OPTIMAL TIME HIDING USING ANT BASED ALGORITHMS IN PRIVACY PRESERVATION IN DATA ITEMS

***P. Tamil Selvan*[1], *Dr. S. Veni*[2]**

Abstract

Privacy preserve data mining (PPDM) have developed into an additional challenging issue in resolve the effect of privatizing user's data. Most of the PPDM method adopting, precise item hiding change the creativity of the dataset and were planned to moderately evaluate the side effects. In this work, we plan to develop an Optimized Social Ant Based Precise Item Hiding (OSA-SIH) method also develop a range of value privacy preservation for ensure point for optimal hiding. Firstly precise data is scattered dataset are evaluate by means of the social ant base qualified item set distribution. Based on the calculate dataset, optimal hiding of precise item be alive with social ant based relative item set distribution silent for superior item sets, make sure moment for optimal hiding. Experiments are then conducted to show the presentation of time for optimal hiding.

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

I. Introduction

One of the widely used approaches by data miners is data perturbation for Privacy Preserving Data Mining (PPDM). Anonymous publication of sensitive transactional data [1] was performed using approximate nearest neighbor addressing the issues related to data utility and execution time (ATD).

In [2], efficient clustering was applied with the objective of improving the computational performance and at the same time reducing the computational cost through Fractional Calculus (*l*-diversity).

In [3], privacy preserving and content protection was performed to address security issues using Oblivious Transfer and Private Information Retrieval (PIR). To address access control mechanism. Reducing Side Effects in Privacy Preserving Data Mining (RSE-PPDM) [4] use hiding missing artificial utility algorithm to minimize the number of deleted transactions and number of side effects. In [5], access control for cloud based on privacy preserving was introduced using group key management scheme. In[6], a new method to preserve privacy for data publishing called slicing was introduced for better data utility. Another method called m-privacy [7] for data publishing was introduced to ensure anonymity. In[8], access control for cloud based on privacy preserving

[1] Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India. E-mail : tmselvanin@gmail.com

[2] Research Supervisor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India. E-mail : venikarthik04@gmail.com

was introduced using group key management scheme. Another method used cryptographic techniques [9] to solve the issues related to confidentiality and security through fine grained attribute based access control policies. In[10], distributed mining of association rules was performed using cryptographic techniques that resulting in reducing the overhead.

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 based Orthogonal Multiplicative and Transformational algorithm.

Experimental results showed that the AOMT algorithm has good performance in optimal time hiding Besides, the proposed algorithm can thus generate minimal side effects on the modified dataset compared to the state-of-the-art works for hiding sensitive item sets.

This paper is organized as follows. The proposed AOMT algorithm to hide the sensitive item sets with optimal side effects is stated in Section 2. Experiments are conducted in Section 3. Analysis of the results and comparison made with the existing works using table and graph is presented in Section 4. Conclusion remarks are provided in Section 5.

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 privacy preservation accuracy and minimizing 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.

2.1 Design of user operational conditions-based sensitive items

The first step in the design of Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is the effective construction of user operational conditions-based sensitive items. In this section, the problem of sensitive item hiding for privacy preservation in distributed dataset is evaluated. It is performed based on the user operational conditions, by proposing a system to measure the global frequent item sets for distributed data item being shared.

This is done by designing an algorithm that hides sensitive frequent items from the global frequent items. The optimal hiding of sensitive item is arrived with social ant based relative item set distribution in the corresponding original dataset even for larger item sets.

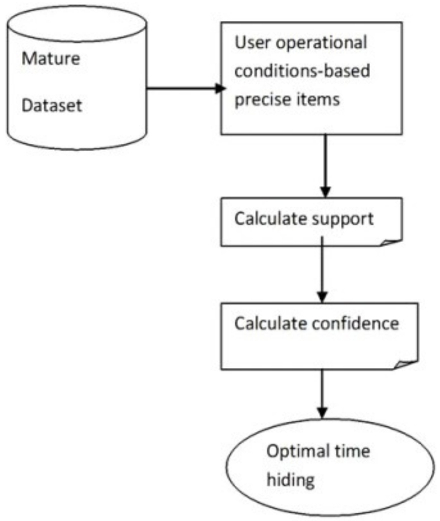


Figure. 1 Block diagram of user operational conditions-based sensitive items

Figure 1 given above shows the block diagram of user operational conditions-based sensitive items. The block diagram includes two main components, where the support and confidence values are evaluated for sensitive item hiding aiming at reducing the time for sensitive hiding.

