International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

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

Pradnya Nagrare[1], Rashmi Tijare[1], Sneha Ghorude[1], Prajakta Chahande[1], Prof. Leena Patil[2]
[1]Student, Department of Computer Science & Engineering, Priyadarshini Institute of Engineering & Technology, Nagpur, India
[2]Professor, Department of Computer Science & Engineering, Priyadarshini Institute of Engineering & Technology, Nagpur, India

Abstract**—***Distributed computing utilizes the perfect model of data mining-as-an administration, utilizing these it is by all accounts a conspicuous decision for organizations saving money on the cost of adding to secure, oversee and keep up an IT foundation. An association/store ailing in mining capacity can outsource its mining needs to specialist co-op on a cloud server. Be that as it may, both the affiliation guidelines and thing set of the outsourced database are seen as private property of the association. The information proprietor encodes the information and sends to the server to safeguard the corporate privacy. Customer sends mining questions to server, and after that server conducts information mining and sends scrambled example to the customer. To get genuine example customer unscrambles scrambled example. In this paper, we consider the issue of outsourcing the affiliation governs mining assignment inside a corporate privacy preserving framework. Consequently Privacy Preserving Data Mining is an examination territory worried with the security decided from by and by identifiable data when considered for information mining. The Rob Frugal encryption procedure is acquainted with beat the security vulnerabilities of outsourced data, which is centered on balanced substitution figures for things also, including fake examples for database. Be that as it may, it contains different fake examples which increment the limit overhead. To beat this issue, the proposed technique incorporates expansion of weighted support in unique support of things to decrease the quantity of fake examples and to upgrade the security level for outsourced data with less multifaceted nature. The fake exchange table information is changed over into lattice configuration to lessen the capacity overhead. Likewise the speculating assault and man in the center assault are conceivable on essential Rob cheap calculation. To defeat these assaults we use Pallier Encryption on after Rob economical encryption plot with a specific end goal to give privacy preserving outsourced mining. In our proposed work we enhanced the security as thing and thing set construct assault are impractical in light of the framework; moreover we decrease the handling time.*

Keywords**—** *Cloud computing, association rule mining, privacy-preserving outsourcing, elliptic curve diffie hellman, rob frugal.*

I. INTRODUCTION

Distributed computing will be processing in which vast gatherings of remote servers are arranged to permit incorporated information stockpiling and online access to PC administrations or assets. With the entry of distributed computing and its model for IT administrations in view of the web and enormous server farms, the outsourcing of information and processing administrations is gaining a novel significance, which is positively required to soar sooner rather than later. In business, outsourcing includes the contracting out of a business procedure to another gathering. Outsourcing plans to give an administration in a corporate privacy preserving structure. Privacy assurance is the primary issue in information mining. Associations, by and large, would prefer not to impart their own private information to different organizations. The thought is that information is distributed by Client for the advantage of permitting experts to mine encoded designs from the scrambled database. As a representation, the value-based database from various associations can be transported to an untouchable which gives mining administrations. The association administration would incline toward not to use an in-house gathering of information mining pros. Also, intermittently information is sent to the server or specialist co-op who is accountable for keeping up the encoded information and directing mining on it because of solicitations from organization examiners of the organization administration. The information proprietor is a customer and the server is alluded to as the specialist co-op. One of the essential issues with this standard is that the server has section to important data of the proprietor and may reveal delicate data from the information. For instance, by taking a gander at the exchange database, the server can infer or reveal which items or things are co-acquired and in this manner, the mined scrambled examples that depict the association clients' subtle elements.

In this paper, we concentrate the issue of outsourcing the affiliation govern mining assignment inside a corporate security preserving structure [7]. Along these lines, Privacy Preserving Data Mining is an investigation range worried with the security decided from really identifiable information when considered for information mining. In this unique circumstance, both the deal exchange database and the mined encoded examples and every one of the points of interest of the organization that can be separated from the information are the property of the organization administration and ought to stay safe from the server and whatever other assailant. In actuality the information mined from the information can be utilized from the organization administration in essential advertising choices to enhance their administrations. An organization or information proprietor needs their information to be mystery however an organization does not have adequate digging mastery for information mining, for this we make the accompanying commitments. We build up an encryption plot, called Improved RobFrugal in that the Encrypt/Decrypt module can utilize to change customer information before it is

Pradnya Nagrare, Rashmi Tijare, Sneha Ghorude, Prajakta Chahande, and Prof. Leena Patil, "Paillier based privacy-preserving mining of association rules from outsourced transaction databases," *International Research Journal of Advanced Engineering and Science*, Volume 2, Issue 1, pp. 186-189, 2017.

