Paillier based Privacy-Preserving Mining of Association Rules from Outsourced Transaction Databases

Asavari Smart
Dept of computer Engineering, DCOER,
Savitribai Phule Pune University, Pune, India
asavari_smart@yahoo.co.in

Prof. P. M. Mane
Dept of computer Engineering, DCOER,
Savitribai Phule Pune University, Pune, India
prashantmane15@gmail.com

Abstract— The Cloud computing is computing in which massive assembling of remote servers are managed to authorized centralized data storage and online access to computer resources , while Privacy-preserving data mining (PPDM) is one of the latest inclination in privacy and security studies. It is determined by one of the important positioning issues of the information era - the right to privacy. With the use of cloud computing services, an organization lack in computational resources can deploy its mining requires to an outsider service provider. However, both the elements and the association rules of the deployed database are observed as private property of the organization. The data owner converts its data and sends it to the server, ships mining queries to the server, and recoup the actual design from the extricate designs received from the outsider server for corporate privacy prevention. In this theory, we study the problems of outsourcing the association rule mining mechanisms within a corporate privacy-preserving framework. The Rob Frugal method is founded with defeat the security obligations of outsourced data. This method is an encryption plan which is based on one to one substitution ciphers for items and fake pattern from the database. In this system attacker discovers data by guessing attack, also man in the middle attack which is possible on Rob Frugal encryption to conquer this problem, the proposed technique encompasses Paillier encryption for enhancing the security level for outsourced data with the less complexity and to protect against the forging the contents of the correspondence. FP-growth algorithm is used for generating association rules for improving the performance and for preserving a homomorphic encryption algorithm Paillier cryptosystem is being used.

Keywords*- Association rules mining, privacy-preserving outsourcing, Paillier Encryption, FP-Growth algorithm.*

*****

I. INTRODUCTION

Cloud computing provides an uncommonly changeable surrounding that allow a immensely flexible environment that validate on-demand infrastructure abilities over the internet to ascendable computing resources (systems, servers, storage, services, and applications)that can be delivered quickly with expense proficiency and minimum attempts. Prototype of mining and administration of information as service will probably rise as popularity of cloud computing develops. Privacy preserving data mining; operates data mining on union of two parties. Data remains private that is no party analyzes nothing but results. Consideration is build that it consist massive databases-Generic provision which is impossible and little loyal parties is exist. PPDM works with preserving the privacy of single data or delicate knowledge without dodged the advantage of the information.

The data mining service structures, focused at enabling enterprises with restricting computational resources and data mining prowess to outsource their data mining requirements to a third party service contributor[2][3]. In this theory, we analyzed in deploying the association rule mining work between a corporate privacy preserving patterns [8]. A real body of task has been complete on privacy-preserving data mining (PPDM) in a various contexts. To protect corporate privacy, the data owner will operates Rob Frugal encryption design [12] and converts it to the server. Data holder will ships the mining queries to the server and recover, true structure from the extricated structure collected from the server, but on Rob Frugal encryption approximation attack and man in the middle attack are possible, and to conquer these attacks Homomorphic encryption algorithm- Paillier cryptosystem is implemented.

First, we used an encryption scheme; rob frugal method as available. Over that we used Paillier encryption. Encrypted data and then send to server for processing. Encrypt/Decrypt module can utilize to transform client data to server and vice versa. To allow the E/D module to recover the true patterns and their correct support, it makes and keeps a compact structure, called synopsis. FP-Growth algorithm is used to generate association rule which has better performance than Apriori algorithm. At the end, we conduct experimental analysis of our schema using a large real dataset. Our results demonstrate that our encryption pattern is successful, scalable, and achieve the desired level of privacy.

II. LITERATURE SURVEY

A. Related Work

A. Association rule mining: - An immense execution of the secure of data and knowledge are implemented in association rule mining techniques. Presently, privacy preserving association protocol mining algorithms are separated into three techniques as per privacy protection techniques [4].

