ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Privacy Preservation through a Blocking Based Rule Hiding Method

Madhuree Thakar[1], Ashutosh Abhangi[2]

M.E. Student, Dept. of Computer Engineering, Noble Group of Institutions, Junagadh, Gujarat, India[1]

Assistant Professor, Dept. of Computer Engineering, Noble Group of Institutions, Junagadh, Gujarat, India[2]

ABSTRACT: Data mining is the process of extracting valuable from data. As more data is gathered, with the amount of data doubling every three years, data mining is becoming an increasingly important tool to transform this data into information. Large repositories of data contain sensitive information which must be protected against unauthorized access. Various data mining techniques can be used to discover useful knowledge from large collections of data. As the data mining technology has rapidly progressed, getting user's sensitive information through data mining technology has become very easy. This led to increasing concerns about the privacy of the underlying data. Now a days, the privacy preserving data mining has become an essential concern due to the rapid growth of electronic data in governments, corporations and different organizations. Such data may implicitly contain sensitive information and can lead to privacy or security threats if they are misused. Association rule hiding is a subarea of privacy preserving data mining that studies the side effects of data mining methods that generated from the disclose the sensitive information belong to individuals or organizations. However, there is a risk of disclosing sensitive information when data is shared between different organizations. The balance between appropriate mining needs and the protection of confidential knowledge when data is released or shared must be carefully managed. In this thesis we focused on privacy preservation for sensitive data using Blocking Based method, here sensitive data is hide using Cryptographic technique instead of deletion of data by Distortion Based method. So that if we want to reuse that sensitive items in future, we can. . We proposed system that implemented in JAVA and also achieve satisfactory results with fewer side effects and data loss.

KEYWORDS: Data Mining Privacy preserving, Apriori algorithm, Association rule hiding, Encryption

I. INTRODUCTION

Data mining: Data mining tools perform data analysis and may uncover important data patterns, contributing greatly to business strategies, knowledge base and scientific and medical research. The widening gap between data and information calls for a systematic development of data mining tools. So simply you can say **data mining** refers to extracting or "mining" knowledge from large amount of data.. The term data mining also called as KDD (‘Knowledge Discovery in Database Processes’). Data mining work with the five major elements: Extract, transform, and load transaction data onto the data warehouse system, Store and manage the data in a multidimensional database system, analyse the data by application software, Present the data in a useful format, such as a graph or table.

Association Rule: Finding frequent patterns, associations, correlations, or causal structures among sets of items in transaction databases. Which items (itemsets) that have support greater than the minimum support and then using the large itemsets to generate the desired rules that have confidence greater than the minimum confidence. Understand customer buying habits by finding associations and correlations between the different items that customers place in their “shopping basket”.AprioriAlgorithm:The Apriori algorithm takes advantage of the fact that any subset of a frequent itemset is also a frequent itemset. The algorithm can therefore, reduce the number of candidates being considered by only exploring the itemsets whose support count is greater than the minimum support count. All infrequent itemsets can be pruned if it has an infrequent subset.

DOI: 10.15680/IJIRCCE.2017. 0504299

Data Blocking: This technique is using the maximum confidence or not reduces the sensitive rule. In database there are D's and I's must be hidden during blocking, because D's or 1's replace with "?". In some applications where publishing wrong data is not acceptable, then unknown values may be inserted to blur the rules. So, that support of certain items goes down to certain level and rule mining algorithm nit able to mine the sensitive rules. Advantages: It Maintain database, instead of inserting false value to block the original value. Disadvantages: Difficult to reproduce the original database. And it is the various side effects like lost rule, ghost rule, false rule etc.

Border Based Approaches:Border based approach is hide the sensitive association rules by modifying the border in the lattice of the frequent and the infrequent item sets in original database. This approach is make the border between the frequent and infrequent items. That way this border is divided the frequent and in frequent item sets. The first frequent item set hiding methodology that is based on the notion of the border. It maintains the quality of database by greedily selecting the modifications with minimal side effect. Advantages: It maintain the database quality by selecting the modification with the minimal side effect.Disadvantages: The border is not easily identify. Then it is difficult to understand based on the heuristic approaches.

