Reinforced Social Ant with Discrete Swarm Optimizer for Sensitive Item and Rule Hiding

P.Tamil Selvan

Abstract ***In data mining Privacy Preserving Data mining (PPDM) of the important research areas concentrated in recent years which ensures ensuring sensitive information and rule not being revealed. Several methods and techniques were proposed to hide sensitive information and rule in databases. In the past, perturbation-based PPDM was developed to preserve privacy before use and secure mining of association rules were performed in horizontally distributed databases. This paper presents an integrated model for solving the multi-objective factors, data and rule hiding through reinforcement and discrete optimization for data publishing. This is denoted as an integrated Reinforced Social Ant and Discrete Swarm Optimization (RSA-DSO) model. In RSA-DSO model, both Reinforced Social Ant and Discrete Swarm Optimization perform with the same particles. To start with, sensitive data item hiding is performed through Reinforced Social Ant model. Followed by this performance, sensitive rules are identified and further hidden for data publishing using Discrete Swarm Optimization model. In order to evaluate the RSA-DSO model, it was tested on benchmark dataset. The results show that RSA-DSO model is more efficient in improving the privacy preservation accuracy with minimal time for optimal hiding and also optimizing the generation of sensitive rules.***

Keywords: Privacy preserving data mining, perturbation-based, data and rule hiding, reinforcement, discrete optimization.

I. INTRODUCTION

In this age of information, various organizations have started collecting and organizing the data for data analysis. To facilitate data analysis, care should be taken during the information sharing. Many researchers concentrate on data analysis and data publishing by hiding the data or rules. In this paper, the investigation ofthe data and rule hiding for preserving privacy of transactional database and therefore ensuring smooth data publishing is called Reinforced Social Ant and Discrete Swarm Optimization (RSA-DSO) model is presented.

In [1] a data perturbation model using Privacy Preserving Data Mining (PPDM) was presented. Multilevel Trust in Privacy Preserving Data Mining(MT-PPDM) [1] used data aggregation concept without the possibility of accessing the information by the third parties and hence, ensuring higher amount of trust

level. However, security aspects were note covered. Protocol for secure mining of association rule was developed in [2] for providing secured mining of association rules using two secure multi-party algorithms.

With the recent developments in cloud computing, data mining-as-a-service receives great attention. In [3], the problem of outsourcing association rule mining taskwithin a corporate privacy-preserving framework was presented to protect the privacy of data using a conservative frequency-Based attack model. Yet another model to provide access control mechanism for relational data using k-anonymity was presented in [4]. The model, not only anonymized the data to meet privacy requirements but also used access control mechanism for imprecise constraints. Another delegation of access control policy using novel optimization algorithms [5] resulted in the privacy of data owners in cloud environment.

In recent years, privacy-preserving publishing of microdataisresearched in a wide manner. Microdatainclude information relating to a person, a household, or an organization. Many research works are conducted in the field of microdataanonymization. In [6], a novel technique called slicing was presented for preserving data utility against privacy threats. However, collaborative form of privacy implications remained unaddressed. To solve this issue, collaborative data publishing using m-privacy [7] was designed using heuristic algorithms that not only achieved comparable utility but also ensured m-privacy efficiently. Fine grained access control was provided in [8] through well known cryptographic techniques.

With the extensive growth in data mining technologies, meaningful information can be acquired for decision making in several domains. A compact pre-large GA-based (cpGA2DT) [9] algorithm was constructed with the objective of hiding sensitive items by deleting the transactions.

The cpGA2DT also solved the evolutionary process by applying both the compact GA-based (cGA) mechanism and the pre-large concept. Minimal side effects were also ensured by applying a flexible fitness function. However, rule hiding remain unsolved. In [10], a heuristic method was designed to minimize the number of hidden sensitive rules and number of modified entries, ensuring scalability.

In this paper, an integrated model called, Reinforced Social Ant and Discrete Swarm Optimization is presented to provide data hiding and rule hiding for data publishing. A reinforcement model designed in RSA-DSOensured data hiding in such a manner that sensitive data item does not occur more than once. Next, a discrete optimization model is RSA-DSO ensures optimized rule hiding for data publishing. This proposed model provides a secured way of privacy preserving of data mining for data publishing, and it achieves higher privacy preservation rate reducing the optimal data hiding time.

