International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Privacy-preserving Security using Mining Algorithm for Encrypted Data in Cloud Environment

[1*]Raghi K.R, [2]Dr. Arjun Paramarthalingam
[1*]Department of Computer Science and Engineering, Anna University Chennai, Tamil Nadu, India,[2]Assistant Professor, Department of Computer Science and Engineering, University College of Engineering,Villupuram, Tamil Nadu, India
Email Id: [1]*raghikr@cs.annauniv.edu, [2]arjun_ucev@ymail.com

***Abstract* — Privacy-preserving association rule mining algorithms have been proposed in this paper to support data privacy.However, the existing algorithms have an additional overhead to insert fake items (or fake transactions) and cannot hide data frequency. We propose a data security and privacy data mining algorithm in cloud environment for encrypted data. The association rule mining utilize the Apriori algorithm with Elgamal and pallier cryptosystem, without additional fake transactions. Thus the proposed algorithm can guarantee the database security and query privacy, while concealing frequency. The proposed idea provides better performance than the existing methods, in terms of association rule mining efficiency.**

Keywords- Data mining; Apriori algorithm; cloud security; Elgamal cryptosystem;

I. INTRODUCTION

Data privacy and security in outsourced data in cloud got more significance in cloud computing applications. Because the outsourced database may include sensitive information, it should be protected against the cloud server. Therefore, the encrypted data before being outsourced to the cloud. It is widely used data mining techniques in the cloud; the association rule mining analyzes the specific data of a company and the association of various information. Mostly, privacy reserving association mining techniques used to support data security [1-3]. Thus, the algorithms depend upon by adding the unnecessary items and it will hide the data frequency in the cloud. At the query processing time, the sensitive information of original data can be analyzed when both the data and the query are encrypted [3].

In this paper, we propose Apriori algorithm based association rule mining algorithm. This algorithm analysis the data generated in cloud environment. The association rule mining is performed over transaction databases [4]. The secure plaintext equality test protocol verifies the input ciphertexts are of identical values. By doing this the proposed approach provides the data security and privacy, while concealing query frequency.

The reminder of the paper is organized as follows. The related literature study associated with privacy-preserving mining techniques is discussed in section 2. The section 3 provides working information about Apriori algorithm based association rule mining algorithm. The experimental study and performance analysis are performed in section 4. The section 5 presented summary about this paper.

II. RELATED WORK

Security of data and privacy-preserving mining technique algorithms based study was performed by various authors [5-6] First, Wong et al [1] proposed a one-to-many item mapping that transforms transactions non-deterministically. However, there is a disadvantage that fake items are easily distinguished from the original data because the probability of fake items in the transaction database is the same. Second, Giannotti et al. [2] proposed an association rule mining algorithm using *k*-anonymity. This algorithm adds fake transactions to the transaction database so that each item can have *k*-1 frequency. However, the original data can be exposed if fake transaction is known. Also, additional operations are neededto remove the frequency of fake transactions. Similarly, Xun etal. [3] proposed an association rule mining algorithm that supports *k*-anonymity on an encrypted database. This algorithm supports data protection and query protection by using Elgamal encryption system. However, it has an additional overhead for adding encrypted fake transactions.To compute the frequency of candidate set, it uses a conditional gate based on the binary array of ciphertext. However, the original data can be inferred if an attacker has some knowledge about data frequency because it does not encrypt the data frequency in query processing.

III. PROPOSED ASSOCIATION RULE MINING ALGORITHM

A. System architecture

The typical types of adversaries are semi-honest and malicious [5]. The cloud is consider as the insider attackers, outside attacker have large number of authority than the adversaries. In Semi-honest adversarial model, the cloud follows the rules correctly, then it may try to obtain the more information not allowed. The cloud can deviate from protocol in the malicious model. We adopt a semi-honest adversarial model by following the earlier work [3]. The proposed system architecture is shown in the Figure 1.

The architecture consists of Owner of the data (*DO*), A-Cloud (C_A), B-Cloud (C_B), and Authorized User (*AU*). The database is owned by the data owner, and User is the service recipient who gains accesses to the cloud. The two cloud servers with two-party computation protocols perform computations securely on C_A and C_B,

The procedure for building systems is as follows.

Firstly from the original database, an Elgamal encryption key pair is generated. Then the information about public key and encrypted database are forwarded to cloud server C_A. the encryption key pair generated from Elgamal approach is forwarded to the cloud server C_B. At last, the query is encrypted by *AU* which is forwarded to C_A. Because the original data can be exposed using the plaintextequality test protocol [6]. This privacy preserving

plaintext equality test protocol (SPET) checks identity of input encrypted data from cloud servers C_A and C_B using Apriori algorithm.

