International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 8 | ISSN : 2456-3307

Secure Outsourcing Association Rule Mining in Horizontally and Vertically Partition Database Using Eclat and Double Encryption Technique

Rutuja Thite, Dr. M. U. Kharat
Department of Computer Engineering, Bhujbal Knowledge City, Nashik, Maharashtra, India

ABSTRACT

Cloud computing uses the ideal model of information mining-as-an service, utilizing these it seems to be an obvious choice for companies saving on the cost of contributing to secure, manage and keep up an IT infrastructure. An association or company who is lacking in mining capacity can outsource its mining needs to third party service providers. Be that as it may, each the association rules and item-set of the outsourced database are seen as private property of the association (company). The data owner encrypt the data and sends to the server to protect the corporate security. Data owner or client transfers its mining queries to server, and afterward server conducts mining task & encrypt rules and sends generated association rules to the data owner or client. To get genuine pattern client decrypts the received rules. Paper focuses on the issue of outsourcing the rule mining task inside a corporate privacy preserving framework. It additionally shows the core idea of privacy preserving association rule mining on vertically partitioned data with utilization of enhanced cryptographic technique. The strategies incorporate cryptographic techniques to minimize the data shared, while adding minimal overhead to the mining task. This research tries to propose desirable algorithm for both vertically as well as horizontally partitioned data. A technique for solving a main problem of privacy preserving association rule mining in two party databases is proposed. To improve the performance of system horizontal partitioning as well as vertical partitioning of data is performed, also double encryption technique is used to increase the security of dataset which includes homomorphic encryption algorithm followed by asymmetric algorithm

Keywords : Data Mining, Association Rules Generation, Vertically and Horizontally Partition Data, Encryption Techniques.

I. INTRODUCTION

Data mining advancement has created as techniques for identifying patterns and patterns from huge amounts of data. Mining incorporates distinctive algorithms, for instance, classification, clustering, rule mining and sequence detection, etc. Previously, each of these algorithms have been made inside a centralized model, with all data accumulated into a central site, and algorithms being keep running against that data.

Association rule mining goes for the disclosure of Itemset that co-occur regularly in transactional data. Centralized mining has been well concentrated beforehand. The issue has an extensive most pessimistic scenario unpredictability of large worst-case complexity, a reality that moves business to outsource the mining procedure to third party service

providers, who have made an efficient, productive and specific solutions. The data owner, beside the mining cost relief, has additional intentions for outsourcing the data mining task. In the first place, it requires insignificant computational resources, since the owner is only required to produce and to exchange the transactions to the miner. This impacts the outsourcing model additionally attractive to applications in which data owners creates transactions as streams and they have constrained resources to look after them. Second, assume that the owner has numerous creation sources of transactions, for e.g., consider a chain of business sectors which deliver exchanges or transactions at different areas. All these transactions can be sent to a single service provider for mining purpose. The service provider could process association rules that are local to the individual stores or global for the whole association. Because of this, the cost of exchanging the transactions among the parties and performing the global mining in a distributed manner is spared.

Then again, the third party service provider becomes only a single point of security attack. On the off chance that third party service provider is not trust worthy, he should be kept from accessing the original data as the data is sensitive to the organization. Likewise, paying little respect to whether the items are public, the newly generated association rules are the private property of the owner and they are proposed to be known only to data owner. In this manner, providing security to both raw data as well as generated association rules by service provider is a key issue in outsourcing of rule mining task. There are different methodologies that can ensure the sensitive information to be safe. The first is to apply an encryption techniques that transforms the original data to another format. On the other hand, utilization of encryption enables the exact rules to be recovered. The another approach is to divide data vertically or horizontally and send it to different servers so any of the outsource server didn't get the complete pattern as data is distributed among the parties, so here the idea to evaluate appropriate encryption techniques for outsourcing of association rules mining is proposed as we combine both the approaches and also increase the security by providing cryptographic techniques to data which is transferred between sender and receiver.