The paper is structured as follows. In Section 2, the basic concepts in data and association rule hiding are reviewed. In Section 3, principles and algorithms for data hiding using reinforced learning model and discrete optimization model for sensitive rule hiding are proposed, respectively. In Section 4, the experimental evaluation and the analysis of results is discussed in detail. Section 5 concludes the paper.

Revised Manuscript Received on July 05, 2019.

Dr. P. TamilSelvan, Assistant Professor in Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore.

II. RELATED WORKS

Through data mining, useful and the most requisite information of the organizations and companies are extracted in a wide manner. However, the organizations do not want to disclose the sensitive data to the public. Therefore, several Privacy Preserving Data Mining (PPDM) techniques were employed to preserve such confidential information through data and rule hiding.

Online Analytical Processing (OLAP) is one of the most eminent decision support and knowledge discovery models. Issues related for protecting private information in OLAP was in [11]. However, with the increase in scalability, the side effects also get increased. To reduce the side effects, in [12], Hiding-Missing-Artificial Utility (HMAU) algorithm was designed and it also reduced the execution time and the number of transactions generated. A data publication model for multiple sensitive attributes was presented in [13] using high-quality released information. A personal metadata management framework was designed in [14] to protect the privacy of metadata.

Data sharing between different users though results in many advantages, privacy laws in existence makes such data sharing becomes difficult. In [15], an integrated model using Elliptic CurveCryptography (ECC) public key cryptosystem and Diffie-Hellman key exchange was presented to solve privacy preserving problem. A comprehensive review on privacy preserving data mining was presented in [16]. Genetic algorithm was applied in [17] to minimize the side effects while extracting rules.

With the increasing development in the field of data mining extraction required knowledge from large data collection, is the most required area to be analyzed. However, with the disclosure of sensitive data released to other users in an inappropriate manner poses severe threats. An evolutionary multi-objective optimization model was designed in [18] with the objective of reducing the side effects. On the contrary, sensitive rules were hidden using Evolutionary Multi-Objective (EMO) [19] mechanisms with the objective of deleting identified transactions for the purpose of hiding sensitive rules. Another method based on EMO was designed in [20] to identify promising transactions to minimize the side effects. In [21] the scope of quality privacy preservation for distributed data mining with optimal side effects on the original datasets. To improve the efficiency of the privacy preserving association rule mining with the constraint minimization in [22].

Based on the aforementioned methods and techniques, several methods were presented to hide data and rules, an integrated model to hide both data and rules is the most requisite form. With this objective, an integrated Reinforced Social Ant and Discrete Swarm Optimization (RSA-DSO) model is presented to minimize the optimal data hiding time and therefore improve the privacy preservation accuracy for data publishing.

III. METHODOLOGY

Sensitive data and rule hiding discover sensitive data items and sensitive rules occurring in a transactional database with the objective of producing significant rules that hold for the data. To preserve the privacy of an individual or an organization, sensitive data items and rules are identified in the transactional database. In this work, an integrated framework using Reinforced Social Ant model for sensitive data item hiding and Discrete Optimization model for sensitive rule hiding is presented in the forthcoming section.

1.1 Integration of Reinforced Social Ant and Discrete Swarm Optimization model

An integrated Reinforced Social Ant and Discrete Swarm Optimization (RSA-DSO) model is proposed to solve multi-objective factor of data hiding and rule hiding which helps in achieving theaccuracy on population census process. In the proposed RSA-DSO model, both Reinforced Social Ant and Discrete Swarm Optimization perform with the same particles. Initially, Reinforced Social Ant (RSA) model is designed to ensure the sensitive data item hiding is executed through the proposed RSA-DSO framework. Next, Discrete Swarm Optimization (DSO) model is introduced to identify the sensitive rules and further hidden for enhancing the accuracy of Privacy preserving data publishing.

In order to improve the computational performance, in this work, an integrated Reinforced Social Ant and Discrete Swarm Optimization (RSA-DSO) model is presented. This RSA-DSO model possess the advantage of not only hiding the sensitive data item but also rules are said to be hidden that serves for data publishing. With the design structure of RSA-DSO model convergence characteristics for data item hiding is strictly performed once. This possesses the advantages of performing data item hiding without redundancy and therefore reduces the computational performance during the data publishing. Next, the hiding of sensitive rules that are performed using the sigmoid function instead of the traditional support-confidence evaluation which evaluates optimal solutions with discrete nature of database. This in turn reduces the computation performance in evaluating the sensitive rule for data publishing.

