Association Rules Hiding for Privacy Preserving Data Mining: A Survey

Gehad Ahmed Sultan Abd El_Aleem
Information Systems Department
Faculty of Computers and Information, Helwan University,
Beni Suef University, Egypt

Laila Abd_Ellatif, PhD
Information Systems Department
Faculty of Computers and Information, Helwan University,
Helwan, Egypt

Ahmed Sharaf
Information Systems Department
Faculty of Computers and Information, Helwan University,
Helwan, Egypt

ABSTRACT

(PPDM) privacy preserving data mining is recent advanced research in (DM) data mining field; Many efficient and practical techniques have been proposed for hiding sensitive patterns or information from been discovered by (DM) data mining algorithms. (ARM) Association rule mining is the most important tool in (DM) data mining, that is considered a powerful and interested tool for discovering relationships between items, which are hidden in large database and may provide business competitors with an advantage, thus the hiding of association rules is the most important point in (PPDM) privacy preserving data mining for protecting sensitive and crucial data against unauthorized access; Many Practical techniques and approaches have been proposed for hiding association rules for (PPDM) privacy preserving data mining; In this paper the current existing techniques and algorithms for all approaches for (ARH) association rule hiding have been summarized.

General Terms

Privacy Preserving Data Mining; Association Rules Mining; Association Rules Hiding;

Keywords

(DM) Data Mining; (PPDM) Privacy Preserving Data Mining; (ARM) Association Rules Mining; (ARH) Association Rules Hiding; (MST) minimum support threshold; (MCT) minimum confidence threshold; (SE) Side Effects; and (SAR) Sensitive Association Rules .

1. INTRODUCTION

All major international treaties and agreements define the Privacy as is essential human rights. Every country in the world recognizes privacy as an essential human right in their constitution, either explicitly or implicitly [1].

Privacy and Emerging Technologies

New technologies, and techniques such as smart cameras, data mining (DM) , and DNA research, etc. became easier than ever before to process and store large amounts of personal and sensitive information. Ethical direction for new and emerging fields of science and technology presented techniques for protecting the problems of disclosure sensitive information; this emerges in privacy of new technologies is very important in various sectors, such as justice and homeland security and health sector, and new technologies, such as biometrics, biomedical technology, information technology, security technology and the prospective applications of particular nanotechnologies [2].

Some Technological Solutions for Protecting Privacy

Privacy preserving has been handled by different methods

1) Privacy preserving data mining (PPDM) [3, 4, 5, 6, 7, and 8].
2) Hippocratic databases (10 fundamental for database's privacy principles) [9].
3) Information sharing across private repositories [10, 11, 12].
4) Privacy-preserving search [13, 14].
5) Information fusion in data privacy [15].

Recent advances in security technologies and in data mining (DM) have given rise to a new stream of research, known as privacy preserving data mining (PPDM). PPDM technologies allow us to extract relevant knowledge and patterns from a large amount of data, but hide sensitive data or information from revelation [16]. Several questions have often being asked in this direction of Privacy Preserving Data mining (PPDM): (1) what kind of algorithms for privacy preserving data mining (PPDM)? (2) Which method is more popular and more effective? (3) How the performance of these algorithms can be measured? and (4) how effectiveness of these algorithms in preserving privacy?

The association rule mining (ARM) is one of the more popular and important problems in data mining; association analysis is considered a powerful tool for discovering relationships which are hidden in large database thus association rules hiding (ARH) algorithms get strong and efficient performance for protecting confidential and crucial data; Thus in this paper association rule hiding (ARH) algorithms for privacy preserving data mining have been reviewed.

The reminder of this paper is organized as follows; Section 2 the privacy preserving data mining (PPDM) have been described, Section 3 the association rule mining (ARM) have been described, In section 4 the association rule hiding (ARH) have been explained, In Section 5 the approaches and techniques algorithms of association rule hiding (ARH) and related work in every approach have been explained, in Section 6 contains Analysis and evaluation of existing approaches and algorithms on association rule hiding, and Section 7 contains the conclusion of the paper.

2. PRIVACY PRESERVING DATA MINING (PPDM)

A. Data Mining (DM)

Data mining (DM) deduce its name from the similarities between searching for valuable knowledge in a large database such as Prospecting for precious metals in rocks usually called "gold mining" not "rock mining",. Thus by analogy data mining (DM) should have been called "knowledge mining". Nevertheless, data mining (DM) have many terminologies such as knowledge discovery in databases (KDD) that describe a more complete process. Other similar terms referring to data mining (DM) are: knowledge extraction, data dredging, and pattern discovery, this is the approach taken in these books [17, 18, 19, and 20].

Data mining (DM) is an important issue for extracting the knowledge from a huge amount of data which can be applied to various domains, such as Web commerce, crime reconnoitering, health care, and customer's consumption analysis, engineering design, business, bioinformatics, scientific exploration, etc.. It have a lot of techniques to extract these knowledge that are expressed in decision trees, clusters, or association rules for example extracting patterns or association rules or correlation, that is very useful for decision making in most organizations, especially in market basket analysis so that the association rule mining (ARM) technique is considered most interested technique in data mining (DM).

B. Privacy preserving Data Mining (PPDM)

Privacy preserving data mining (PPDM) is a new research direction in data mining and statistical databases, the field of large number of data rich environments take the interest of many researchers and administrators in many fields, such as biomedicine (e.g., electronic health records), the Internet (e.g., Web commerce" customer's consumption analysis"), wireless networks (e.g., mobility data from sensors) , and crime reconnoitering, where data mining (DM) algorithms are analyzed for the side effects (SE) they incur in data privacy For example, through data mining (DM), one is able to infer sensitive information, including personal information or even patterns from non-sensitive information or unclassified data For example, consider Egypt supermarket like car-fore, Suppose shopkeeper of supermarket mines the association rules of marketing, where he found that most of the customers who buy chepsi also buy Pepsi-Cola, The Manager of sale can consider grouping these items to increase sales and puts some discount on the cost of chepsi with another sale. This is how customers of car-fore will now move to Reliance for this each supermarket is ready to hide sensitive association rules of its own sensitive products against unauthorized access that may provide business competitors with an advantage, This scenario leads to the research of sensitive knowledge hiding in database to solve these inference of sensitive information.

F. Bonchi, B. Malin, and Y. Saygin wrote an article [23], that has discussion of privacy issues for data mining researchers, also The most important study, status, and main research methods of Privacy preserving data mining (PPDM) have been introduced in [3, 4, 5, 6, 7, 8, 21, 22, and 24].

3. ASSOCIATION RULE MINING

3.1 Definition

The most efficient data mining technique is Association rule mining; that is strong tool for discovering relationships which are hidden in large database, That have more interested in market data analysis and many other fields, in 1993 R. Agarwal introduced the first algorithm for association rule mining [25], Association rule mining algorithms have two metrics, which scan the database of transactions and first calculate the support and then confidence.

