

## REDUCING SIDE EFFECTS OF ANT BASED ORTHOGONAL MULTIPLICATIVE AND TRANSFORMATIONAL ALGORITHM WITH SENSITIVE DATA ITEMS

*P. Tamil Selvan<sup>1</sup>, Dr. S. Veni<sup>2</sup>*

### ABSTRACT

Privacy preserving data mining (PPDM) has become a more demanding issue in resolving the effects of privatizing user's data. Most of the PPDM technique adopting, sensitive item hiding changes the originality of the dataset and were designed to partially evaluate the side effects. Perturbation of sensitive item set is, however, not considered in the evaluation process, thus raising the probability of artificial item sets. In this work, we plan to develop an Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique and expand the scope of quality privacy preservation for distributed data mining with optimal side effects on the original dataset. The sensitive item hiding is performed through multiplicative and transformational data perturbation. This data perturbation is based on socially cohesive relational rate between sensitive and non sensitive item sets, ensuring rate of side effects. The side effects on the modified dataset are checked for several users' requested item set distribution. Experiments are then conducted to show the performance of the proposed technique in rate of side effects.

**Keywords:** Privacy Preserving Data Mining, Perturbation, Sensitive Item Hiding, Social Ant,

Multiplicative Data Perturbation, Transformational Data Perturbation

### I. INTRODUCTION

In this paper, an Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is proposed for distributed data mining to obtain quality privacy preservation with optimal side effects on the original dataset. This is performed using user operational conditions-based sensitive items, social ant-based relative item set distribution and Ant-based Orthogonal Multiplicative and Transformational algorithm.

### II. OPTIMIZED SOCIAL ANT BASED SENSITIVE ITEM HIDING

Some applications require protection against the disclosure of private, confidential, or secure data. In this section, an efficient technique called Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) for data publishing is designed with the objective of improving the rate of side effects on the modified dataset at relatively lesser amount of time. The elaborate design of OSA-SIH technique is given below.

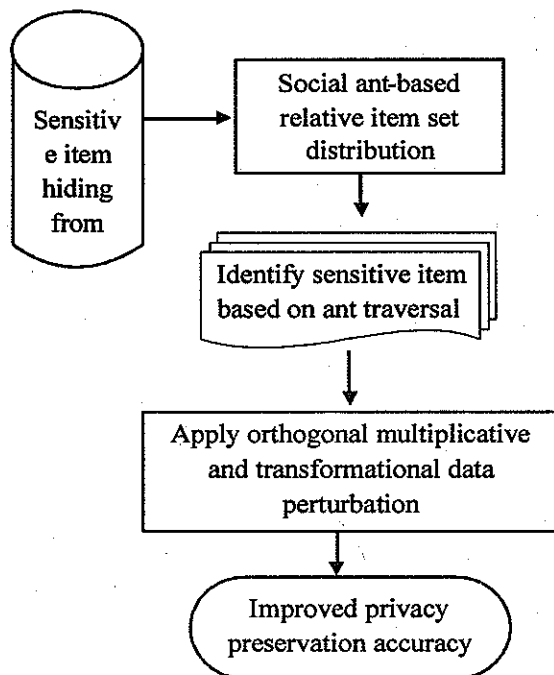
<sup>1</sup>Research Scholar, Department of Computer Science Karpagam University, Coimbatore, India. E-mail : tmselvanin@gmail.com

<sup>2</sup>HOD, Department of Computer Science Karpagam University, Coimbatore, India. E-mail : venikarthik04@gmil.com

**2.1 Design of social ant-based relative item set distribution**

The second step in the design of Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is the construction of social ant-based relative item set distribution. In order to arrive at optimal hiding of sensitive item and hide sensitive data item, the proposed OSA-SIH technique uses social ant based relative item set distribution in the corresponding original dataset even for larger item sets.

Hiding of sensitive item is done through multiplicative and transformational data perturbation technique based on socially cohesive relational rate between sensitive and non sensitive item sets of the original dataset to generate a modified dataset 'MD'



**Figure 1** Block diagram of Ant-based Orthogonal Multiplicative and Transformational Data Perturbation

Figure 1 given above shows the block diagram of Ant-based Orthogonal Multiplicative and Transformational Data Perturbation. The basic concept of ant principle is that the random wandering nature and upon successful identification of food return to their colony while laying down pheromone trails. On the other hand, if other ants identify those paths, the ants again do not traverse at random manner, but it blindly follows the trail provided by the earlier ants. In a similar manner, if the items in the transaction is said to occur repeatedly, then it is said to be sensitive item. By changing the item in a random manner with the aid of probability functions, frequent sensitive items are hidden in an efficient manner.

