Cloud Data Seclusion Scheme (DSS) and Partitioning Using Arm

B.Arulselvi[1] , B.Chandra[2], K.Jagestiya[3], R.Sivaranjini[4]
[1, 2, 3] Dept of Computer Science and Engineering
[4]Assistant Professor, Dept of Computer Science and Engineering
[1, 2, 3, 4] Krishnasamy College of Engineering and Technology, Cuddalore, Tamil Nadu, India.

Abstract- *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. This paper focused on the problem of accurate association rule mining over outsourced database with achieving the security and privacy of data. To enhance security for the data cryptographic encryption techniques is used.*

I. INTRODUCTION

Data mining is the process of analyzing hidden patterns of data according to different perspectives for categorization into useful information, which is collected and assembled in common areas, such as data warehouses, for efficient analysis, data mining algorithms, facilitating business decision making and other information requirements to ultimately cut costs and increase revenue. Data mining is also known as data discovery and knowledge discovery. The major steps involved in a data mining process are

- Extract, transform and load data into a data warehouse
- Store and manage data in a multidimensional databases
- Provide data access to business analysts using application software
- Present analyzed data in easily understandable forms, such as graphs.

Data mining involves exploring and analyzing large blocks of information to glean meaningful patterns and trends. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam Email filtering, or even to discern the sentiment or opinion of users.

KEY TAKEAWAYS

- Data mining is the process of analyzing a large batch of information to discern trends and patterns.
- Data mining can be used by corporations for everything from learning about what customers are interested in or want to buy to fraud detection and spam filtering.
- Data mining programs break down patterns and connections in data based on what information users request or provide.

The data mining process breaks down into five steps. First, organizations collect data and load it into their data warehouses. Next, they store and manage the data, either on in-house servers or the cloud. Business analysts, management teams and information technology professionals access the data and determine how they want to organize it. Then, application software sorts the data based on the user's results, and finally, the end-user presents the data in an easy-to-share format, such as a graph or table.

Organization

The remainder of this paper is organized as follows. Section2 introduces related works. Section 3 briefly presents the system analysis. Section 4 describes the system architecture. Section 5 is the conclusion.

II. RELATED WORK

Privacy-preserving Association Rule Mining and Frequent Itemset Mining on Vertically Partitioned Databases. In [9], the first work to identify and address privacy issues in vertically partitioned databases, a secure scalar product protocol is presented and used to build a privacy-preserving frequent itemset mining solution. Association rules can then be found given frequent itemsets and their supports.

The most relevant work is the privacy-preserving association rule mining solution presented in [11]. In this solution, a data owner known as the master is responsible for the mining. The other data owners (known as slaves) insert fictitious transactions to their respective datasets, and send the datasets to the master. Each data owner will also send his set of real transactions' IDs to a semi-trusted third-party server. Privacy-preserving data mining (PPDM) (either perturbation or secure multi-party computation (SMC) based approach) cannot solve the DMED problem. Perturbed data do not possess semantic security, so data perturbation techniques cannot be used to encrypt highly sensitive data. Also the perturbed data do not produce very accurate data mining results. Secure multi-party computation based approach assumes data are distributed and not encrypted at each participating party. In addition, many intermediate computations are performed based on non-encrypted data. As a result, in this paper, we proposed novel methods to effectively solve the DMED problem assuming that the encrypted data are outsourced to a cloud. Specifically, we focus on the classification problem since it is one of the most common data mining tasks. Because each classification technique has their own advantage, to be concrete, this paper concentrates on executing the k-nearest neighbor classification method over encrypted data in the cloud computing environment.

III. SYSTEM ANALYSIS

PROPOSED SYSTEM:

In this paper, we propose a cloud-aided privacy-preserving frequent itemset mining solution for vertically partitioned databases, which is then used to build a privacy-preserving association rule mining solution. Both solutions are designed for applications where data owners have a high level of privacy requirement. The solutions are also suitable for data owners looking to outsource data storage. Here, a secure k-NN classifier works over semantically secure encrypted data. Once the encrypted data are outsourced to the cloud, the user does not participate in any computations Therefore, each of the businesses (i.e. data owners) will own some transaction partitions in the joint database. However, these businesses may not wish to disclose such data, which include trade secrets (e.g. there may be other competing businesses sharing the same joint database) and customer privacy (e.g. due to regulations in existing privacy regime). Therefore, a privacy-preserving mining solution must be applied.

IV. SYSTEM ARCHITECTURE

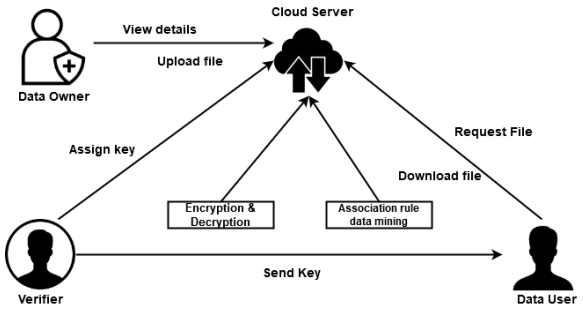

Fig. System Architecture

The steps involved in this process are given below:

- Data Owner
- Query Processing over Encrypted Data,
- Association Rule and Apriori Algorithm,
- Encryption and Decryption module,
- Verifier

Data Owner

In this module, privacy preserving data mining has considered two related settings. One, in which the data owner and the data miner are two different entities, and another, in which the data is distributed among several parties who aim to jointly perform data mining on the united corpus of data that they hold.

