A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms

Mohammad Naderi Dehkordi
Ph.D Student, Science & Research Branch, Islamic Azad University (IAU)
Department of Computer Engineering, Tehran, Iran
E-mail: naderi@iaun.ac.ir

Kambiz Badie, Ahmad Khadem Zadeh
Iran Telecom Research Center, Tehran, Iran
E-mail: {k_badie, zadeh}@itrc.ac.ir

***Abstract*—Extracting of knowledge form large amount of data is an important issue in data mining systems. One of most important activities in data mining is association rule mining and the new head for data mining research area is privacy of mining. Today association rule mining has been a hot research topic in Data Mining and security area. A lot of research has done in this area but most of them focused on perturbation of original database heuristically. Therefore the final accuracy of released database falls down intensely. In addition to accuracy of database the main aspect of security in this area is privacy of database that is not warranted in most heuristic approaches, perfectly. In this paper we introduce new multi-objective method for hiding sensitive association rules based on the concept of genetic algorithms. The main purpose of this method is fully supporting security of database and keeping the utility and certainty of mined rules at highest level.**

***Index Terms*—Data Mining, Privacy Preserving, Sensitive Association Rules, Genetic Algorithms**

I. INTRODUCTION

Today transactional databases have been used in great number of organization and businesses. Association rule mining is the special activity in data mining for processing and extracting knowledge from these transactions. The customer behavior and his/her shopping patterns can be comprehended with applying proper data mining algorithms. This knowledge is very useful for enhancement of business and suitable decision making.

However the irrefutable profit of analyzing business data, there always lurks the fear of unauthorized access to sensitive knowledge which are stored in databases or deduced from them. One of important classical aspects in the process of association rule mining is true mining of real world knowledge. Recent issue in association rule mining is keeping the confidence of data [14,15]. Most of information systems contain private information, such as social security numbers, income, disease type, etc. therefore these information should be correctly protected and hided from unauthorized access. Although the security of data has been permanent goal in database management systems, mining of knowledge and preventing of sensitive knowledge disclosure becomes the most important and highest priority goal in data mining process. Basically the sharing of data between businesses in purpose of reaching valuable information is useful but it can bring a lot of disadvantages.

Recent advances in data mining algorithms increased the risk of information leakage and its confidence issue. Because of this progress, the parallel research area has been started to over come the information leakage risks and immunization of mining environment**.** Privacy preserving against mining algorithms is a new research area that investigates the side-effects of data mining methods that derive from the privacy diffusion of persons and organizations.

In this paper we are studying the privacy breaches which incurred from certain type association rules. In doing so we suppose that a certain subset of association rule, which is extracted from specific datasets, is considered as sensitive/critical rules. Our major goal then is modification of original data source in such a way that it would be impossible for the adversary to mine the sensitive rules from the modified data set as long as all the remaining non sensitive information and/or knowledge remains as close as possible to this of the original set, as our minor goal.

The method developed in this paper uses binary transactional dataset as an input and modifies the original dataset based on the concept of genetic algorithms in such a way that all of sensitive rules become hide and minimum modification performed in original dataset. The most famous possible style for transaction modification is distortion of original database (i.e., by replacing 1's by 0's and vise versa). We select this style of modification in our method. Modification of the dataset causes so many side-effect problems. The modification process can affect the original set of rules, that can be mined from the original database, either by hiding rules which are not sensitive (*lost rules*), or by introducing rules in the mining of the modified database, which were not supported by the original database (*ghost rules*). We have tried to minimize these unpleasant results by minimum and suitable modification of original dataset.

II. RELATED WORK

The problem of privacy preserving in association rule mining was first addressed in [1]. After this beginning, researchers conduct so many methods to solve the privacy issue of mining results. Generally, modification/sanitization techniques can be categorized into two groups: data blocking and data distortion approaches.

Some blocking-based techniques are addressed in [1,2,3]. The major concept of blocking approaches, are replacing the actual values of the items with "unknown" symbols in the proper transactions. The main reason of using blocking techniques is that algorithms do not add artificial information in the database. This is so important when the source database contains critical information that extracting wrong known will consequences dangerous effects. More specifically, in [1] a number of algorithms are presented, each of which blocks in a different way, either 1's or 0's in order to achieve the best possible results. Blocking also deals with the so-called database inference problem, a problem which is already addressed in [5,6]. In this problem we want to prevent an adversary from inferring a hidden value of an item in a specific transaction of the database, and in [5,13] Bayesian techniques are used in order to eliminate the inference of the hidden value by the adversary.

