An Optimization based Modified Maximum Sensitive Item-Sets Conflict First Algorithm (MSICF) for Hiding Sensitive Item-Sets

D.Jaya Kumari
Assoc.Professor,
Department of IT,
RRS College of Engg. & Tech.
JNTUH, Hyderabad-502300, India

Prof. Nistala.V.E.S.Murthy
Department of CS & SE
Andhra University
Vishakapatnam-530003, India

Prof.S.Srinivasa Suresh
Department of AS & CT
I^{2}IT, Pune – 411057, India

ABSTRACT

In privacy preserving data mining, utility mining plays an important role. In privacy preserving utility mining, some sensitive itemsets are hidden from the database according to certain privacy policies. Hiding sensitive itemsets from the adversaries is becoming an important issue nowadays. Also, only very few methods are available in the literature to hide the sensitive itemsets in the database. The existing paper utilized two algorithms; such as HHUIF and MSICF are conceal the sensitive itemsets, so that the adversaries cannot mine them from the modified database. To accomplish the hiding process, this method finds the sensitive itemsets and modifies the frequency of the high valued utility items. But, the performance of this method lacks if the utility value of the items are same. To solve this problem, in this paper a modified MSICF algorithm with Item Selector (MMIS) is proposed. The MMIS algorithm computes the sensitive itemsets by utilizing the user defined utility threshold value. The proposed MMIS reduces the computation complexity as well as improves the hiding performance of the itemsets. The algorithm is implemented and the resultant itemsets are compared against the itemsets that are obtained from the conventional privacy preserving utility mining algorithms.

General Terms

Data Mining, Privacy

Keywords

Utility Mining, Privacy Preserving Utility Mining, Sensitive Itemsets, Utility Value, Frequency Value, Maximum Sensitive Itemsets Conflict First (MSICF)

1. INTRODUCTION

The collection of digital information by governments, corporations, and individuals has created tremendous opportunities for knowledge-based decision making. Driven by mutual benefits, or by regulations that require certain data to be published, there is a demand for the exchange and publication of data among various parties [1].Some of the organization needs privacy for the original data. So recently all the organizations are utilizing the Privacy Preserving Utility Mining (PPUM) for the security purpose. Many data mining applications deal with privacy- sensitive data. It is randomly perturbing the data while preserving the underlying probabilistic properties [2]. Privacy is usually measured using some form of disclosure risk, while the data utility is traditionally measured as information loss between the original data and the transformed sanitized data [3].

The problem of privacy preserving data mining has become more important in recent years because of the increasing ability to store personal data about users and the increasing sophistication of data mining algorithm to leverage this information. A number of techniques have been suggested in recent years in order to perform privacy preserving data mining [4].PPUM research usually takes one of the three philosophical approaches: (i) Data hiding (ii) Rule hiding (iii) Secure Multiparty Computation. Its main goal is to develop efficient algorithm that allow one to extract relevant knowledge from a large amount of data, while prevent sensitive data and information from disclosure or inference [5]. Moreover, some of those algorithms can be computationally very expensive and thus cannot be used when very large sets of data need to be frequently released. Therefore, in addition to data quality, performance also needs to be carefully assessed [6]. The first type of privacy, termed as output privacy, is that the data is altered so that the mining result will conserve certain privacy. Many modification techniques such as perturbation, blocking, aggregation, swapping and sampling are used for this type of privacy [7] [8]. The second type of privacy, labeled as input privacy, is that the data is manipulated so that the mining result is not affected or less affected. The cryptography based and reconstruction based techniques are used for this type of privacy [9] [10].

2. RELATED WORKS

In 2009, Mohammad Naderi Dehkordi *et al.* [11] have presented the Extracting of knowledge form large amount of data was an important issue in data mining systems. One of most important activities in data mining was association rule mining and the new head for data mining research area was privacy of mining. A lot of researches have done in the 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 was privacy of database that is not warranted in most heuristic approaches, perfectly. They introduced new multi-objective method for hiding sensitive association rules based on the concept of genetic algorithms. The main purpose of the method was fully supporting security of database and keeping this utility and certainty of mined rules at highest level.

In 2011, Vijayarani *et al.* [12] have discussed about the association rule hiding problem. Association rule mining, one of the very important data mining techniques. The process of discovering itemsets that frequently co-occur in a transactional database so as to produce significant association rules that hold for the data was known as Association rule mining. This process was modifying the original database by hiding the sensitive data to protect the sensitive association rules. In the paper, they have proposed Artificial Bee Colony optimization algorithm for hiding the sensitive association rules. They analyzed the efficiency of the Artificial Bee Colony optimization technique by using various performance factors.

