DOI: https://doi.org/10.17762/turcomat.v11i3.13610

Research Article

Secure and Efficient Association Rule Mining over Encrypted Cloud Data

Aditya Agnihotri

Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand, India 248002,

Abstarct: Association rule mining and frequent itemset mining are two approaches to data analysis that have garnered a lot of interest and have been subjected to a lot of research. These approaches can be used for a range of applications. In this effort, our primary focus is on mining encrypted cloud databases while maintaining strict adherence to privacy standards. The owners of the data in this scenario are interested in discovering the association rules or frequent item sets derived from a collective data collection, but they want to do so while limiting the amount of their sensitive raw data that is made available to other owners of the data and to third parties as much as is practically possible. This study looks at the topic of outsourcing jobs related to association rule mining and doing it in a cloud environment. The study does so within the context of various frameworks that are supposed to preserve corporate privacy. According to the findings of the study, there is a novel approach that can be used to guarantee that every item that is updated and sent from the data owner to the server can be interchanged without the attackers needing any further prior knowledge. Our methods are efficient, scalable, and safeguard users' privacy when applied to a transaction database that is both large and authentic. This database is a representation of our complete algorithm. In this technique, the final protocol for mining association rules is proposed, in addition to mining that protects users' privacy by using a cloud infrastructure that is both safe and resistant to data breaches.

Keywords: Database, Cloud, Association rule mining, Privacy-preserving, Cloud Security.

I. Introduction

In a few years, it is expected that outsourcing of data and computing services would draw a huge number of researchers as internet-based and data centre-based IT services, including cloud computing, become more prevalent. Services such as business intelligence and knowledge discovery, both of which comprise complex analytics that are based on data mining, will be among those eager to be outsourced to the cloud because of the data-intensive nature of these services and the sophistication of the data mining algorithms. As a result, the trend of data mining and data management is likely to increase as interest in the cloud increases [1]. With the help of a third-party service provider, enterprises with limited computational resources and data mining skills should be able to outsource their data mining needs [2], [3]. The most serious breach in security is caused by the fact that the server may always access vital information from the owner data and it can obtain information. Yet, the transactions and their extracted techniques or patterns both belong to the data owner and always will. And these have to stay secure on the server [4]. In the context of our research, knowledge protection and privacy preservation are two key objectives of sensitive information protection. The final refers to privacy-preserving clustering, whereas the first is about privacy-preserving association rule mining. The similarities between knowledge protection and privacy preservation are enticing features. For instance, in knowledge protection, an organisation owns the data and must safeguard any sensitive information that might be learned from it, whereas in privacy preservation, individuals own their personal data [5]. In cloud computing applications, data privacy and security in outsourced data have gained significant importance. The outsourced database should be safeguarded from the cloud server because it might contain sensitive data. Thus, the encrypted data must first be exported to the cloud. The association rule mining method is one of the most popular strategies in the cloud; it examines the associations between numerous pieces of information and the specific data of an organisation. Techniques for mining associations that protect privacy are mostly employed to promote data security[6-8]. In order to mask the data frequency in the cloud, algorithms rely on introducing extraneous things. Sensitive information from the original data can be analysed during query processing if both the data and the query are encrypted [8].

In order to create an encryption scheme, we have suggested a way in this work With the intention of creating an encryption technique that will offer formal privacy assurances, we have presented a method in this work. We will validate this method using sizable real-world transaction databases (TDBs). Based on an encrypt-decrypt (E/D) module that is necessary to be viewed as a "black box," the client encrypts the data. This module changes the input data so that it can be stored in an encrypted database. The server performs data mining and sends encryption patterns to the owner. The returned patterns' real identities and supporters are ascertained using the E/D module. When data are encrypted utilizing 1-1 substitution ciphers without the utilization of fake transactions, the fact that many ciphers and consequently the transactions as well as prototypes can indeed be broken down by the server with a strong likelihood the with start of the frequency-based assault is irrelevant. This is because the server can break down the transactions and prototypes. So, the primary objective of this research is to present innovative encryption methods to illustrate formal privacy guarantees against server-based attacks. The server may make advantage of the background information while regulating the resource needs. In recent years, research on Privacy-Preserving Database Mining (PPDM) has received a lot of interest. The main technique discussed here is the gathering of private information from various owners by a collector known as a server with the main objective of integrating the information and performing mining on the resulting gathered information. Because the collectors cannot be trusted to maintain the privacy of the data as it is being acquired, the data are subject to arbitrary disruption. It has been possible to disturb data using a variety of methods while still guaranteeing that the systematic attributes and mined patterns adequately resemble the original data's mined patterns.