Let us set ' $\alpha(x_a)$ ' as the pheromone intensity of the 'D' ant (i.e. Dataset) which is at position ' $x_i$ ' and be initialized as a constant. Then the probability of the dataset '' that hide sensitive item from '' to '' is mathematically formulated as given below.

$$p_{ab} = \left( \frac{\alpha(x_b)}{\alpha(x_c)} \right), \text{ where } x_b, x_c \in D \tag{1}$$

From (1), the position of the dataset provides a solution of the problem for optimal hiding of sensitive item based on social ant based relative item set. Once the optimal sensitive item is obtained, the proposed OSA-SIH performs the task of hiding sensitive item through multiplicative and transformational data perturbation technique. This multiplicative and transformational data perturbation technique is based on sensitive and non sensitive item sets of the original dataset through which a modified dataset is generated.

The objective behind the use of orthogonal multiplicative and transformational data perturbation is to improve the rate of side effects during the data perturbation process. The orthogonal multiplicative and transformational data perturbation in proposed OSA-SIH technique uses orthogonal transformation.

Let us consider two datasets 'L' and 'M' of size 'i \* n matrix' and 'j \* n matrix' respectively with orthogonal matrix represented as 'O'. Now the mathematical formulation for the orthogonal multiplicative and transformational data perturbation for two datasets '' and '' is as given below

$$A = LO; \quad B = MO \quad (2)$$

$$AA^T = LL^T; \quad BB^T = MM^T \quad (3)$$

$$AB^T = L O O^T B^T = LM^T \quad (4)$$

From (2), (3) and (4), by applying an orthogonal matrix based on socially cohesive relational rate between sensitive and non sensitive item sets, all the pair distances and similarities from column vectors 'A and B' are preserved in an efficient manner in the perturbed data. At the same time, both the sensitive and non sensitive items and the transformation process are kept secret, whereas only the perturbed data is viewed by the third user. As a result, the rate of side effects is improved in a significant manner.

Figure 2 given below shows the ant-based orthogonal multiplicative and transformational algorithm.

