general, data modification [1] is used in order to modify the original values of a database that needs to be released to the public and in this way to ensure high privacy protection. It is important that a data modification technique should be in concert with the privacy policy adopted by an organization. The third dimension refers to the data mining algorithm, for which the data modification is taking place. This is actually something that is not known beforehand, but it facilitates the analysis and design of the data hiding algorithm. Among them, the most important ideas have been developed for classification data mining algorithms, like decision tree inducers, association rule mining algorithms, clustering algorithms, rough sets[4] and Bayesian networks. The fourth dimension refers to whether raw data or aggregated data should be hidden.

The complexity for hiding aggregated data in the form of rules is of course higher, and for this reason, mostly heuristics have been developed. The lessening of the amount of public information causes the data miner to produce weaker inference rules that will not allow the inference of confidential values. This process is also known as “rule confusion”. The last dimension, which is the most important, refers to the privacy preservation technique [2] used for the selective modification of the data. Selective modification is required in order to achieve higher utility for the modified data given that the privacy is not jeopardized.

Problem Statement

Association Rule Mining:

The problem of mining association rules was introduced in [2]. Let I = { i,, i,;.., i, } be a set of literals, called items. Given a set of transactions D, where each transaction T is a set of items such that T   I , an association rule[1] is an expression X   Y where x   I, Y   I , and X   Y =Φ .The X and Y are called respectively the body (left hand side) and head (right hand side) of the rule. An example of such a rule is that 90% of customers buy hamburgers also buy Coke. The 90% here is called the confidence of the rule, which means that 90% of transaction that contains X also contains Y.

The confidence is calculated as IX   YI / IxI.The support of the rule is the percentage of transactions that contain both X and Y. which is calculated as IX   YI /N . In other words, the confidence of a rule measures the degree of the correlation between itemsets, while the support of a rule measures the significance of the correlation between itemsets. The problem of mining association rules is to find all rules that are greater than the user-specified minimum support and minimum confidence

Description of Problem

Sensitive Rules Hiding (By Changing the Support or Confidence of the Rules).

The main objective of rule hiding is to transform the database such that the sensitive rules are masked, and all the other underlying patterns can still be discovered. For doing this the support or the confidence of the large item sets or the association rule is changed which helps in hiding them. In this regard, the minimum support and confidence will be altered into a minimum interval. As long as the support and/or the confidence of a sensitive rule lie below the middle in these two ranges, the confidentiality of data is expected to be protected. The rule hiding method hides a group of association rules, which is characterized as sensitive. One rule is characterized as sensitive if its disclosure risk is above a certain privacy threshold. Sometimes, sensitive rules should not be disclosed to the public since, among other things, they may be used for inferring sensitive data, or they may provide business competitors with an advantage

Related Work

Following section discusses two methods of hiding rules technique along with their complete analysis.

One Rule at A Time (Proposed By Veryki- os Et Al, Etc.) Distortion Based Technique (Sanitization)

In this work [1] authors propose strategies and a suite of algorithms for hiding sensitive knowledge from data by minimally perturbing their values. The hiding strategies proposed are based on reducing the support and confidence of rules that specify how significant they are .In order to achieve this, transactions are modified by removing some items, or inserting some new items depending on the hiding strategy. The constraint on the algorithms proposed is that the changes in the database introduced by the hiding process should be limited, in such a way that the information incurred by the process is minimal. Selection of the items in a rule to be hidden and the selection of the transactions that will be modified is a crucial factor for achieving the minimal information loss constraint.

On the basis of sensitive item (proposed by shyue-liang wang et al.) Distortion based Technique (sanitization)

Technique proposed in this work tries to hide certain specific items that are sensitive and proposes two algorithms to modify data in the Dataset so that sensitive items cannot be inferred through association rule mining algorithms. Concept of this paper says that if the sensitive item is on the LHS of the rule then increase its support and if the sensitive item is on the right of the rule then decrease its support. This work is in contrast with previous work as approach in [1] hides a specific rule and the approach in [2] tries to hide all the rules containing sensitive items (either in the right or in the left)

Analysis of Existing Techniques

Existing approaches have some problems. Data perturbation [5] considers the applications where individual data values are confidential rather than the data mining results and concentrated on a specific data mining model, namely, the classification by decision trees.
Additive noise can be easily filtered out in many cases that may lead to compromising the privacy.