Near the beginning, data distortion techniques take on initial heuristic-based sanitization strategies like Algo1a, Algo1b, Algo2a, Algo2b and Algo2c [7]. The major difference between approaches is heuristic determination of selection strategies on which transactions should be sanitized and which items selected for modification. Following techniques like WSDA, PDA [8] and Border-Based [9] improved the initial heuristic algorithms to greedy algorithms (based on finding local optimal modification). WSDA technique is reached through the use of priority values assigned to transactions based on weights. In theses approaches tried to greedily select the modifications with minimal side effects on data utility and accuracy.

In the rest of this paper, we present our efficient and novel method for sanitization of database and rule hiding, as well as implementations and experiment results with binary data sets. One of the most interesting parts in our paper is evaluation of hiding performance in our work. In order to analysis the performance we suggest some criterion. In Sect. 3, the general problem formulation and the basic definitions regarding sensitive association rules are given. In Sect. 4, we present the algorithm which implements the distortion technique based on Genetic Algorithm approach. In Sect. 5 we evaluate the effectiveness of the algorithm based on some criterion and we present experimental results from their implementation, also providing a qualitative analysis of the proposed techniques. We conclude the paper in Sect. 6, also giving hints for future work.

III. PROBLEM FORMULATION

Let $I = \{i_1, i_2, ..., i_m\}$ be a set of items and let D is the dataset of transactions that the goal of sanitization is its modification in order to no sensitive rule disclosed. Any $X \subseteq I$ is an itemset. Each itemset which contains *k* items called *k-itemset.* Let $D = \{T_1, T_2, ..., T_n\}$ be a set of transactions. The well known measure in frequent itemset mining is support of itemset. The support measure of an item $X \subseteq I$ in database D, is the count of transactions contain X and denoted as *Support_count(X)*. An itemset *X* has support measure *s* in dataset D if *s%* of transactions support *X* in dataset *D*. Support measure of *X* is denoted as *Support(X)*.

$$Support(X) = \frac{Support\_count(X)}{n} \times 100$$

(where n is number of transactions in dataset *D*).
Itemset *X* is frequent itemset when Support(X) $\geq MST$ where *MST* is Minimum Support Threshold that is predefined threshold. After mining frequent itemsets, the association rule is an implication of the form $X \rightarrow Y$, where $X, Y \subset I$ and $X \bigcap Y = \phi$. The *Confidence* measure for rule $X \rightarrow Y$ in dataset *D* is defined

$$Confidence(X \rightarrow Y) = \frac{Support(XY)}{Support(X)} \times 100 \cdot$$

Note while the support is a measure of the frequency of a rule, the confidence is a measure of the strength of the relation between sets of items. Association rule mining algorithms scan the database of transactions and calculate the support and confidence of the candidate rules to determine if they are considerable or not. A rule is considerable if its support and confidence is higher than the user specified minimum support and minimum confidence threshold. In this way, algorithms do not retrieve all possible association rules that can be derivable from a dataset, but only a very small subset that satisfies the minimum support and minimum confidence requirements set by the users. An association rule-mining algorithm works as follows. It finds all the sets of items that appear frequently enough to be considered relevant and then it derives from them the association rules that are strong enough to be considered interesting. The major goal here is to preventing some of these rules that we refer to as "sensitive rules", from being revealed. The problem of privacy preserving in association rule mining (so called association rule hiding) focused on this paper can be formulated as follows:

Given a transaction database D, minimum support threshold "MST", minimum confidence threshold "MCT", a set of significant association rules R mined from D and a set of sensitive rules $R_{Sen} \subseteq R$ *to be hided, generate a new database* D'*, such that the rules in* $R_{non-Sen} = R - R_{Sen}$ *can be mind from* D' *under the same "MST" and "MCT". Further, no normal rules in* $R_{non-Sen}$ *are falsely hidden (lost rules), and no extra fake*

rules (ghost rules) are mistakenly will mined after the rule hiding process.