In 2010, Nissim Matatov *et al.* [13] have proposed a different approach for achieving k-anonymity by partitioning the original dataset into several projections such that each one of them adheres to k-anonymity. Moreover, any attempt to rejoin the projections, results in a table that still complies with k-anonymity. A classifier was trained on each projection and subsequently, an unlabelled instance was classified by combining the classifications of all classifiers. Guided by classification accuracy and k-anonymity constraints, the proposed data mining privacy by decomposition (DMPD) algorithm uses a genetic algorithm to search for optimal feature set partitioning. Ten separate datasets were evaluated with DMPD in order to compare to that classification performance with other k-anonymity-based methods. The results show that DMPD performs better than existing k-anonymity-based algorithms and there was no necessity for applying domain dependent knowledge.

In 2012, Ziauddin *et al.* [14] have presented the scope of Association Rule Mining and KDD was very broad. Over the last fifteen years it has been developed at a dynamic rate. Although it has been emerged as a new technology but Association Rule Mining was still in a stage of exploration and development. In the paper they presented a survey of research work carried by different researchers since has beginning. Of course, a single article cannot be a complete review of all the research work, yet we hope that it would provide a guideline for the researcher in interesting research directions that have yet to be explored.

In 2012, Guillermo Navarro-Arribas *et al.* [15] have proposed the anonymization of query logs was an important process that needs to be performed prior to the publication of such sensitive data. It ensures the anonymity of the users in the logs, a problem that has been already found in released logs from well known companies. In the paper presented the anonymization of query logs using micro aggregation. This technique ensures the k-anonymity of the users in the query log, while preserving its utility. They provide the evaluation of this proposal in real query logs, showing the privacy and utility achieved, as well as providing estimations for the use of such data in data mining processes based on clustering.

In 2009, Keke Chen *et al.* [16] have proposed an approach based on geometric data perturbation and data-mining-service oriented framework. The key problem of applying geometric data perturbation in multiparty collaborative mining was securely unifying multiple geometric perturbations that were preferred by different parties, respectively. They have developed three protocols for perturbation unification. Thisapproach has three unique features compared to the existing approaches. (1) With geometric data perturbation, these protocols could work for many existing popular data mining algorithms, while most of other approaches were only designed for a particular mining algorithm. (2) Both the two major factors: data utility and privacy guarantee were well preserved, compared to other perturbation-based approaches. (3) Two of the three proposed protocols also have great scalability in terms of the number of participants, while many obtainable cryptographic approaches consider only two or a few more participants.

In 2010, Jieh-Shan Yeh *et al.* [17] have proposed the privacy preserving utility mining (PPUM) with two novel algorithms, HHUIF and MSICF, to achieve the goal of hiding sensitive itemsets so that the adversaries could not mine them from the modified database. The work also minimizes the impact on the sanitized database of hiding sensitive itemsets. The experimental results show that HHUIF achieves lower miss costs than MSICF on two synthetic datasets. On the other hand, MSICF generally has a lower difference ratio than HHUIF between original and sanitized databases.

3. The Proposed Modified Privacy Preserving Utility Mining Algorithm

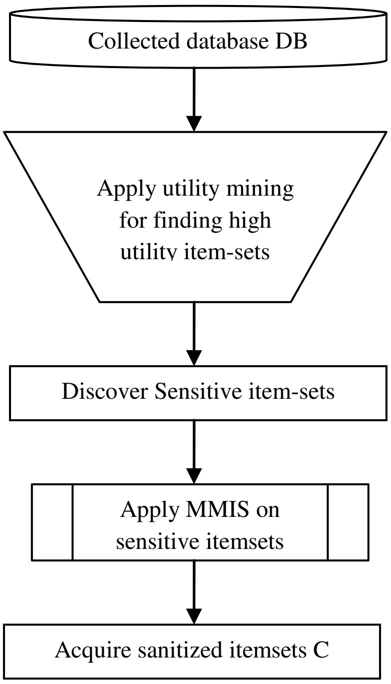
Collected database DB
Apply utility mining for finding high utility item-sets
Discover Sensitive item-sets
Apply MMIS on sensitive itemsets
Acquire sanitized itemsets C