1.2 Reinforced Social Ant Sensitive Data Item Hiding

The idea behind the design of Reinforcement Learning (RL) model for data item hiding is the search for minimum cost path while obtaining optimal path in a graph that involves finding a balance between exploration (searching optimal sensitive item) and exploitation (given dataset). Figure 1 illustrates the Reinforced Social Ant Sensitive Data Item Hiding model.

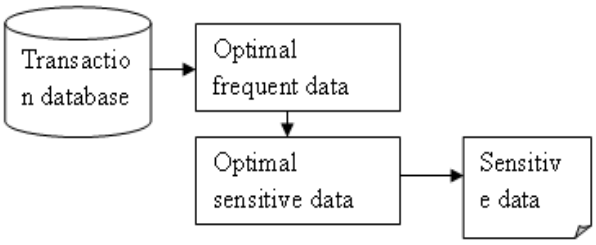

Figure 1 Illustration of Reinforced Social Ant Sensitive Data Item Hiding

As illustrated in the figure, the Reinforced Social Ant Sensitive Data Item Hiding obtains a transaction database (i.e. Adult dataset) as input. From the obtained transaction database, optimal frequent data and sensitive data identification are made using a reinforcement model. With the resultant obtained value, the sensitive data are hidden.

Let the database be considered $D = \{t_1, t_2, \dots, t_n\}$, where $t_1, t_2, \dots, t_n$ represents the set of transactions with each

transaction comprising of a set of items '$I = i_1, i_2, \ldots, i_n$'. Here '$i_1$' represents the item number 1, 'i_2' represents the item number 2 and so on. A minimum support threshold is assigned to be '$mint$', on the other hand, support of an item is represented by '$\sup(i_p)$', with frequent item denoted by '$freq(i_p)$', and is mathematically formulated as given below.

$$\sup(i_p) = \{t_1(p) + t_2(p) + \cdots t_n(p)\} \quad (1)$$

$$freq(i_p) = \frac{\sup(i_p)}{|D|} \geq mint \quad (2)$$

Optimal (i.e. sensitive item) path using social ant is said to arrive at with the convergent result of cooperation between ants (i.e. transactions) in the colony (i.e. database). However, it is highly required that the traversal should in no way include a sensitive data item more than once. With this objective, a data hiding model using Reinforced Social Ant Sensitive (RSA-S) not only identifies the minimum cost involved in obtaining sensitive data item, but also the extraction of sensitive data item only once. The data hiding model using Reinforced Social Ant Sensitive (RSA-S) consists of two steps, obtaining sensitive data item through reinforcement learning and hiding sensitive data item.

Let '$r(Cost_a, Cost_b)$', represents a measure of cost for obtaining sensitive data item from transaction 'T_a' to 'T_b'. Then, the total cost involved for obtaining sensitive data item for 'n' transactions '$p_1, p_2, \ldots, p_n$' respectively which is mathematically obtained as given below.

$$Cost(p_1, p_2, \ldots, p_n) = \sum_{q=1}^{n} r(Tp_q, Tp_{q+1}) + r(Tp_{n-1}, Tp_n) \quad (3)$$

The Reinforced Social Ant Sensitive algorithm is employed to identify optimal sensitive data from a given database 'D', of 'n' transaction with a local heuristic '$\beta_{pq} = \frac{1}{r}(p, q)$' and '$\alpha$' representing the mathematical formulation for pheromone formation. Then, the probability that an item 'k', in the transaction 'p' present in 'q' is formulated as given below.

$$Prob_{pq}^{k}(t) = \frac{[\beta_{pq}] * [\alpha_{pq}(t)]}{[\beta_{pn}] * [\alpha_{pn}(t)]} \quad (4)$$

From (4), shorter edges with greater amount of pheromone (i.e. frequent data item) are favoured by multiplying the pheromone on edge 'p, q' by the corresponding heuristic value '$\beta(p, q)$'. This is analogous to reinforce model where optimal frequent items are reinforced and selected for identification of sensitive data hiding. With the obtained frequent data item, sensitive data item is obtained through '$Dfactor$' and is formulated as given below.

$$Dfactor = \left[\frac{Max\left(Sup(i_p)\right) - Min\left(Sup(i_p)\right) * |D|}{1 - Min\left(Sup(i_p)\right)}\right] \quad (5)$$

From (5), the optimal sensitive data item to be obtained is denoted as '$Dfactor$' and is defined as the difference between thelargest support count among all sensitive data items in '$Max\left(Sup(i_p)\right)$', and the minimum support count '$Min\left(Sup(i_p)\right)$', respectively. The process of identifying sensitive data item using reinforcement model is illustrated below.