- *The support (SUPP):* is the frequent of an item(s) in all data base transactions.
- *The confidence (CONF):* is the degree of strength or correlation between items.

1. *"Calculate SUPP"* First find all frequent item sets which occur at least as frequently as a pre-determined minimum support count.
2. *" Calculate CONF"* Generating strong association rules from the frequent item sets based on user defined minimum confidence, that is a very important step to determine whether a rule is interesting or not.

support and confidence of the rules calculated by Association rules algorithms, then retrieve only those rules having support and confidence higher than or equal the user specified minimum support threshold (MST) and minimum confidence threshold (MCT) where (MST) minimum support threshold and (MCT) minimum confidence threshold are two given or a user-specified minimum thresholds. Figure (1) verifies the flow chart of association rule mining (ARM).

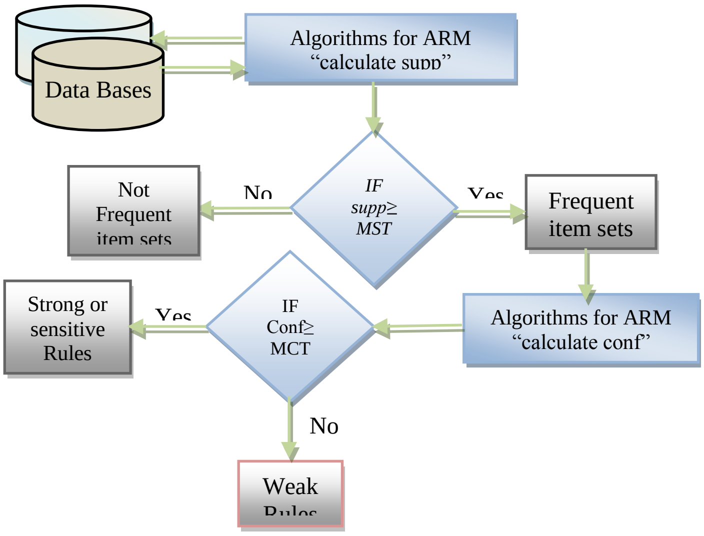
Data Bases
Algorithms for ARM "calculate supp"
IF supp≥ MST
No
Not Frequent item sets
Yes
Frequent item sets
Algorithms for ARM "calculate conf"
IF Conf≥ MCT
Yes
Strong or sensitive Rules
No
Weak Rules

Fig 1. Flow Chart of Association Rule Mining (ARM)

3.2 Association rule mining strategy (calculate Supp & Conf)

The problem of Association rule mining can be formulated as follows. Let I = $\{i_1, i_2... i_n\}$ be a set of items, D = $\{t_1, t_2...t_n\}$ be a set of transactions where $t_i \subseteq I$ and each transaction t Є D, A unique identifier, TID, is associated with each transaction. A transaction t supports K, a set of items I, $K \subseteq t$, an item set K contains m items is called a m-item set, for example, an item set {A, B} as AB where $A \cap B = \emptyset$, and a TID set {1, 4} as 14 that considered 2-item set. The rule of (A → B) where A is called left-hand-side (LHS) or the antecedent and B is the right hand side (RHS) or consequent to discover this rule is strong or not. The support and confidence should be calculated, the support of an item set AB is the number of transactions in which A and B matches this transactions or sub transactions. An item set AB is frequent if its support is greater than or equal to a user specified (MST) minimum support threshold value that calculated by the following equation:

$$\textbf{Support } (\mathbf{A \rightarrow B}) = |\mathbf{A \cap B}| / |\mathbf{D}| \qquad \mathbf{(1)}$$

Where |D| is all transactions in data base. The confidence of an item set A and B is calculated by following equation:

Confidence (A → B) = |A ∩ B| / |A| (2)

where |A ∩ B| is the number of transactions in database D that contains item set A and B and |A| is the number of transactions that contains A . A rule (A → B) is strong if two equations (1) and (2) ≥ given two thresholds, support (A ᴜ B) ≥ MST and Confidence (A ᴜ B) ≥ MCT. The most popular algorithms for association rule mining (ARM) are A priori algorithm [26], it discovers meaningful item sets and constructs association rules within large databases, but the generation of candidate item sets needs to perform contrasts against the whole database, level by level, in the process of creating association rules. Performance is considerably affected, as the database is repeatedly scanned to contrast each candidate item set with the database, FP-growth algorithm [27] which is an extended prefix-tree structure for storing compressed, crucial information about frequent patterns and developed an efficient, FP-growth method, for mining the complete set of frequent patterns by pattern fragment growth. FP-growth method is faster than the A priori algorithm, efficient, and scalable for mining both long and short frequent patterns, and also several methods have been proposed in association rule mining to discover all the strong rules [27, 28, 29, 30, and 31].

More advanced and recent association rule mining (ARM) algorithms have been proposed such as “RMAIN” algorithm which works repeatedly on subsequent portions of new transactions, After a portion has been analyzed, the new rules are combined with the old ones, so that no reruns through the processed transactions are performed in the future [32], also another algorithm for association rules with multiple constraints which enables users to concentrate on mining their interested association rules instead of the complete set of association rules [33], Also more efficient algorithm with multi-objective have been proposed, that based on genetic algorithm and Euclidean distance formula [34].

4. ASSOCIATION RULE HIDING (ARH)

4.1 Definition

The Association rule hiding (ARH) is a subfield of Privacy Preserving Data Mining (PPDM); that the process of sanitization has been happened that transforms the source database (D) into a released or sanitized or perturbed database (D\) so that the sensitive rules cannot be extracted from released or sanitized or perturbed database (D\). a set (R) of Strong rules that are mined from (D) and (RH) a subset of R, where (RH) is the set of sensitive rules (RH ⊆ R), this sensitive rules are part or all strong rules that have been mined from Data base.

The objective of the association rule hiding (ARH) algorithms is to hide or cover sensitive information (rules hiding ”RH”) from unauthorized access so that they cannot be discovered through association rule mining (ARM) algorithms. The data that be sanitized are either non distributed (localization) data or distributed data bases over several sites for example patient data may belong to two hospitals. It may be unethical or even illegal to distribute the patient data to either site. The rules hiding (RH) process has done by decreasing support or confidence below the minimum threshold as the following example.

Hiding a rule (e.g. X→ Y), can be done either by decreasing the support of the item set X and Y below (MST) minimum support threshold or decreasing the confidence of the item set X and Y below (MCT) minimum confidence threshold as follow.

1) **Either decreasing the support of a rule X →Y**
   can be done by decreasing the support of the corresponding large item set XY.
2) **Or decreasing the confidence of a rule X → Y**
   can be done by either increasing the support of X in transactions and not of Y or by decreasing the support of Y in transactions supporting both XY. Decreasing the confidence as follows.

a) Either decreaseing the nominator while keeping the denominator fixed.

b) Or increasing the denominator while keeping the nominator fixed.

