International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 06 | June-2017 www.irjet.net p-ISSN: 2395-0072

D-Eclat Association Rules on Vertically Partitioned Dynamic Data to Outsourced Securely

Rasika Khairnar[1], Prof. P. D. Lambhate[2]

[1] *Student, Department of Computer Engineering, JSPM College of Engineering, Pune, Maharashtra, India*
[2] *Asst. Professor, Department of Computer Engineering, JSPM College of Engineering, Pune, Maharashtra, India*

Abstract - *Cloud computing is enhanced with the information mining-as-a service model. This service becomes popular choice among various kind of companies. This service is the cost effective, secure, time efficient and reliable. The various organizations does not have mining abilities, therefore they can outsource their mining need on cloud server. But the association rules and item sets of database are the private properties of organization. It suffers from the problems of security and authorization. This paper focused on the problem of accurate association rule mining over outsourced database with achieving the security and privacy of data. For this, initially data owner encrypt the data before outsourcing. After this, when client requesting for mining to the cloud server, it returns the results in encrypted format. For association rule mining system makes use of D-Eclat algorithm and proves that it is more time and memory efficient than Eclat algorithm. Also, association rule mining is applied on both horizontally and vertically partitioned data and prove that vertically partitioned outperforms in terms of time efficiency and system utilization. The performance of the system is tested on dynamically created dataset.*

Key Words: *Data mining, data outsourcing, privacy preserving, database partitions, association rule mining.*

1. INTRODUCTION

Paragraph comes content here. Paragraph comes content Information mining method has emerged as methods for identifying patterns as well as patterns from extensive amounts of information[1]. Mining incorporates different algorithms, for example, clustering, classification, association rule mining as well as sequence detection. Generally, every one of these algorithms have been produced in a centralized model, with all information being accumulated into a central site, as well as algorithms being keep running against that information. Association rule mining discovers all principles in the databases that fulfill some minimum support as well as minimum confidence constraints[2]. Numerous algorithms are utilized to improve the protection and security of information. Vertically partitioned imply that each site contains a few components of an transaction. Utilizing the conventional market basket case, one site may contain basic supply buys, while another has apparel buys. Utilizing a key, for example, MasterCard number and date, we can join these to recognize connections between buys of dress and basic supplies. In any case, this unveils the individual buys at each site, perhaps damaging purchaser security agreements. There are more reasonable cases. In the process of sub-assembly manufacturing, diverse makers give segments of the completed item. Cars incorporate a few subcomponents; tires, electrical hardware, and so on; made by independent producers. Once more, we have restrictive information gathered by a few parties, with a single key joining every one of the informational sets, where mining would help distinguish/foresee breakdowns. A real life example is the cur-rent trouble in Ford Motor as well as Firestone Tire. Ford Explorers with Firestone tires from a particular factory had tread separation issues in specific conditions, which have caused 800 injuries. Due to the tires did not have any issues on other cars, as well as other tires on Ford Explorers did not have any issue, neither one of the sides felt capable. The time taken to find the main issue prompted an advertising bad dream and the inevitable substitution of 14.1 million tires. A large number of these were most likely fine Ford Explorers represented just 6.5 million of the supplanted tires. Manufacturers had their own particular information early era of association rules in light of the greater part of the information may have empowered Ford and Firestone to determine the security issue before it turned into an public relations bad dream. Casually, the issue is to mine association rules crosswise over two databases, where the columns in the table are at various sites, splitting each row. One database is assigned the primary as well as is the initiator of the protocol. The other database is the responder. There is a join enter exhibit in both databases. The rest of the attributes are available in one database or the other, however not both. The objective is to discover association rules including attributes other than the join key.

The main contribution of this systems enlists here:

- Dynamic data encryption and outsourcing.
- Horizontal and vertical database partitioning
- Association rule mining over encrypted data by using Eclat and D-Eclat
- Secure Outsourcing of association rules over cloud server.

In this paper we study about the related work done, in section II, the proposed approach modules description, mathematical modeling, algorithm and experimental setup

in section III .and at final we provide a conclusion in section IV.

