ECLAT Algorithm for Encrypted Files in the Cloud for Fast Association Rule Mining

Lisha Purbey
Dept. Of Computer Science and Engineering
Cambridge Institute of Technology
India

Samhitha N
Dept. Of Computer Science and Engineering
Cambridge Institute of Technology Bengaluru,
Bengaluru, India

Shreya M K
Dept. Of Computer Science and Engineering
Cambridge Institute of Technology
Bengaluru, India Bengaluru, India

Teja P K
Dept. Of Computer Science and Engineering
Cambridge Institute of Technology
Bengaluru, India Bengaluru, India

Dr Ganesh D R
Dept. Of Computer Science and Engineering
Cambridge Institute of Technology Bengaluru, India

***Abstract*— To support data privacy, association rule mining algorithms are preferred. We have used an association rule mining algorithm to safeguard the encrypted data in the cloud in this paper. We have used the ECLAT algorithm for association rule mining and the AES algorithm for encryption purposes. The ECLAT algorithm alongside the Greedy Depth First Search algorithm has been used instead of the Apriori algorithm. We use the AES algorithm rather than the ElGamal cryptosystem. Thus, the algorithms used in this paper could guarantee both security of information and protection of the query whilst harboring the frequency of data.**

Keywords — Cloud Computing; Privacy Preserving; Advanced Encryption Standard (AES) algorithm; Equivalence Class Clustering and bottom-up Lattice Traversal (ECLAT) algorithm; Greedy Depth First Search algorithm; Encrypted files

I. INTRODUCTION

As the framework organization is progressing quickly, we are going up against a dangerous situation of freely upholding data from various customers. Several of the current systems revolve around using cryptography to trouble rough data on the data benefactor, which can guarantee the real data freely is not sensible for the protection of complete estimations over openly upheld data , since the disturbance of unrefined data on each customer would not impact the estimation over openly available data.

The issue of protecting the mining information has got continuously worthy of considering the extending capability of storing the singular data related to the customers, and the growing data mining estimations to utilize this information. As the improvements are made in cloud computing the research on protecting the data security in the databases that are redistributed has been highlighted. The databases that are redistributed may contain sensitive information therefore the safeguarding of data should be

ensured from external enemies which also includes the server of a cloud. Thus, before outsourcing the database to the cloud the data base should be encrypted. The specific data of an association and the relationship that deals with information are inspected by the association rule mining, one of the largely used data mining techniques. To safeguard data association rule mining algorithms are being used recently. There are additional problems in these estimations. The problems are of adding fake items and unable to hide the frequency of the data. Whilst handling the query, the sensitive information could be accessed by the cloud by noticing the frequency of the information despite the data and query being encoded. In the paper, we have used the ECLAT algorithm for association rule mining and the AES algorithm for encryption purposes. The ECLAT algorithm is used along with the Greedy Depth First Search algorithm instead of the Apriori algorithm. We use the AES algorithm rather than the ElGamal cryptosystem. Hence, the algorithms used in this paper could guarantee both security of the data and protecting the query privacy in faster manner.

II. RELATED WORK

For providing privacy for the data, the algorithm privacy - preserving association rule mining has been used. Firstly, [2] The mapping of one - to - many data is shown in this paper, inconvenience shown are the fake data can be detected by referring previously sent information. [3] The algorithm suggested will make use of K - frequency, however the disadvantages are the inadequate data gets included onto the database. [4] Here the algorithm put forward will support the K - Anonymity which can have additional overhead because of adding encoded transactions which are not real. In this paper, we can take a scenario in which the data owner encodes its information and the information will be stocked in cloud repositories. For mining the rules taken from the information, the data owner will hand over the task to n ≥ 2 servers which are called semi - honest, the servers will cooperate and performs association rule mining over the encoded data which will be in the cloud and also returns encoded rules to the client. Downside