II. Related Work

First, Clifton et al. [4] the server can always get crucial information from the owner and can use it to learn crucial information, which is the biggest security drawback. Yet, the transactions and their extracted patterns both belong to the data owner and always will. And these must stay secure on the server. Second, Chaurey et al. [5] in the context of our research, knowledge protection, and privacy preservation are two key objectives of sensitive information protection. The final refers to privacy-preserving clustering, whereas the first is about privacy-preserving association rule mining. The similarities between knowledge protection and privacy preservation are enticing features. For instance, in knowledge protection, an organisation owns the data and must safeguard any sensitive information that might be learned from it, whereas in privacy preservation, individuals own their personal data. Security of data and mining method that protects privacy [9-10] Third, Wong et al. [6] algorithms-based research offered a one-to-many item mapping that changes transactions in an ad hoc manner. This has the drawback that since there is the same possibility of false items in the transaction database, phony items may be easily distinguished from the real data. Fourth, Giannotti, et al. [7] suggested a k-anonymity-based association rule mining algorithm. For each item to have a k-1 frequency, this technique inserts fictitious transactions into the transaction database. But, if the fake transaction is knownmay make the original data available.

Similarly, Xun et al. [8] suggested a technique that works with an encrypted database and enables k-anonymity is called as association rule mining. This technique uses Elgamal encryption technology to enable data protection and query protection. But, adding encrypted phony transactions comes with an extra cost. A conditional gate based on the binary array of the ciphertext is used to calculate the frequency of the candidate set. Nevertheless, since the data frequency isn't encrypted during query processing, the original data can be guessed if an attacker is aware of the data frequency.

The Bresson, Catalano, and Pointcheval (BCP) encryption system [15], in recent years, a homomorphic cryptographic system featuring a twofold decryption method has been suggested by Liu et al. [14]. It is an association rule mining system that protects privacy. It works well in secure multiparty computing in the presence of multiple keys [16]. In a multi-key system, each DO is an authority on their own set of public-secret keys. The cloud database can allow DOs to access their itself data at any time.

We offer a versatile and safe cloud-assisted association rule mining method for datasets that have been horizontally partitioned. In contrast to the majority of current studies, our suggested technique is impervious to collusion assaults and permits dispersed association rule mining to be completed by data owners while maintaining their and the mined findings' privacy.

III. Problem Statement

3.1 System Model

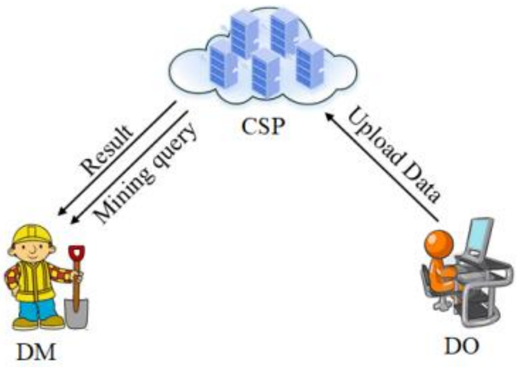


Fig. 1. System Model

We are dedicated to providing privacy-preserving association rule mining through the use of the single-cloud option whenever a miner transmits a mining query to a cloud server that has amassed a substantial number of encrypted transaction records provided by data owners. A cloud service provider (CSP), a data owner (DO), and a data miner are the three components that make up the system paradigm, which is depicted in Figure 1. (DM). In particular, the honest yet odd CSP model that we use in our system is known as the semi-honest CSP model.

- Data Owner (DO): To carry out association rule mining, data owners submit encrypted transaction records to the cloud server.
- Data Miner (DM): By outsourcing mining queries to a cloud server, the goal of data mining is to unearth any possible hidden association rules within the objects.
- Cloud Service Provider (CSP): After receiving a mining query from a data miner, the cloud server executes also fig. 2. The system model citation rule mining results from the encrypted transaction records provided by the data owner are sent to the data miner.

3.2 Attack Model

We primarily take into account internal and external adversaries.

The active attack strategy is quite expensive for external attackers, whereas the passive assault cost is incredibly minimal and leaves no evidence. Hence, the passive attack is the primary attack strategy used by the external adversary we are considering.

Moreover, inside attackers have the ability to gather compute process intermediate results. Attacker A's aptitude is described as follows:

- By listening in on the public channel, a can get communication information from all parties.
- A may attempt to get privacy or mining data from DO and DM by corrupting the CSP.
- In order to try to acquire the privacy of DO, A can simultaneously corrupt the DM and the CSP.