2. REVIEW OF LITERATURE

M. N. Kumbhar and R. Kharat, have the several of technique for PPARM[1] is performed also their outcomes are thought about. For fulfilling the privacy constraints in vertically partitioned databases, algorithm in view of cryptography methods, Homo-morphic encryption, Secure Scalar product as well as Shamir's secret sharing strategy are utilized. For horizontal Partitioned databases, algorithm that consolidates advantage standpoint of both RSA public key cryptosystem as well as Homomorphic encryption system as well as algorithm that utilizations Paillier cryptosystem to calculate worldwide backings are utilized.

In paper [2] , D. H. Tran, W. K. Ng and W. Zha, have designed CRYPPAR. CRYPPAR is a full-fledged system for privacy preserving association rule mining depending on a cryptographic. Authors utilize secure scalar product protocols as well as public key cryptosystems in CRYPPAR for effectively mining of association rules on vertically partitioned information. They also acquaint a partial topology with lower correspondence cost however much as could reasonably be expected. Also conducted several test runs. Test outcomes demonstrate that the system is proficient in privacy preserving association rules as well as may turn into a general structure for PPDM frameworks.

D. Trinca and S. Rajasekaran, have concentrated on the issue present in of privately mining association rules in vertically distributed Boolean databases [3]. At start, they designed an efficient multiparty protocol for computing item sets which provide privacy of the particular parties. The designed protocol is algebraic as well as recursive in nature, as well as depends on an as of late proposed two-party protocol for a similar issue. It is not just appeared to be considerably speedier than comparable protocols, additionally more secure. Next, they exhibited a variation of the extended protocol that is impervious to collusion among parties. As future work, it is fascinating to plan as well as test parallel variations of the developed multi-party convention.

Yiqun Huang, Zhengding Lu and Heping Hu, gave the secure scalar product of two parties from the perspective of matrix computation[4]. They also provided a way of security maximizing for both two parties. The securities in the two groups would be adjusted. Sensitive factors that impact the security of the two groups are additionally broke down.

In paper [5], L. Li, R. Lu, K. K. R. Choo, A. Datta and J. Shao have developed an efficient homomorphic encryption technique also a secure comparison system. After that they also developed a cloud-aided frequent itemset mining resolution that is utilized to develop an association rule mining arrangement. Developed method is intended for outsourced databases that enable various information owners to effectively share their information safely without bargaining on information security. Developed method release less data about the raw information than most existing arrangements. In contrast with the main known arrangement accomplishing a comparative protection level as our proposed arrangements, the execution of our proposed arrangements is 3 to 5 orders of magnitude higher.

F. Liu, W. K. Ng and W. Zhang have developed a protocol for outsourced rule mining known as PORM[6]. PORM does association rule mining in supervision of the outsourced model in which information is in encrypted format as well as outsourced. They also confirmed that PORM can return the frequent rules as well as check if a rule holds properly. They also confirmed that PORM satisfies two security properties: user server security as well as user-user security.

F. Giannotti, L. V. S. Lakshmanan, A. Monreale, D. Pedreschi and H. Wang, have analyses the issue of outsourcing the association rule mining process in a corporate privacy-preserving system[7]. They have developed an attack model depending on back-ground information as well as devise a method for protection preserving outsourced mining. Designed method guarantees that each changed thing is unclear concerning the aggressor's background information, from in any least k1 other changed items.

In paper [8], A. K. Sahu, R. Kumar and N. Rahim, have utilized idea of distributed database that splits the centralize database in distributed database environment, database may be partitioned in various aspects like horizontally partitioned, vertically partitioned as well as mixed mode. The papers given privacy preserving information mining algorithms working over vertically partitioned database utilizing the ideas of distribution privacy preservation as well as furthermore lessen the time and space complexity nature with zero rate of information leakage.