faced here are decreased level of accuracy. [5] The mining of the database is persuaded by the issue of decision support which has affected most of the huge retail companies. The technology which uses the bar - code has made progression which has made it feasible for huge corporations to gather and stock the enormous amount of information, which is mentioned as the basket information. Downside of this paper is locating every single rule which are significant for seller's market and for including applications such as mailing. [6] Here the algorithm can carry KNN grouping by utilizing the encoded index fund and grabble arithmetic circuits. Disadvantages are it is an indolent learner, it doesn't utilize everything from the provided training information and for the classification the entire training data is taken into consideration. [7] The usage of private key which will be hidden and the cipher texts will be divided among the users during the calculations. Drawback in this paper is it doesn't provide proper identity verification service, reliability is completely built on privacy along with efficiency of the keyword. Finally, [8][9] The rich volume and the Smartphone's usage can have a hold on power of the crowd for the mobile clients. Inconvenience in this paper is, it takes over imbalance and changeability of the users will be disregarded. Here we are making use of AES algorithm which makes it hard to take out information. Greedy algorithm assists to approve the specific document conforming to the client.

III. SYSTEM ARCHITECTURE

A. *Architecture diagram*

Most familiar opponents (adversaries) are pernicious and semi-honest. We contemplate that clouds are inside adversaries than the outsiders who could compromise the information's stored in the cloud. Moreover the insider adversaries have more authority to the information stored. There are generally two types of adversarial models. One is semi-honest and the other is malicious adversarial models. In semi - honest model, the clouds appropriately obey the protocol given to it and yet may make an effort to acquire auxiliary data that is not being permitted. In malicious model, the cloud may not follow the protocol given to it. By adhering the prior work, we have adopted a semi – honest model [4]. The proposed system architecture is shown in figure 1.

The proposed system comprises of Data Owner(DO), Cloud A(CA), Cloud B(CB) and the Authorized User (AU). The database is in possession of the Data owner, and the one who gain access to the cloud is service beneficiary AU. CA and CB are two cloud servers which performs the required computations securely. CA stores the files in an encrypted format. Therefore, the files would be secure from cloud adversaries as it is stored in an encrypted format. Moreover, CB stores the key pairs which help in maintaining the integrity of the files.

The approach in order to build the proposed system is demonstrated below. Firstly, the DO upload the encoded file to Cloud A. Secondly, the AU send the query regarding a file request in an encrypted format to CA. Thirdly, CA sends the approval so that the user can access the encrypted files user and then sends a file request access to CB asking to grant access to he key pairs. Fourthly, CB grants the access for key pairs so that the AU can view the required files. The request related to the key and contents are approved in both the clouds. So, now the user would be able to download the file in a decrypted format.

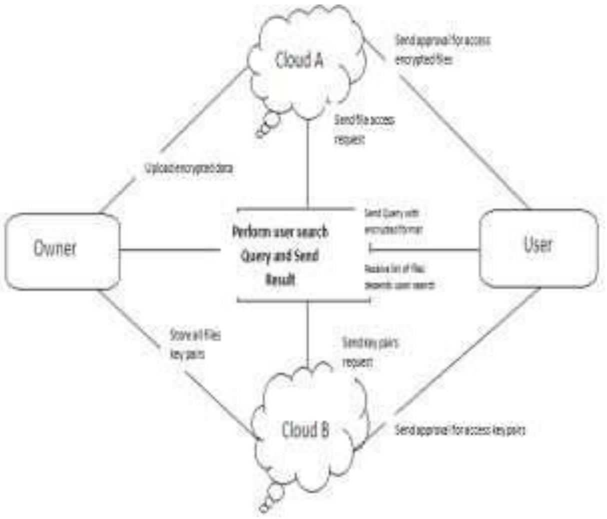

Cloud A
Send approval for access encrypted files
Send file access request
Upload encrypted data
Owner
Perform user search Query and Send Result
Send Query with encrypted format
Receive list of files depends user search
User
Store all Files key pairs
Send key pairs request
Cloud B
Send approval for access key pairs

Fig. 1. Architecture of the System

B. *Process of searching*

The procedure altogether comprises of five steps as depicted in the figure 2. Firstly, user would log - in and enters the keyword that he/she wants to search. While the file is being searched in the repository, the algorithm ECLAT is carried out on the keyword being entered. The frequent itemsets are collected by the algorithm ECLAT along with the Greedy Depth First Search algorithm. The tf - idf (i.e. term frequency- inverse term frequency) are being calculated for the content as soon as the search button is clicked by the user. The completion of the order of tf - idf is succeeded by arranging the files that are highly recommended in the first to the last order. All of these files contain the keyword entered by the user. The sole purpose of Greedy Depth First Search algorithm is that it would recommend the exact files required by the user. Finally, the files that are finalized by the system would be displayed to the user.