IV. Association Rule Mining on Encrypted Cloud Data

Since the identical item was replaced in the encrypted transaction database by a single encryption, the database server may now count the support of any encrypted itemset. According to the minimal support offered by the client, the DB server can therefore execute the Apriori algorithm in the replacement of the encrypted database.

The Apriori approach allows the database server to locate all frequently occurring encrypted item sets with the least amount of support before constructing the association rules. The client is then given strong encrypted association rules that satisfy the required levels of support and confidence by the DB server.

The client decrypts the encrypted strong association rules and searches for the items in the association rules using the standard item table.

All elements in the transaction database are constantly encrypted during association rule mining. This method so achieves item privacy. Nevertheless, since the cloud-based DB server is aware of the supports for encrypted itemsets in the encrypted transaction database, this approach cannot withstand the background knowledge-based assault.

The pattern mining task, privacy model, and encryption/decryption strategy can all be used to classify privacy-preserving mining of association rules.

4.1 Pattern Mining Task

According to reference [13], the most common and pervasive problem associated with pattern mining is as follows:

Find all of the item sets that support the transaction database "D" and have a support threshold of "x" when you are given these two pieces of information. In this essay, the sole purpose is to focus on the investigation of a privacy-protecting outsourced framework for the mining of frequent pattern data.

4.2 Privacy Model

The original dataset is encrypted by the owner of the data, who then converts it into an encrypted database to safeguard the identity of the particular data pieces. Cipher items are those found in an encrypted database, whereas Plain things are those found in the original database.

The encrypted database may be known to the server or an intrusive party that attacks it. Due to this, the proposed strategy is built on two key ideas: first, adding some fictitious transactions to the encrypted database, and second, replacing every item in the database with 1 to 1 substituted ciphers.

4.3 Encryption/ Decryption Scheme

4.3.1 Encryption

This phase includes the addition of the "RobFrugal" encryption algorithm, which is used to convert a transaction database into its encrypted form. There are three main phases to it:

1. Make use of a substitute cypher text with a ratio of 1 to 1 for each plain item.
2. The manner of item grouping that is used must be a specific one.
3. An approach is necessary in order to include fictional transactions.

4.3.2 Decryption

Once the client requests that the server execute a pattern execution query with the specified support threshold, the server will always deliver the computed frequency patterns for the encrypted databases. Our proposed E-D system is a workable alternative to an outsourced transaction database for privacy-preserving pattern mining, but a proper and effective implementation was still required. On the other hand, it is impractical to save the support for every cipher pattern.

4.4 Grouping items for privacy

Given that the items are in a supported table, some classification techniques can be used to group the objects into fixed-size groups. To begin, the Frugal approach is applied. We further assume that the cipher items are referenced in the descending order that the item support table is sorted. Given this information, the items' support will strictly decrease monotonically. Moreover, thrifty grouping is optional if the item support table is sorted using this approach in the descending order of the support.

4.5 Constructing Fake Transaction

If a noise table is provided, it should include the noise requirements for each cipher component. The following is how the phony transactions are produced:

1. We will discard the last items that have support equal to or greater than the maximum support of the group or the last items that have the most in common with each group in order to get rid of all the rows that have no noise at all.
2. The noise level created will be used to sort the succeeding rows in decreasing order.

V. Proposed System

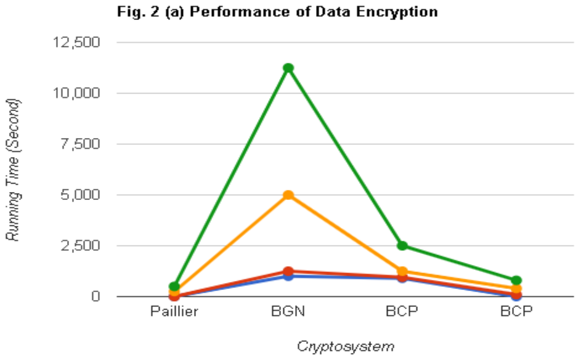
Rob-Frugal as compared to current systems, a novel system that successfully maintains Time and Space complexity in a practical condition is developed using Frugal Algorithms as the fundamental concept.

We are concerned about the security of external servers and are probably thinking about using data mining as a service, so we will address cloud data security as a top priority. So, it is important to take into account how to address various problems with the security and privacy of previously outsourced data with a new, enhanced system that has been proposed in order to elaborate on an attacker's behaviour pattern.