Figure 1: Structure of the Proposed MMIS algorithm

Let C be the transaction database, containing a set of transactions $C=\{T_1,T_2,T_3,\cdots T_Y\}$, where y is the total number of transactions. The database D contains a set of items, which is denoted as $J=\{i_1,i_2,i_3,\cdots i_X\}$, where x is the total number of items in the database. Each transaction T_y is a set of items and a set of items is termed as an itemset. Moreover, the external utility value of each item in the database is stored in the external utility table, which is referred as, $E=\{e(i_1),e(i_2),e(i_3),\cdots e(i_X)\}$ where $e(i_X)$ is the external utility value of an item i_X ; $i_X \subseteq J$ the frequency value of each item i_X in transaction T_Y is $v(i_X,T_Y)$ is the number of items i_m acquired in transaction T_Y .

3.1 Utility Mining Algorithm

Utility mining is used to find all the itemset's utility values. The utility value of item i_X in transaction T_Y is defined as,

$$a(i_X, T_Y) = e(i_Y) \times b(i_X, T_Y) \qquad (1)$$

The utility value of an itemset P in transaction T_Y is denoted as,

$$a(P, T_Y) = \sum_{i_X \in P} a(i_X, T_Y) \qquad (2)$$

Then, find the itemsets whose utility value is higher than the user specified threshold value ε , where, the minimum utility threshold is ε . The itemset P is a high utility itemset, if $a(P) \geq \varepsilon$. These high utility itemsets are stored in $K = \{s_1, s_2, \cdots s_m\}$ and such itemsets are sensitive itemsets. The sensitive itemsets should be concealed according to some security strategies. To perform the sanitizing process, the existing method [14] has utilized two algorithms: HHUIF and MSICF. Amongst these two algorithms, MSICF produces lower DS than the HHUIF and the MSICF algorithm is described below.

3.2 Obtainable MSICF Algorithm

The main objective of the MSICF algorithm is to diminish the utility value of each sensitive itemset by modifying the quantity values of items which has the maximum conflict count among items in the sensitive itemsets. The pseudo code of the MSICF algorithm is given below:

Algorithm 1: MSICF

Input: The original database C ; the minimum utility threshold ε ; the sensitive itemsets $K = \{s_1, s_2, \cdots s_m\}$

Output: The sanitized database C' so that s_m cannot be mined

```
1. Calculate $Icount_{i_m}(K)$ for all S
2. Arrange $i_m$ by decreasing order of $Icount_{i_m}(K)$
3. for each sensitive itemset $s_m \in C$
4.   $diff = a(s_m) - \varepsilon$ // the utility value needs to be reduced
5.   while $(diff > 0)$ {
6.     $o(i_X, T_Y) = \arg\max_{(i \in s_m, s_m \subseteq T)} (a(i,T))$
7.     modify $o(i_X, T_Y)$ with
         $o(i_X, TY) = \begin{cases} 0 & ,if \;\; a(i_X,T_Y) < diff \\ o(i_X,T_Y) - \left\lceil \frac{diff}{s(i_X)} \right\rceil & ,if \;\; a(i_X,T_Y) > diff \end{cases}$
8.     $diff = \begin{cases} diff - a(i_X, T_Y) & ,if \;\; a(i_X,T_Y) < diff \\ 0, & ,if \;\; a(i_X,T_Y) > diff \end{cases}$
     }
9. return the sanitized database $C'$
```

This MSICF process continues until the utility value of each sensitive itemset becomes lower than ε . This existing MSICF privacy preserving utility mining algorithm has some drawbacks in the hiding process, and such drawback is formulated in the following section.

3.3 Problem Formulation

The MSICF algorithm hides the sensitive itemsets having high utility value. But, the drawback of this algorithm is that, if the items in the sensitive itemsets having same utility value, then it will decrease the hiding performance. For example, $\{A, B\}$ is a sensitive itemset $(\varepsilon = 120)$, having utility value $u(A, B) = 200$. To hide the itemset $\{A, B\}$, here the frequency value of item in the itemset having high utility value is changed. In case, if both A, B have the same utility value as 100 then, any one of the item's value is modified randomly. Hence, this process creates an impact between these items. To solve this drawback, in this paper a Modified MSICF algorithm with Item Selector (IS) (MMIS) is proposed. The Item Selector is used to select the high utility value itemset by using the following algorithm, and subsequently the frequency value of the selected items is modified. The developed IS will reduce the computation complexity as well as improves the hiding performance of the itemsets.