4.2 Side Effects (SE) of Association Rule Hiding (ARH)

Also the process of sanitization data is very important for Privacy of sensitive rules, it have some side effects (SE) on the non sensitive rules that should not be hidden (or lost rules), and some new wrong rules (or ghost rules) may be generated, which were not previously existing that cause an undesirable side effect so that Association rule hiding (ARH) must satisfy following conditions:

1. Sensitive Rule (SR) : should not be generated from Sanitized database.

2.Non sensitive rule (NSR): must be generated that may be lost along with sensitive rules.

3.New Ghost rules (GR) or wrong rules*:* may be created that should not be generated from Sanitized database.

In Figure (2): the Flow Chart verifies the Steps of association rule hiding (ARH).

There are many methods and algorithms to reduce the loss of non-sensitive rules or the creation of ghost rules during the rule hiding process, an interesting set of techniques for association rule hiding with limited side effects (SE), which has been discussed in [35].

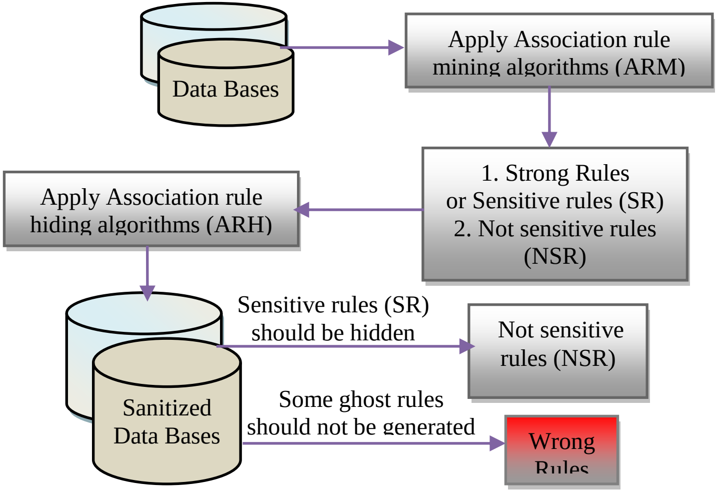
Data Bases
Apply Association rule mining algorithms (ARM)
1. Strong Rules or Sensitive rules (SR) 2. Not sensitive rules (NSR)
Apply Association rule hiding algorithms (ARH)
Sensitive rules (SR) should be hidden
Not sensitive rules (NSR)
Sanitized Data Bases
Some ghost rules should not be generated
Wrong Rules

Fig. 2. Flow Chart of Association Rule Hiding

5. ASSOCIATION RULE HIDING APPROACHES & RELATED WORK

There are five approaches for Association Rule Hiding: heuristic approaches, Border-based approaches, cryptographic approaches, exact approaches, and reconstruction approaches.

the common strategy adopted by the majority of researchers is Hiding the sensitive association rules by hiding their generating item sets; the Table 5 (summary of survey) summarizes all algorithm(s) and method(s), authors, and year of each algorithm for five approaches of association rule hiding (ARH), which are Consequent from the oldest to the most recent and are grouped according to the approach of every algorithm.

5.1 Heuristic approach

Majority of researchers concentrate on sanitization of original database heuristically, as this approach have efficient, fast and scalable algorithms that select the appropriate data sets for modification, heuristic algorithms are based on mainly two techniques Data distortion technique and data blocking technique.

5.1.1 Data distortion technique:

Data distortion is done by the alteration of an attribute value. It changes a selected set of 1-values to 0-values (delete item(s)) or 0-values to 1- values (add item(s)). There are two basic approaches for rule hiding in data distortion based technique: Reduce the support of rules below the minimum support threshold or reduce the confidence below the minimum confidence threshold of rules as follow.

1) *Reduce Support:* by changing some of the 1-values to 0-values so that the support of the corresponding sensitive rules is appropriately lowered below the minimum support threshold.

2) *Reduce confidence:* by changing some of the 0-value in transactions that support left rule's or antecedent support (left side) to 1-values so that the support of the corresponding sensitive rules is appropriately increase rule's antecedent support (left side) of the sensitive rules until the rule confidence decreases below the minimum confidence threshold or by decreasing the support of consequent support (right side) in transactions supporting both left and right of sensitive rules.

Example: Consider sample database given in Table 1, Selecting minimum support = 20% and minimum confidence = 80% and applying association rule mining algorithm, two association rules XY→Z (confidence = 100%) and YZ→X (confidence= 100%) are mined, now suppose rule XY→Z is sensitive and needs to be hidden, decreasing the confidence of a rule XY→Z can be done by either increasing the support of XY in transactions not supporting Z (as shown in Table 2) or by decreasing the support of Z in transactions supporting both XY and Z (as shown in Table 3). Decreasing support of rule XY→Z can be done by decreasing the support of the corresponding large item set XYZ (as shown in Table 4) [36].

Table 1 Sample Database

TID	Items		
1	X, Y, Z		
2	X, Y, Z		
3	X, Z	**Rule**	**Confidence**
4	X, E	XY→Z	100%
5	Z, D	YZ→X	100%

Table 2 Hiding XY→Z by Increasing Support of XY

TID	Items		
1	X, Y, Z		
2	X, Y, Z		
3	X, Z	**Rule**	**Confidence**
4	X, Y, E	XY→Z	66%
5	Z, D	YZ→X	100%

Table 3 Hiding XY→Z by Decreasing Support of Z

TID	Items		
1	X, Y		
2	X, Y, Z		
3	X, Z	**Rule**	**Confidence**
4	X, E	XY→Z	50%
5	Z, D	YZ→X	100%

Table 4 Hiding XY→Z by Decreasing Support of XYZ

TID	Items		
1	X, Y		
2	X, Y		
3	X, Z	**Rule**	**Confidence**
4	X, E	XY→Z	0%
5	Z, D	YZ→X	0%

In [37] The authors proposed three single rule heuristic hiding algorithms (1a, 1b, 2a), that are based on the reduction of either the support or the confidence of the sensitive rules, but not both The first two algorithms reduce the confidence of the sensitive rule either (ISL), or by (DSR) until the confidence lies below the minimum threshold, the third algorithm decreases the frequency of a sensitive rule, by decreasing the support (DS) of either the antecedent or the rule consequent, until either the confidence or the support lies below the minimum threshold but this algorithms have some side effects , non-sensitive rules may be lost.

In [38] Four Algorithms (Minimum Frequency Item Algorithm) Min FIA, (Maximum Frequency Item Algorithm) Max FIA, (Item Grouping algorithm) IGA, and Naïve have been proposed, the first three algorithms depends on Item Restriction but the Fourth one (Naïve) depends on Pattern Restriction, Naive Algorithm delete all items of selected transaction except the item with the highest frequency in the database, Min FIA, algorithm Max FIA selects the item with the maximum support in the restrictive pattern as a victim item Unlike Min FIA algorithm selects item with the smallest support in the pattern and it removes this selected item from the sensitive transactions. Algorithm IGA collects restricted patterns in groups of patterns so that all sensitive patterns in the group will be hidden in one step.