Let us consider a dataset 'D ' where '$I = I_1, I_2, \dots, I_n$ ' represents the items, consisting of 'n ' transaction comprises of the set of items in such a way that '$T \in I$ '. Then, the association rule is of the form

$$P \rightarrow Q, where\ P \in I \ \&\ Q \in I \qquad (1)$$

Where 'P ' and 'Q ' are said to be the antecedent and consequent of rule respectively. Relative strength of an item with respect to its strong or weak nature is evaluated using two factors namely, support and confidence of the item. The first factor to be measured for sensitive item hiding is the support and is mathematically formulated as given below.

III. Existing Techniques

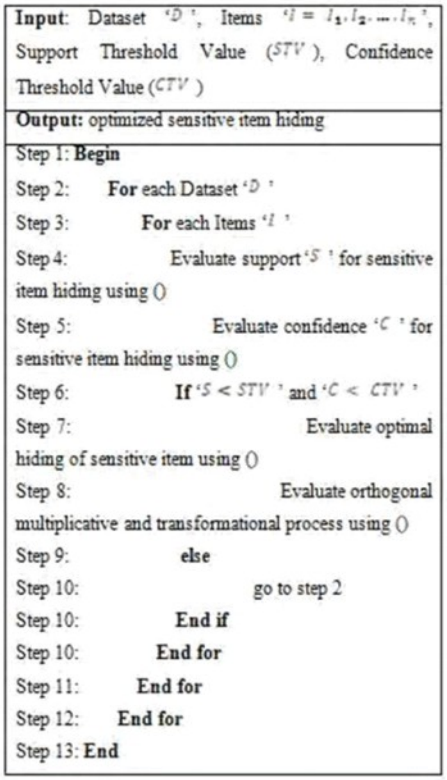
$$S(P \rightarrow Q) = S(P \cup Q) = \left(\frac{(P \cup Q)}{n}\right) \qquad (2)$$

From (2), support 'S ' measures the proportion of transactions that includes both 'P ' and 'Q ' respectively, with 'n ' denoting the total number of transactions involved during sensitive item hiding. The second factor to be measured for sensitive item hiding is the confidence formulated as given below.

$$C(P \rightarrow Q) = \left(\frac{(P \cup Q)}{P}\right) = \left(\frac{S(P \cup Q)}{S(P)}\right) \qquad (3)$$

From (3), the confidence 'C ' is the percentage for a transaction that contains 'P ' also contains 'Q '. A rule is significant if its support and confidence are higher than the user designated Support Threshold Value (STV) and Confidence Threshold Value (CTV). As a result, using the ant-based relative item set distribution algorithm not all the items are retrieved, but only a very small member that satisfies the '' and ''are retrieved. In this way, the time for optimal hiding is significantly reduced.

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



```
Input: Dataset ‘$D$’, Items ‘$I = I_1, I_2, \dots, I_n$’, Support Threshold Value ($STV$), Confidence Threshold Value ($CTV$)
Output: optimized sensitive item hiding
Step 1: Begin
Step 2:     For each Dataset ‘$D$’
Step 3:         For each Items ‘$I$’
Step 4:             Evaluate support ‘$S$’ for sensitive item hiding using ()
Step 5:                 Evaluate confidence ‘$C$’ for sensitive item hiding using ()
Step 6:                 If ‘$S < STV$’ and ‘$C < CTV$’
Step 7:                         Evaluate optimal hiding of sensitive item using ()
Step 8:                         Evaluate orthogonal multiplicative and transformational process using ()
Step 9:                 else
Step 10:                        go to step 2
Step 10:                End if
Step 10:            End for
Step 11:        End for
Step 12:    End for
Step 13: End
```

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 ‘’ are 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, privacy preservation accuracy 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 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 and time for optimal hiding.

The time for optimal hiding is measured based on the total number of transactions and the time required for single transaction. The time for optimal hiding is measured in terms of milliseconds (ms) and is formulated as given below.

Time = n*Time (item hiding for single transaction) (4)

IV. Discussion

The Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is compared against the existing Anonymous Transaction data (ATD-PPDM) [1] and L-Diversity in Privacy Preserving Data Mining (*l*-Diversity-PPDM) [2].The experimental results using JAVA are compared and analyzed through table and graph form as given below.

4.1 Impact of time for optimal hiding

The comparison of time for optimal hiding is presented in table 1 with respect to the total number of transactions in the range of 5 – 35 collected at different time stamps from the adult dataset records. With increase in the number of transactions, the time for optimal hiding is also increased though not observed to be linear. This is because of the different types and nature of the transaction, the time for optimal hiding also gets varied.