J. Ren and B. Zhang, [9] developed an efficient non-deterministic one-to-n substitution encryption transformation. Contrasted and the first algorithm, developed algorithm accomplished the non-determinacy by choosing appropriate E specifically and avoided doing redundant repetitive operations that have over half likelihood in the original one. They additionally promise it sufficiently secure not to be secured by a balanced mapping. Additionally, our E-era is unessential with the source things set I which makes our algorithm more adaptable to scramble diverse itemsets.

In paper [10], M. S. Joyce and V. Nirmalrani, developed a novel method which minimizes the leakage of the data as well as maximizes the security of the horizontally distributed databases. The results of the developed system gives technique that has no need of trusted third party, the sites themselves communicate each other for a secure mining and increase the security. This proposed architecture covers every one of the disadvantages

happened in the past algorithms as well as has been executed in the synthetic employment office database.

3. SYSTEM ARCHITECTURE/SYSTEM OVERVIEW

Detailed description of the proposed system is discussed in this section.
The architectural view of proposed privacy preserving association rule mining system is presented in figure 1. The system consists of three entities named as, data owner, system server and cloud server. Initially, multiple data owners send their private data to server. For security purpose, database is encrypted and then store on server. After receiving databases, cloud server combined all database. This combined version of database contains either original or fake data. For performance point of view, combined database is partitioned horizontally as well as vertical manner. After this, association rule mining procedure is applied on partitions of database. For association rule mining two algorithms are used and compare their performance. These algorithms are named as E-clat and D-Eclat algorithm. These rules further outsourcing to cloud server. Cloud server provides encrypted rules only when user requesting for the same. At the user side, rules are decrypted locally.

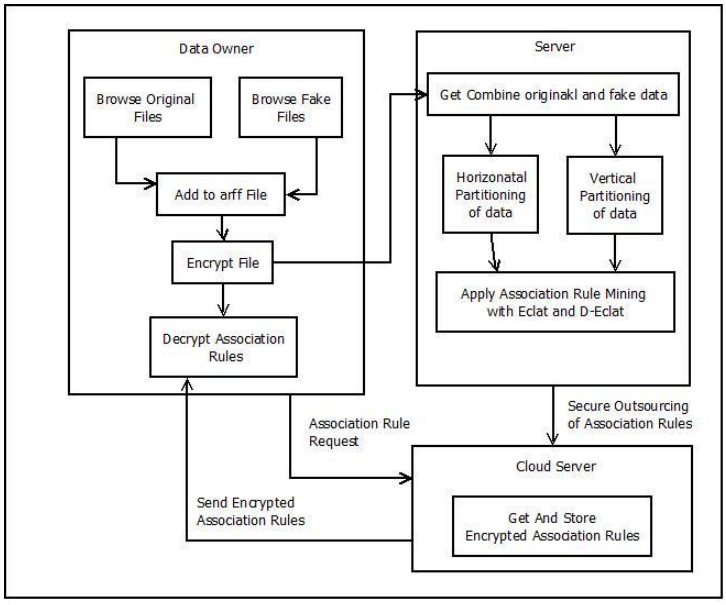

Fig -1: System Architecture

Module Description:

1. Dynamic databases creation by data owner To create database, initially owners collect product list of number of users. This product list is collected and converted into arff file. The arff file is then encrypted. For encryption three approaches are used named as, hash function, probabilistic homomorphic encryption, and substitution cipher. For rule mining purpose, arff file is sending to cloud server.
2. Association rule mining at server: Browse and combined the data: All data is collected from multiple data owners and combined into single database. This database contains either original or fake data.

Horizontal and vertical partitioning of data:

The combined database is partitioned as horizontal or vertical manner. For vertical partitioned data, rule mining is depend on support count of itemsets. And for horizontal partitioning, transactions are distributed among item sets.

Association Rule Mining:

Association rule mining is applied on partitioned database. For this purpose two algorithms are used names as Eclat and D-Eclat algorithm. This rule mining approach is applied on encrypted data files so that the generated rules are also in the encrypted format. These rules are then sending and storing at cloud server.

3. Fetching of rules from cloud server:

The data owners requesting to cloud server for association rules. In returns, cloud server provides encrypted association rules to data owners in encrypted format. After receiving these encrypted rules, data owner decrypt those association rules locally.