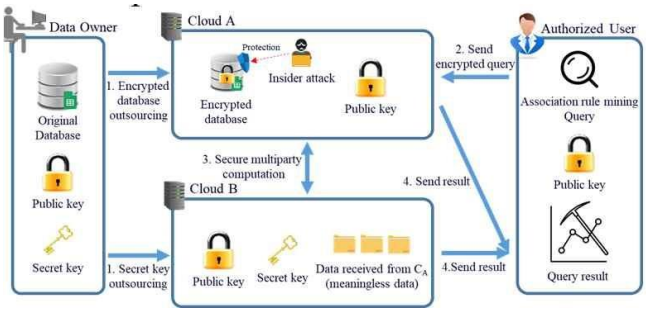

Figure 1. System architecture of proposed work

B. Secure protocol

The proposed SPET protocol returns 1 if the plaintexts of two ciphers are equal and returns 0 otherwise. The working of SPET protocol is shown in Algorithm 1. First, C_B generates a composite number t and send E(t) to C_A(line 1~2). Second, C_A multiplies E(t) by E($cipher_1$) and E($cipher_2$), respectively. C_A sends g^{r1} and g^{r2} to C_B, where g^{r1} and g^{r2} represent the front of E($t \times cipher_1$) and that of E($t \times cipher_2$), respectively(line 3~6). Third, C_B returns $g^{r1} \times g^x$ and $g^{r2} \times g^x$, where x is the secret key(line 7~8). Fourth, C_A computes $\alpha = \frac{t \times m_1 g^{r1x}}{t \times m_2 g^{r2x}} \times \frac{g^{r1x}}{g^{r1x}} = \frac{m_1}{m_2}$(line 10). Finally C_A returns 1 if α is 1 and returns 0 otherwise (line 11~12).

Algorithm 1. Secure plaintext equality test protocol

Input: E($cipher_1$), E($cipher_2$)
Output: if $cipher_1$= $cipher_2$ return $\alpha = 1$ else $\alpha = 0$

C_B
01: generate t(t is composite number)
02: send E(t) to C_A
C_A
03: receive E(t) from C_B $r1$ $r1x$
04: E($cipher_1$)*E(t) = (g_{r2}, $t \times m_1 g_{r2x}$)
05: E($cipher_2$)*E(t) = (g , $t \times m_2 g$)
06: send to g^{r1}, g^{r2} to C_B
C_B
07: calculate g^{r1x}, g^{r2x} ($x = secret$ ke)
08: send to g^{r1x}, g^{r2x} to C_A
C_A
09: receive g^{r1x}, g^{r2x} from C_B g^{x}
10: calculate $\alpha = \frac{t \times m_1 g}{t \times m_2 g^{r2x}} \times \frac{g}{g^{r1x}}$
11: if α ==1, then return result = 1
12: else return result = 0

For association rule mining, we propose a privacy-preserving Apriori algorithm by using SPET protocol in cloud computing. The proposed algorithm consists of candidate set generation and frequency set calculation.

C. Candidate set generation

The candidate set generation step generates a candidate set containing many patterns, each of which has multiple items. The procedure of the candidate set generation step is as follows. First, one pattern pair <p_1, p_2> is selected in the k-1 frequent set, where p_1 and p_2 are different patterns. Second, we perform a join operation between p_1's items and p_2's items, and insert the joined result into the candidate set, i.e., S_k, if the result consists of k items. Finally, we perform a join operation for all pairs except <p_1, p_2> and return S_k to C_A.

D. Frequent set calculation

The frequent set calculation step calculates the frequency of S_k, as shown in Algorithm 2. First, one pattern of S_k is selected (line 1~2). Second, the SPET protocol is performed between the items of the selected pattern and the items of the transaction. If the result of SPET protocol is 1, the number of the matched items(*match*) is incremented by 1 (line 3~8). Third, when *match* is equal to k, E($x.sup$) is multiplied by g, where g is an arbitrary integer that is not included in a cyclic group of the encryption key (line 9~10). Fourth, the SPET protocol is performed between E($x.sup$) and E(g^{minsup}). If the result of SPET is 1, the frequent attribute of x is included in the frequent set (line 11~14). Finally, the frequent set calculation for the remaining patterns of S_k is

Performed in the same way (line 15~16)

Algorithm 2. Frequent set calculation

Input: Candidate k-item set S_k
Output: Frequent set L_k

01: for(all $x \in S_k$)
02: for(all $y \in$ E(T))
03:match=0
04: for(i = 0 to k)
05: for(j = 0 to $y.NumItem$)
06: if(SPET(x_i, y_j)) $match$++
07: end for
08: end for
09: if($match$==k){
10: enc_mul($x.sup$, g)
11: if(SPET($x.sup$, E(g^{minsup})) $x.freq$ = true }
12: end for
13: if($x.freq$ == true) $L_k \cup x$
14: end for
15: return L_k

E. Proposed privacy-preserving mining Apriori algorithm

The proposed Apriori algorithm is shown in Algorithm 3. First, we set L_1 to 1-item sets which are received from the data owner (line 1). Second, we perform the candidate set generation algorithm of 4.1, called Candidate_set_generation(E(L_{k-1})), where E(L_{k-1}) represents the k-1 frequent set (line 4). Third, the frequency of S_k is calculated (line 6). Finally, if the k frequent set is no longer generated, the k-1 frequent set is returned (line 5)