The suggested system will create cloud association privacy-preserving rules according to recently developed standards for:
1) Job of Pattern Matching
2) Privacy Model No. 2
3) Encryption and decryption method
4) Future Development of Attack Models

Performance of Data Uploading and Encryption: Take note that the DOs perform the data encryption offline. The DOs are typically resource-constrained customers. Figure 2(a) displays the data encryption performance, and Figure 2(b) displays the upload communication costs (b).

As can be seen in Fig. 2, BCP's running time for data encryption is substantially less than BGN's, and even shorter for the BCP variation, but both are longer than Paillier's running time. Paillier and BGN have lower communication costs than BCP and BCP variant, which are practically equal. Since the majority of DOs are resource-constrained, our technique significantly reduces the calculation cost of the DOs compared to [1]'s protocol 2, albeit at the cost of somewhat increased communication.



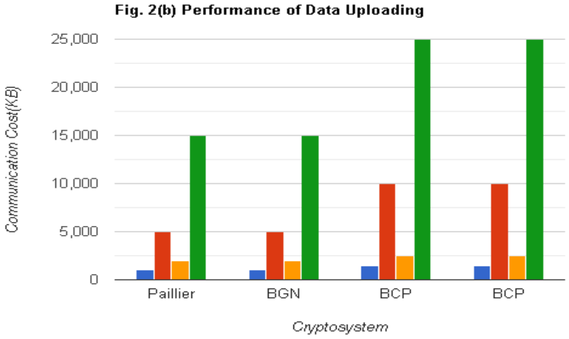


Fig. 2. Performance of Data

Association rule mining performance: With the chess dataset, we evaluate how quickly our system will run in the cloud. Figure 2 shows the total uploading and encrypting time. We can get the conclusion that our protocol is slower if the BCP variation is used as the primary cryptosystem in our design. Additionally, because the cloud typically has "infinite" computer resources and power, our technique can operate much more quickly in a true cloud environment.

VI. Conclusion

In many applications, maintaining privacy in data mining operations is a crucial concern. The majority of randomization-based strategies are probably going to be crucial in this field. In this study, An original strategy has been used to find a solution to the problem of data mining while protecting users' privacy while working within the setting of an outsourced business transaction database. In comparison to many other perturbation and anonymity strategies, this strategy is effective and superior. The proposed approach will significantly cut down on execution time and space requirements as well as false rule issues from the earlier work.

V. References

[1] R. Buyya, C. S. Yeo, and S. Venugopal, “Marketoriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities”, in Proc. IEEE Conf. High Performance Comput. Commun., 2008.

[2] 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, 2007.

[3] L. Qiu, Y. Li, and X. Wu, “Protecting business intelligence and customer privacy while outsourcing data mining tasks”, Knowledge Inform. Syst., 2008.

[4] C. Clifton, M. Kantarcioglu, and J. Vaidya, “Defining privacy for data mining”, in Proc. Nat. Sci. Found. Workshop Next Generation Data Mining, 2002.

[5] V. Richhariya, P. Chaurey, “A Robust Technique for Privacy Preservation of Outsourced Transaction Database”, IJRET, 2014.

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

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

[8] 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.

[9] 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.

[10] 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.

[11] D. Agrawal, C.C. Aggarwal, “On the Design and Quantification of Privacy Preserving Data Mining Algorithms”, In Proceedings of the 20th ACM SIGMOD-SIGACTSIGART Symposium on Principles of Database Systems, 2001.

[12] A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrk, “Privacy Preserving Mining of Association Rules”, In Proceedings the 8th

[13] ACM SIGKDD International Conference on Knowledge Discovery in Databases and Data Mining, 2002.

[14] S. L. Warner, “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias”, J. Am. Stat. Assoc., 1965.

[15] Agrawal, R., Imieli´nski, T., Swami, A.: Mining association rules between sets of items in large databases. In: ACM SIGMOD Record, vol. 22, pp. 207–216. ACM (1993)

[16] Cheng, K., Wang, L., Shen, Y., Wang, H., Wang, Y., Jiang, X., Zhong, H.: Secure k-NN query on encrypted cloud data with multiple keys. IEEE Trans. Big Data (2017)

[17] Liu, X., Deng, R.H., Choo, K.-K.R., Weng, J.: An efficient privacy-preserving outsourced calculation toolkit with multiple keys. IEEE Trans. Inf. Forensics Secur. 11(11), 2401–2414 (2016)

[18] Peter, A., Tews, E., Katzenbeisser, S.: Efficiently outsourcing multiparty computation under multiple keys. IEEE Trans. Inf. Forensics Secur. 8(12), 2046–2058 (2013)