. Verkios, A.K. Elmagarmid, and E. Bertino. Hiding Association Rules by using Confidence and Support (2001), In proceedings of the 4th International Workshop on Inforamtion Hiding , pp 369-383.

[3] S.R.M. Oliveria, O.R. Zaiane, Privacy Preserving Frequent Item Set Mining (2002), In proceedings of the 2002 IEEE International Conference on Privacy, Security and Data Mining (CRPITS 2002)., PP 43-54.

[4] E. D.Pontikakis, A.A.Tsitsonis, V.S. Verykios, An Experimental Study of Distortion-based Techniques for Association Rule Hiding (2004), In Proceedings of the 18th Conference on Database Security (DBSEC'2004), PP 325-339.

[5] Yiicel Saygen, Vassilios, S. Verykios, Ahmed K. Elmagramid, Privacy preserving association rule mining(2002), Proceedings of the 12th International Workshop on Research issues in Data Engineering: Engineering E-Commerce/ E-Business Systems (RIDE.02), IEEE.

[6] X.Sun, P.S. Yu, A Border Based Approach for Hiding Sensitive Frequent Item sets (2005), In Proceedings of the 5th IEEE international Conference on Data Mining (ICDM' 2005), PP 426-433.

[7] G.V. Moustakides, V.S. Verykios, A max-min Approach for Hiding Frequent Item sets (2006), In workshops proceedings of the 6th IEEE International Conference on Data Mining (ICDM' 2006), PP 502-506.

[8] Menon, S.Sarkar, S.Mukherjee , Maximizing accuracy of shared databases when concealing sensitive patterns, Information Systems Research (2005), 16(3), PP 256-270.

[9] A. Gkoulalas Divanis, V.S. Verykios,An Integer Programming Approach for Frequent Item set Hiding (2006), In proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM), pp 748-757.

[10] A. Gkoulalas Divanis, V.S. Verykios, Association Rule Hiding for Data Mining (2010), Advances in Database Systems, Volume 41.