Table 1 Tabulation for time for optimal hiding

IV. Existing Tools And Techniques - Pros & Cons

V. Conclusion :

Amount of transactions (n)	Time for optimal hidden values (ms)		
	OSA-SIH	Anonymous Transaction data	L-Diversity(PPDM)
5	3.33	3.66	4.66
10	4.96	5.26	6.41
15	6.67	6.96	8.01
20	9.03	9.31	10.44
25	6.36	6.59	7.81
30	9.46	9.76	10.86
35	12.32	12.64	13.72

To ascertain the performance of the time for optimal hiding, comparison is made with two other existing methods Anonymous Transaction data (ATD-PPDM) [1] and L-Diversity in Privacy Preserving Data Mining (*l*-Diversity-PPDM) [2].

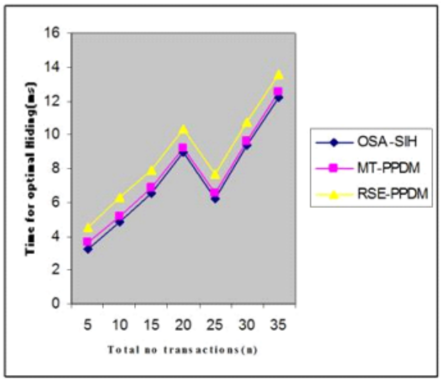


Figure 3 Measure of time for optimal hiding

The time for optimal hiding is reduced by applying of user operational conditions-based sensitive items in OSA-SIH. With the application of user operational conditions-based sensitive items, where the support and confidence values are evaluated for sensitive item hiding that provides the results with respect to total number of transactions reducing the time for optimal hiding by 6.85% compared to ATD-PPDM. Besides, by applying the association rules and comparing with the user designated Support Threshold Value ($_{STV}$) and Confidence Threshold Value ($_{CTV}$) minimizes the time for optimal hiding by 28.09% compared to *l*-Diversity-PPDM.

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. A user operational conditions-based sensitive item are designed as a measure for identifying the support and confidence value and proposed a proposed a system to measure the global frequent item sets for distributed data item being shared based on user query. Experimental evaluation is conducted with the Adult Data Set extracted from UCI repository to provide optimal time hiding on answering user query requests. Performances results reveal that the proposed OSA-SIH technique strengthen the optimal time hiding on high dimensional dataset. Compared to the privacy preserving data mining techniques, the proposed OSA-SIH technique provides 12.85% high rate of privacy preservation accuracy and minimizes the time for optimal hiding by 17.47% compared to ATD-PPDM and *L*-Diversity PPDM respectively.

REFERENCES

[1] Gabriel Ghinita, Panos Kalnis, and Yufei Tao, *"Anonymous Publication of Sensitive Transactional Data",* IEEE Transactions on Knowledge and Data Engineering, Volume 23, Issue 2, September 2014, Pages 161-174.

[2] Pawan R. Bhaladhare and Devesh C. Jinwala *"A Clustering Apporach for the l-Diversity Model in Privacy Preserving Data Mining Using Fractional Calculus-Bacterial*

Foraging Optimization Algorithm", Advances in Computer Engineering, September 2014 , Pages 1-13.

[3] 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.

[4] 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.

[5] Mohamed Nabeel, Elisa BertiIssue , *"Privacy Preserving Delegated Access Control in Public Clouds",* IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 9, September 2014, Pages 2268-2280.

[6] Li T, Li N, Zhang J, Molloy I. Slicing: A new approach to privacy preserving data publishing. IEEE Transactions on Knowledge and Data Engineering. 2012 Mar; 24(3):561–74.

[7]. Goryczka S, Xiong L, Fung BCM. m-Privacy for Collaborative Data Publishing. 7th International Conference on Collaborative Computing: Networking, Applications and Work sharing Collaborate.Com). 2011 Oct 15-18. p. 1–10.

[8]. Nabeel M, BertiIssue E. Privacy preserving delegated access control in public clouds. IEEE Transactions on Knowledge and Data Engineering. 2014 Sep; 26(9):2268–80.

[9]. Nabeel M, Bertino E. Privacy-preserving fine-grained access control in public clouds. IEEE Computer Society Technical Committee on Data Engineering. 2012 Dec; 35(4):1–10.

[10]. Glu MK, Clifton C. Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data. 12th International Conference on Hybrid Intelligent Systems (HIS). 2012. p. 2–13.