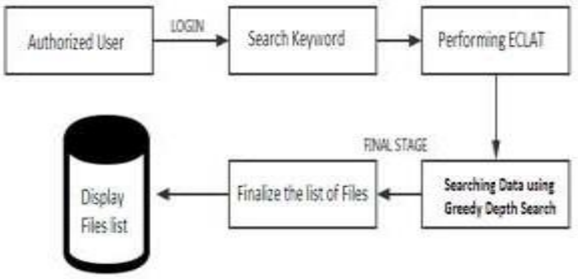


Fig. 2. Architecture of the Searching Process

IV. PROPOSED ALGORITHM

We have used three algorithms in our proposed system. The proposed system would be able to get fine search results at speed. Firstly, for encryption, the AES algorithm has been used. Secondly, for Association Rule Mining the algorithm named ECLAT has been used. Thirdly, for recommending the appropriate requested file as per the user's search keyword greedy depth first search algorithm has been used. With the help of these three algorithms, we ensure a system that outputs faster results along with fine results

A. AES Algorithm

The Advanced Encryption Standard (AES) algorithm has been used for the encryption of the key-pairs. This algorithm is a symmetric block cipher which implies that for both encryption and decryption the same key is being used [9]. The other name of this algorithm is Rijndael, the original name of this algorithm. This algorithm has some attributes. They are immune to the attacks known, speed, compactness of the code, and simplicity of the design. We have used the algorithm that has key size-256 having 14 rounds. The architecture of the this algorithm is shown in the fig- 3.

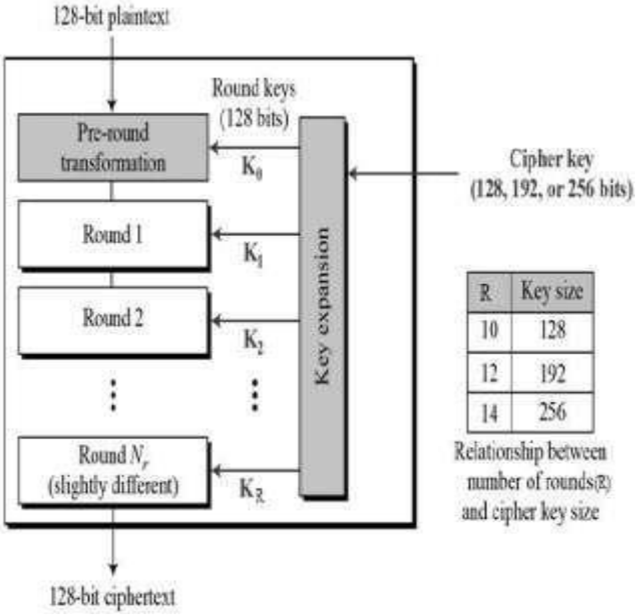


Fig. 3. Architecture of AES.

B. ECLAT Algorithm

ECLAT is short for "Equivalence Class Clustering and bottom - up Lattice Traversal". For Association Rule Mining, the ECLAT algorithm is among the most popular methods. The ECLAT algorithm is more well-structured and a scalable form of the algorithm Apriori. Whilst the Apriori algorithm do the task in the horizontal fashion similar to the breadth first search of a graph traversal, the ECLAT algorithm performs the task in a vertical fashion similar to the depth first search of a graph traversal. The ECLAT algorithm is faster than the Apriori algorithm because of this vertical behavior. The working of Eclat algorithm is depicted in figure 4.