Exact Approaches: Another name of the exact approach is the non-heuristic algorithm which is formulated to constrain satisfaction problem (CSP) and solve by using the binary integer programming (BIP). It Provide the optimal solution to all constrain. It is first used the exact approach for hide the rules. And it provides an optimal solution of rule hiding problem. In also used to hide sensitive rules by formulating constraint satisfaction problem without any side effects with the concepts of positive and negative border sets. By using adopting divide and conquer technique on constraints. Advantages: it gives guarantees to provide the optimal solution without any side effect. Disadvantages: the approach is require the high complexity due to the binary integer programming.

Reconstruction Based Approaches: Reconstruction based approaches is generate privacy for database by using sensitive characteristic from the original database it produce lesser side effect in database. It define the FP tree based algorithm which reconstruct the original database and efficient generate number of secure database. Advantages: Reconstruction based approaches is create the privacy of database and lesser side effect than heuristic based approaches. Disadvantages: The problem is number of transaction is restricted in new database.

Cryptographic Based Approaches: Cryptographic based approaches used in multi-party computation .In which data in distributed from different location. The owner of the database is want to share their data, and at the same time they also want the privacy at their end. Cryptographic Based approaches can be classified two way that is Horizontal partition distributed data and vertical partition distributed data. In horizontal partition distributed data is provide the different rows are placed in different tables that are distributed in different locations. In vertical partition data some column keeps in one table and remaining column in another table. The basic rule of the association rule over horizontal partition distributed database. Efficiently mine association rules over vertically partitioned data. They introduce a partial topology to lower communication cost as much as possible. Empiric al results. Advantages: It provide the security in multi-party computation over the partition database. Disadvantages: It does not provide the security of the output of a computation. Communication and Computation cost should be low

II. RELATED WORK

ShymaMogtaba, EimanKambalet. al [2] In this paper describe overviewed of privacy preservation Data Mining issues and its different techniques. The purpose of the association rule hiding algorithm for privacy preserving data mining is to hide certain crucial information so they cannot be discovered through association rule. And described that the existing association rule hiding algorithm (FHSAR) developed in java code was integrated with Weka in order to balance the trade-off between utility and privacy preservation. This is based on the association rule hiding approach and modifying the database transactions so that the confidence of the association rule can be reduced. Advantages of this paper is the performance of FHSAR improved because of two reasons: First, a heuristic function is used to obtain a prior weight for each transaction, by which the order of transactions modified can be efficiently decided. Secondly, the correlations between the sensitive association rules and each transaction in the original database are analyzed, which can effectively select the proper item to modify. And number of new rules is minimized and independent of the size of database, which can be discovered.

MasoodaModakaet. al [3] In this paper research work focuses on how implemented procedures to mine distributed association rules on horizontally and vertically partitioned data extracting. And Contribution of participating parties individual parties responsible for just encrypting the data and sending it to the controller, the controller generates the large itemsets and generating the rules. A secure protocol and algorithm find the global association rule using homomorphic encryption. Disadvantage if it is used the Apriori algorithm which tends to be slow with very large datasets.

Peng Cheng et. al [4] In this paper, overview of how a hiding method based on evolutionary multi-objective optimization(EMO) is proposed, which performs the hiding task by selectively inserting items into the database to decrease the confidence of sensitive rules bellow specified thresholds. The side effects generated during the hiding process are taken as optimization goals to be minimized. Hype, a recently proposed EMO algorithm, is utilized to identify promising transactions for modification to minimize side effects. Results on real datasets demonstrate that the proposed method can effectively perform sanitization with fewer damages to the non-sensitive knowledge in most cases. To find the optimal subset of transactions for sanitization to hide all sensitive rules and simultaneously minimize side effects accompanied. Because this problem includes four optimization goals, HypE, a fast hyper volume - based algorithm dedicated to many-objective optimization, is utilized to drive the evolution process forward. In this paper EMO-Add Item can perform the sanitization task with fewer knowledge distortions for most test cases.

Bhoomika R Mistry et. al [5] In this research paper reviewed a various approach of association rule hiding techniques. Also describe in this paper but used the Heuristic approach in Data Distortion Technique. MDSRRC algorithm is hide the sensitive association rule with few modification on database. Which maintain the data quality and reduced the side effect of database. MDSRRC algorithm, increase the efficiency and reduced the side effect by minimizing the modification on database and used distributed database to hide the association rule. In this paper to extend the MDSRRC algorithm in distributed database, they tried to make association rule more secure. And I will also try to the MDSRRC algorithm, increase the efficiency and reduced the side effect by minimizing the modification on database. and will used distributed database to hide the association rule.