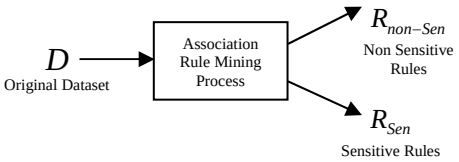

Figure 1. Association Rule Mining process input and outputs

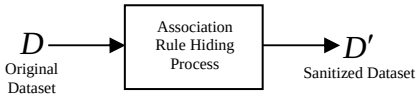

Figure 2. Association rule hiding process input and outputs

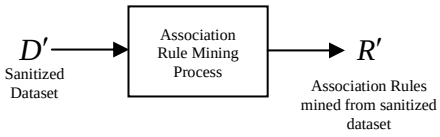

Figure 3. Association Rule Mining after Association Rule Hiding

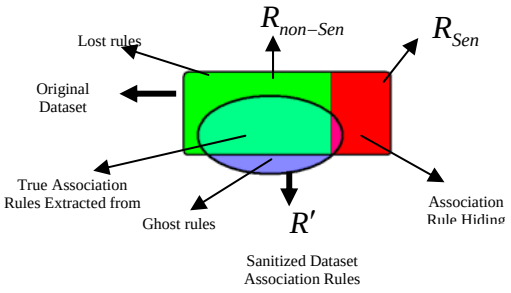

Figure 4. Side effects of association rule hiding

In [1] proved that solving above problem by sinking the support of the large itemsets via removing items from transactions or adding fake item into the transactions (also referred to as "sanitization" problem) are an NP-hard problem. Therefore, we are looking for a special modification of D (the source dataset) in D' (sanitized dataset which is going to be released) that *maximizes* the number of rules in $R_{non-Sen}$ (*minimizing* number of lost rules) that can still be mined. Therefore we involve specific optimization problem. In one side we must conceal the sensitive association rule, thus it is necessary to modify the dataset and in the other side we should keep the utility of modified dataset to extracting useful information and rules. Therefore we have selected the genetic algorithm approach to solving this optimization problem.

Problem formulation elements are depicted in Figures 1 to 3 and side effects of sanitization problem is shown in Figure 4.

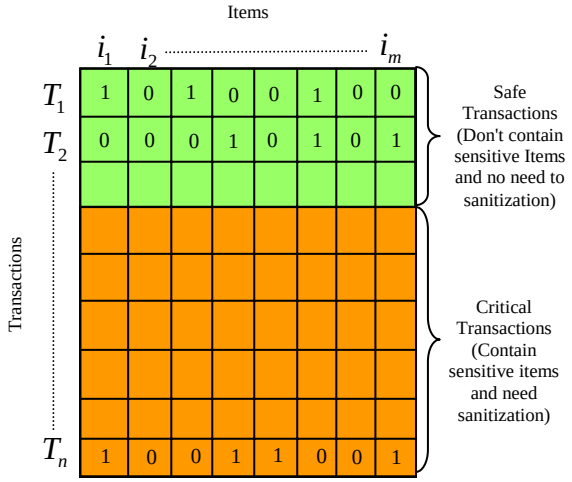

Figure 5. Preprocessing phase of our approach

IV. PROPOSED SOLUTION

In the following section we will explain our approach specifically. The most important parts in this work are preprocessing phase and the specification of our fitness function in Genetic Algorithm method.

A. Preprocess of Original Dataset

Dataset Pre-Sanitization Process (DPSP) is first step of our method that involves preprocess the original database. For the reason that sensitive items are limited to some transactions, therefore there is no need to modify all of the transactions in our algorithms. So our algorithms in preprocessing phase select the all transactions that support sensitive items. With this critical phase of our algorithm we can reach to better performance of sanitization speed and less number of modification needed in hiding process. Further, by preprocessing of original dataset we will see that the size of each chromosome decreases significantly. This phase is depicted in Figure 5.