A. E-Clat Method

It takes a depth-first search and adopts a vertical layout to represent databases, in which each item is represented by a set of transaction IDs (called a tidset) whose transactions contain the item.
However, using tidsets has an advantage that there is no need for counting support, the support of an itemset is the size of the tidset representing it.
The main operation of Eclat is intersecting tidsets, thus the size of tidsets is one of main factors affecting the running time and memory usage of Eclat.
The bigger tidsets are, the more time and memory are needed

B. D-Eclat Method

The diffset format (the difference of two sets) has drastically reduced the running time and memory usage of the Eclat algorithm and the Eclat algorithm using diffset format is called dEclat algorithm.
It was the same as the Eclat, except that it sorted tidsets in ascending order and diffsets in descending order according their size.
By sorting diffsets and tidsets the memory usage and running time of dEclat could be reduced significantly.

C. Algorithm Used :
For association rule mining, Eclat and D-eclat algorithm used to find frequent itemset and strong rule generation. Following is algorithm

Input: $E((i_1,t_1),\dots(i_n,t_n))|P), s_{min}$
Output: $F(E, s_{min})$
1: *for all* i_j *occuring in* E *do*
2: $P \coloneqq P \cup i_j$ // *add* i_j *to create a new prefix*
3: $init(E')$ // *initialize a new equivalence class with the new prefix P*
4: *for all* i_k *occuring in* E *such that* $k > j$ *do*
5: $t_{tmp} = t_j \cap t_k$
6: *if* $|t_{tmp}| \geq s_{min}$ *then*
7: $E' \coloneqq E \cup (i_k, t_{tmp})$
8: $F = F \cup (i_k \cup P)$
9: *end if*
10: *end for*
11: *if* $E' \neq \{\}$ *then*
12: $Eclat(E', s_{min})$
13: *end if*
14: *end for*

4. RESULT AND DISCUSSION

4.1 Experimental Setup

The system is built using Java (JDK Version 1.6) framework on any Windows platform. The Net Beans (Version 8.1) IDE are used as a development tool. The system doesnt require any specific hardware to run; any standard machine is capable of running the application.

4.2 Database

Database contain arff file of product list. This file is dynamically created by data owners. This file contain transaction ID, transactions and ERV. For each customer, unique ID is allocated. Transaction contain list of all products purchase by customer. ERV represent the reality of data that is whether it is original or fake. 1 represent original transaction and 0 represent fake transaction.

4.3 Result and Discussion

In this system, dataset is either horizontally or vertically partitioned and for association rule mining Eclat or D-Eclat algorithm is used. In proposed system we are used the D-Eclat algorithm for association rule generation.
Figure 2 and 3 represent the graphical view of comparison of Memory required to generate association rules with Different Ts, constant Tc and Constant TS, different Tc, described in table 2. D-Eclat on vertically portioned data is more efficient that other system.
Ts=Threshold Support and
Tc=Threshold Confidence

Table -1: Memory Comparison

	Association rule		MEMORY IN BYTES	
Sr No	Ts	Tc	Ecalt	Declat
1	3	60	78514088	52342726
2	4		73350648	48900482
3	5		100390256	66926838
4	3	50	131631231	87754150
5		60	83683016	55788678
6		70	76311336	50874224