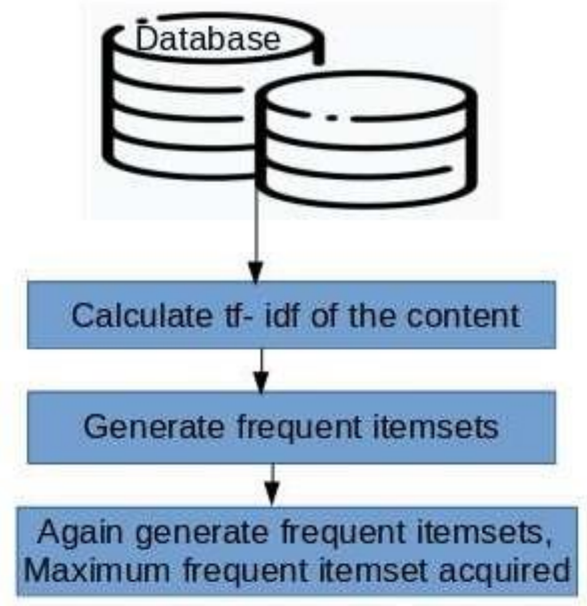
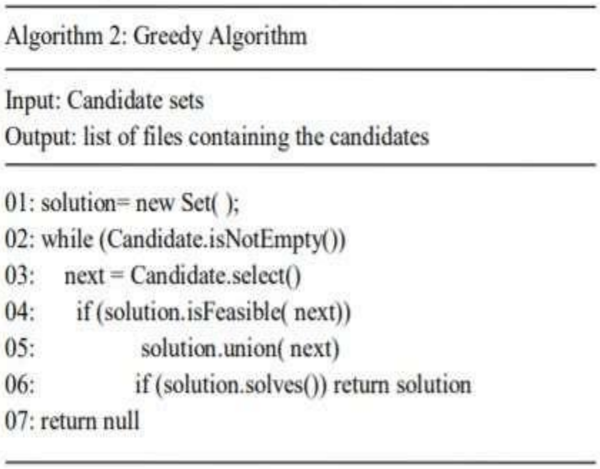

Database
Calculate tf- idf of the content
Generate frequent itemsets
Again generate frequent itemsets,
Maximum frequent itemset acquired

Fig. 4. Working of ECLAT.

C. Greedy Depth First Search Algorithm

The algorithm, Greedy searches the leaf of the graph for local maxima based on the function assigned for it's assessment. The motive of Greedy algorithms is to choose optimally at each and every step in an attempt of finding the altogether optimum at all steps to find an optimal choice globally. The greedy algorithms are simple, easy to implement, and have less time complexity problems. Moreover, greedy algorithms are intuitive and mostly used in systems to solve optimization problems. A depth first search algorithm occupies very less memory and attains the goal in less amount of time making the searching process faster. Hence we propose a Greedy Depth First Search algorithm that makes the system optimized.



Algorithm 2: Greedy Algorithm

Input: Candidate sets
Output: list of files containing the candidates

```
solution= new Set( );
while (Candidate.isNotEmpty())
    next = Candidate.select()
     if (solution.isFeasible( next))
          solution.union( next)
          if (solution.solves()) return solution
return null
```

V. PERFORMANCE ANALYSIS

The proposed system is known as F-ARM (fast association rule mining). The proposed system's performance is measured against the S-ARM(secure association rule mining) algorithm submitted in the prior work [1]. To analyze the performance, the time required to fetch the document that is being stored by the cloud has been used. By changing the number of data, the performance of the algorithms S-ARM and F-ARM are measured. The graph of the analysis of the performance is depicted in fig 5.

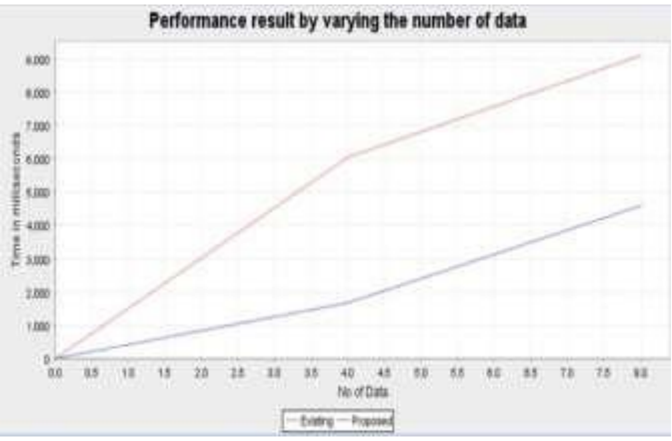


Fig. 5. Performance by ranging the number of the data.