B. GA Proposed Solution for Privacy Preserving

1) Genetic Algorithm Background

A Genetic Algorithm performs fitness tests on new structures to select the best population. Fitness determines the quality of the individual on the basis of the defined cost function. Genetic Algorithms are meta-heuristic search methods that have been developed by John Holland in 1975. [10,11] GA's applied natural selection and natural genetics in artificial intelligence to find the globally optimal solution to the optimization problem from the feasible solutions. In nature, an individual's fitness is its ability to pass on its genetic material. The fortune of an individual chromosome depends on the fitness value; the better the fitness value, the better the chance of survival. Genetic Algorithms solve design problems similar to that of natural solutions for biological design problems [12].

2) Population Generation and Chromosome Presentation

In Genetic Algorithms, a population consists of a group of individuals called chromosomes that represent a complete solution to a defined problem. Each chromosome is a sequence of 0s or 1s. The initial set of the population is a randomly generated set of individuals. A new population is generated by two methods: steady-state Genetic Algorithm and generational Genetic Algorithm. The steady-state Genetic Algorithm replaces one or two members of the population; whereas the generational Genetic Algorithm replaces all of them at each generation of evolution. In this work a generational Genetic Algorithm is adopted as population replacement method. In this method tried to keep a certain number of the best individuals from each generation and copies them to the new generation (this approach known as elitism).

The each transaction is represented as a chromosome and presence of an i^{th} item in transaction showed by 1 and absence of the item by 0 in i^{th} bit of transaction. The fitness of a chromosome is determined by several factors and different strategies. Each population consists of several chromosomes and the best chromosome is used to generate the next population. For the initial population, a large number of random transactions are chosen. Based on the survival fitness, the population will transform into the future generation.

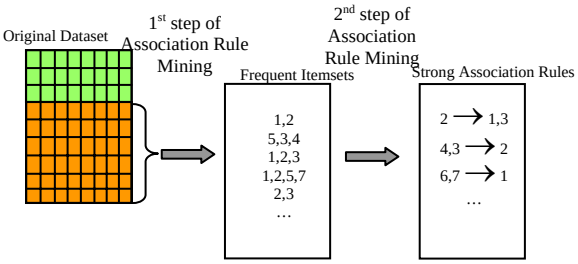

Figure 6. Association Rule Mining Phases

3) Fitness Strategies

Based on our sanitization method, we have conducted four fitness evaluation strategies in this paper. We will discuss these strategies in following sub sections.

a) Confidence-based Fitness Strategy

First fitness strategy relies on both hiding all sensitive rules and minimum number of modification in original dataset. We design this fitness strategy based on weighted sum function as follows:

minimize: $cost\_function\_1 = W_1 \times$ *Rules Hiding Distances* $+ W_2 \times$ *Number of Modifications*

where:

- $W_1 + W_2 = 1$ (where is the necessary condition for weighted sum optimization problem and their values specified based on their costs)
- $Rules\ Hiding\ Distances = \sum_{i=1}^{Number\ of\ sensitive\ Rules} Rule_i\ Hiding\ Distance$
- $Rule_i\ Hiding\ Distance = \begin{cases} 0 & if\ Confidence(Rule_i) \le MCT \\ Confidence(Rule_i) - MCT & otherwise \end{cases}$
- $Number\ of\ Modifications = \sum_{j=1}^{|Critical\ Transactions| \times |I|} D'_j \oplus D_j$

Where: $|Critical\ Transactions|$ is number of critical transactions (in Figures 5 and 6 colored by orange) and $|I|$ is number of items in original database (denoted by *D*). And finally D'_j and D_j are j^{th} item of each dataset after and before sanitization respectively.

Association rule mining process depicted in Figure 6. In this fitness strategy we are trying to filter sensitive rules in 2nd step of mining process. Further, this strategy tried to apply minimum modifications in original dataset.

b) Support-based Fitness Strategy

Second fitness strategy relies on both hiding all sensitive itemsets and minimum number of modification in original dataset. We design this fitness strategy based on weighted sum function as follows:

minimize $cost\_function\_2 = W_1 \times$ *Itemsets Hiding Distances* $+ W_2 \times$ *Number of Modifications*

where:

- $W_1 + W_2 = 1$ (where is the necessary condition for weighted sum optimization problem and their values specified based on their costs)
- $Itemset\ Hiding\ Distances = \sum_{i=1}^{Number\ of\ sensitive\ Itemsets} Itemset_i\ Hiding\ Distance$
- $Itemset_i\ Hiding\ Distance = \begin{cases} 0 & if\ Support(Itemset_i) \le MST \\ Support(Itemset_i) - MST & otherwise \end{cases}$
- $Number\ of\ Modifications = \sum_{j=1}^{|Critical\ Transactions| \times |I|} D'_j \oplus D_j$

Where: $|Critical\ Transactions|$ is number of critical transactions (in Figure 5 colored by orange) and $|I|$ is number of items in original database (denoted by *D*). And finally D'_j and D_j are jth item of each dataset after and before sanitization respectively.