delivered to the server. Second, to permit the E/D module to recuperate the genuine examples and their right support, we recommend that it makes and keeps a minimized structure, called outline. Third, we present expansion of weighted support in unique support of things and framework development of fake exchange to lessen the capacity overhead. Fourth with a specific end goal to give privacy preserving outsourced mining, we use Pallier Encryption after Improve Rob Frugal encryption plot. With utilization of ECDH calculation speculating assault and man in the center assault are unrealistic on our proposed framework. Fifth, for better execution Enhance FP-Growth calculation is utilized rather than Apriori calculation [13] for affiliation control era. Finally, we coordinate test examination of our example using an expansive genuine dataset, our outcomes show that our encryption composition is compelling, adaptable, and achieve the coveted level of security.

II. Literature Survey

A. Substitution cipher techniques:

W. K. Wong et al. [1] proposed substitution figure strategies in the encryption of value-based information for outsourcing affiliation run mining. In the wake of perceiving the non-unimportant risks to the unmistakable coordinated thing mapping substitution figure, we propose a more secure encryption arrange in view of a one-to-n thing mapping that changes exchanges non deterministically, yet ensures adjust unscrambling. They build up a successful and productive encryption calculation dependant on this technique.

B. Data Perturbation:

There are two methodologies that can secure touchy data. Is to an encryption capacity that changes the first information to a totally new organization [4], [2]. The second could be to apply information irritation, which alters the first crude information haphazardly [3]. The bother approach is less appealing since it could just give surmised comes about; by the by, the work of encryption permits similar principles that they are recuperated.

C. k-Support Anonymity

The foundation learning, for example, the backings of incessant thing sets can be used to acquire privacy data in the outsourcing of regular thing set mining. In this paper [5], C. Tai, P. Yu proposed k-bolster namelessness to give security against an educated aggressor with correct bolster data. To accomplish k-bolster secrecy, they present a pseudo-scientific categorization tree and host the third gathering mine the summed up incessant thing sets. The build of the pseudo-scientific classification tree encourages stowing away of the first things and limits the fake things presented in the encoded database. The test comes about demonstrated that the strategies for k-bolster secrecy accomplish great privacy protection with direct stockpiling overhead.

D. Corporate Privacy Preserving Mining

In this paper [6], F. Giannotti, A. Monreale concentrated the issue of privacy preserving mining of continuous examples on an encoded outsourced exchange database. We acknowledge a dynamic model where the foe knows the area of things and their correct recurrence and can use this figuring out how to distinguish figure things and figure thing sets. We proposed an encryption plot, called Rob Frugal that depends on 1-1 substitution figures for things and adding fake exchanges to make each figure thing offer a similar recurrence. It utilizes the reduced summation of the fake exchanges from which the genuine support of mined examples from the server can be effectively recuperated.

E. Association Rule Mining by Evmievski

Evmievski et al. [8] proposed an approach, for directing the privacy preserving affiliation govern mining. Kargupta et al. [9] proposed a technique in view of arbitrary lattice ghastly separating to recuperate unique information from the annoyed information. Huang et al. [10] proposed assist, the two information reproduction strategies, first PCA-DR and second, MLE-DR.

F. Randomized Response

The principal individual to propose the Randomized Response (RR) was Warner [14]. The RR plan was at first created in the insights group. It used to gather the data from people with the end goal that, the overview questioners and the information processors don't know which of the two option inquiries are respondent have replied. In information mining, the strategy for randomization is a basic method, can be effortlessly connected at information gathering time. It was a valuable method for concealing individual information in privacy holding information mining. The randomization technique is more productive [11]. However, it brings about high data misfortune. The writing on Privacy-Preserving Mining of Association principles can be arranged into Pattern mining assignment, privacy demonstrate, lastly Encryption/Decryption plot.

III. Methodology

We displayed Homomorphic Paillier encryption and FP-Growth affiliation run creation strategies for privacy driving forward mining of affiliation standards from outsourced exchange database. The execution points of interest of proposed framework are demonstrated in figure 1.

We embrace a moderate recurrence based assault display in which the server knows the correct arrangement of things in the proprietor's information and moreover, it likewise knows the correct support of ever thing in the first information. It was one of the early chips away at protecting against the recurrence based assault in the information mining outsourcing situation. It has been presented utilizing fake things to safeguard against the frequency based assault it was deficient with regards to a formal hypothetical examination of protection assurances and has been appeared to be defective as of late in, the technique for breaking the proposed encryption is given. Along these lines, in our past and preparatory work, we proposed to tackle this issue by utilizing k-protection, i.e. in that everything in the outsourced dataset ought to be indistinguishable from at any rate k − 1 things with respect to their support. The engineering behind our model is represented in figure 1.