1) Heuristic based technique: - Heuristic based techniques normally implemented for the complex event. This method addresses the issue of dataset marking to needs data restoration. This method is frequently recommended for the data alteration. For that, data distortion techniques are recommended which implemented data distortion methods for the alteration of private data [5].

2) Re-construction based association rule mining: - For operating the association rule mining various techniques have issues of privacy preserving. Hence these algorithms are used for scattered data and then rebuild the dispensation. To operate

types of data, Agrawal et al. in 2002 presented this concept on restructured-based association rule [7].
3) Cryptography-Based techniques:-This method is frequently implemented for data encryption. There are multiple Cryptography-based concepts were presented in the account of privacy securing data mining algorithms. Cryptography based concepts like Secure Multiparty Computation (SMC) are protected at the end of the computations.

B. *Challenges in Privacy preserving Technique Algorithms*

The provocation in privacy preserving association algorithm for association rule concealing are data deficit, expensive, regain original data after concealing and should be effective enough for massive datasets [1].
1) Loss of information: It is described as the ratio between the total of without an error the failing made in organize the Prevalence of the matters from a refined database and the wholly of each last one of prevalence of elements in the initial database is called as the data loss [1].
2) Expensive Protocols: It is concentrated within encryption are initially costly of the fact that they need senile encryptions for every single bit [1].
3) Regain authenticate information after concealing Privacy preserving mining contains of many of policies to recuperate the information from the massive database which also contains of sensitive information [1].
4) Reinforce of massive datasets a many of information can now be fluently acquire to and keep away in focus of huge advances throughout there. Currently days, databases with data warehouse saves and supervise measures of data which are desert immense. Subsequently, a PPDM computation must be scheduled and accomplish with the capability of taking care of massive datasets that may at current ongoing expansion [1].

C. *Requirements of a PPDM algorithm*

1) Accuracy - Because of the concealing the accuracy is nearly recognized with the data loss technique: PPDM algorithm has required keeping up high precision to lessen data loss [1]. The less is the data loss; the better is the information quality.
2) Completeness and Consistency - Completeness evaluates the stages of fumbled information in the fumigate database. Inadequate information has a great impact on data mining comes about and enfeeble the information mining computing from provided an exact demonstration of the basic information [1].
3) Scalability -It is a substitute censorious aspect to research the operation of a PPDM algorithm. Actually, ascendable displays the expertise structure when data volume exceeds [1].
4) Data quality - It is an important block of PPDM. Superior quality data that has been ordered specifically for data mining tasks will escort about valuable data mining models and produce. On the other hand, low quality data has a significant negative impact on the benefits of data mining output [1].
5) Security - It is the stages of assurance against anguish, loss of data, and wrong doing. There are two principle theories with respect to how to arrange the problems of protection that appear today. The primary is a lawful and administration concept whereby consortiums are constrained by the way they save and utilization data concentrated around security law and open strategy. [1].

D. *k-support anonymity*

The framework knowledge such as the supports of repeated elements groups can be performed to gain privacy data in the outsourcing of repeated element set mining. In this paper [11], k-support anonymity to avail security against knowledgeable intruders with actual support data is presented. To acquire k-support anonymity, they present a pseudo taxonomy tree and have the third party mine the normal repeated elements sets instead. The fabricate of the pseudo taxonomy tree provides concealing of the authentic elements and restricts the forge elements presented in the encrypted database. The experimental output presented that the techniques of k-support anonymity gain very good privacy insurance with average storage overhead.

III. IMPLEMENTATION DETAILS

We presented homomorphic Paillier encryption and FP Growth association rule creation methods for privacy persevering mining of association rules from outsourced transaction database. The implementation details of proposed system are shown below.