The Main objective of this undertaking is to provide communication complexity, computational complexity and less storage cost of our association rule mining and frequent item set mining solutions and with the help of E-clat algorithm. With the help of Double Encryption we are processing frequent Itemset mining and association rule mining for high privacy requirements. To maintain a strategic distance from disclosure of supports / confidences, this paper proposes an efficient homomorphic encryption scheme and a secure outsourced comparison scheme for comparing supports / confidences with thresholds. The newly introduced encryption scheme is custom fitted for the proposed comparison. The introduced homomorphic encryption scheme only requires modular multiplications and additions, and is more productive than the existing homomorphic encryption schemes used in other rule mining algorithms. The Input for Project is real time dataset such as retail dataset from UCI Machine Learning Repository, the dataset includes transactions and while the single transaction contain Item set, in which each item is comma or space separated. The output of project is Association Rules based on the Mining query sent by the user. The mining query is threshold value which varies from 0 to 1. Frequent item set mining and association rule mining have been utilized in applications such as web usage mining, market basket analysis, health care, bioinformatics and prediction. So the owners of Chains of retail shop or retail shop, Chief Website

Developer of an E-commerce business etcetera could make use of this very project.

Literature review is described in the section II. Section III presents the proposed system implementation details which includes homomorphic encryption, ECC encryption / decryption algorithm and E-clat rule generation algorithm. Section IV presents experimental analysis, results and discussion of proposed system. Section V concludes our proposed system. While at the end list of references paper are presented.

II. LITERATURE REVIEW

Lichun. Li, R. Lu, K. K. R. Choo, A. Datta and J. Shao [1] for vertically partitioned databases author proposed a privacy-preserving outsourced frequent item set mining solution. This can help the data owners in outsource mining assignments, without trading off data privacy assignments are performed securely on their joint data. By comparing with the existing system proposed system free less information about the raw data. In experimental results using distinct parameters and datasets at run time in each solution is only one order higher than that in the best non-privacy-preserving data mining algorithms. Hence the utilization of resource consumption at the data owner end is very low and the data and computing work are outsourced to the cloud servers.

In paper [2] authors introduced productive result integrity verification approach that can give deterministic certification for outsourced frequent item set mining. The key thought of the approach is to develop cryptographic confirmations of all (in) frequent item sets. They discussed how to streamline the quantity of verifications to enhance the execution. For experimental results generation IBM generator is used to generate four synthetic datasets S1, S2, S3, and S4 of various sizes. Also real-world retail dataset is used and results are calculated on various performance parameters such as time, scalability, etc.
In paper [3] authors tackled issues of privacy preserving association rule mining are tended here. Specifically, privacy preserving algorithms over horizontal and vertical partitioned databases are examined and results are compared. For vertically partitioned of data more than one attribute is required in transaction while for horizontally partitioned dataset only one attribute is sufficient which generates the association rule which is discussed in the paper also how to partition and merge data horizontally and vertically is discussed by authors.

D. H. Tran, W. K. Ng and W. Zha [4], discussed CRYPPAR, is a compulsory system for privacy preserving association rule mining which is dependent on cryptographic approach over vertically partitioned data. Results obtained by performing experimental test indicates the strategy of building is effective and may become a general way to do PPDM in real life. Authors also use different methodology to acquire better performance such as secure scalar product for two parties, partial topology generator for association rule mining, support computation of an Item set.

D. Trinca and S. Rajasekaran [5], here more focus is towards problems occurred while privately mining association rules in vertically distributed Boolean databases. For estimating item sets which preserves the privacy of individual gatherings author proposed an efficient multiparty convention. The selected convention is algebraic and recursive in nature, and it is based on two-party protocol for the same problem. It isn't just shown to be much quicker than similar protocols, yet additionally more secured. Also it is a variant of the extended protocol that is impervious to collusion among parties.