Algorithm 3. Proposed Apriori Algorithm

Input: Encrypted transaction database E(T)
Item set length k
Candidate pattern set S_k
Output: Frequency pattern set L_{k-1}

01: L_1 = {l_1,…,l_n | *E(T)}
02: k = 2
03: while(TRUE)
04: E(S_k) = Candidate_set_generation(E(L_{k-1}))
={c_1,…,c_p|$c \in k$ candidate set}
05: if(E(S_k)=Ø) return E(L_{k-1}) to AU
06: E(L_k)=Frequent set calculation(E(T),E(S_k))

07:end While

IV. EXPERIMENTAL STUDY & PERFORMANCE ANALYSIS

The data security of the algorithm. Is the viewpoint of CA, the proposed algorithm encrypts the data frequency and the encrypted database consists of unidentifiable encrypted transactions. Because the Elgamal cryptosystem returns different ciphertexts for the same plaintext, there is no leakage of the original data. In the viewpoint of CB, the data cannot be exposed because the front of the ciphertext is not contained in the original data. Therefore, the proposed Apriori algorithm proves that it safe the data in semi model.

We evaluate the performance of the proposed Apriori algorithm, called S-ARM (Secure Association Rule Mining). The performance analysis was done under Intel Xeon E3- 1220v3 3.10GHz, 32GB RAM. The proposed algorithm usesGMP library to represent a big integer in an Elgamal cryptosystem. The proposed algorithm is compared with theDP-ARM (Data Privacy Association Rule Mining) algorithm proposed by Xun et al.[3] because DP-ARM is the only existing algorithm to support both data privacy and query privacy. For performance analysis, we use the retail dataset collected from the Belgian market [7], and the performance measure of S-ARM and DP-ARM by varying the number of data. We also measure their performances by varying support changes (*minsup*) from 5% to 30% of data. Table 1 shows parameters for our performance analysis.

TABLE I. PARAMETERS FOR PERFORMANCE ANALYSIS

The Size of dataset	2k, 4k, 6k, 8k, 10k
Fake transaction ratio(Q)	50%, 100%
Minimum support	5%, 10%, 15%, 20%, 25%, 30%
Key Size	1024

A. *Performance analysis varying the number of data*

The performance result of S-ARM and DP-ARM by varying the number of data is shown in Figure 2. When *minsup* is 10% and Q is 50%, S-ARM shows 205% performance improvement on the average, compared with DP-ARM, and when Q is 100%., S-ARM shows 405% performance improvement. The reason is why S-ARM requires no additional operation for fake transactions unlike DP-ARM. In addition, S-ARM requires no binary operation by using Elgamal cryptosystem through SPET protocol.

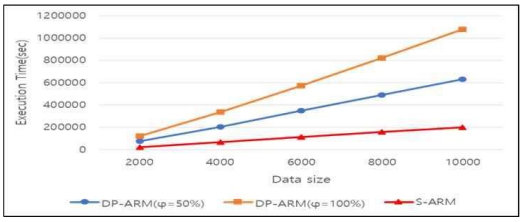

Figure 2. Performance result by varying the number of data

B. *Performance analysis varying minsup*

The performance result of S-ARM and DP-ARM according to *minsup* is shown in Figure 3. When the number of data is 10,000 and Q is 50%, S-ARM shows 216% performance improvement on the average, compared with DP-ARM, and when Q is 100%, S-ARM shows 429% performance improvement on the average. The reason is why S-ARM does not require no additional operation for the fake transactions unlike DP-ARM.

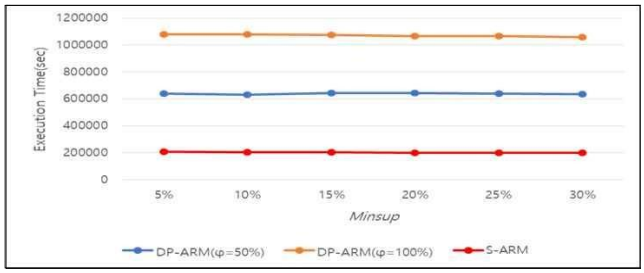

Figure 3. Performance result by varying minsup

V CONCLUSIONS AND FUTURE WORK

Privacy-preserving Apriori algorithm using the Elgamal cryptosystem, encrypted data with additional fake transaction is discussed in this paper. When hiding the frequency of data in cloud, it supports privacy of database system. The algorithm shows better performance in privacy and security about 3 to 5 times based on the association rule mining. As a future work, we plan to study on the parallel execution of the proposed algorithm for fast processing.

REFERENCES

[1] 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.

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

[3] 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.

[4] Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." Proc. 20th International conference on Very large data bases, VLDB. Vol. 1215. 1994.

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

Conference on Cloud Computing (CLOUD), 2017.

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

[7] Brijs, Tom. "Retail market basket data set." Workshop on Frequent Itemset Mining Implementations (FIMI'03). 2003.