In this fitness strategy we are trying to filter sensitive itemsets in 1st step of mining process (showed in Figure 6). Further, this strategy tried to apply minimum modifications in original dataset.

c) Hybrid Fitness Strategy

Third fitness strategy relies on hiding all sensitive rules and items. Further, minimum number of modification in original dataset is applied. We design this fitness strategy as hybrid of first and second strategies.

$$\text{minimize } cost\_function\_3 = W_1 \times Hiding\ Distances + W_2 \times Number\ of\ Modifications$$

where:

- $W_1 + W_2 = 1$ (where is the necessary condition for weighted sum optimization problem and their values specified based on their costs)
- $Hiding\ Distances = \sum_{i=1}^{Number\ of\ sensitive\ Itemsets\ /\ Rules} Itemset_i\ Hiding\ Distance + Rule_i\ Hiding\ Distance$
- $Itemset_i\ Hiding\ Distance = \begin{cases} 0 & if\ Support(Itemset_i) \le MST \\ Support(Itemset_i) - MST & otherwise \end{cases}$
- $Rule_i\ Hiding\ Distance = \begin{cases} 0 & if\ Confidence(Rule_i) \le MCT \\ Confidence(Rule_i) - MCT & otherwise \end{cases}$
- $Number\ of\ Modifications = \sum_{j=1}^{|Critical\ Transactions| \times |I|} D'_j \oplus D_j$

Where: $|Critical\ Transactions|$ is number of critical transactions (in Figure 5 colored by orange) and $|I|$ is number of items in original database (denoted by *D*). And finally D'_j and D_j are jth item of each dataset after and before sanitization respectively.

In this fitness strategy we are trying to filter sensitive itemsets/rules both in 1st and 2nd steps of mining process (showed in Figure 6). Further, this strategy tried to apply minimum modifications in original dataset.

d) Min-Max Fitness Strategy

Fourth fitness strategy relies on minimizing number of sensitive rules and maximizing number of non-sensitive association rules that can be extracted from sanitized dataset. (See Figures 1 to 4 again). We design this fitness strategy as follows:

$$\text{minimize } cost\_function\_4 = W_1 \times \min(|R' \cap R_{Sen}|) + W_2 \times \max(|R' \cap R_{non-Sen}|)$$

or

$$\text{minimize } cost\_function\_4 = W_1 \times |R' \cap R_{Sen}| - W_2 \times |R' \cap R_{non-Sen}|$$

where:

- $W_1 + W_2 = 1$ (where is the necessary condition for weighted sum optimization problem and their values specified based on their costs)

In this strategy tried to balance hiding all sensitive rules and keeping non-sensitive information.

4) Selection

After evaluation of population's fitness, the next step is chromosome selection. Selection embodies the principle of "survival of the fittest". Satisfied fitness chromosomes are selected for reproduction. Poor chromosomes or lower fitness chromosomes may be selected a few or not at all. There are several selection methods, such as: "Roulette-Wheel" selection, "Rank" selection and "Tournament" selection. In *Tournament* selection, which is used in this paper, two chromosomes are chosen randomly from the population. First, for a predefined probability *p*, the more fit of these two is selected and with the probability *(1-p)* the other chromosome with less fitness is selected [19].

5) Crossover

Main function of crossover operation in Genetic Algorithms is combination two chromosomes together to generating new offspring (child). Crossover occurs only with some probability (crossover probability). Chromosomes are not subjected to crossover remain unmodified. The intuition behind crossover is exploration of new solutions and exploitation of old solutions. Better fitness chromosomes have a prospect to be selected more than the worse ones, so good solution always alive to the next generation. There are different crossover operators that have been developed for various purposes. Single-point crossover and multi-point are the most famous operators. In this paper single-point crossover has been applied to make new offspring. Normally high value of crossover probability is used (between 0.80 and 0.90).