In [39] the author introduced an efficient Sliding Window Algorithm (SWA) that are applied to every group of K transactions (thus formulating a window of size K) , that computes the number of supporting transactions that need to be sanitized for each rule and then sorts them in ascending order of size. For each selected transaction, the corresponding item is removed and then the transaction is copied to the sanitized dataset. This algorithm improves the balance between protection of sensitive knowledge and pattern discovery, However this algorithm doesn't take the effect of non-sensitive rules into consideration, so that the author proposed three algorithms (Aggregate, Disaggregate, Hybrid) [40], that out-perform SWA by offering higher data utility and lower distortion to solve this problem.

The works in [37] have been improved in [41], that introduced three strategies and five algorithms (1a, 1b, 2a, 2b, and 2c) using both the approaches reducing the support or confidence (ISL or DSR); the first three algorithms are rule-oriented. In other words, they decrease either the confidence or the support of a set of sensitive rules, until the rules are hidden; the last two proposed (added) algorithms are item set-oriented. They decrease the support of a set of large item sets until it is below a user-specified threshold; these algorithms finally proved that there is no optimal solution.

Downright Sanitizing Algorithm (DSA) [42] aims at balancing privacy and disclosure of information by blocking some inference channels to block forward inference attack and backward inference attack to hide sensitive rules.

In [43] Two algorithms have been proposed, The first algorithm, called Priority-based Distortion Algorithm (PDA), reduces the confidence of a rule by converting 1's to 0's in items belonging in its consequent, the second algorithm Weight-based Sorting Distortion Algorithm (WDA), which assigns transactions a priority weight and sorts them ascending, Then it uses these weights to compute the priority value for each transaction, Then hide transactions which support a sensitive rule.

Maximum item conflict first (MICF) algorithm [44], this algorithm is effective, has a low sanitization rate, and can generally achieve a significantly lower misses cost than those achieved by the Min FIA, Max FIA, IGA and Algorithm 2b.

In [45] two algorithms have been proposed, ISL (Increase Support of LHS) and DSR (Decrease Support of RHS), for hiding informative association rule sets without pre-mining and selection of hidden rules, In this two algorithms, assuming that predicting items are given.

In [46] (FHSAR) fast hiding sensitive association rules. This algorithm can completely hide any given SAR by scanning database only once.

In [47] DSR algorithm was proposed, to hide the sensitive rules that contain sensitive items in right hand side only, so that sensitive rules containing specified sensitive items on the right hand side of the rule cannot be inferred.

In [48] DSRRC (Decrease Support of R.H.S. item of Rule Clusters) algorithm, which hide sensitive rules at certain level, it clusters the sensitive association rules based on R.H.S.

(WBSD) Weight Based Sorting Distortion algorithm [49], it alter a particular data that match a particular sensitive rules, Then hide those transactions which support a sensitive rule by alteration of some items, and assigns them a priority weight and sorts them in ascending order according to the priority value of each rule.

Algorithm in [50] can hide the generated crucial association rule on the both side (LHS "ISL" and RHS "DSR"), so it reduce the number of modification, hide more rule in less time.

Representative rules (RR) [51], this algorithm distorts the position of the sensitive items where these items are altered but its support is never changed, it uses (RR) to prune the rules first and then hides the sensitive rules.

ISLRC (Increase Support of L.H.S. item of Rule Clusters) [52], based on ISL approach this algorithm hides only rules that contain single item on L.H.S of the rule.

(IHC) Increasing hiding counter [53], that have Modified definition of confidence and support, that computes confidence and support as follows Conf (X→Y) = (X U Y) / (X+ counter of rule) and Supp (X→Y) = (X U Y) / (N+ counter of rule), actually To hide the rule X→Y (containing sensitive element X on LHS), this algorithm repeatedly increases the hiding counter of the rule X→Y until conf (X→Y) goes below (MCT).

Algorithm in [54] can hide the generated crucial association rule on the both side (LHS "ISL" and RHS "DSR") correspondingly, so it reduce the number of modification, hide more rule in less time.

Two algorithms (Advanced Decrease Support of R.H.S items of Rule Cluster) ADSRRC and (Remove and Reinsert L.H.S of Rule) RRLR have been proposed for solving the problems of DSRRC algorithm, In ADSRRC algorithm the sensitive rules have been clustered like DSRRC algorithm, but the time of ADSRCC algorithm is faster than the DSRRC algorithm because it needs two sorting acts only. RRLR algorithm has been proposed to hide association rules with multiple RHS as it reduces the confidence of the sensitive rules for hiding these sensitive rules, in this algorithm two sorting operations are done so that the runtime is less than DSRRC algorithm. In addition, RRLR algorithm is more effective than DSRRC algorithm as the number of lost rules and the numbers of database changes have been decreased [55]. In [56] uses RR technique correspondingly,

MDSRRC [57] (Modified Decrease Support of R.H.S. item of Rule Clusters) have been proposed to hide association rules; can hide rules by clustering RHS and LHS. At first, sensitivity of items in sensitive rules' RHS calculated and LHS then it select the higher cluster to delete. MDSRRC is more efficient than DSRRC as it reduces database modification and side effects by deleting the effective candidate item.

An Improved APRIORI algorithm has been presented in [58] that generates strong association rules, this algorithm decreases unnecessary database scan in the time of generating frequent large item sets, then hide sensitive association rule by using an improved APRIORI algorithm.

DCL algorithm has been proposed in [59] that make clustering in double two directions (left and right) then select the minimum cluster to choose the deleted item from heavy transaction in the case of right cluster or add the item in the light transaction in the case of left cluster, in this algorithm the author proved that DCL is more efficient than MDSRRC, ADSRRC, and DSRRC. But DCL may suffer from a problem of lost time as it has double sort of transactions according to heavy weight then according to length of transaction.

HSARWI has been proposed in [60] this algorithm look like FHSAR [46] algorithm, but this algorithm when choosing deleted item concentrates on maximum weight, that calculated by different method from the method of FHSAR this maximum weight is concentrating on right hand side in all cases this may cause a lot of modifications in the case of items that have maximum weight in left hand side with big average.

DCMHAR [61] have made enhancements in DCL algorithm by calculating the number of deletion in the case of right clusters and the number of addition in the case of left clusters to reduce the side effects.

5.1.2 The blocking technique:

It replaces a value with an unknown notation (often represented by '?') instead of adding or removing item sets. In this regard, the definition of minimum support and minimum confidence will be altered into interval minimum support (Min supp and Max supp), and interval minimum confidence (Min conf and Max conf) correspondingly this interval called The safety margin, so the support and/or the confidence of a sensitive rule should lies between of these two ranges of values. the first Work related to the Blocking technique was in [62] that introduced three algorithms (GIH, CR, CR2) The first algorithm, relies on the reduction in the support of the generating item sets of the rule, while the other two rely on the reduction of the rule confidence of the rule, below the minimum thresholds, the Second Work related to the Blocking technique have proposed in [63], an efficient approach of [62, 63] have proposed in [64].