Peng Cheng et. al [1] In this paper paper studied that privacy preservation in association rule mining. A new distortion- based method is proposed which hides sensitive rules by removing some items in a database to reduce the support or confidence of sensitive rules below specified thresholds. In order to minimize side effects on knowledge, the informa- tion on non-sensitive itemsets contained by each transaction is used to sort the supporting transactions. The candidates that contain fewer non-sensitive itemsets are selected for modification preferably. In order to reduce the distortion degree on data, the minimum number of transactions that need to be modified to conceal a sensitive rule is derived. And it hide sensitive rules with fewer side effects and data loss. sorts the supporting transaction in accordance with their relevance Advantage of this paper is Provides protection of confidential knowledge Using distortion based method hides sensitive rule in order to minimize side effects on knowledge, the information on non-sensitive item sets contained by each transaction is used to sort the supporting transactions. The candidates contain fewer non-sensitive item sets. It can achieve satisfactory results with fewer side effects and data loss. Disadvantage of it is for the hiding sensitive rule by removing or destroying items from database.

III. Proposed Algorithm

A. *Description of the Proposed Algorithm:*

Proposed Work I am trying to hide sensitive rules by encrypt that items using Blocking Based Method. Here I will use RSA Cryptographic algorithm for hide sensitive items and to reduce the support or confidence of sensitive rules using cryptographic. So whenever we need in future we can reuse it by decrypt the items.

RSA is an algorithm used by modern computers to encrypt and decrypt messages. It is an asymmetric cryptographic algorithm. Asymmetric means that there are two different keys. This is also called public key cryptography, because one of them can be given to everyone. The other key must be kept private.

B. *RSA Algorithm steps with example:*

- Choose p = 3 and q = 11
- Compute n = p * q = 3 * 11 = 33
- Compute φ(n) = (p - 1) * (q - 1) = 2 * 10 = 20
- Choose e such that 1 < e <φ(n) and e and φ (n) are coprime. Let e = 7
- Compute a value for d such that (d * e) % φ(n) = 1. One solution is d = 3 [(3 * 7) % 20 = 1]
- Public key is (e, n) => (7, 33)
- Private key is (d, n) => (3, 33)
- The encryption of $m = 2$ is $c = 2^7 \% 33 = 29$
- The decryption of $c = 29$ is $m = 29^3 \% 33 = 2$

IV. PSEUDO CODE

```
Step1:  Database
Step2:  Find frequent itemsets and association rules by applying Apriori algorithm
Step3:  Find out relevance value and filter sensitive rules
Step4:  Sort sensitive rules in descending order by its relevance value
Step5:  Perform Encryption operation for hiding sensitive items from top most
Step6:  Update support and confidence of non-sensitive rules. Also update database.
```

V. SIMULATION RESULTS

Here we compare the results of time with respect to the randomly selected transactions, where time is vary because in the existing system there is remove sensitive items and in proposal system, it converts sensitive items into encrypted form.