6) Mutation

After performing crossover operation, the new introduced generation will only have the character of the parents. This behavior can lead to a problem where no new genetic material is introduced in the offspring and finding

better population has been stopped. Mutation operator permits new genetic patterns to be introduced in the new chromosomes (random changed in random gene of chromosome). Mutation introduces a new sequence of genes into a chromosome but there is no guarantee that mutation will produce desirable features in the new chromosome. The selection process will keep it if the fitness of the mutated chromosome is better than the general population, otherwise, selection will ensure that the chromosome does not live to mate in future. Same as crossover operator, the mutation rate (mutation probability) is defined to manage how often mutation is applied. Contrasting crossover, the mutation rate is very low, about 0.005 to 0.01.

V. PERFORMANCE EVALUATION

To illustrate our proposed approach for the association rule hiding problem, validation of its feasibility and discussion about sanitization performance, let us consider an example.

Example

In this example we have original dataset and some sensitive association rule (See tables I to III).

TABLE I. ORIGINAL DATASET

T1	1 3 4
T2	1 2 3 5
T3	2 3 5
T4	2 5
T5	1 2 3 6

TABLE II. SENSITIVE RULES

R1	1 → 2
R2	2,5 → 3

TABLE III. ASSOCIATION RULES EXTRACTED FROM ORIGINAL DATASET WITH MCT=0.58 AND MST=0.25

Rule	Confidence	Support
1,2 → 3	1	0.4
3,5 → 2	1	0.4
1 → 2,3	0.66	0.4
1,3 → 2	0.66	0.4
2,3 → 1	0.66	0.4
5 → 2,3	0.66	0.4
2,3 → 5	0.66	0.4
2,5 → 3	0.66	0.4
1 → 3	1	0.6
5 → 2	1	0.6
3 → 1	0.75	0.6
2 → 3	0.75	0.6
3 → 2	0.75	0.6
2 → 5	0.75	0.6
5 → 3	0.66	0.4
1 → 2	0.66	0.4

TABLE IV. GENETIC ALGORITHM PARAMETERS SPECIFICATIONS

Population Size	20
Mutation Rate	0.01
Crossover Probability	0.80
Chromosome Length	30
Number of Generations	50

As we can see in table I, there are five transactions in original dataset and assumed two sensitive rules form sanitization problem. Both rules are strong (Their value of Suppose and Confidence are greater than thresholds). So problem is modification of dataset in order to concealing both rules. We consider four sanitization strategies to solve this problem separately. The specifications of our Genetic Algorithm for privacy preserving in association rule mining is depicted in table IV.

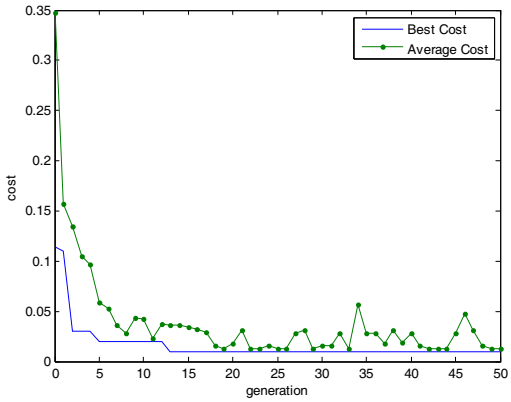

Figure 7. Confidence-Based fitness strategy with respect to generations

TABLE V. SANITIZED DATASET IN CONFIDENCE-BASED FITNESS STRATEGY

T1	1 3 4
T2	1 3 5
T3	2 3 5
T4	2 5
T5	1 2 3 6

As we can see in figure 7, there is a trend of convergence into finding best solution in according to confidence of association rules. The approach here is finding the solution which satisfies concealing rules and minimum modification needed. By comparing sanitized solution in table V and original dataset in table I, we will find out that modification applied in T2 and in item 2 (i.e. item 2 has been eliminated). In this solution only one modification needed and after sanitization both of rules will not extracted again.