Input: database 'D', Transaction $\{t_1, t_2, \ldots, t_n\}$, minimum support threshold '$mint$',
Output: Optimized sensitive data hiding

```
1: Begin
2:    For each database 'D'
3:        For every transaction 't1, t2, ..., tn',
4:            Measure support count using (1)
5:            Measure frequent item using (2)
6:            Obtain sensitive data item using (5)
7:        End for
8:    End for
9: End
```

Algorithm 1: Reinforcement-based Sensitive Data Hiding

As shown in algorithm 1, given a database 'D' and for every transaction '$t_1, t_2, \ldots, t_n$', the objective of Reinforcement-based Sensitive Data Hiding algorithm is to identify the optimal sensitive data to be hidden (with the probability that the occurrence of sensitive data item should be once). For this, the support count is measured, followed by which the optimal frequent items are obtained. With the optimized frequent data items, reinforcement based model aims in maximizing the rewards (i.e. identification of sensitive data items) and data hiding is measured using the '$Dfactor$'.

1.3 Discrete Swarm Optimized Sensitive Rule Hiding

The task of sensitive rule hiding in the proposed work is performed using Discrete Swarm Optimization (DSO) model. The Discrete Swarm Optimized Sensitive Rule Hiding in conjunction to the traditional PSO obtains a sigmoid function to identify the sensitive rules. To start with the DSO model, initially identifies the association rule through which the sensitive rules are obtained and finally they are hidden to improve the efficiency of privacy preservation between different users.

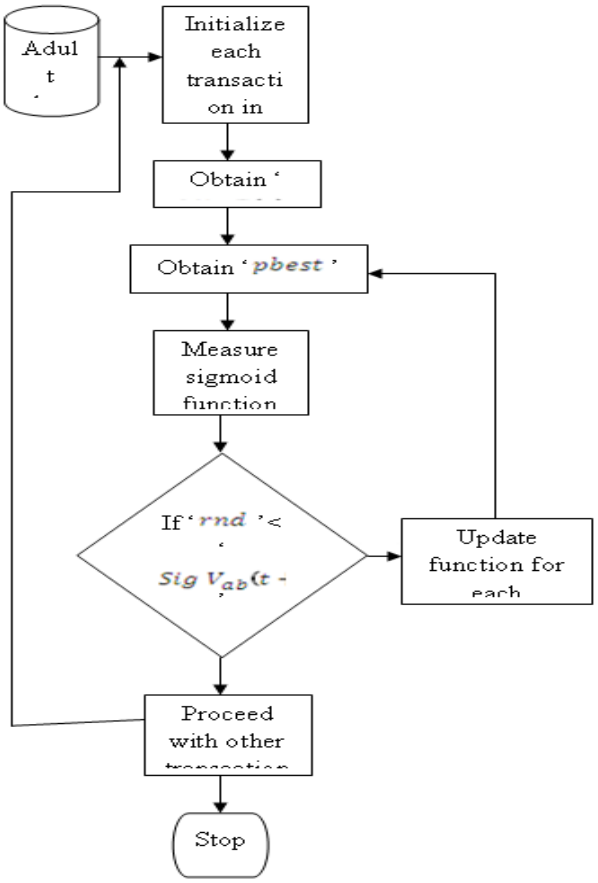

Figure 2 Flow chart of DSO model

As shown in the figure (i.e. figure 2), in DSO model, particles (i.e. transactions) are used torepresent problem solutions, where each particle has a velocity. Followed by which an iterative evolution processis performed. During each iteration and particle in the DSO model is updatedwith the aid of its personal best value '$pbest$'and global best value '$gbest$' using the designed fitness function to derive the association rule.