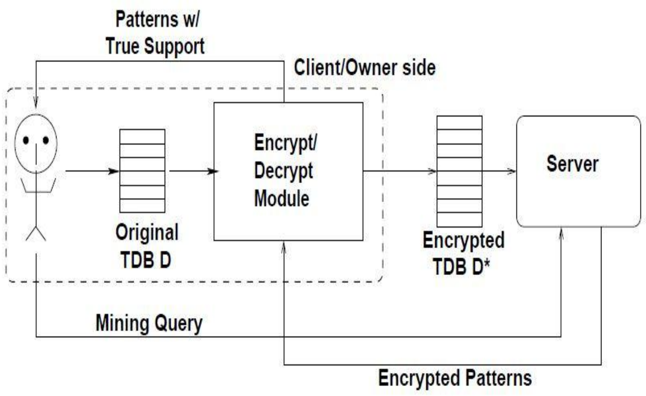

Fig. 1. System architecture.

The customer scrambles its information utilizing an encode/unscramble module in protection saving, this module can be fundamentally regarded as a "black box" from its point of view. The server conducts information mining and sends the (encoded) examples to the proprietor. The propose encryption plot has the property that the returned backings are not genuine backings. In the propose framework the E/D module recoups the genuine personality of the returned designs too their actual backings. The (E/D) module insignificant to demonstrate that if the information is scrambled utilizing 1-1 substitution figures, In the figure content many figures and subsequently the exchanges and examples can be broken by the server with a high likelihood by propelling the recurrence based assault. In the propose strategy devise encryption plans with the end goal that formal protection certifications can be demonstrated against assaults led by the server utilizing foundation information. Initially, we formally characterize an assault demonstrate for the enemy and make exact the foundation learning the foe may have. Our idea of security requires that for each figure message thing, there are in any event k−1 unmistakable figure things that are indistinct from the thing with respect to their backings Second, we make an encryption conspire, called RobFrugal that the E/D module can utilize to change customer information before it is sent to the server. Third, to permit the E/D module to recuperate the genuine examples and their right support of information thing, we suggest that it makes and keeps a smaller structure, called synopsis. We additionally give the E/D module with an effective system for incrementally keeping up the outline against updates as appends.

IV. Algorithms

A. *Rob Frugal Encryption*

1) 1 to 1 substitution cipher :

The method which transformed original transaction database D into its encrypted version D^*. To improve the security fake transaction are added with encrypted database. Table 1(a) shows original transaction while table 1(b) shows transaction after one to one substitution (encrypted)

Table 1 (a): TDB

TDB
Soda Nuts
Soda Milk
Milk Soda
Nuts Milk
Soda Dates
Nuts Soda
Soda Egg
Nuts Cake
Cake

Table1 (b): TDB*

TDB*
e6 e5
e6 e4
e4 e6
e5 e4
e6 e2
e5 e6
e6 e3
e5 e1
e1

2) Support calculation :

This approach was started with calculation of support of the items. Support count is the number of time the items occurred in the original transaction database.

3) Frugal grouping :

Table 2: Descending order of items based on their item support.

Item	Support
e6	6
e5	4
e4	3
e1	2
e3	1
e2	1

4) Robust k-Grouping method (Rob Frugal Grouping) – Where k be the group size (i.e 2 or 3), Here we consider the group size as 2.

Given the items support table, from a group of size k such that no two items from any original transaction comes adjacent to each other i.e. we can't group e6,e5 or e6,e2 as they occurs adjacent in original transaction.

After K-grouping method we get output as:

Table 3: Rob Frugal with K-robust grouping.

Item	Support	Noise (Difference)
e6	6	0
e1	2	4
e5	4	0
e3	1	3
e4	3	0
e2	1	2

5) Fake transaction construction:

We can construct Fake transaction by adding Noise in to original transaction i.e. we can add e1 4 times, similarly e3 3times and e2 2 times in original transaction, so we generate transaction from given noise as {e1, e3, e2}, {e1, e3}, {e1}

B. *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 = (αn + 1) $\beta^n mod\ n^2$.

d) Compute the following modular multiplicative inverse μ = (L (g^λmod $n^2)^{-1}$ 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:

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

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

c. The cipher text can be calculated as:

$$\text{Cipher} = g^{mess} \cdot r^n.\text{mod } n^2.$$

3) Decryption:

a. Cipher text c ∈ $\mathbb{Z}^*_n{}^2$

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

C. 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 β = comb. of nodes in P do
    3) pattern = β ∪ α
       sup= min (sup of the nodes in β )
    4) else
       for each $a_i$ in the header of Tree do {
    5) generate pattern = β ∪ α
       sup= $a_i$.support
    6) construct β's conditional pattern base
       FPTree = construct β's conditional FP-tree
    7) If Tree β = null
       Then call FP-growth (Tree β, β)}
}
```

V. Experimental Result

Following are the results obtained during the implementation phase:

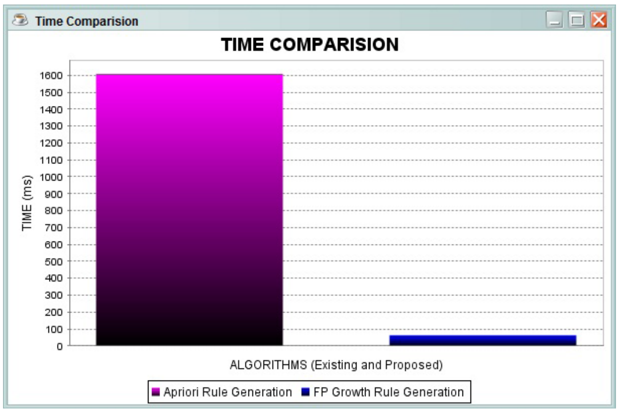

Fig. 2. Time comparison for association rule generation.

Figure 2 compares the performance between FP-Growth, and Apriori Algorithm. Graphs show the execution time of implementations over the various instances.

VI. Conclusion

System represent a set of encryption methods of encryption strategies for Transactional databases that are appropriate for outsourcing affiliation govern mining. Beginning from a straightforward balanced substitution figure, which is powerless to assaults, we use Paillier Homomorphic encryption calculation which gives preferable security over existing loot thrifty calculation. Likewise for affiliation govern era FP-Growth calculation is utilized which has better execution than Apriori. Represents about demonstrate that our encryption system is extremely vigorous to assaults instead of basic coordinated figure, which can be effortlessly broken with the assistance of foundation information. Additionally man in the center assault and speculating assault are unrealistic as framework uses Paillier encryption strategies. At long last, through experimentation the proposed framework has better execution regarding time and security and lead era.

References

[1] 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*, pp. 111-122, 2007.

[2] G. I. Davida, D. L. Wells, and J. B. Kam, "A database encryption system with sub keys," ACM TODS, vol. 6, issue 2, pp. 312-328, 1981.

[3] J. He and M. Wang, "Cryptography and relational database management systems," in *IDEAS*, 2001.

[4] B. Iyer, S. Mehrotra, E. Mykletun, G. Tsudik, and Y. Wu, "A framework for efficient storage security in RDBMS," in *EDBT*, 2004.

[5] C. Tai, P. S. Yu, and M. Chen, "K-support anonymity based on pseudo taxonomy for outsourcing of frequent item set mining," in *Proc. Int. Knowledge Discovery Data Mining*, pp. 473-482, 2010.

[6] F. Giannotti, L. V. Lakshmanan, A. Monreale, D. Pedreschi, and H.Wang, "Privacy preserving data mining from outsourced databases," in *Proc. SPCC2010 Conjunction with CPDP*, pp. 411-426, 2010.

[7] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," *IEEE Trans. Knowledge Data Eng.*, vol. 16, no. 9, pp. 1026-1037, 2004.

[8] S. J. Rizvi and J. R. Haritsa, "Maintaining data privacy in association rule mining," in *Proc. Int. Conf. Very Large Data Bases*, 2002.

[9] A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrke, "Privacy preserving mining of association rules," *Information System*, 2004.

[10] H. Kargupta, S. Datta, Q. Wang, K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques," in *Proceedings of the 3rd International Conference on Data Mining*, 2003.

[11] Z. Huang, W. Du, and B. Chen, "Deriving private information from randomized data," in *Proceedings of the ACM SIGMOD Conference on Management of Data*, 2005.

[12] S. L. Warner, "Randomized response: A survey technique for eliminating 999999999999Evasive Answer Bias," *J. Am. Stat. Assoc.*, 1965.

[13] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrk, "Privacy preserving mining of association rules," in *Proceedings the 8th ACM SIGKDD International Conference on Knowledge Discovery in Databases and Data Mining*, 2002.

[14] R. R. Ahirwal, M. A. Samrat Ashok, "Elliptic curve diffie- hellman key exchange algorithm for securing hypertext information on wide area network," *(IJCSIT) International Journal of Computer Science and Information Technologies*, vol. 4, issue 2, pp. 363-368, 2013.