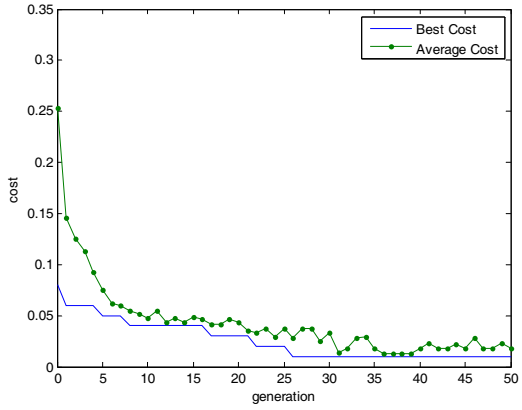

Figure 8. Support-Based fitness strategy with respect to generations

TABLE VI. SANITIZED DATASET IN SUPPORT-BASED FITNESS STRATEGY

T1	1 3 4
T2	1 2 5
T3	2 3 5
T4	2 5
T5	1 2 3 6

As we can see in figure 8, there is a trend of convergence into finding best solution in according to confidence of association rules. The approach here is finding the solution which satisfies concealing rules and minimum modification needed. By comparing sanitized solution in table VI and original dataset in table I we will find out that modification applied in T2 and in item 3 (i.e. item 3 has been eliminated). In this solution only one modification needed and after sanitization both of rules will not extracted again.

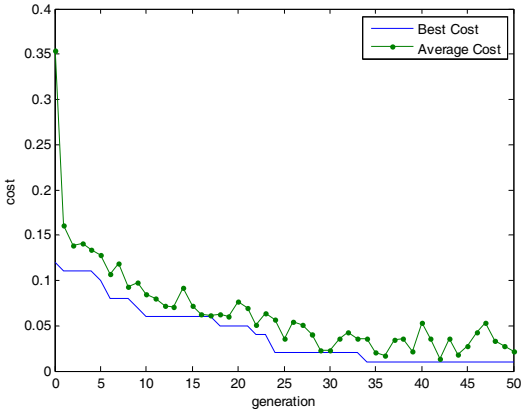

Figure 9. Hybrid fitness strategy with respect to generations

TABLE VII. SANITIZED DATASET IN HYBRID FITNESS STRATEGY

T1	1 3 4
T2	1 3 5
T3	2 3 5
T4	2 5
T5	1 2 3 6

As we can see in figure 9, there is a trend of convergence into finding best solution in according to confidence of association rules. The approach here is finding the solution which satisfies concealing rules and minimum modification needed. By comparing sanitized solution in table VII and original dataset in table I we will find out that modification applied in T2 and in item 2 (i.e. item 2 has been eliminated). In this solution only one modification needed and after sanitization both of rules will not extracted again.

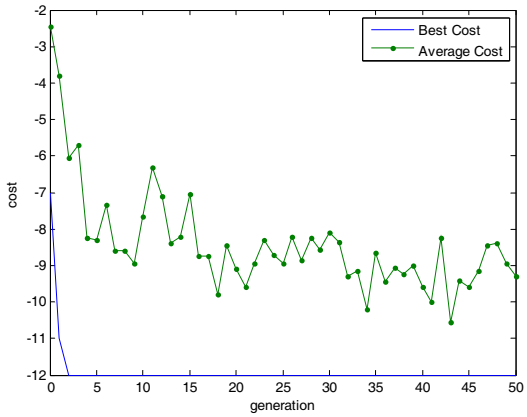

Figure 10. Min-Max fitness strategy with respect to generations

TABLE VIII. SANITIZED DATASET IN MIN-MAX FITNESS STRATEGY

T1	1 3 4
T2	1 2 3 5
T3	2 5 4 6
T4	1 2 3 5 6
T5	2 3 4

As we can see in figure 10, there is a trend of convergence into finding best solution in according to confidence of association rules. The approach here is finding the solution which satisfies concealing rules and minimum modification needed. By comparing sanitized solution in table VIII and original dataset in table I we will find out that modification applied in T3, T4 and T5 (i.e. some real items are eliminated and number of unreal items added into the transactions). The main criterions in this solution are number of sensitive rules and number of non-sensitive rules that can be mined from sanitized dataset. In this strategy tried to maximizing number of non-sensitive rules and minimizing number of sensitive rules.