Privacy considerations [6] regularly constrain data mining projects. This paper tends the problem of association rule mining where transactions are distributed across various sources. Every site holds few attributes of each transaction, and the sites wish to work together to identify globally valid association rules. In any case, the goals must not reveal individual transaction information. Creators demonstrate a two-party algorithm for proficiently finding frequent Itemset with minimum support levels, without either site revealing individual transaction values.

Tassa, Tamir. [7], for secure mining of association rules in horizontally distributed databases author proposed protocol. The present driving protocol is that of Clifton and Kantarcioglu, the protocol proposed by them are relies on Fast Distributed Mining (FDM) algorithm of Cheung et al, which is an unsecured distributed variation of the well-known Apriori algorithm. The required fixings in their protocol are two novel secure multi-party algorithms — one that registers the union of private subsets that each of the associating partner players hold, and another that tests the incorporation of an element held by one player in a subset held by another.

Bettahally N. Keshavamurthy, Asad M. Khan,Durga Toshniwal [8] goal is to build up a global association rules model on the basis of the genetic algorithm (GA) and also find the frequent items. Because of GA inherent features it is used with respect to local maxima/minima and domain-independent nature for extensive space search technique to find exact or estimated solutions for optimization and search problems. These features are like robustness. The idea of trusted outsider with two offsets has been used for privacy preservation of the data. The data are first anonymized at local party end, and after that, the aggregation and global association is done by the trusted outsider.

In paper [9] creators stated as because of the encouraged development in the various fields, for example Cloud Computing. A third party service provider, the server comes in the frame. When an organization, the data owner, who require in expertise or the resources, outsources its mining needs. However the data owner thinks that both the item sets and the rules of the outsourced database as a classified property. The server stores the data and ships transformed by the data owner. At that point the data owner sends mining queries to the server, and the server returns the extracted patterns. From these examples or retrieved patters, the owner recoups the genuine patterns. Inside corporate privacy-preserving structures, the problem of outsourcing the association rule mining responsibilities in the outsourced environment is studied. In paper [9] to improve the security, rob frugal algorithm is used in which one to one substitution is performed & then fake transaction are added before sending the original dataset to third party server.

III. PROPOSED APPROACH

Problem Statement

In order to provide secure mining implementing twofold Encryption Technique (homomorphic and elliptic curve cryptography). Currently for frequent Itemset mining Apriori Algorithm, FP-Growth is been used but people have failed to notice that these algorithm required more time period and they works on un-encrypted data. Association rule mining as well as frequent Itemset mining are the two broadly utilized data analysis techniques which are generally utilized for finding frequently co-occurring data items and interesting association relationships between data items respectively in substantial exchange databases. These two strategies have been utilized in applications such as market basket analysis, health care, web usage mining, bioinformatics and Prediction. Doing in that capacity we assess the

computational complexity, communication complexity and storage cost of our association rule mining and frequent item set mining solutions concentrating on how our solutions can protect data owner's data from the outsourced servers and the other data owners.

Proposed System Overview

The Fig. 1 shows the proposed system architecture. The system model consist of two or more data owners and a server. Every data owner have their own sensitive information, the data owners encrypt their sensitive information prior outsourcing to the third party server. Data owner can likewise ask the server to mine rules or frequent Itemset from the joint database for their sake, for these the data owner has to send mining query to server. The server is entrusted with the gathering and putting away of databases got from different data owners, at server side data is partially decrypted & then association rules are generated over those data using E-clat algorithm, then at the end those encrypted association rules are send to relevant data owners. At owner side the owner has to decrypt those encrypted rules to get original association rules.