4. Modified MSICF Algorithm with Item Selector (MMIS)

The main objective of the MMIS algorithm is to select the best items from the sensitive itemsets having same high utility value. The high utility value itemsets are hidden by modifying the frequency values of items contained in the sensitive itemsets based on the minimum utility threshold value ε . This hiding process is repeated until all the sensitive itemsets utility value become lower than the threshold value ε .The proposed MMIS algorithm is described below.

Algorithm 2: Modified MSICF

Input: the original database C ; the minimum utility threshold ε ; the sensitive itemsets $K = \{s_1, s_2, \cdots s_m\}$

Output: the sanitized database C' so that s_m cannot be mined. Finding threshold value

$$\varepsilon = \alpha \sum_{n=0}^{X} \sum_{m=0}^{Y} (T_n(m) * T_n(m)')\beta$$

The aforementioned threshold computation process is done with the aid of hybrid technique of artificial bee colony (ABC) and genetic algorithm (GA),

```
1. Calculate $Icount_{i_m}(K)$ for all S
2. Arrange $i_m$ by decreasing order of $Icount_{i_m}(K)$
3. for each sensitive itemset $s_m \in K$
```

```
4.  $diff = a(s_m) - \varepsilon$ // the utility value needs to be reduced
5.  while $(\text{diff} > 0)$ {
6.      if $s_m$ contains two items $s_m \subseteq (i_Z, i_X)$
7.      Compare $a(i_Z, T_Y)$ and $a(i_X, T_Y)$
8.      if $a(i_X, T_Y) = a(i_Z, T_Y)$ go to step 9 otherwise go to step 17
9.      Select $i_Z, i_X$ items frequency values $b(i_X, T_Y)$ and $b(i_Z, T_Y)$
10.     Sort $b(i_X, T_Y)$ and $b(i_X, T_Y)$ frequency values in descending order and sored in $S_X$ and $S_Z$
11.     Select top $n^{b(i_X,T_Y)}, n^{b(i_Z,T_Y)}$ values from $S_X$ and $S_Z$
12.     Compute frequency $f^{n^{b(i_X,T_Y)}}, f^{n^{b(i_Z,T_Y)}}$ value for each $n^{b(i_X,T_Y)}, n^{b(i_Z,T_Y)}$ value
13.     compute $\gamma^{i_X}, \gamma^{i_Z}$
            $\gamma^{i_X} = \sum_{n=1}^{l} n^{b(i_X,T_Y)} * f^{n^{b(i_X,T_Y)}}$
            $\gamma^{i_Z} = \sum_{j=1}^{p} n^{b(i_Z,T_Y)} * f^{n^{b(i_Z,T_Y)}}$
14.     If $\gamma^{i_X} \geq \gamma^{i_Z}$ then change $b(i_X, T_Y)$ otherwise change $b(i_Z, T_Y)$
15.     $o(i_X, T_Y) = \max_{(i_X \in s_m, T_Y \in n^{b(i_X,T_Y)})} (a(i,T))$
16.     modify $o(i_X, T_Y)$ with
            $o(i_X, T_Y) = \begin{cases} 0 & , if \;\; a(i_X, T_Y) < diff \\ o(i_X, T_Y) - \left\lceil \frac{diff^2}{s(i_X) * \varepsilon} \right\rceil & , if \;\; a(i_X, T_Y) > diff \end{cases}$
17.     $o((i_X, T_Y), (i_Z, T_Y)) = \max_{(i \in s_m, s_m \subseteq T)} (a(i,T))$
        repeat again 16.
18.     $diff = \begin{cases} diff - a(i_X, T_Y) & , if \;\; a(i_X, T_Y) < diff \\ 0, & , if \;\; a(i_X, T_Y) > diff \end{cases}$
19. return the sanitized database $C'$
```

In this threshold value formula, where
$T_n(m)$ - Transaction value,
$T_n(m)'$ - Utility value, α, β - Weight age value

The proposed algorithm shows that the item-set s_m Contains two items and their utility values are found and check for higher utility value. If suppose both items utility values are same, then find frequency value for each item-sets otherwise it will move to the final condition modified step. The same utility items frequency values are sorted in descending order and top most value is selected. The top most items frequency values are determined and two parametric values are found by multiplying the top most items frequency with top most items value. After that, compare both parametric values based on the item chosen and change that item frequency value.