VI. CONCLUSIONS

In this paper, we have introduced a novel method for concealing sensitive association rules. Our offerings in this paper can be summarized as follows: First, a pre-sanitization process called Dataset Pre-Sanitization Process (DPSP). DPSP select which transaction(s) and which item(s) in each transaction should be changed in order to all association rules concealed safely and minimum side effect accrues. Second, four sanitization strategies proposed that these strategies are the hearth of our approach. Different criterion introduced in these sanitization strategies. Some of them are: number of modifications (in original dataset), hiding distances both in frequent itemsets and association rules (in sanitized dataset), number of sensitive rules (in sanitized dataset), number of non-sensitive rules (in sanitized dataset), number of lost rules (in sanitized dataset) and number of ghost rules (in sanitized dataset). The work presented here introduces the idea of both rule and itemset sanitization, which complements the old idea behind data sanitization. At present, we are looking for new aspects of sanitization and proposing new fitness functions according to new types of sanitization. Our permanent goal in this area is keeping privacy and accuracy of dataset as more as possible.

REFERENCES

[1] M. Atallah, E. Bertino, A. Elmagarmid,M. Ibrahim and V. Verykios. Disclosure limitation of sensitive rules. *Proc. of IEEE Knowledge and Data Engineering Exchange Workshop (KDEX)*, November 1999.

[2] Y. Saygin, V. Verykios and C. Clifton. Using unknowns to prevent discovery of association rules. *ACM SIGMOD Record*, vol. 30, no. 4, 2001.

[3] V. Verykios, A. Elmagarmid, E. Bertino, Y. Saygin and E. Dasseni. Association Rule Hiding. *IEEE Trans. on Knowledge and Data Engineering*, 16(4), 2004.

[4] Y. Saygin, V. Verykios and A. Elmagarmid. Privacy preserving association rule mining. *Proc. of 12th Intl. Workshop on Research Issues in Data Engineering (RIDE)*, February 2002.

[5] L. Chang and I. S. Moskowitz. Parsimonious downgrading and decision trees applied to the inference problem. In *Workshop on New Security Paradigms*, 1998.

[6] T. Johnsten, V. Raghavan, K. Hill: The security assessment of association mining algorithms. In: *Proceedings of the 16th Annual IFIP WG* 11.3 Working Conference on Database Applications Security, pp. 163–174 ,2002.

[7] E. Dasseni,V.S. Verykios, A. Elmagarmid, and E. Bertino. Hiding association rules by using confidence and support. In: *Proc. of the 4th Int'l Information Hiding Workshop (IHW'01)*. Springer-Verlag, 2001. 369-383.

[8] E.D. Pontikakis, A. Tsitsonis, and V.S. Verykios. An experimental study of distortion-based techniques for association rule hiding. In *Proc. of the 18th Annual IFIP WG 11.3 Working Conf. on Data and Applications Security*. 2004.

[9] X. Sun, and P.S. Yu, A border-based approach for hiding sensitive frequent itemsets. In: *Proc. of the 5th* P *IEEE Int'l Conf. on Data Mining (ICDM'05)*. IEEE Computer Society, 2005. 426-433.

[10] L. David, *Handbook of Genetic Algorithms***.** New York: Van Nostrand Reinhold. 1991.

[11] D.E. Goldberg, *Genetic Algorithms: in Search, Optimization, and Machine Learning***.** New York : Addison-Wesley Publishing Co. Inc. 1989.

[12] D. Goldberg, B. Karp, Y. Ke, S. Nath, and S. Seshan, *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, 1989.

[13] I.S. Moskowitz, L. Chang: A Computational Intelligence Approach to the database Inference Problem. IOS, Amsterdam, 2000.

[14] Y.-H.Wu, C.-M. Chiang, and A. L. P. Chen. Hiding sensitive association rules with limited side effects. *IEEE Transactions on Knowledge and Data Engineering*, 19(1):29–42, 2007.

[15] J. Natwichai, X. Li, and M. Orlowska. A reconstruction-based algorithm for classification rules hiding. In *Proceedings of the 17th Australasian Database Conference (ADC 2006)*, pages 49–58, 2006.