5.2 Border-based approaches

In this approach the borders of the original data set are perturbed in the lattice of the frequent and the infrequent patterns, border approach are in [65] the authors presented a heuristic approach that uses the notion of the border of the non-sensitive frequent item sets to follow the impact of altering transactions in the database. The proposed algorithm first computes the positive and the negative borders in the lattice of all item sets and then focuses on preserving the quality of the computed borders during the hiding process that lead to minimal side-effects To reduce the support of a sensitive item set from the negative border, the algorithm calculates the impact of the possible item deletions by computing the sum of the weights of the positive border elements that will be affected. Then, it proceeds to delete the candidate item that will have the minimal impact on the positive border. also another algorithms of boarder approaches rely on the max min criterion for the hiding of sensitive item have been proposed in [66, 67] both algorithms apply the idea of the max min criterion in order to minimize the impact of the hiding process to the revised positive border which is produced by removing the sensitive item sets and their super item sets from the lattice of frequent item sets, by restricting the impact on the border. The recent algorithm in border approach is algorithm of association rule hiding based on intersection lattice (AARHIL) [68] this an efficient algorithm for hiding a specified set of sensitive association rules based on intersection lattice of frequent item sets. It specify the victim items based on the characteristics of the intersection lattice of frequent item sets and identify transactions for data sanitization based on the weight of transactions. The AARHIL is new algorithm for hiding a specific set of sensitive association rules, this algorithm have minimum side effects, less CPU-Time, and low complexity.

5.3 Cryptographic Approach

This approach uses encryption the database instead of distortion it for sharing sensitive data, and used in multiparty computation, If the database of the organization is partitioned between several sites, then secure computation between them is needed. For securing the data that partitioned horizontally or vertically, in vertically [69] Scalar product protocols were used, where transactions are distributed across sources, where each site holds some attributes of each transaction and the sites wish to collaborate to identify globally valid association rules. However, the sites must not discover individual transaction data, two party algorithms for generating frequent item sets with minimum support levels without revealing individual transaction values. In [70] securing method for computing the size of the intersection of sets of items held by different parties, another two algorithms [71] for both vertically and horizontally partitioned data, with cryptographically strong privacy, and also another algorithm in [72] privacy preserving association rule mining on distributed homogenous database algorithm, which modified with preserving privacy and accurate results, this algorithm, is based on a semi-honest model with negligible collision probability, and have flexibility to extend to any number of sites without any change in implementation can be achieved, and also less time as any increase doesn't add more time to algorithm because all sites of client perform the mining technique in the same time so the overhead in communication time only, and The cost of total bit-communication for the algorithm is function in (N) sites. Fully homomorphism encryption scheme [73] used a secure comparison technique. In [74] a new protocol is proposed which combines the advantages of the two approaches (the Randomization approach and the Cryptographic approach). Recently In [75] the author proposed an algorithm (IPPM) Improved Privacy Preserving Mining; this algorithm is a good techniques with security that hides logical instances from others.

5.4 Exact Approach

The Exact approaches are non-heuristic algorithms which use satisfaction problem of finding an optimal sanitization method such as linear programming or integer programming; these algorithms can produce optimal hiding solution or exact solution with ideally no side effects (loss rules or ghost rules) but these approaches need several orders of magnitude slower than the heuristic ones, especially due to the runtime of the linear or integer programming solver need more time. The first NP-hard problem for optimal hiding the association rules are proposed in [76]. In [77] the authors initially made use of border revision theory introduced by Sun and Yu [65] so as to achieve optimal solution as compared to previous approaches, this exact algorithm is proposed to minimize the distance between the original database and its sanitized version for association rule hiding. In [78] the author proposed an exact border based approach to achieve optimal solution as compared to previous approaches.

Recently Tabu search technique [79] uses binary transactional dataset as an input and modifies the original dataset for hiding sensitive association rules without any loss of data.

5.5 Reconstruction Approach

Reconstruction approach is a recent and advanced approach for hiding association rules, which first performs classification on rules of the original dataset to enable the owner of the data to identify the sensitive rules then; they proceed to construct a decision tree that is designated only on non-sensitive rules approved by the data owner. a fp-tree based method is presented in [80] for inverse frequent set mining which is based on reconstruction technique. this approach consists of three phases, the figure 3 shows this three phases: 1) the first phase generates frequent item set with their supports from original database D by mining algorithms, 2) the second phase runs sanitization algorithm over frequent item set fs and get the sanitized frequent item sets of FS$^{\backslash}$, and 3) the third phase is to generate sanitized database D$^{\backslash}$ from FS$^{\backslash}$ by using inverse frequent set mining algorithm.

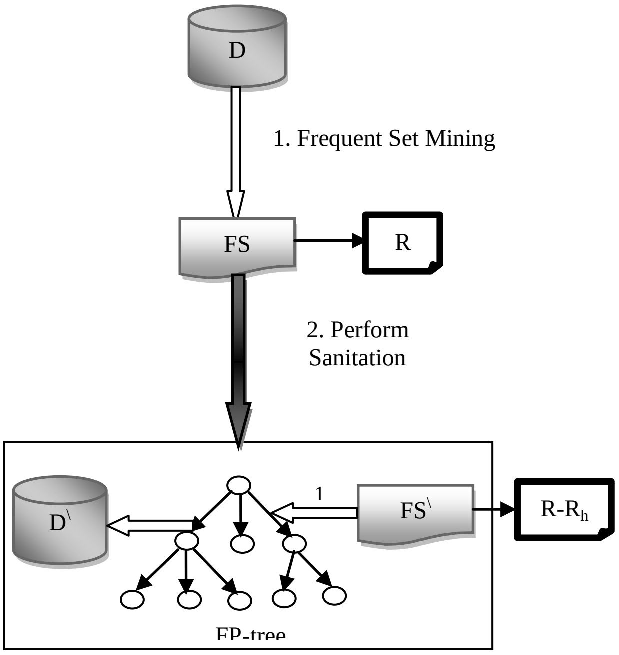

Fig. 3. Framework of reconstruction approach

But this algorithm is very complex as it involves generation of modified dataset from frequent set. In [81] that use Bayesian algorithms (Uniform and Gaussian perturbations) for privacy levels for distribution reconstruction in numerical data. In [82] derive formulae for an unbiased support estimator and its variance, which recover item sets supports from randomized datasets to preserve the categorical data before sending a transactions to the server.in[83] Give a coarse Constraint-based Inverse Item set Lattice Mining procedure (CIILM) for hiding sensitive frequent item sets. In [84] randomization transactions have been introduced by adding some items to each transaction, but by removing no item from any transaction. This approach first uses any association rule mining tool with the original minimum support to filter out all possible frequent item sets from the randomized transactions. Then, it reconstructs the support of each possible frequent item set level by level to find frequent item sets. In [85] attributes are first masked using aggregation (for numeric data) and swapping (for nominal data), without considering the k-anonymity constraint A genetic algorithm technique is then applied to the masked data to find a good subset of it. This subset is then reproduced to form the sanitized dataset that satisfies the k- anonymity constraint. Finally the following table 5 has summarization of all approaches and techniques of Association rule hiding with algorithms, methods, and years.