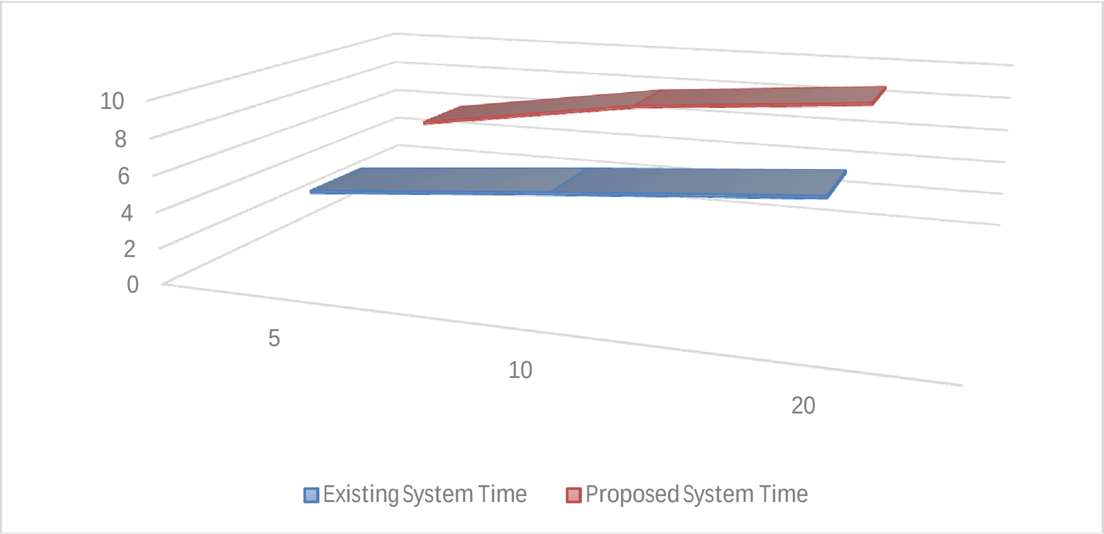

Fig.1.Comparison of Existing System Time v/s Proposal System Time with respect to the Transactions

DOI: 10.15680/IJIRCCE.2017. 0504299

VI. CONCLUSION AND FUTURE WORK

For the privacy preservation, using cryptographic algorithm can help to hide all sensitive rules without removing any items from database. In this thesis we provide more security to the sensitive items without removing items from the database to reduce the support or confidence of sensitive rules below specified thresholds. Encryption operation performs on sensitive items for hiding and if we need in future then perform decryption operation on that then we can reuse it. After hide sensitive rules, it contains fewer non-sensitive itemsets are selected for modification and after that we can also disclose information without risk. So the side effects will be minimized and fewer data loss. In future, hide sensitive data with different types of encryption methods also with the less time for same or more confidentiality.

REFERENCES

1. Peng Cheng, John F. Roddick, Shu-Chuan Chu·, Chun-Wei Lin, " Privacy preservation through a greedy, distortion-based rule-hiding method", © Springer Science+Business Media New York 2015, © Springer Science+Business Media New York 2015
2. ShymaMogtaba(&) and EimanKambal," Association Rule Hiding for Privacy Preserving Data Mining", © Springer International Publishing Switzerland 2016 P. Perner (Ed.): ICDM 2016, LNAI 9728, pp. 320–333, 2016. DOI: 10.1007/978-3-319-41561-1_24
3. Bhoomika R Mistry, Amish Desa " Privacy preserving heuristic approach For Association Rule Mining in Distributed Database",IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIIECS'15, 978-1-4799-6818-3/15/$31.00 © 2015 IEEE
4. Peng Cheng, Chun-Wei Lin, Jeng-Shyang Pan "Use HypE to Hide Association Rules by Adding Items", PLOSONE|DOI:10.1371/journal.pone.0127834
5. MasoodaModaka and RizwanaShaikhb, "Privacy Preserving Distributed Association Rule Hiding Using Concept Hierarchy", www.sciendirect.com Elsevier B.V. , 1877-0509 © 2016
6. PengChng, Ivan Lee LiLi, Kuo-Kun Tseng, Jeng-Shyang Pan 'BRBA:ABlocking-BasedAssociationRuleHidingMethod',Copyright @2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved
7. Mrs. S. Vasanthi, Ms.S. Nandhini, 'Privacy Presrving using Association Roe in Data Mining Techniques', International Jouranal of Advanced Research in Computer and Communication Engeeniring Vol.4, Issue 8, August 2015,DOI 10.17148/IJRCCE.2015.48106
8. https://www.tutorialspoint.com/data_mining/dm_overview.htm
9. https://github.com/kylepolich/apriori-algo-presentation/blob/master/google-scholar-crawl.ipynb
10. http://t0.gstatic.com/images?q=tbn:ANd9GcRzXWBm65nz1ng_u_R-z5cwki1Z-A0-0ymi8HSq_wdskMFqaLI6
11. https://en.wikipedia.org/wiki/RC4
12. *Zhang G, Yang Y, Chen J (2012a) A historical probability based noise generation strategy for privacy protection in cloud computing. J ComputSystSci 78(5):1374–1381.* http://doi.org/10.1016/j.jcss.2011.12.020

BIOGRAPHY

MadhureeThakaris a Research in the area of privacy preservation and study Master of Computer Engineeringin Noble College of Engineering, Junagadhand she completed her Bachelor of Engineering from Dr.Subhash Technical Campus, Junagadh.