VI. SPECIFICATION AND EFFICIENCY

A. *Hardware and Software Specifications*

a. HARDWARE SYSTEM CONFIGURATION:-
Processor - Pentium – IV
RAM - 8 GB
Hard Disk - 25 GB

b. SOFTWARE SYSTEM CONFIGURATION:-
Operating System - Windows 10
Programming Language - Java
Documentation - MS-Office
Cloud - Everdata

B. *Efficiency*

Any algorithm's efficiency is examined by calculating the complexity of its time and space as depicted in equation (1). The ideal (optimal) solution is given by the algorithm that has the least time complexity. Preferably, the solution that is optimal is always used.

$$\text{Efficiency} = \frac{\text{Time taken for existing}}{\text{Time taken for proposed}} \times 100\% \qquad (1)$$

$$\text{Efficiency} = \frac{325}{1456} \times 100\% = 22.32\% \text{ efficiency}$$

CONCLUSION

In this paper, we show how the data files can be safely protected in clouds. We also show how fast the files are being retrieved from the clouds along with recommending the exact files required by the user based on the user's search keyword. In our system, for association rule mining the ECLAT algorithm has been used in place of the Apriori algorithm. Moreover, we have used the Advanced Encryption Standard (AES) algorithm instead of the ElGamal Cryptosystem. The algorithm responsible to find the frequent itemsets is ECLAT. The union of both the Greedy Depth First Search and the ECLAT Algorithm gives good yield in the process of mining.

ACKNOWLEDGMENT

This work have been supported by the Cambridge Institute of Technology, Bengaluru (Computer Science Department). We would like to extend our gratitude to our HOD Dr. Shashikumar D. R. for his encouragement throughout the span of work.

REFERENCES

[1] Hyeong-Jin Kim, Jae-Hwan Shin, Young-ho Song, et al. "Privacy-preserving Association Rule Mining Algorithm for Encrypted Data in Cloud Computing." IEEE 12th International Conference on Cloud Computing (CLOUD), 2019.

[2] Wong, Wai Kit, et al. "Security in outsourcing of association rule mining." Proceedings of the 33rd international conference on Very large data bases. VLDB Endowment, 2007.

[3] Giannotti, Fosca, et al. "Privacy-preserving mining of association rules from outsourced transaction databases." IEEE Systems Journal 7.3 (2013): 385-395.

[4] Yi, Xun, et al. "Privacy-preserving association rule mining in cloud computing." Proceedings of the 10th ACM symposium on information, computer and communications security.ACM, 2015.

[5] Rakesh Agrawal & Ramakrishnan Srikant "Fast algorithms for mining association rules." Proc. 20th int. conf. very large data bases, VLDB. Vol.1215. 1994.

[6] Kim, Hyeong-Jin, Hyeong-Il Kim, and Jae-Woo Chang. "A Privacy- Preserving KNN Classification Algorithm Using Yao's Garbled Circuit on Cloud Computing." Cloud Computing (CLOUD), 2017 IEEE 10th International Conference on.IEEE, 2017.

[7] Jakobsson, Markus, and Ari Juels. "Mix and match: Secure function evaluation via cipher texts." International Conference on the Theory and Application of Cryptology and Information Security. Springer, Berlin, Heidelberg, 2000.

[8] Bhargav Hegde, Dayananda P, Mahesh Hegde, Chetan C, " Deep Learning Technique for Detecting NSCLC", International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-3, September 2019, pp. 7841-7843. DOI: 10.35940/ijrte.C6540.098319

[9] Zhibo Wang, Jiahui Hu, Jing Zhao. "Pay On-demand: Dynamic Incentive and Task Selection for Location-dependent Mobile

[10] Crowd sensing Systems." IEEE 38th International Conference on Distributed Computing Systems (ICDCS). 2018.

[11] www.informatik.uni-freiburg.de

[12] ijartet.com