Table 5 Summary of Survey

		Algorithm(s) or Method(s)	**Year**
Heuristic Approach	Distortion	Three algorithms (1a, 1b, 2a)	2001
		Four algorithms (Min FIA ,Max FIA, IGA ,Naïve)	2002
		SWA	2003
		five algorithms (1a, 1b, 2a, 2b, 2c)	2004
		DSA	2004
		Two algorithms PDA,WDA	2004
		MICF	2006
		Aggregate, Disaggregate, Hybrid methods	2007
		ISL, DSR of predicted items	2007
		FHSAR	2008
		DSR	Aug 2008
		DSRRC	2010
		WBSD	2011
		ISL , DSR	July 2011
		RR technique	Jan 2012
		ISLRC	June 2012
		IHC	Oct 2012
		ISL and DSR correspondingly	Nov 2012
		ADSRRC and RRLR	2012
		RR technique correspondingly	Jan 2013
		MDSRRC	May 2013
		Wight item technique with new efficient results	July 2014
		An Improved APRIORI algorithm	Dec 2014
		DCL	Jan 2015
		HSARWI	Jan 2016
		DCMHAR	Feb 2016
	Blockin	Three algorithms GIH, CR, CR2	Dec 2001
		Improved of GIH, CR, CR2	2002
		ISL , DSR Using unknowns	2005
Border		Two algorithms (positive and negative borders)	2005
		A Max Min approach	2006
		Advanced Two algorithms Max Min1, Max Min 2	2008
		AARHIL algorithm	June 2013
Cryptographic Approach		Scalar product protocols in vertical data	2002
		Two Protocol with Three steps	Mar 2004
		Two algorithms for vertically & horizontally	2007
		Efficient algorithm	Apr 2010
		Fully homomorphism encryption	2012
		Randomization & Cryptographic technique	2012
		IPPM	Mar 2012
Exact Approach		NP-hard problems	1999
		An Integer programming approach	2006
		Exact border based approach	2009
		Tabu search	2011
Reconstruction Approach		Bayesian algorithms	2002
		Advanced Bayesian algorithms	2004
		CIILM	2004
		A FP-tree based method	June 2007
		Randomization then Reconstruction	2009
		swapping and genetic algorithm	2009

6. ANALYSIS AND EVALUATION OF FIVE APPROACHES

It is observed that the heuristic algorithms have a lot of researches as it is efficiency, scalability, and quick responses, but the degree of certainty are not complete that may suffer from undesirable side effects on the non-sensitive rules in the data that lead them to identify approximate hiding solutions, some of the non-sensitive rules may be lost along with sensitive rules, and new ghost rules may be created because of the distortion or blocking process.

Border based approaches provide an enhancement more than heuristic approaches, they are dependent on heuristics to decide the item modifications; that they apply on the original database has less side effects than heuristic approach , but not produce optimal solution.

Cryptographic approaches can secure mining of partitioned data, but it very complicated and more time because of encryption methods that need to be decrypted in another side.

Exact approaches provide an exact (optimal) hiding solution that satisfies all the constraints with ideally no side effects, but have very long time due to integer programming solver to solve the optimization problem.

Reconstruction approaches Create privacy aware database by exacting sensitive characteristic from the original database, and lesser side effects in database than heuristic. But it restricts the number of transactions in the new database.

7. CONCLUSION

Hiding the sensitive association rules is a common strategy adopted by the majority of researchers; so that in this paper we have collected and summarized all algorithm(s), method(s), and year(s) of five approaches of association rule hiding (ARH), which are Consequent from the oldest to the most recent. We need to further perfect those approaches and merge some benefits between them for developing some efficient methods for more enhancement results.

In future, hybrid technique can be found to reduce the side effects and increase the efficiency by reducing the modifications on database, while hiding the association rules. Parallel algorithm can be developed to hide sensitive rules and also improve the performance of the algorithms for large database. An algorithm for incremental environment can also be developed, as most of the current frequent hiding algorithms are designed for static database.

8. ACKNOWLEDGEMENT

The author would like to acknowledge the helping of Dr. kareem kamal, Dr. Saber Abd-Allah, Eng. Ahmed Ezzat.

9. REFERENCES

[1] B. Davies, D. BANisAR Global Trends in Privacy Protection: An International Survey of Privacy, Data Protection, and Surveillance Laws and Developments (August 30, 2012). John Marshall Journal of Computer & Information Law, Vol. XVIII, No. 1, Fall 1999.

[2] F. Berlin Germany International Conference of the PRESCIENT Project 27 – 28 November 2012.

[3] D. O'Leary, Knowledge Discovery as a Threat to Database Security, in: G. Piatetsky-Shapiro, W.J. Frawley (Eds.), Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, 1991, pp. 507–516.

[4] C. Clifton, D. Marks, Security and privacy implications of data mining, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 1996, pp. 15–19.

[5] R. Agrawal, R. Srikant "Privacy Preserving Data Mining," Proc. 2000 ACM SIGMOD Int'l Conf. Management of Data, ACM Press, 2000, pp. 439-450.

[6] A. Evfimievski, J. Gehrke, and R. Srikant "Limiting Privacy Breaches in Privacy Preserving Data Mining," Proc. 22nd ACM SIGACT-SIGMOD-SIGART Symp. Principles of Database Systems, ACM Press, 2003, pp. 211-222.

[7] V. Verykios, E. Bertino, I. Fovino, L. Provenza, Y. Saygin, and Y. Theodoridis, State-of-the-art in privacy preserving data mining, ACM SIGMOD Record 33 (1) (2004) 50 – 57.

[8] W. Bart Schermer, The limits of privacy in automated profiling and data mining, computer law & security review 27 (2011) 45 – 52.

[9] R. Agrawal, M. Kaufmann"Hippocratic Databases," Proc. 28th Int'l Conf. Very Large Databases, 2002, pp. 143-154.

[10] Y. Lindell, B. Pinkas "Privacy Preserving Data Mining," Proc. Crypto 2000, Springer Verlag, 2000, pp. 37-55.

[11] R. Agrawal, R. Srikant "Information Sharing across Private Databases," Proc. 2003 ACM SIGMOD Int'l Conf. Management of Data, ACM Press, 2003, pp. 86-97.

[12] J. Dyer "Building the IBM 4758 Secure Coprocessor," Computer, Oct. 2001, pp. 57-66.

[13] M. Bawa, R. Bayardo, and R. Agrawal "Privacy Preserving Indexing of Documents on the Network," to appear in Proc. 29th Int'l Conf. Very Large Databases, Morgan Kaufmann, 2003.

[14] B. Chor "Private Information Retrieval," IEEE Symp. Foundations of Computer Science, IEEE CS Press, 1995, pp. 41-50.