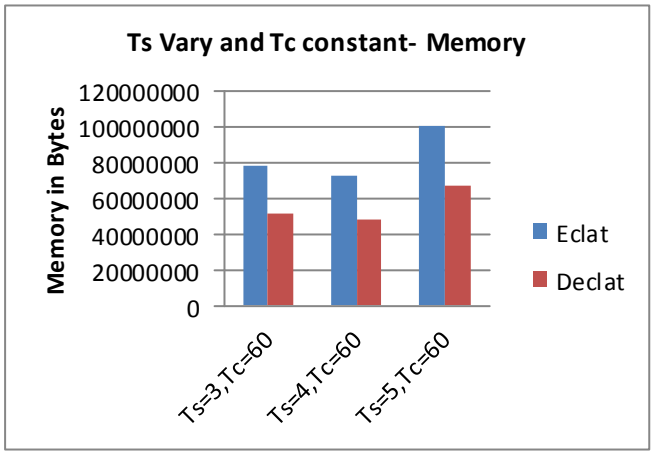

Fig -2: Memory Comparison when Ts in vary and Tc constant

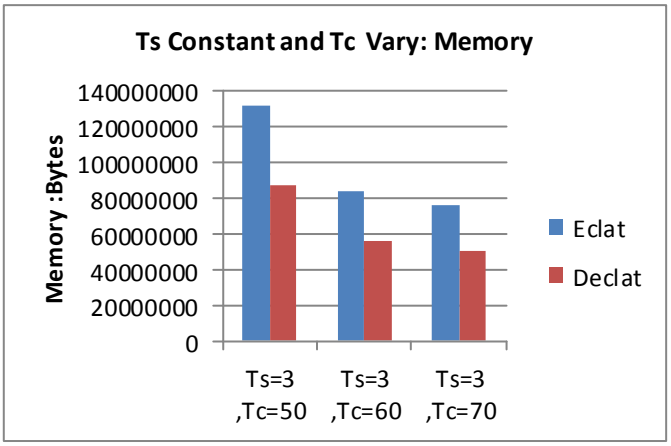

Fig -3: Memory Comparison when Ts is constant and Tc Vary.

Table II describes the Time analysis of existing system and proposed system. Figure 4 and 5 represent the graphical view of comparison of Time required to generate association rules with Different Ts, constant Tc and Constant TS, different Tc , described in table 3. Time required for the proposed system is less as compare to existing system.

Table -2: Time Comparison

	Association rule		TIME IN MILISECONDS	
Sr No	Ts	Tc	Ecalt	Declat
1	3	60	29659	19733
2	4		35657	21769
3	5		16067	10712
4	3	50	23860	15907
5		60	14724	9816
6		70	15331	10221

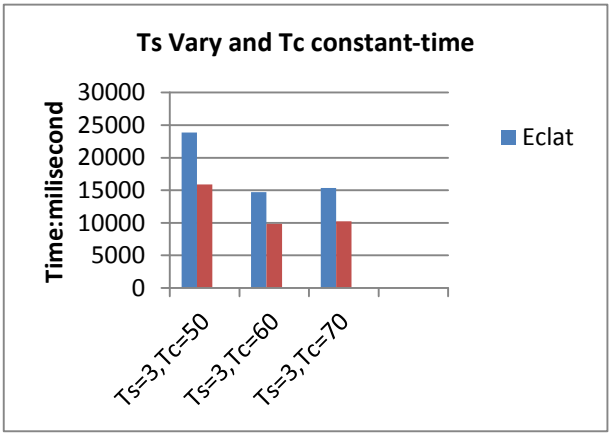

Fig -4: Time Comparison when Ts in vary and Tc constant.

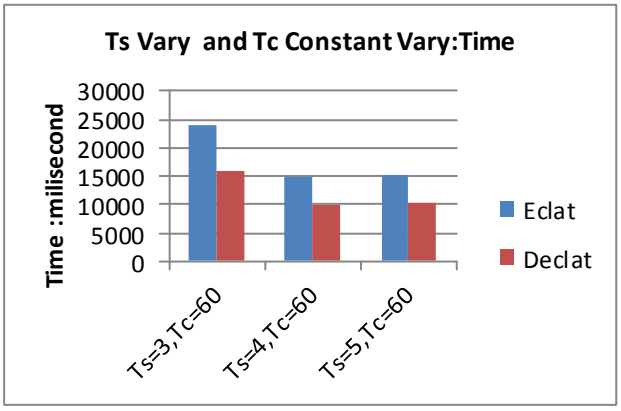

Fig -5: Time Comparison when Ts in vary and Tc constant is vary.