A. *System Architecture*

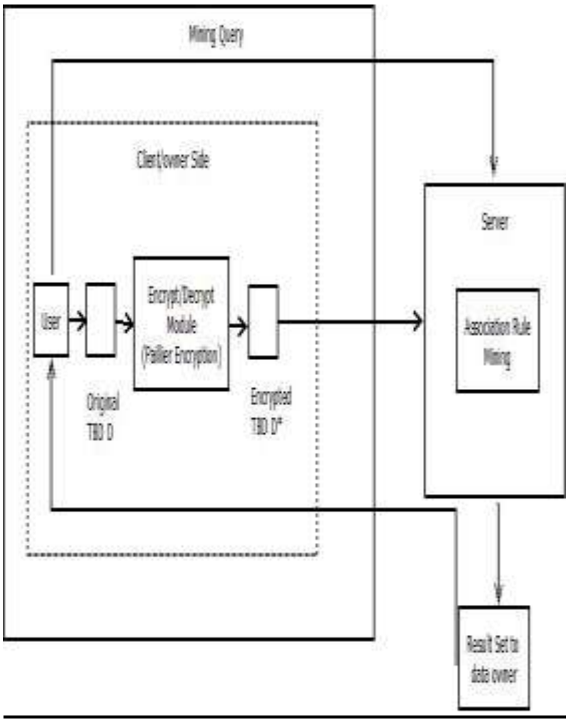

Fig.1: System Architecture

B. *Encryption:*

The section shows the view of rob frugal encryption patterns it uses 1-1 substitution cipher technique which converts authenticate transaction database D into its encrypted version D*.To enhance the privacy, rob frugal encrypted data further double encrypted with Paillier encryption with the help of fake transactions of rob frugal encryption. Paillier algorithm allows computations to be performed on data without decryption. This method provides maximum privacy for man in middle attack and guessing attack and replay attack. Paillier algorithm for encryption and FP growth algorithm for generating rules which gives better performance than existing system are given below.

- *Paillier Encryption*

1) Key generation :

a) Select two large prime numbers a and b arbitrary and independent of each other such that gcd (n, Φ (n)) = 1, where Φ (n) is Euler Function and n=ab.

b) Calculate RSA modulus n = ab and Carmichael's function is given by λ = lcm (a-1, b-1).

c) Select g called generator where $g \in \mathbb{Z}^*_{n2}$
Select α and β randomly from a set $\mathbb{Z}_n^*$
Then calculate $g = (\alpha n + 1)\ \beta^n mod\ n^2$.

d) Compute the following modular multiplicative inverse μ = (L (g^λ mod n^2)$^{-1}$ mod n. Where the function L is defined as L (u) = $(u-1)/n$.

The public (encryption) key is (n and g).
The private (decryption) key is (λ and μ).

2) Encryption:

Let mess be a message to be encrypted
Where mess $\in \mathbb{Z}_{n.}$

a. Select random r where $r \in \mathbb{Z}^*_{n2.}$

b. The cipher text can be calculated as:
$\text{Cipher} = g^{mess} \cdot r^n.\text{mod } n^2.$

3) Decryption:

a. Cipher text c $\in \mathbb{Z}^{*\ 2}_{n}$

b. Original message: mess = L (cipher$^\lambda$ mod n^2).μ mod n.

- *Association Rule Generation(FP-Growth)*

Input: Built FP-tree
Output: complete set of frequent patterns
Method: Call FP-growth (FP-tree, null).
Procedure FP-growth (Tree, α)

```
{
    1)  If the event that Tree contains a single path P then
    2)  **For each** $\beta$ = comb. of nodes in P **do**
    3)  pattern = $\beta \cup \alpha$
        sup = min(sup of the nodes in $\beta$ )
    4)  **else**
        **for each** $a_i$ in the header of Tree **do {**
    5)  generate pattern = $\beta \cup \alpha$
        sup = $a_i$.support
    6)  construct $\beta$'s conditional pattern base
        FPTree = construct $\beta$'s conditional FP-tree
    7)  If Tree $\beta$ = null
        Then call FP-growth (Tree $\beta$, $\beta$)}
}
```