The proposed MMIS algorithm hides the sensitive item-sets having high utility values, so the adversaries cannot mine such sensitive item-sets from the database.

5. EXPERIMENTAL RESULTS

The proposed MMIS algorithm is implemented in the working platform of MATLAB version 7.12. The performance of the proposed MMIS algorithm is measured by conducting experiments on one dataset. In the dataset, the proposed MMIS algorithm finds the sensitive item-sets that have high utility than the specified minimum utility threshold value. The sensitive item-sets are mined from the dataset and the corresponding item-sets items utility value is changed by utilizing IS.

5.1 Dataset Description

In this paper, only one dataset is utilized for the performance analysis of proposed MMIS algorithm. The Dataset I contains 200 transactions with 10 different items. Dataset is described in Table I.

Table I: Dataset Description

Dataset	Number of transactions	Distinct items
Dataset I	200	10

5.2 Performance Analysis

The effectiveness of proposed technique is analyzed by invoking some performance measures given in [18]. Moreover, the proposed MMIS algorithm performance is compared with the conventional MSICF algorithm. The performance analysis is carried out by changing the minimum utility threshold as 1000, 1500, 2000, 2500 and 3000. The performance measures of the proposed and conventional algorithms are shown in the following Table II. The performance measures are described below,

(i) Hiding Failure (HF):

Hiding failure measures the percentage of sensitive itemsets discovered from D. The HF is measured by the sensitive itemsets of both the original database and the sanitized database, which is stated as follows,

$$HF = \frac{|H(C')|}{|H(C)|} \qquad (3)$$

In Eqn. (3), $H(C)$ and $H(C^{'})$ represents the sensitive itemsets from original database C and the sensitive itemsets from sanitized database $C^{'}$, respectively.

(ii) Miss Cost (MC):

Miss cost measures the difference ratio of valid itemsets presented in the original database and the sanitized database. The Miss Cost value is computed as,

$$MC = \frac{|\,\delta(C) - \delta(C^{'})\,|}{|\,\delta(C)\,|} \qquad (4)$$

Where, $\delta(C)$ and $\delta(C^{'})$ denotes the non-sensitive itemsets discovered from the original database C and the sanitized database $C^{'}$, respectively.

(iii) Dissimilarity (Diff):

The dissimilarity between the original database D and the sanitized database $D^{'}$ is calculated as,

$$DS = \frac{1}{\sum_{m=1}^{X} \varphi_C(m)} \left(\sum_{m=1}^{X} [\varphi_C(m) - \varphi_C^{'}(m)] \right) \qquad (5)$$

Where, $\varphi_C(m)$ and $\varphi_{C^{'}}(m)$ represents the frequency of the m^{th} item in the database C and the frequency of the m^{th} item in the database $C^{'}$.

As can be seen from Table II, the performance measure shows that the proposed algorithm has offered higher performance compared to the conventional algorithm. The hiding failure value of MMIS algorithm is lower than the conventional MSICF algorithm. The low value of HF shows that the proposed technique hides the sensitive items more efficiently than the conventional MSICF algorithm [17]. Similarly, the dissimilarity values of the proposed MMIS algorithm are also lower than the conventional MSICF algorithm. But Miss Costs of the proposed system are higher when compared to conventional MSICF algorithm.

The following figures 2, 3 and 4 shows the graphical representation of the proposed and conventional techniques performance in HF, MC and DS performance measures for different minimum threshold values.

Data set	Threshold value (ε)	MSICF [17]			MMIS		
		HF	MC	DS	HF	MC	DS
Data Set I	1000	0.810219	3.080292	1.61E-116	0.510949	3.270073	2.43E-214
	1500	0.633803	6.873239	3.45E-50	0.464789	7.239437	6.11E-106
	2000	0.071429	18.96429	3.04E-08	0.071429	19.89286	1.64E-12
	2500	0	20.5	1.55E-07	0	21.5	1.56E-09
	3000	0.038462	20.53846	1.50E-07	0	21.5	7.57E-08

Table II: Performance comparison between proposed MMIS algorithm and Conventional MSICF algorithm