3. CONCLUSIONS

This paper solves the problem of privacy-preserving based association rule mining over outsourcing data. This paper presents the privacy preserving outsourcing of association rule mining on dynamic dataset. For association rule mining system makes use of EClat and D-Eclat Algorithm. This association rule mining is applied on horizontal and vertical partitioning database. To provide security, combination of 3 different encryptions is used. Cryptographic hash function, Substitution cipher and probabilistic homomorphic encryption function, are used to encrypt ID, Transaction, ERV value .Experimental results prove that the combination on D-Eclat algorithm on vertical partitioning data produces more accurate rules in minimum amount of time..

REFERENCES

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

[2] 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.

[3] 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, 402-409

[4] Yiqun Huang, Zhengding Lu and Heping Hu, "A method of security im-provement for privacy preserving association rule mining over vertically partitioned data," 9th International Database Engineering and Application Symposium (IDEAS'05), 2005, pp. 339-343

[5] L. 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.

[6] F. Liu, W. K. Ng and W. Zhang, "Encrypted Association Rule Mining for Outsourced Data Mining," 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, Gwangiu, 2015, 550-557

[7] F. Giannotti, L. V. S. Lakshmanan, A. Monreale, D. Pedreschi and H. Wang, "Privacy-Preserving Mining of Association Rules From Out-sourced Transaction Databases," in IEEE Systems Journal, vol. 7, no. 3, 385-395, Sept. 2013

[8] A. K. Sahu, R. Kumar and N. Rahim, "Mining Negative Association Rules in Distributed Environment," 2015 International Conference on Computa-tional Intelligence and Communication Networks (CICN), Jabalpur, 2015, 934-937.

[9] J. Ren and B. Zhang, "An Improvement on a Non-deterministic One-to-n Substitution Scheme in Outsourcing Association Rule Mining," 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA, 2009, pp. 43-47.

[10] M. S. Joyce and V. Nirmalrani, "Privacy in horizontally distributed databases based on association rules," 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil, 2015, pp. 1-6.

[11] Zhao, Chunye, et al. "Efficient association rule mining algorithm based on user behavior for cloud security auditing." Online Analysis and Computing Science (ICOACS), IEEE International Conference of. IEEE, 2016.

[12] Tran, Duc H., Wee Keong Ng, and Wei Zha. "CRYPPAR: An efficient framework for privacy preserving association rule mining over vertically partitioned data." TENCON 2009-2009 IEEE Region Conference. IEEE, 2009.

[13] Ren, Jinghan, and Baowen Zhang. "An Improvement on a Nondeterministic One-to-n Substitution Scheme in Outsourcing Association Rule Mining." Computer Science and Information Engineering, 2009 WRI World Congress on. Vol. 4. IEEE, 2009.

[14] Liu, Jie, Xiufeng Piao, and Shaobin Huang. "A privacy-preserving mining algorithm of association rules in distributed databases." First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06). 2006.

[15] Mr. Nitin J.Ghatge, Prof. Poonam D. Lambhate "An Effective Use of Meta Information for Text Mining International Journal of Advanced Research in Computer Engineering Technology (IJARCET) Volume 4, Issue 6, June 2015.

[16] Creighton, Chad, and Samir Hanash. "Mining gene expression databases for association rules." Bioinformatics 19.1 (2003): 79-86., Nov. 1999.

BIOGRAPHIES

Rasika Khairnar, is currently pursuing M.E (Computer) from Department of Computer Engineering, Jayawantrao Sawant College of Engineering,Pune, India. Savitribai Phule, Pune University, Pune, Maharashtra ,India -411007. She received her B.E. (Computer) Degree from MET'S BKC IOE, Nashik, Savitribai Phule Pune University, Pune, Maharashtra, India - 422003. Her area of interest is programming languages & data mining.

Prof. P.D.Lambhate, received her Degree from WIT, solapur, ME(Comp) from BVCOE Pune, Pursing PhD. In computer Engineering. She is currently working as Professor at Department of Computer and IT , Jayawantrao Sawant College of Engineering, Hadapsar, Pune, India 411028, affiliated to Savitribai Phule Pune University, Pune, Maharashtra, India - 411007. Her area of interest is Data mining, search engine.