C. *Decryption:*

After encrypted data sending to server, server performs required actions like rule generation and send it back to client. While in network man in middle attack was possible but Paillier encryption resolves that problem. Client decrypts the rules generated by server, removes the effect of fake transaction using compact synopsis and gets the original rules

IV. RESULT AND DISCUSSION

a. *Experimental Setup*

The system is built using Java framework (version jdk 8) on Windows platform. The Net bean (version 8.0) is used as a development tool. The system doesn't require any specific hardware to run; any standard machine is capable of running the application.

b. *Results*

In this section we discussed the results obtained for the proposed system. Here we discussed the comparison graph between the existing and proposed system.

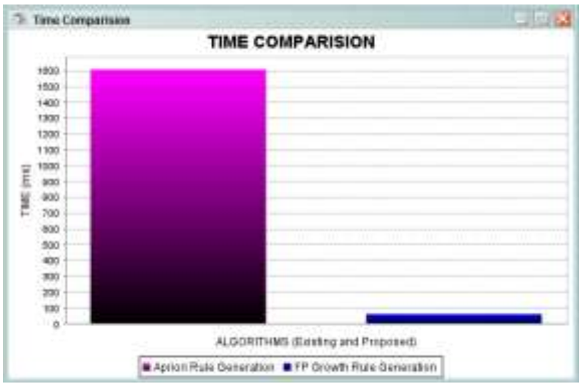

Figure 2: Time Comparison Graph

The above graph shows the time comparison graph of the proposed system. It shows the time required for rule generation for both the algorithm FP growth and Apriori algorithm. From the above graph it is conclude that time required for FP-growth is less than the time required for Apriori algorithm.

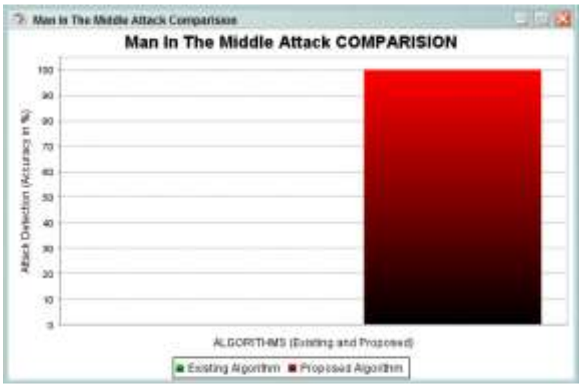

Figure 3: Man in Middle attack Comparison Graph

The above graph shows the man in middle attack comparison graph. Existing graph shows attack detected by the existing algorithm and proposed algorithm. Comparison graph shows that proposed algorithm detect 100% man in middle attack.

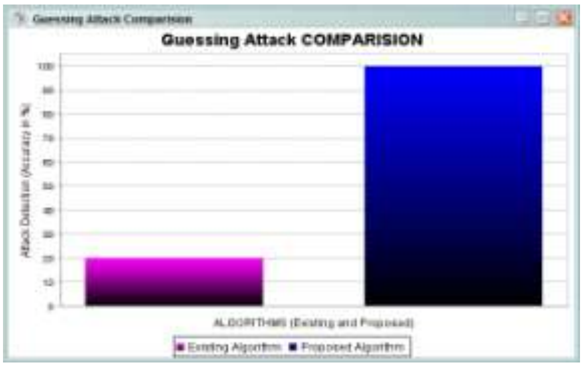

Figure 4: Guessing Attack Comparison Graph

The above graph shows the guessing attack comparison graph. Existing graph shows attack detected by the existing algorithm and proposed algorithm. Comparison graph shows that proposed algorithm detect 100% man in middle attack.