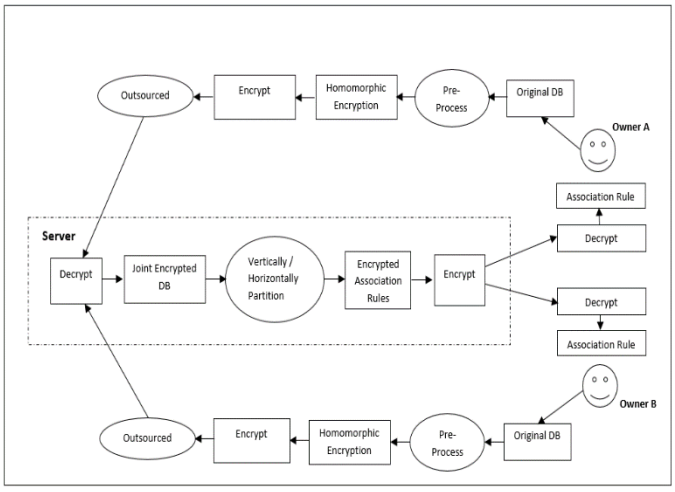

Figure 1. Proposed System Architecture

Algorithm

1) ECC Encryption Algorithm

Key Generation:

Public key and secret key both are generated by Key Generation. The Alice encrypts the message with bob's public key and the bob decrypts with secret key. After this we need to pick a quantity d inside the variety of n. utilizing next equation we may make the general public key.

PUK = n * q

n be the random number which is selected within (1 to n-1).

q be the any point on the curve.

PUK is public key and d be the secret key.

Encryption:

Let, Msg be the message that we're sending. We will put this message on the curve. Assume on the curve E; m has the point Msg. Select k randomly within [1 - (n-1)].

CYT1 and CYT2 be the cipher text.

CYT1 = k*PUK

CYT2 = Msg + k*Q

CYT1 and CYT2 will be send.

Decryption

Msg = CYT2 - d * CYT1

Where Msg is the original message.

Msg = CYT2 - d * CYT1

Msg is denoted as CYT2 - d * CYT1

CYT2 - d * CYT 1= (Msg + k * Q) - d * (k * PUK)

(CYT 2= Msg + k * Q and CYT 1= k * P) = Msg + k * d * P - d * k * P (cancelling out k * d * PUK) = Msg

2) Homomorphic Encryption Algorithm

KeyGeneration

$(s, q, p) \leftarrow$ Keygen()

Where, s is random number of cyclic order Z*p

Both p and q are large prime numbers.

Encryption EC ( )

$$EC(SKey, msg, d) = s^{d}(rq + msg) \bmod p$$

Where, SKey be the secret key,
msg be the original message which is to be encrypt.
d be the small positive integer which also called cipher text degree,
r be the big random positive integer.

Decryption algorithm DEC ( )

$$DEC(SKey, c, d) = \left(ci * s^{-d} \bmod p\right) \bmod q$$

Where, SK be the secret key,
ci be the cipher text or encrypted text.
d be the small positive integer which also called cipher text degree,
Both p and q are large prime numbers.

3) Eclat Algorithm

Step 1. Get td_list for every item in database with one DB scan operation.
Step 2. td_list of {x} is exact similar to the list of transactions
Which contains {x}
Step 3. Intersect td_list of {x} with the td_list of all other items.
Step 4. Items, resulting in td_list of {x,y}, {x,z}, {x,w},.= {x}
Conditional database (if {x} removed)
Step 4. Repeat from 1 on {x}-conditional database
Step 5. Repeat for all other items.

Mathematical Model

SYSTEM = {INPUT, OUTPUT, FUNCTIONS, RESULT}
Where

INPUT (I) = {{Dataset} min support}
Where

Dataset = transactions, item-set
t0, t1, tn where t= transaction
i0, i1 in where i = item
min support = is in 0 to n where n is positive integers

OUTPUT (O) = {T, I}
Where
T is a set of generated pattern of Transaction set.
I is set of generated pattern of Item-sets.

FUNCTIONS(Fn)={DatabaseConnector(), ruleGeneration(), Association rule gen(),FrmMain(), SendReceive(), MergeData(), ECCEncryption(), HomomorphicEncryp-
tion()}

RESULT (Rt) = {0, 1}
Where
0 = Failure of generating patterns for Item-set.
1 = Success for generating patterns.