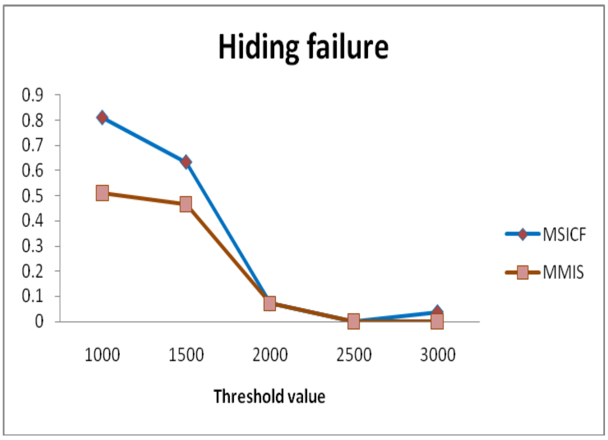

Figure 2: Graphical representation of proposed MMIS and existing MSICF algorithms performance in terms of Hiding Failure (HF)

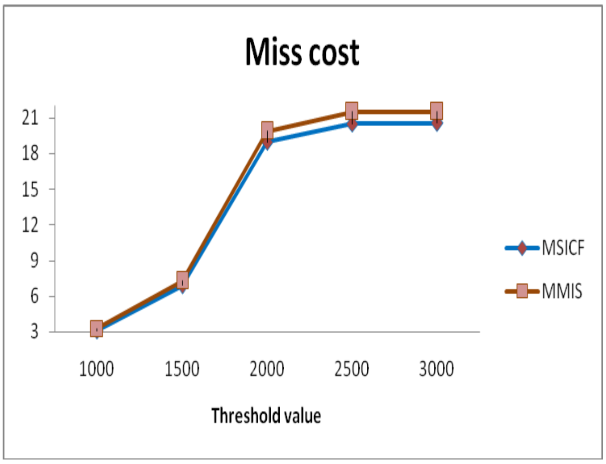

Figure 3: Graphical representation of proposed MMIS and existing MSICF algorithms performance in terms of Miss Cost (MC)

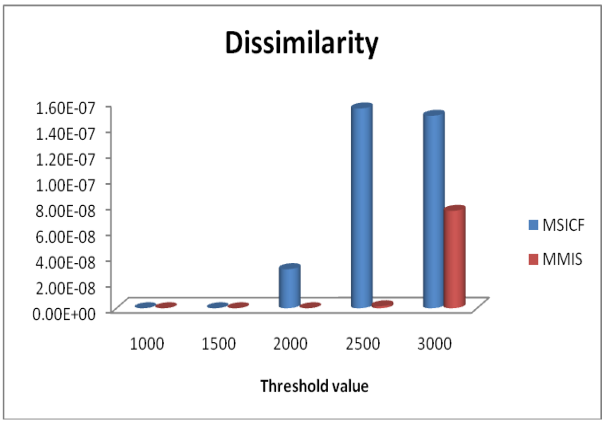

Figure 4: Graphical representation of proposed MMIS and existing MSICF algorithms performance in terms of DS

The fig. 2, 3, and 4 are shows the performance of MMIS and MSICF algorithms in different minimum utility threshold values with different performance measures. The performance measure HF value of the proposed technique is low when compared to MSICF. This low value illustrates that the MMIS algorithm performs the sanitization process perfectly than the MSICF. Moreover, the high MC value shows that the sanitization database contains more valid items than the original database. The MC value is low for the MSICF algorithm. Also, the DS measure shows that the MMIS algorithm removes the sensitive items and its corresponding sensitive transactions. As it is known from these graphs, the proposed technique has offered high performance in different minimum utility threshold values with different performance measures. Thus, the proposed MMIS algorithm efficiently hides the sensitive itemsets from the original database and provides a database with the non sensitive itemsets.

6. CONCLUSION

In this proposed technique, MMIS privacy preserving utility mining algorithm for hiding the high utility sensitive item sets by utilizing exploiting the Item Selector (IS). The enhanced MMIS algorithm successfully hides the sensitive item sets from the adversaries even though the items utility value is similar or non similar. Initially the proposed technique presents a privacy preserving utility mining (PPUM) model and builds up an MMIS algorithm to reduce the impact on the source database of privacy preserving utility mining. This algorithm modifies the database transactions containing sensitive item sets to minimize the utility value below the given threshold while preventing reconstruction of the original database from the sanitized one. The experimental results proved that the performance of the proposed MMIS algorithm was better than the conventional MSICF algorithm. In future, by making small modifications in computing threshold process or by changing the optimization algorithm these results can be improved. This in case reduces the computation time taken for the whole process and retrieve better results.