Here, the '$pbest$'value represents the personalbest solution of a transaction with respect to the fitness function so far. On the other hand, the '$gbest$' value is the best solution among all '$pbest$' values inthe population (i.e. database). The transactions and their corresponding velocities arethus updated using these two '$pbest$'and '$gbest$' values. The updating function of each transaction in a database is as given below.

$$V_a(n+a) = V_a(n) + pcons\,(pbest - p_i(n)) + gcons\,(gbest - p_i(n))$$

(6)

From (6), 'V_a' denotes the velocity of the 'ath' particle (i.e. transaction) in a population (i.e. database), where '$pcons$' and '$gcons$' are the constants determining the personal best and global best solutions. Here '$pbest$' and '$gbest$' represents the personal best and global best values respectively. The proposed fitness function is the weighted sum of ratio of the transaction with respect to the support of frequent item sets and is as given below.

$$Min\ F(x) = [f_1(x), f_2(x), \ldots, f_n(x)], where\ x = x_1, x_2, \ldots, x_n$$

(7)

From (7), the function '$F(x)$' represents the fitness function with '$f_i(x)$' representing the fitness value and 'x' representing the vector for 'n' decision variables. With the obtained '$pbest$' and '$gbest$' values, the sensitive rules are identified as an alternatives to traditional support and confidence factors by applying a sigmoid function as given below.

$$Sig\ V_{ab}(t+1) = \frac{1}{1 + e^{-V_{ab}(t+1)}}$$

(8)

From (8), '$V_{ab}(t+1)$' denotes the 'bth' of velocity generated by (6). Whenever a function gets updated (from 6), a random number 'rnd' lying between '0' and '1' is generated for the purpose of comparison. If the random number is lessthan '$Sig\ V_{ab}(t+1)$', then '$V_{ab}(t+1)$', then the identified rule is the sensitive rule. On the other hand, if the random number is greater than '$Sig\ V_{ab}(t+1)$', then the identified rule is not a sensitive rule and process continues with other transactions in the database. The process of identifying sensitive rule and hiding using discrete model is illustrated below.

Input: database 'D', Transaction $\{t_1, t_2, \ldots, t_n\}$, minimum support threshold '$mint$', random number 'rnd',

Output: Optimized rule hiding

1: **Begin**
2: **For** each database 'D'
3: **For** every transaction '$t_1, t_2, \ldots, t_n$'
4: Generate association rule using (6)
5: Measure fitness function using (7)
6: Extract sensitive rules using (8)
7: Generate random number between '0' and '1'
8:**If** 'rnd' < '$Sig\ V_{ab}(t+1)$',
9: Then '$V_{ab}(t+1)$' = 1
10: Sensitive rule identified
11: Perform rule hiding
12: **End if**
13: **If** 'rnd' > '$Sig\ V_{ab}(t+1)$',
14: Then '$V_{ab}(t+1)$' = 0
15: Not a sensitive rule
16: **End if**
17: **End for**
18: **End for**
19: **End**

Algorithm for 2 DSO algorithm

As illustrated in the algorithm 2, for every database and every transaction, to start with, the association rules are created. Followed by which, a fitness function is measured to extract the sensitive rules. All possible sensitive rules are extracted from (8). Due to the discrete nature, the RSA-DSO model uses a sigmoid function to find a solution to discrete optimization problem. With this the optimized sensitive rules are identified.

IV. EXPERIMENTAL SETUP

An integrated reinforced social ant and discrete swarm optimization (RSA-DSO) model is developed with the objective of improving the sensitive data and rule hiding for achieving data publishing. RSA-DSO model is implemented in java language.An adult dataset is used The RSA-DSOmodel uses the Adult data set fromtheUniversityofCaliforniaIrvinedatarepository that includes information regarding age, level and current employment type.

The final nominal attributes arecountry of residence, gender and race. The continuous attributes are age, hours workedper week, education number, capital gain and loss and a survey of weight attribute assigned to an individual depends on information such as area ofresidence and type of employment.The performance of the RSA-DSO model is evaluated with metrics such as privacy preservation accuracy, time for optimal data hiding, number of sensitive rules, overall processing time with respect to size of transaction and total number of transactions.

1.4 Privacy preservation accuracy

Three different experiments are conducted. The first experiment shows the relationship between the number of perturbed copies and the privacy preservation accuracy by applying RSA-DSO in Adult dataset for varying number of perturbed copies. In table 1, experimental result are reported with respect to number of perturbed copies for adult dataset. Figure 3 shows the privacy preservation accuracy when the number of perturbed copies are increased.