IV. RESULTS AND DISCUSSION

A. Experimental Setup

All the experimental cases are implemented in Java in congestion with Netbeans tools and MySql as backend, algorithms and strategies, and the competing rule generation approach along with various encryption technique, and run in distributed environment with Master System having configuration of Intel Core i5-6200U, 2.30 GHz Windows 10 (64 bit) machine with 8GB of RAM and Slave System with configuration of Intel Core i5-2430M, 2.40 GHz Windows 7 (64 bit) machine with 4GB of RAM.

B. Dataset Description

The Input for Project is real time dataset such as retail dataset or mushroom dataset from the UCI Machine Learning Repository, Retail Dataset is a

transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers. The dataset includes transactions and while the single transaction contain Item set, in which each item is comma or space separated.

C. Result and Discussion

Here, the performance between existing and proposed system is compare. Fig. 2 shows the time required to run existing algorithm (FP- Growth) and proposed algorithm (E-clat with vertically & horizontally partitioned data). In Figure 2 X-axis show algorithms while Y-axis show required time to run the algorithm in milli seconds. Table 1 shows the reading from which the below graph is generated.

Dataset	Existing System (Time in ms)	Proposed System (Time in ms)
Retail Dataset (1000 Transactions)	140	27

Table 1 .Time Comparison

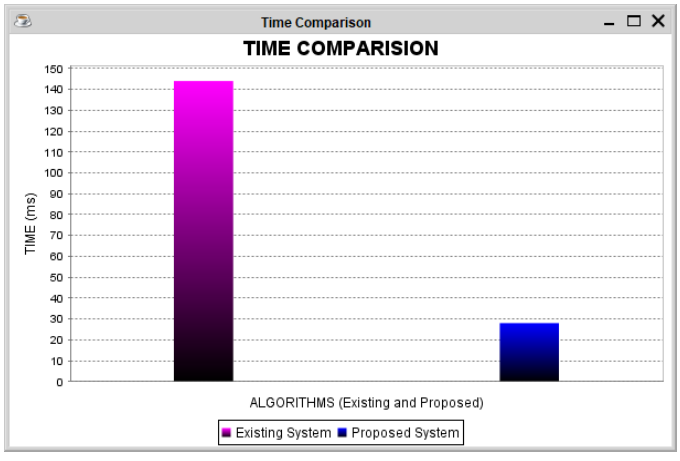

Figure 2 .Time Comparison Graph

Figure 3 shows Memory Comparison of Existing System and Proposed System. X-axis shows algorithms and Y-axis shows Memory in mb. The reading are taken from Table 2.

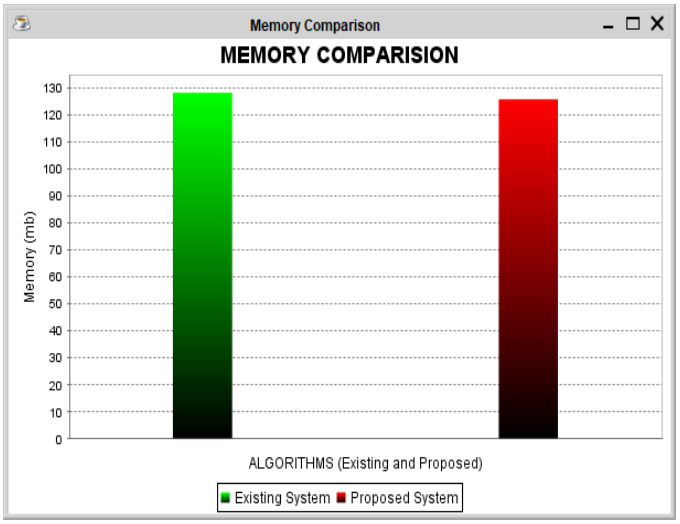

Figure 3. Memory Comparison Graph

Table 2. Memory Comparison

Dataset	Existing System (Memory in mb)	Proposed System (mb)
Retail Dataset (1000 Transactions)	128	124