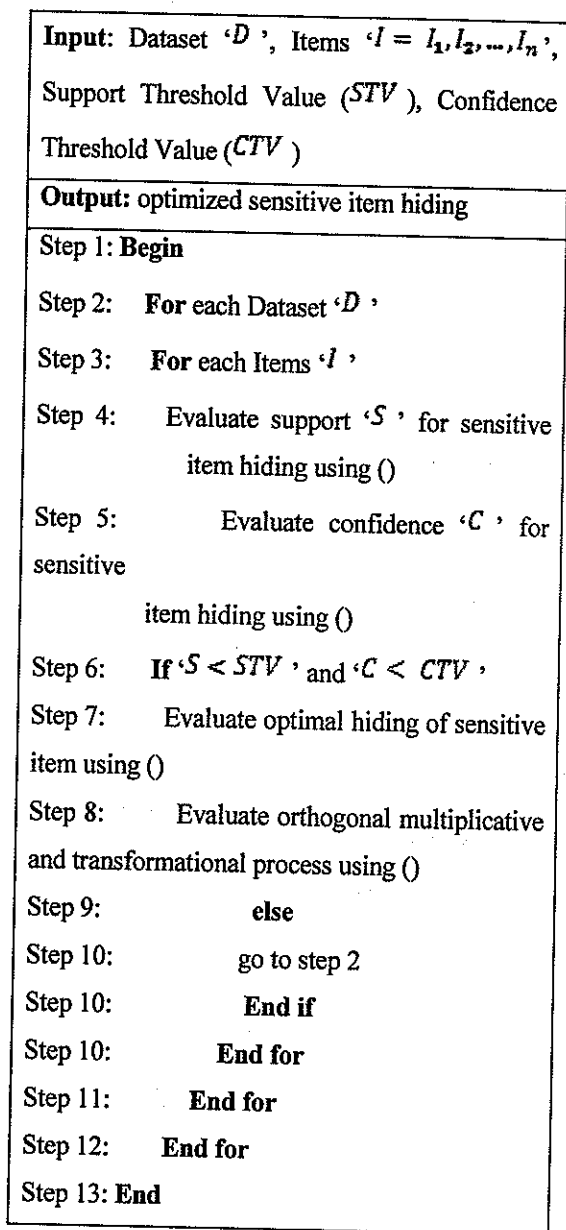


Figure : 2 Ant-based orthogonal multiplicative and Transformational algorithm

The Ant-based based Orthogonal Multiplicative and Transformational (AOMT) algorithm given above includes four main steps. The first step measures the support for sensitive item hiding. The second step evaluates the confidence value for sensitive item hiding. Next, a comparison is made between the

support threshold 'STV' and confidence threshold 'CTV' with the evaluated confidence 'C' and support value 'S'. Followed by this, optimal hiding of sensitive item and orthogonal multiplicative and transformational process is performed. If the values of support " and confidence " is less than the support threshold " and confidence threshold " respectively, item hiding is performed, otherwise, the same operations is performed with other transactions. In this way, rate of side effects is ensured in an efficient manner.

### III. EXPERIMENTAL SETTINGS

Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is developed for data publishing using JAVA platform. The OSA-SIH technique uses the Adult data set from the University of California Irvine data repository that contains information on individuals such as age, level of education and current employment type.

The dataset used in this work has forty nine thousand records and also binomial label that indicates the salary of less or greater than fifty thousand US dollars, referred to as <50K or >50K in this work. The data for experimental purpose has been divided into a training dataset containing thirty two thousand records and a test dataset containing sixteen thousand records.

There are fourteen attributes consisting of seven polynomials, one binomial and six continuous attributes and are used in the OSA-SIH technique to preserve the privacy of certain attributes including salary, relationship and marital status. The

employment class attribute denotes the employer type (i.e. self employed or federal) and occupation refers to the employment type (i.e. farming or managerial). The education attribute comprises of high school graduate or doctorate. The relationship attribute includes the information related to unmarried or married.

The final nominal attributes are country of residence, gender and race. The continuous attributes are age, hours worked per week, education number, capital gain and loss and a survey weight attribute assigned to an individual based on information such as area of residence and type of employment. The performance of the OSA-SIH technique is evaluated for parameters such as number of transactions, size of transaction, rate of side effects on the modified dataset.

The rate of side effects measures the difference between the actual size of transaction and the modified dataset generated during privacy preserving. The mathematical formulation for rate of side effects is given as below.

$$RoSE = (Size - MD) \quad (5)$$

From (5), the rate of side effects 'RoSE' is measured on the basis of size of transaction 'Size', modified dataset 'MD' respectively. It is measured in terms of kilobytes (KB).

### IV. DISCUSSION

The Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is compared against the existing Hierarchical K-means clustering in

Privacy Preserving Data Mining [1] and Mining of Association Rules on Horizontally Partitioned Data in Preserving Data Mining [2]. The experimental results using JAVA are compared and analyzed through table and graph form given below.

#### 4.1 Impact of rate of side effects

The rate of side effects to obtain quality privacy preservation for distributed data mining using OSA-SIH, to Hierarchical K-means clustering in Privacy Preserving Data Mining and Mining of Association Rules on Horizontally Partitioned Data is elaborated in table 1. We consider the technique with differing size of transaction in the range of 100KB to 700KB for experimental purpose using JAVA.

Table 1 Tabulation for rate of side effects

Size of transaction (KB)	Rate of side effects (KB)		
	OSA-SIH	Hierarchical -k-means Clustering of PPDM	Mining of Association Rules on Horizontally Partitioned Data -PPDM
100	68	72	74
200	75	83	87
300	90	98	102
400	103	111	115
500	120	128	132
600	130	138	142
700	145	153	157

In figure 3, we depict the rate of side effects while generating a modified dataset from an original dataset with size of transaction range from 100 KB to 700

KB for the purpose of experiment. From the figure, the rate of side effects resulted using the proposed OSA-SIH technique is lower when compared to two other existing methods Hierarchical K-means clustering in Privacy Preserving Data Mining [1] and Mining of Association Rules on Horizontally Partitioned Data in Preserving Data Mining [2]. Besides it can also be observed that by increasing the size of transaction, the rate of side effects is also increased using all the methods. But comparatively, it is lower using OSA-SIH technique.