7. ACKNOWLEDGMENTS

The author would like to express her heart full thanks to the Prof. N.V.E.S.Murthy, Department of Computer Science and Systems Engineering, AUCE, Andhra University, for his constant encouragement and help throughout the preparation of this paper.

8 .REFERENCES

[1] Benjamin C. M. Fung, Ke Wang, Rui Chen And Philip S. Yu, “Privacy-Preserving Data Publishing: A Survey of Recent Developments”, ACM Computing Surveys, Vol. 42, No. 4, Article 14, pp. 1-53, June 2010.

[2] Hillol Kargupta, Souptik Datta, Qi wang and Krishnamoorthy Sivakumar, “Random Data Perturbation Techniques and A Privacy Preserving Data Mining”, in proceedings of IEEE International Conference on Data Mining, pp. 1-23, 2003

[3] Michal Sramka, "Data Mining as a tool in privacy preserving data publishing", Tatra M t. Math. Publications, Vol.45, pp. 151–159, 2010

[4] Md. Riyazuddin, Dr.V.V.S.S.S.Balaram, Md.Afroze, Md.JaffarSadiq and M.D.Zuber, “An Empirical Study on Privacy Preserving Data Mining”, International Journal of Engineering Trends and Technology, Vol.3, No.6, pp.687-693, 2012

[5] Xiaodan Wu, Chao-HsienChu, Yunfeng Wang, Fengli Liu and Dianmin Yue, "Privacy preserving data mining research: current status and key issues", Springer -Lncs, pp. 762-772, 2007

[6] Elisa Bertino, Igor Nai Fovino and Loredana Parasiliti Provenza, “A Framework for Evaluating Privacy Preserving Data Mining Algorithms”, Data Mining and Knowledge Discovery, Vol.11, pp. 121–154, 2005

[7] C. Clifton, “Using Sample Size to Limit Exposure to Data Mining”, Journal of Computer Security, Vol. 8, pp. 281- 307, Dec. 2000.

[8] Y. Saygin, V.S. Verykios, C. Clifton, “Using Unknowns to Prevent Discovery of Association Rules”, SIGMOD Record, Vol. 30, pp. 45- 54, Dec. 2001

[9] A. Evfimievski, “Randomization in Privacy Preserving Data Mining”, in Proceedings of the SIGKDD Explorations, Vol. 4, pp. 43- 48, Dec. 2002

[10] M. Kantarcioglu, C. Clifton, “Privacy-preserving distributed mining of association rules on horizontally partitioned data”, ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Jun. 2002

[11] Mohammad Naderi Dehkordi, Kambiz Badie and Ahmad Khadem Zadeh, “A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms”, Journal of Software, Vol. 4, No. 6, pp. 555-562, August 2009

[12] M.Sathiya Prabha and S.Vijayarani, “Association Rule Hiding using Artificial Bee Colony Algorithm”, International Journal of Computer Applications, Vol. 33, No.2, pp. 41-47, November 2011

[13] Nissim Matatov, Lior Rokach and Oded Maimon, “Privacy preserving data mining: A feature set partitioning approach”, Information Sciences, Vol.180, pp. 2696–2720, 2010

[14] Ziauddin, Shahid Kammal, Khaiuz Zaman Khan, Muhammad Ijaz Khan, “Research on Association Rule Mining”, Advances in Computational Mathematics and its Applications (ACMA), Vol. 2, No. 1, pp. 226-236, 2012

[15] Guillermo Navarro-Arribas, Vicenç Torra, Arnau Erola and Jordi Castellà-Roca, “User k-anonymity for privacy preserving data mining of query logs”, Information Processing and Management 48, pp. 476–487,2012

[16] Keke Chen and Ling Liu, “Privacy preserving Multiparty Collaborative Minin with Geometric Data Perturbation”, IEEE Transactions on Parallel and Distributed Computing, Vol. Xx, No. Xx, pp. 1-13, January 2009

[17] Jieh-Shan Yeh and Po-Chiang Hsu, “HHUIF and MSICF: Novel algorithms for privacy preserving utility mining”, Expert Systems with Applications, Vol. 37, pp. 4779–4786, 2010

[18] Stanley R. M. Oliveira and Osmar R. Zaiane, "Privacy Preserving Frequent Itemset Mining", In Proceedings of the IEEE international conference on Privacy, security and data mining, Vol. 14, pp. 43-54, 2002