Table 3. Result Comparison with Similar System

Parameters	Proposed System	Base Paper [1]	Paper [2]
Data Partition	Horizontal & Vertical	Vertical	No Partition of Data
Encryption	Double Encryption	Homomorphic Encryption	SHA1 (For Data Verification)
Rule	Eclat	Eclat	Apriori

Generation Algorithm			
Required Time to run Algorithms (Rule Generation)	Less	Moderate	More

V. CONCLUSION AND FUTURE SCOPE

Here secure outsourcing association rule mining in horizontally and vertically partitioned databases using E-clat and double encryption technique is proposed, which enables the data owners to outsource mining task on their joint information in a privacy-preserving manner. Additionally the issue of (corporate) privacy-preserving mining of frequent patterns (From which association rules can easily be computed) on vertically and horizontally partitioned encrypted outsourced transaction database using Eclat Algorithm is studied. We assumed that a moderate model where the adversary knows the domain of items and their correct frequency can utilize this information to figure cipher items and cipher Item sets. We used double encryption scheme, i.e elliptic curve cryptography encryption over homomorphic encrypted transaction to enhance the security. Also to improve the results in terms of time and memory; complex database query are used which removes duplicate transactions rules.

VI. REFERENCES

[1]. Lichun. Li, R. Lu, K. K. R. Choo, A. Datta and J. Shao, "Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 8, pp. 1847-1861, Aug. 2016.

[2]. B. Dong, R. Liu and W. H. Wang, "Integrity Verification of Outsourced Frequent Itemset Mining with Deterministic Guarantee," 2013 IEEE 13th International Conference on Data Mining, Dallas, TX, 2013, pp. 1025-1030.

[3]. M. N. Kumbhar and R. Kharat, "Privacy preserving mining of Association Rules on horizontally and vertically partitioned data: A review paper," Hybrid Intelligent Systems (HIS), 2012 12th International Conference on, Pune, 2012, pp. 231-235.

[4]. D. H. Tran, W. K. Ng and W. Zha, "CRYPPAR: An efficient framework for privacy preserving association rule mining over vertically partitioned data," TENCON 2009 - 2009 IEEE Region 10 Conference, Singapore, 2009, pp. 1-6.

[5]. D. Trinca and S. Rajasekaran, "Towards a Collusion-Resistant Algebraic Multi-Party Protocol for Privacy-Preserving Association Rule Mining in Vertically Partitioned Data," 2007 IEEE International Performance, Computing, and Communications Conference, New Orleans, LA, 2007, pp. 402-409.

[6]. Vaidya, Jaideep, and Chris Clifton. "Privacy preserving association rule mining in vertically partitioned data." Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002.

[7]. Tassa, Tamir. "Secure mining of association rules in horizontally distributed databases." IEEE Transactions on Knowledge and Data Engineering 26.4 (2014): 970-983.

[8]. Bettahally N. Keshavamurthy, Asad M. Khan,Durga Toshniwal “Privacy preserving association rule mining over distributed databases using genetic algorithm” Springer-Verlag London 2013 Neural Computing &

Application (2013) 22 (Supp 1): S351–S364 DOI 10.1007/s00521-013-1343-9

[9]. F. Giannotti, L. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang, "Privacy-preserving mining of association rules from outsourced transaction databases," IEEE Systems Journal, vol. 7, no. 3, pp. 385–395, 2013.

Cite this article as :

Rutuja Thite, Dr. M. U. Kharat, "Secure Outsourcing Association Rule Mining in Horizontally and Vertically Partition Database Using Eclat and Double Encryption Technique ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 3 Issue 8, pp. 384-392, November-December 2018.
Journal URL : http://ijsrcseit.com/CSEIT1838109