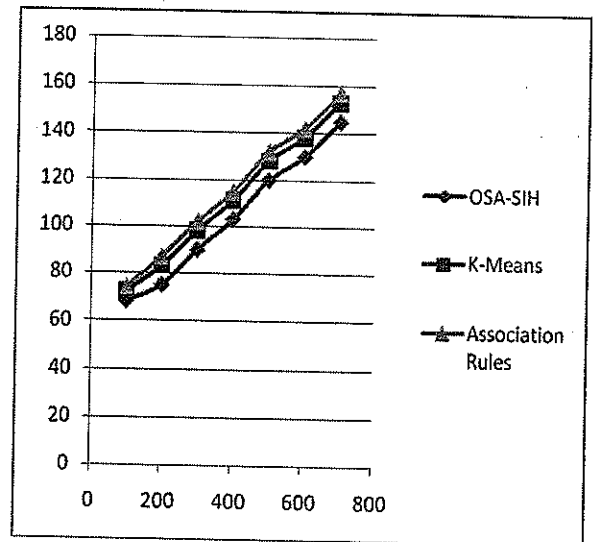


Figure 3 Measure of rate of side effects

Figure 3 as shown above measures the rate of side effects generated when original dataset is transformed to modified dataset with the objective of hiding certain items (i.e. attributes) for the purpose of privacy preservation. The rate of side effects of hiding item sets on the modified dataset are verified for various user requested item set distribution on the privacy preserving distributed data mining, which in turn also improves the user trust level. By applying

correlation based privacy preserving, the rate of side effects is minimized using OSA-SIH by 7.36% compared to Hierarchical K-means clustering in Privacy Preserving Data Mining. In addition, by following the two considerations in hiding a sensitive item using correlation-based approach, the rate of side effects is reduced by 11.04% compared to Mining of Association Rules on Horizontally Partitioned Data in Preserving Data Mining

## V. CONCLUSION

An Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique with scope of quality privacy preservation for distributed data mining with optimal side effects on original dataset has been designed. The objective of providing such a design is to ensure high quality privacy preservation of the data items of corresponding user's privileges for distributed data and to decrease the time for optimal hiding for various user requested item set distribution. The proposed social ant-based relative item set distribution provides rate of side effects for large item sets through multiplicative and transformational data perturbation technique. Experimental evaluation is conducted with the Adult Data Set extracted from UCI repository to provide high quality privacy preservation of data items and measured the rate of side effects on answering user query requests.

## REFERENCES

- [1] Anrong Xue, Dongjie Jiang, Shiguang Ju, Weihe Chen, and Handa Ma, "Privacy-Preserving Hierarchical-k-means Clustering on Horizontally Partitioned Data", International Journal of Distributed Sensor Networks, Volume 5, Issue 1, 2009, Pages 81 – 82.
- [2] Murat Kantarcıo glu and Chris Clifton, "Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", 2012 12th International Conference on Hybrid Intelligent Systems (HIS),Pages 2-13.
- [2] Russell Paulet, Md. Golam Kaosar, Xun Yi, and Elisa Bertino, "Privacy-Preserving and Content-Protecting Location Based Queries", IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 5, May 2014, Pages 1200-1210.
- [3] Dimitrios Karapiperis and Vassilios S. Verykios, "An LSH-Based Blocking Approach with a Homomorphic Matching Technique for Privacy-Preserving Record Linkage", IEEE Transactions on Knowledge and Data Engineering, Volume 27, Issue 4, April 2015, Pages 909-921.
- [4] Leopoldo Bertossi and Lechen Li, "Achieving Data Privacy through Secrecy Views and Null-Based Virtual Updates", IEEE Transactions on Knowledge and Data Engineering, Volume 25, Issue 5, May 2013, Pages 987-1000.
- [5] Sara Hajian and Josep Domingo-Ferrer, "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining",