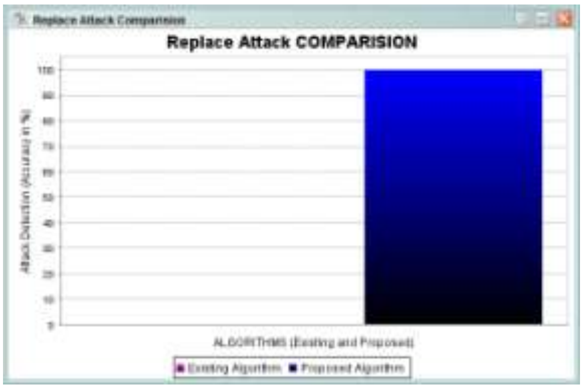

Figure 5: Replay attack Comparison Graph

The above graph shows the replay attack comparison graph. Existing graph shows attack detected by the existing algorithm and proposed algorithm. Comparison graph shows that proposed algorithm detect 100% man in middle attack.

CONCLUSION

System represents a set of encryption methods for Transactional databases that are suitable for outsourcing association rule mining. Starting from a simple one-to-one substitution cipher, which is susceptible to attacks, we utilize Paillier Homomorphic encryption algorithm which provide better security than existing rob frugal algorithm. Also for association rule generation FP-Growth algorithm is used which has better performance than Apriori. Results show that our encryption technique is very robust to attacks as opposed to simple one to- one cipher, which can be easily broken with the help of background knowledge. Also man in the middle attack and guessing attack are not possible as system uses Paillier encryption techniques. Finally, through experimentation the proposed system has better performance in terms of time and security and rule generation.

REFERENCES

[1] R.Natarajan,Dr.R.Sugumar,Mahendran,K. Anbazhagan , “A survey on Privacy Preserving Data Mining”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 1,MARCH 2012.

[2] W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis, “Security in outsourcing of association rule mining,” in Proc. Int. Conf. Very Large Data Bases, 2007, pp. 111-122.

[3] L. Qiu, Y. Li, and X. Wu, “Protecting business intelligence and customer privacy while outsourcing data mining tasks,” Knowledge Inform. Syst., vol. 17, no. 1, pp. 99-120, 2008.

[4] Vassilios S. Verykios, Elisa Bertino1 4 et al., “State-of-the-art in Privacy Preserving Data Mining,” SIGMOD Record, Vol. 33, No. 1, March 2004, pp.50-57.

[5] Agrawal, R., and Srikant (2007), “Privacy Preserving Data Mining”, Proceedings of the 19th ACM International Conference on Knowledge Discovery and Data Mining, Canada, pp. 439-450.

[6] Chris Clifton, Murat Kantarcioglou,XiadongLin,and Michaed Y.Zhu, “Tools for privacy preserving distributed data mining,” a SIGKDD Explorations 4 (2002), no. 2.

[7] Evfimievski A,Srikant R,Agrawal R, et al. , “Privacy preserving mining of association rules,” In: Proc. of t he Eighth ACM SIGK2DD International Conference on Knowledge Discovery and Data Mining, ACM Press,2002, pp.217-a228.

[8] C. Clifton, M. Kantarcioglu, and J. Vaidya, “Defining privacy for data mining,” in Proc. Nat. Sci. Found. Workshop Next Generation Data Mining, 2002, pp. 126133.

[9] Christian Borgelt ,“An Implementation of the FP-growth Algorithm” Department of Knowledge Processing and Language Engineering School of Computer Science, Otto-von-Guericke-University of Magdeburg Universitatsplatz 2, 39106 Magdeburg, Germany.

[10] P. K. Prasad and C. P. Rangan, “Privacy preserving birch algorithm for clustering over arbitrarily partitioned databases,” in Proc. Adv. Data Mining Appl., 2007, pp. 146-157.

[11] C. Tai, P. S. Yu, and M. Chen, “K-support anonymity based on pseudo taxonomy for outsourcing of frequent itemset mining,” in Proc. Int. Knowledge Discovery Data Mining, 2010, pp. 473□482.

[12] F. Giannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang, “Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases”, sept.2013.