A potential problem of traditional additive and multiplicative perturbation is that each data 48(i)-7p(i)-7(e)19(i)--7(4) 528 the RHS i.e. it fails to hide all the rules containing sensitive item and takes more number of passes to prune all the rules containing sensitive items.

The Proposed Approach

This proposed work concentrates on the hiding sensitive rules by changing the support and the confidence of the association rule or frequent item set as data mining mainly deals with generation of association rules. Most of the work done by data miner revolves around association rules and their generation. As it is known that association rule is an important entity, which may cause harm to the confidential information of any business, defense organization and raises the need of hiding this information (in the form of association rules) that association amongst the data is what is understood by most of the data users so it becomes necessary to modify the data value(s) and relationships (Association Rules). Saygin et al [1] and Wang et al [2] have proposed some algorithms which help in reducing the support and the confidence of the rules. Hiding of association rules that expose the sensitive part of the data, researchers are bound to modify the data value(s) and relationships (Association Rules) because association amongst the data is what is understood by most of the data users. Changing the support or the confidence of the sensitive items existing in the database can modify data values. Many methods for hiding of association rules by changing the data values have been proposed in the literature [1, 2]. Existing approaches fail to hide all the rules, which contain sensitive items and even if they do so the number of passes required are many.

The proposed approach neither increases nor decreases the support of the sensitive items rather it just changes the position of the sensitive item in the database and results in hiding more number of association rules which contain sensitive items.

Hiding Association Rules Using R-Rules

The proposed approach selects all the association rules containing sensitive items either in the left or in the right from the set of all association rules generated from a dataset. Then these rules are represented in representative rules (RR) format with sensitive item on the LHS/RHS of the rules. Select a rule from the set of RR's, which has sensitive item on the LHS/RHS of the rule. Select a transaction that completely support RR i.e. it contains all the items in the RR. Now from this selected transaction delete the sensitive item and add the same sensitive item to a transaction which partially supports RR i.e. where items in RR are absent or only one of them is present.

Based on this a new algorithm for modifying database without changing the support of the sensitive item and

(the72.025.672.57 Tmb)25(ti)-4.34 (BER)-30ET giyf an102105-1s 4macontain sensiti pres92] TJBER g72.024 736.06 Tm2( )] TJET EMC /P <MCCID 6

```
If H is empty then EXIT;
Select all the rules with min_supp containing h and store in U//h can either be on LHS or RHS
Repeat {
Select all the rules from U with same LHS
Join RHS of selected rules and store in R; //make representative rules
}Until (U is empty);
Sort R in descending order by the number of supported items;
Select a rule r from R
Compute confidence of rule r.
If conf>min_conf then   {//change the position of sensitive item h.
Find T1={t in D|t completely supports r ;
If t contains x and h then
Delete h from t
Find T1={t in D|t does not support LHS(r) and Partially supports x;
Add h to t
Repeat
{

```

[redacted]

From this rules se [redacted] > A and C=> B can be repr [redacted] ent and add C to a transa [redacted] change transaction T2 to [redacted] ve item without changing [redacted]

And the new set o [redacted]

i.e. all the rules c [redacted] RHS are hidden.

Table references: [redacted]
min_supp = 33%, [redacted]
Sensitive item = B [redacted]
Similarly if H={B} i [redacted] and the modified dataset [redacted]

[redacted]

B => C	50	75
C => B	50	75
B => A	50	75
A => B		

Table.6 Modified Dataset1 for the proposed approach (Sensitive Item ς B)

TID	ITEMS
T1	ABC
T2	ACD
T3	BCE
T4	ACDE
T5	

Table.8 Database for Dataset1 in Table.1 before and after hiding C and B

TID	ITEMS	D1(C sensitive)	D2(B sensitive)