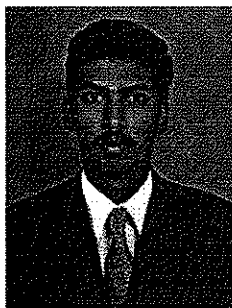
- IEEE Transactions on Knowledge and Data Engineering, Volume 25, Issue 7, July 2013, Pages 1445-1459.
- [6] Fosca Giannotti, Laks V. S. Lakshmanan, Anna Monreale, DiIssue Pedreschi and Hui (Wendy) Wang, "Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases", IEEE Systems Journal, Volume 7, Issue 3, September 2013, Pages 385-395.
- [7] Tamir Tassa, "Secure Mining of Association Rules in Horizontally Distributed Databases", IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 4, April 2014, Pages 970-983.
- [8] Anna Cinzia Squicciarini, Dan Lin, Smitha Sundareswaran, and Joshua Wede, "Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites", IEEE Transactions on Knowledge and Data Engineering, Volume 27, Issue 1, January 2015, Pages 193-206.
- [9] Chun-Wei Lin, Tzung-Pei Hong and Hung-Chuan Hsu, "Reducing Side Effects of Hiding Sensitive Item sets in Privacy Preserving Data Mining", The Scientific World Journal, April 2014, Pages 1-13.
- [10] Zahid Pervaiz, Walid G. Arif Ghafoor and Nagabhushana Prabhu, "Accuracy-constrained Privacy-preserving Access Control Mechanism for Relational Data", IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 4, April 2014, Pages 795-807.
- [11] Tiancheng Li, Ninghui Li, Jian Zhang, Ian Molloy, "Slicing: A New Approach to Privacy Preserving Data Publishing", IEEE Transactions on Knowledge and Data Engineering, Volume 24, Issue 3, March 2012, Pages 561-574.
- [12] Slawomir Goryczka, Li Xiong, Benjamin C. M. Fung, "m-Privacy for Collaborative Data Publishing", 2011 7th International Conference on Collaborative Computing: Networking, Applications and Work sharing (Collaborate.Com), Pages 1-10.
- [13] Mohamed Nabeel, Elisa Bertino, "Privacy Preserving Delegated Access Control in Public Clouds", IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 9, September 2014, Pages 2268-2280.
- [14] Mohamed Nabeel, Elisa Bertino, "Privacy-Preserving Fine-Grained Access Control in Public Clouds", IEEE Computer Society Technical Committee on Data Engineering, 2012, Pages 1 -10.
- [15] Murat Kantarcýo glu and Chris Clifton, "Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", 2012 12th International Conference on Hybrid Intelligent Systems (HIS), Pages 2-13.

- [16] Yaping Li, Minghua Chen, Qiwei Li, and Wei Zhang, "Enabling Multilevel Trust in Privacy Preserving Data Mining", IEEE Transactions on Knowledge and Data Engineering, Volume 24, Issue 9, September 2012, Pages 1598 – 1612.
- [17] Nan Zhang, Wei Zhao, "Privacy-Preserving OLAP: An Information-Theoretic Approach", IEEE Transactions on Knowledge and Data Engineering, Volume 23, Issue 1, January 2011, Pages 122-138.
- [18] Shaopeng Guan, Yuan Zhang, and Yue Ji, "Privacy-Preserving Health Data Collection for Preschool Children", Computational and Mathematical Methods in Medicine, September 2013.
- [19] Pawan R. Bhaladhare and Devesh C. Jinwala, "A Clustering Approach for the 50YÜ-Diversity Model in Privacy Preserving Data Mining Using Fractional Calculus-Bacterial Foraging Optimization Algorithm", Advances in Computer Engineering, September 2014, Pages 1-13.
- a paper in International Conference. His research interests are Data mining and warehousing.



Dr. S. Veni completed her Ph.D in Computer Science from Bharathiar University in 2014. she is working as Associate Professor in Department of Computer Science, Karpagam University, Coimbatore. Her experience is 12 yrs. she has presented various papers in National and International Conference. Her research interests are Computer Networks.

#### AUTHOR'S BIOGRAPHY



P. Tamil Selvan completed his M.Phil in Computer Science from Karpagam University in 2009. He is working as Assistant Professor in Department of Computer Science, Karpagam University, coimbatore. His experience is 8 yrs. He has presented