[15] G. Navarro-Arribas, V. Torra "Information fusion in data privacy: A survey" Information Fusion 13 (2012) 235–244.

[16] W. Xiaodan, C. ChaoHsien, W. Yunfeng, L. Fengli, Y. Dianmin "Privacy Preserving Data Mining Research: Current Status and Key Issues" Computational Science – ICCS, Volume 4489, 2007, pp 762.

[17] J. Han, M. Kamber Data Mining: Concepts and Techniques University of Illinois at Urbana-Champaign / 2006 / pp5.

[18] H. Margaret Dunham Data Mining Introductory and Advanced Topics / 2003 / pp.9-10.

[19] Hand, Mannila, and Smyth Principles of Data Mining / 2001 /pp1.

[20] M. Berry, G. Linoff Data Mining Techniques / Paperback / 2004 /pp2.

[21] J. Yeh, P. Hsu" HHUIF and MSICF: Novel algorithms for privacy preserving utility mining" Expert Systems with Applications 37 (2010) 4779–4786

[22] E. Bertino, I. Nai Fovino" A Framework for Evaluating Privacy Preserving Data Mining Algorithms" Data Mining and Knowledge Discovery, 11, 121–154, 2005

[23] F. Bonchi, B. Malin, Y. Saygin" Recent advances in preserving privacy when mining data" Data & Knowledge Engineering 65 (2008) 1–4.

[24] X. Qi, M. Zong "An Overview of Privacy Preserving Data Mining" Procedia Environmental Sciences 12 (2012) 1341 – 1347.

[25] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases". In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, May 26-28 1993, pp. 207-216.

[26] R. Agarwal, R. Srikant, "Fast algorithm for mining association rules in large databases", Proceedings of 1994 International Conference on VLDB, 1994, pp. 487-499.

[27] J. Han, J. Pei, Y. Yin, R. Mao," Mining Frequent Patterns without Candidate Generation: A Frequent- Pattern Tree Approach", Data Mining and Knowledge Discovery, 8, 2004, pp. 53-87.

[28] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A.I. Verkamo, "Fast Discovery of Association Rules," Advances in Knowledge Discovery and Data Mining, chapter 12, U.M. Fayyad et al., eds., AAAI/MIT Press, pp. 307-328, 1996.

[29] J. Liu, Y. Pan, K. Wang, and J. Han, "Mining Frequent Item Sets by Opportunistic Projection," Proc. ACM Conf. Knowledge Discovery and Data Mining, 2002.

[30] S.J. Yen and A.L.P. Chen, "A Graph-Based Approach for Discovering Various Types of Association Rules," IEEE Trans. Knowledge and Data Eng.,vol. 13, no. 5, 2001.

[31] M.J. Zaki, "Scalable Algorithms for Association Mining," IEEE Trans. Knowledge and Data Eng.,vol. 12, no. 3, pp. 372-390, 2000.

[32] D. Dudek," RMAIN: Association rules maintenance

without reruns through data", Information Sciences 179 (2009) 4123–4139.

[33] L. Guang-yuan, C. Dan-yang, G. Jian-wei" Association Rules Mining with Multiple Constraints" Procedia Engineering 15 (2011) 1678 – 1683

[34] N. Jain,V. Sharma,M. Malviya," Reduction of Negative and Positive Association Rule Mining and Maintain Superiority of Rule Using Modified Genetic Algorithm" International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-2 Number-4 Issue-6 December-2012.

[35] C. Chiang, A. Chen Hiding Sensitive Association Rules with Limited Side Effects. IEEE Transactions on Knowledge and Data Engineering, 19(1), 2007.

[36] K. Shah, A. Thakkar, A. Ganatra "A Study on Association Rule Hiding Approaches"International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-3, Feb 2012.

[37] E. Dasseni, V. S.Verykios, A. K.Elmagarmid, and E. Bertino, "Hiding Association Rules by using Confidence and Support," In Proceedings of the 4th Information Hiding Workshop (2001), pp.369– 383.

[38] S. R. M. Oliveira, O. R. Zaïane, "Privacy Preserving Frequent Itemset Mining", IEEE International Conference on Data Mining Workshop on Privacy, Security, and Data Mining, Maebashi City, Japan. Conferences in Research and Practice in Information Technology, Vol. 14.2002.

[39] S. R. M. Oliveira, O. R. Zaïane,"Protecting sensitive knowledge by data sanitization", In: Proc. of the 3rd IEEE Int'l Conf. on Data Mining (ICDM'03), IEEE Computer Society, USA, 2003, pp. 613-616.

[40] A. Amiri, D. toshare: Protecting sensitive knowledge withdata saniti-zation. Decision Support Systems, 43(1):181–191, 2007.

[41] V.S. Verykios, A. Elmagarmid, E. Bertino, Y. Saygin, , and E. Dasseni, "Association rule hiding", IEEE Transactions on Knowledge and Data Engineering, 2004, 16(4): pp. 434-447.

[42] S.R.M. Oliveira, O.R. Zaıane, Y. Saygin, "Secure association rule sharing, advances in knowledge discovery and data mining, in: Proceedings of the 8th Pacific-Asia Conference (PAKDD2004), Sydney, Australia, 2004, pp.74–85.

[43] E. D. Pontikakis, A. A. Tsitsonis, and V. S. Verykios. An experimental study of distortion-based techniques for association rule hiding. In Proceedings of the 18th Conference on Database Security (DBSEC 2004), pages 325–339, 2004.

[44] Y. Chiang Li a, J. Shan Yeh b, C. Chang "MICF: An effective sanitization algorithm for hiding sensitive patterns on data mining" Advanced Engineering Informatics 21 (2007) 269–280.

[45] S. Liang Wang, B. Parikh, A. Jafari "Hiding informative association rule sets" Expert Systems with Applications 33 (2007) 316–323.

[46] C. Weng, S. Chen, H. Che Lo," A Novel Algorithm for Completely Hiding Sensitive Association Rules" Eighth International Conference on Intelligent Systems Design and Applications 2008 IEEE.

[47] K. Duraiswamy, D. Manjula, N. Maheswari "A New Approach to Sensitive Rule Hiding" computer and information science Vol. 1, No. 3 August, 2008.

[48] Modi, C.N.; Rao, U.P.; Patel, D.R., "Maintaining privacy and data quality in privacy preserving association rule mining", IEEE 2008 Seventh International Conference on Machine Learning and Applications, pp 1-6, 2010.

[49] R. Sugumar, C. Jayakumar, A. Rengarajan," Design a Weight Based sorting distortion algorithm using Association rule Hiding for Privacy Preserving Data mining" R Sugumar et al, International Journal of Computer Science & Communication Networks,Vol 1(3), 270-276 (2011).

[50] Y. Kumar Jain, V. Kumar Yadav, G. S. Panday "An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining" International Journal on Computer Science and Engineering (IJCSE) Vol. 3 No. 7 July 2011.