Table 1 Tabulation for privacy preservation accuracy

Number of perturbed copies	Privacy preservation accuracy (%)		
	RSA-DSO	MT-PPDM	Protocol for secure mining of association rule
10	79.38	64.48	59.36
20	81.52	70.48	60.45
30	83.58	72.54	62.51
40	80.31	69.26	62.24
50	84.32	73.96	68.93
60	88.35	77.43	70.38
70	92.48	81.44	78.41
80	85.14	82.17	80.37
90	90.28	85.11	83.32
100	94.17	87.23	84.37

While generating privacy preserving accuracy rate, with discrete optimized model, the personal best and global best are compared with the fitness function, through which the privacy preserved perturbed copies are obtained. Followed by this, by applying a sigmoid function instead of measuring the traditional support and confidence for measuring the sensitive rules, the privacy preserving accuracy gets increased.

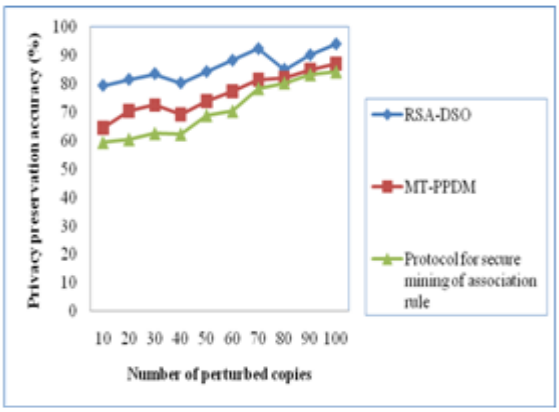

Figure 3 Measure of privacy preservation accuracy with respect to the number of perturbed copies

Figure 3 shows the measure of privacy preservation accuracy with respect the varying number of perturbed copies using RSA-DSO, MT-PPDM and Protocol for secure mining of association rule respectively. From the figure, it is observed that the privacy preservation accuracy shows ascending tendency with the increase in the number of perturbed copies. This is due to the application of DSO algorithm applies a fitness function to extract the sensitive rule to find a solution to discrete optimization problem. This in turn improves the privacy preservation accuracy for sensitive data and rule hiding by 11% than when compared to MT-PPDM. In addition, with the application of DSO algorithm, the transactions and their corresponding velocities are updated using '$pbest$ 'and '$gbest$ ' values with respect to random number, the privacy preservation accuracy is improved by 18% than when compared toProtocol for secure mining of association rule.

1.5 Time for optimal hiding

The second experiment shows the relation between the time for optimal hiding and the total number of transactions (i.e. varying number of transactions) from Adult datasetof UCI repository as shown in table 2. The abalone dataset contains fourteen attributes, six integer and eight categorical attributes. Only the integer attributes were considered for data item and rule hiding to preserve privacy on transaction database for data publishing. The number of instances is 48842.

Table 2 Tabulation of time for optimal hiding

Total number of transactions	Time for optimal hiding (ms)		
	RSA-DSO	MT-PPDM	Protocol for secure mining of association rule
5	3.14	3.85	5.14
10	4.28	5.15	7.32
15	6.89	7.39	9.15
20	9.59	10.25	13.23
25	7.23	8.25	11.43
30	10.43	11.35	13.43
35	13.14	14.82	15.90
40	17.32	20.42	24.13
45	19.25	23.24	28.14
50	25.13	28.90	31.43

In order to measure the time for optimal hiding, both the data hiding time and rule hiding time were considered and their summation was evolved for optimal hiding. It is measured in terms of milliseconds (ms) with

respect to the total number of transactions in the range of 5 and 50.

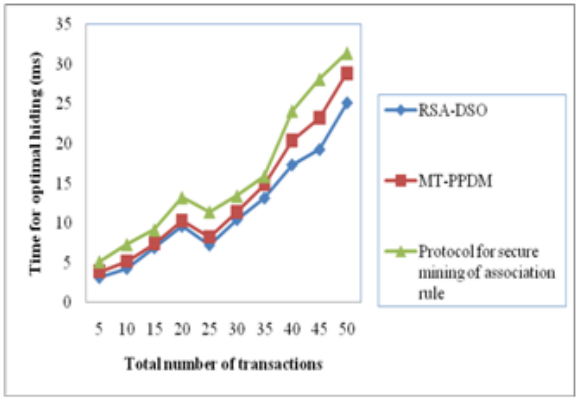

Figure 4 Time for optimal hiding with respect to total number of transactions

Figure 4 depicts the relationship of the time for optimal hiding and the total number of transactions while keeping the population size (i.e. data items) constant. The time for optimal hiding is generated by applying Reinforcement-based Sensitive Data Hiding and DSO algorithm. Figure 4 illustrates that if the total number of transactions is increased, the time for optimal hiding is also increased, however, betterment achieved using RSA-DSO model. Also, the time for optimal hiding was seen as be reduced in between the transactions 20 and 30. This is because, though the total number of transactions increases, the data item to be hidden does not remains the same and differs in a non-linear manner. Therefore, the time for rule generation also varies this in turn has an influential factor for optimal hiding. From Figure, it is seen that by incorporating reinforcement model for data item hiding with the probability of occurrence of sensitive data item to be appeared only once. This in turn results in the improvement of time for optimal hiding. Moreover, RSA-DSO model evaluates and separates the sensitive data items where optimal frequent items are reinforced with the aid of '$Dfactor$', therefore improves the time for optimal hiding by 15% compared to MT-PPDM [1] and 42% compared to Protocol for secure mining of association rule [2] respectively.

1.6 Number of sensitive rules

The third experiment shows the time for number of sensitive rules with respect to different number of items. In table 3, experimental results are reported with respect to number of items for Adult dataset. The figure shows the number of sensitive rules generated by changing the number of items. Sensitive rules are the rules to be hidden. The number of sensitive rules is obtained by the ratio of number of association rules generated to number of items.

Table 3 Tabulation of number of sensitive rules

Number of items	Number of sensitive rules (%)		
	RSA-DSO	MT-PPDM	Protocol for secure mining of association rule
1	61	71	79
2	65	75	82
3	69	79	85
4	65	75	83
5	68	78	86
6	72	82	90
7	67	77	84
8	69	79	87
9	61	81	89
10	63	83	91

The efficiency of the model is measured on the basis of minimum sensitive rules generated and is measured in terms of percentage (%). The table shows the tabulation results for the sensitive rule generation using RSA-DSO, MT-PPDM [1] and Protocol for secure mining of association rule [2] respectively. From the table, it is observed that the RSA-DSO model had comparatively lesser number of sensitive rules generated than other methods,and the RSA-DSO model proves outperforming others.Furthermore, the comparison results also suggest thatthe RSA-DSO represents a new model for data and rule hiding to preserve privacy for data publishing. The result for number of sensitive rules using RSA-DSO, MT-PPDM [1] and Protocol for secure mining of association rule [2] is depicted in figure 5.

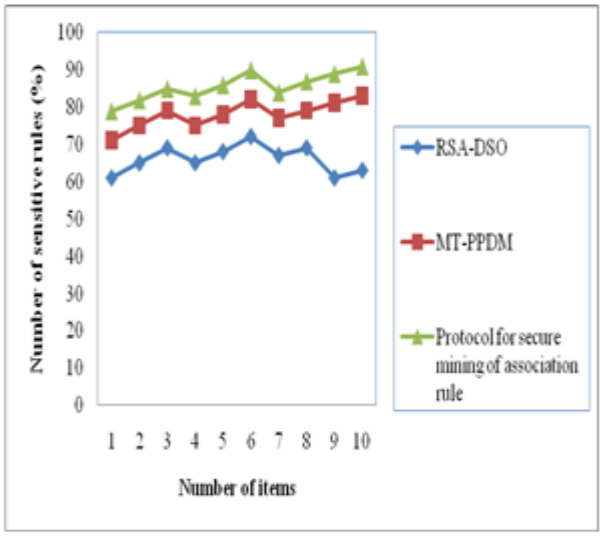

Figure 5 Sensitive rule generation by varying the number of items in Adult dataset

As illustrated in figure 5, the results show that number of sensitive rules generated after construction of an integrated reinforced social ant and discrete swarm optimizationisslightly less when compared with MT-PPDM [1] and Protocol for secure mining of association rule [2] and tested with the Adult dataset. This is because with the integration of social ant and discrete optimization model, optimal number of sensitive data items (with only one occurrence) and discretization of rule generation. This in turn reduces the number of sensitive rules to be generated using RSA-DSO model with respect to vary the number of items. Moreover with the convergence of sigmoid function, increases the randomization of evolution approach and therefore reduces the sensitive rules generated by 18% compared to MT-PPDM and 30% compared to Protocol for secure mining of association rule respectively.