[51] D. Jain, P. Khatri, R. Soni, and B. Kumar Chaurasia" Hiding Sensitive Association Rules without Altering the Support of Sensitive Item(s)" Second International Conference, CCSIT 2012 Bangalore, India, January 2- 4, 2012, pp. 500–509, 2012"

[52] S. keer, A. Singh "Hiding Sensitive Association Rule Using Clusters of Sensitive Association Rule" International Journal of Computer Science and Network (IJCSN) ISSN 2277-5420 Volume 1, Issue 3, June 2012.

[53] S. Sharma, P. Jain," A Novel Data Mining Approach for Information Hiding" International Journal of Computers and Distributed Systems Vol. No.1, Issue 3, October 2012.

[54] M. Mahendran,R. Sugumar, K.Anbazhagan, R. Natarajan"An Efficient Algorithm for Privacy Preserving Data Mining Using Heuristic Approach" International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 9, November 2012 .

[55] Sh. Komal, and T, Amit, and G. Amit, "Association Rule Hiding by Heuristic Approach to Reduce Side Effects & Hide Multiple R.H.S. Items", International Journal of Computer Applications, Vol. 45, No. 1, 2012, pp. 0975-8887.

[56] Kasthuri S and Meyyappan T," Hiding Sensitive Association Rule Using Heuristic Approach" International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.1, January 2013.

[57] Nikunj H. Domadiya, "Hiding sensitive association rules to maintain privacy and data quality in database"; Advance Computing Conference (IACC), 2013 IEEE 3rd International.

[58] Sh. S. Sambhaji, K. Pravin P , " Study of Mining and Hiding of Sensitive Association Rule " International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 3, Issue 6, November-December 2014 ISSN 2278-6856F.

[59] Z. K. Abari1, M. N. Dehkordi, "Privacy Preserving in

Association Rule Mining" ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 1, No.13 , January 2015.

[60] M. SakenianDehkordi, and M. NaderiDehkordi " Introducing an algorithm for use to hide sensitive association rules through perturb technique" Journal of AI and Data Mining, Published online. Accepted 18 January 2016.

[61] Z. Kiani Abari, M. Naderi Dehkordi, "Double Clustering Method in Hiding Association Rules" Journal of Advances in Computer Research Quarterly pISSN: Sari Branch, Islamic Azad University (Vol. 7, No. 12345-606x eISSN: 2345-6078, Sari, I.R.Iran February 2016), Pages: 67-8.

[62] Y.Saygin, V. S. Verykios, and C. Clifton, "Using Unknowns to Prevent Discovery of Association Rules," ACM SIGMOD, vol.30(4), pp. 45–54, Dec. 2001.

[63] Y. Saygin, V. S. Verykios, and A. K. Elmagarmid, "Privacy preserving association rule mining," In Proceedings of the 12th International Workshop on Research Issues in Data Engineering (2002), 151–158.

[64] S.L.Wang and A. Jafari, "Using unknowns for hiding sensitive predictive association rules", In Proc. IEEE Int'l Conf. Information Reuse and Integration (IRI 2005), Aug. 2005, pp. 223-228.

[65] X. Sun and P. S. Yu. A border-based approach for hiding sensitive frequent item sets. In Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM 2005), pages 426–433, 2005.

[66] V. Moustakides and V. S. Verykios. A max min approach for hiding frequent item sets. In Workshops Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 2006), pages 502–506, 2006.

[67] G. V. Moustakides, V. S. Verykios, "A Max Min approach for hiding frequent item sets" Data & Knowledge Engineering 65 (2008) 75–89.

[68] H. Quoc Le, S. Arch-int, and N. Arch-int "Association Rule Hiding Based on Intersection Lattice" Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2013, Article ID 210405, 11 pages.

[69] J. Vaidya, C. Clifton "Privacy Preserving Association Rule Mining in Vertically Partitioned Data" SIGKDD '02 Edmonton, Alberta, Canada Copyright 2002 ACM 158113567.

[70] J. Vaidya, C. Clifton "Secure Set Intersection Cardinality with Application to Association Rule Mining" Department of Computer Sciences Purdue University 250 N University St West Lafayette, IN 47907-2066 March 15, 2004

[71] S. Zhong, "Privacy preserving algorithms for distributed mining of frequent item sets" Information Sciences 177 (2007) 490–503

[72] A. 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.

[73] M. G. Kaosar , R. Paulet, X. Yi" Fully homomorphic encryption based two-party association rule mining" Data & Knowledge Engineering 76 – 78 (2012) 1– 15.

[74] S. Z. Alborzi, A. Raj, and M. H. Saraee, "Privacy Preserving Mining of Association Rules on Horizontally distributed Databases" IPCSIT vol. 41 (2012) .

[75] A. Tomar, V. Richhariya, M. Ku. Mishra, "A Improved Privacy Preserving Algorithm Us-Ing Association Rule Mining In Centralized Da-Tabase" International Journal of Advanced Technology & Engineering Research (IJATER) VOLUME 2, ISSUE 2, MARCH 2012.

[76] M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim, and V. S. Verykios, "Disclosure limitation of sensitive rules", In: Scheuermann P, ed. Proc. of the IEEE Knowledge and Data Exchange Workshop (KDEX'99). IEEE Computer society, 1999.pp. 45-52.

[77] A. Gkoulalas-Divanis and V.S. Verykios, "An Integer Programming Approach for Frequent Itemset Hiding," In Proc. ACM Conf. Information and Knowledge Management (CIKM '06), Nov. 2006.

[78] A. Gkoulalas-Divanis and V.S. Verykios, "Exact Knowledge Hiding through Database Extension," IEEE Transactions on Knowledge and Data Engineering, vol. 21(5), pp. 699–713, May 2009.

[79] S.Vijayarani, A. Tamilarasi, R. SeethaLakshmi "Tabu Search based Association Rule Hiding" International Journal of Computer Applications (0975 – 8887) Volume 19– No.1, April 2011.

[80] S. Rizvi, J. Haritsa Maintaining Data Privacy in Association Rule Mining. VLDB Conference, 2002.

[81] A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrke," Privacy preserving mining of association rules",Information Systems 29 (2004) 343–364.

[82] X. Chen, M. Orlowska, and X. Li "A new framework for privacy preserving data sharing" In: Proc. of the 4P th P IEEE ICDM Work shop: Privacy and Security Aspects of Data Mining. IEEE Computer Society, 2004. 47-56.

[83] Y. Guo "Reconstruction-Based Association Rule Hiding" Proceedings of SIGMOD2007 Ph.D Workshop on Innovative Database Research 2007(IDAR2007), June 10, 2007, Beijing, China.

[84] J. Lin , Y. Cheng, " Privacy preserving item set mining through noisy items " Expert Systems with Applications 36 (2009) 5711–5717.

[85] D. Zhu, X. Li, S. Wu," Identity disclosure protection: A data reconstruction approach for privacy-preserving data mining" Decision Support Systems 48 (2009) 133 –140.