V. CONCLUSION

In this paper, an integrated model with the objective of solving multi-objective factors, namely, data and rule hiding through reinforcement and discrete optimization model called, integrated Reinforced Social Ant and Discrete Swarm Optimization (RSA-DSO) model is presented. The Reinforced Social Ant Sensitive Data Item Hiding model is proposed based on the concept of reinforcement learning to satisfy two different requirements, optimal sensitive data time and high privacy preservation accuracy for data publishing. Next, the Discrete Swarm Optimized Sensitive Rule Hiding is designed based on the sigmoid function to derive the sensitive rules instead of using traditional support and confidence values, ensuring minimum sensitive rule generation. Finally, an integrated Reinforced Social Ant and Discrete Swarm Optimization model results in the advantages of increasing the privacy preservation accuracy for data publishing. Moreover, the experiment results on a small testbed and the simulation results on RSA-DSO model Reinforcement-based Sensitive Data Hiding and DSO algorithm significantly improve the privacy preservation accuracy and also reduces the optimal hiding time to a greater extent.

Future enhancement will focus on overcoming the limitations that are discussed earlier. In future, the particle swarm optimization mechanisms are developed in order to improve the privacy preservation and hence reduce the side effects while data publishing. In addition, DSO algorithm is extended to further optimize the data hiding on the large datasets in PPDM. Finally, the security will be ensured with the data distribution process to achieve a secure data publishing in PPDM.

REFERENCES

1. 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.
2. TamirTassa, "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.
3. FoscaGiannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino 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.
4. ZahidPervaiz, Walid G. ArifGhafoor, NagabhushanaPrabhu, "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.
5. Mohamed Nabeel, Elisa Bertino, "Privacy Preserving Delegated Access Control in Public Clouds", IEEE Transactions on Knowledge & Data Engineering, Volume 26, Issue 9, September 2016, Pages 1-14.
6. 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, 2012, Pages 1-12.
7. SlawomirGoryczka, Li Xiong, Benjamin C. M. Fung, "m-Privacy for Collaborative Data Publishing", IEEE Transactions on Knowledge & Data Engineering, Volume 26, Issue 10, October 2014, Pages 1-10.
8. 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.
9. Chun-Wei Lin, Binbin Zhang, Kuo-Tung Yang and Tzung-Pei Hong, "Efficiently Hiding Sensitive Itemsets with Transaction Deletion Based on Genetic Algorithms", Hindawi, The Scientific World Journal, 2014, Pages 1-13.
10. Yi-Hung Wu, Chia-Ming Chiang, and Arbee L.P. Chen, "Hiding Sensitive Association Rules with Limited Side Effects", IEEE Transactions On Knowledge and Data Engineering, Volume 19, Issue 1, January 2007, Pages 29-42.
11. Nan Zhang, Wei Zhao, "Privacy-Preserving OLAP: An Information-Theoretic Approach", IEEE Transactions On Knowledge And Data Engineering, Volume 23, January 2011, Pages 122-138.
12. Chun-Wei Lin, Tzung-Pei Hong and Hung-Chuan Hsu, "Reducing Side Effects of Hiding Sensitive Itemsets in Privacy Preserving Data Mining", Hindawi, Scientific World Journal, 2014, Pages 1-12.
13. [13] Tong Yi andMinyong Shi, "Privacy Protection Method for Multiple Sensitive
14. Attributes Based on Strong Rule", Hindawi, Mathematical Problems in Engineering, 2015, Pages 1-14.
15. Yves-Alexandre de Montjoye, ErezShmueli, Samuel S. Wang, Alex Sandy Pentland, "openPDS: Protecting the Privacy of Metadata through SafeAnswers", Plos one, Research Article, Volume 9, Issue 7, July 2014, Pages 1-9.
16. Mohamed Ouda, Sameh Salem, Ihab Ali, and El-SayedSaad, "Privacy-Preserving Data Mining in Homogeneous Collaborative Clustering", The International Arab Journal of Information Technology, Volume 12, Issue 6, November 2015, Pages 604-612.
17. Yousra Abdul Alsahib S. Aldeen1, MazleenaSalleh and Mohammad AbdurRazzaque, "A comprehensive review on privacy preserving data mining", Springer, Aldeen et al. SpringerPlus open journal, November 2015, Pages 1-36.

AUTHORS PROFLE

Dr.P.TamilSelvan, completed his Ph.D in Computer Science from Karpagam Academy of Higher Education in 2017. He is working as Assistant Professor in Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore. His experience is 10 yrs. He has presented a paper in International Conference. His research interests are Data mining and warehousing.