Query Processing over Encrypted Data:

Various techniques related to query processing over encrypted data have been proposed. However, we observe that is a more complex problem than the execution of simple queries over encrypted data. For one, the intermediate in the classification process, should not be disclosed to the cloud or any users. We emphasize that the recent method in reveals to the user.

Association Rule:

Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys a dozen eggs, he is 80% likely to also purchase milk."

Apriori Algorithm:

Apriori is designed to operate on databases containing transactions. The purpose of the Apriori Algorithm

is to find associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items and is called a transaction. The output of Apriori is sets of rules that tell us how often items are contained in sets of data.

Encryption Module:

Encryption is a process which transforms the original information into an unrecognizable form. This new form of the message is entirely different from the original message. That's why a hacker is not able to read the data as senders use an encryption algorithm. Encryption is usually done using key algorithms.

Data is encrypted to make it safe from stealing. However, many known companies also encrypt data to keep their trade secret from their competitors.

Decryption Module:

Decryption is a process of converting encoded/encrypted data in a form that is readable and understood by a human or a computer. This method is performed by un-encrypting the text manually or by using keys used to encrypt the original data.

Process

1. Key generation algorithm The key generation algorithm KeyGen() is a probabilistic algorithm, which takes a security parameter λ as input and outputs a secret key.
2. Encryption algorithm The encryption algorithm E() is a probabilistic algorithm, which takes a secret key SK, a plaintext m ∈Fq and a parameter d as inputs. The algorithm outputs a cipher text c ← E(SK, m, d). T.
3. The decryption algorithm D() is a deterministic algorithm, which takes a secret key SK, a cipher text c ∈Fp and the cipher text's degree d as inputs. The algorithm outputs a plaintext m ← D(SK, c, d).
4. Proposed Secure Outsourced Comparison Scheme- The proposed secure comparison scheme is based on the symmetric homomorphism encryption scheme discussed in Section

Equipment performance:

In addition to equipment condition related data, variety of performance related data is available for each piece of mining equipment. This data is collected though fleet dispatch systems now used by a majority of surface mines and some underground mines. Alternatively, this data can be collected by on-board monitoring systems, an example being Caterpillar VIMS system discussed above. If installed on a mining truck the VIMS collects data on truck load size, truck speeds, and the like. It also calculates cycle times and other truck performance related data, and stores all for downloading or transmittal to mine databases

V. CONCLUSION

To protect user privacy, various privacy-preserving classification techniques have been proposed over the past decade. The existing techniques are not applicable to outsourced database environments where the data resides in encrypted form on a third-party server. This paper proposed a novel privacy-preserving data over encrypted data in the cloud with association rule . Our protocol protects the confidentiality of the data, user's input query, and hides the data access patterns. We also evaluated the performance of our protocol under different parameter settings.

REFERENCES

[1] P. Mell and T. Grance, "The NIST definition of cloud computing (draft)," NIST Special Publication, vol. 800, p. 145, 2011.
[2] S. De Capitani di Vimercati, S. Foresti, and P. Samarati, "Managing and accessing data in the cloud: Privacy risks and approaches," in Proc. 7th Int. Conf. Risk Security Internet Syst., 2012, pp. 1–9.
[3] P. Williams, R. Sion, and B. Carbunar, "Building castles out of mud: Practical access pattern privacy and correctness on untrusted storage," in Proc. 15th ACM Conf. Comput. Commun.Security, 2008, pp. 139–148.
[4] P. Paillier, "Public key cryptosystems based on composite degree residuosity classes," in Proc. 17th Int. Conf. Theory Appl. Cryptographic Techn., 1999, pp. 223–238.
[5] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, "k-nearest neighbor classification over semantically secure encrypted relational data," eprint arXiv:1403.5001, 2014.
[6] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. 41st Annu. ACM Sympos. Theory Comput., 2009, pp. 169–178.
[7] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in Proc. 30th Annu. Int. Conf. Theory Appl. Cryptographic Techn.: Adv. Cryptol., 2011, pp. 129–148.
[8] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, pp. 612–613, 1979.
[9] D. Bogdanov, S. Laur, and J. Willemson, "Sharemind: A framework for fast privacy-preserving computations," in

Proc. 13th Eur.Symp. Res. Comput. Security: Comput. Security, 2008, pp. 192–206.
[10] R. Agrawal and R. Srikant, "Privacy-preserving data mining," ACM Sigmod Rec., vol. 29, pp. 439–450, 2000.
[11] Y. Lindell and B. Pinkas, "Privacy preserving data mining," in Proc. 20th Annu. Int. Cryptol. Conf. Adv. Cryptol., 2000, pp. 36–54.
[12] P. Zhang, Y. Tong, S. Tang, and D. Yang, "Privacy preserving Naive Bayes classification," in Proc. 1st Int. Conf. Adv. Data Mining Appl., 2005, pp. 744–752.
[13] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules," Inf. Syst., vol. 29, no. 4, pp. 343–364, 2004.
[14] R. J. Bayardo and R. Agrawal, "Data privacy through optimal kanonymization," in Proc. IEEE 21st Int. Conf. Data Eng., 2005, pp. 217–228.
[15] H. Hu, J. Xu, C. Ren, and B. Choi, "Processing private queries over untrusted data cloud through privacy homomorphism," in Proc.IEEE 27th Int. Conf. Data Eng., 2011, pp. 601–612.
[16] M. Kantarcioglu and C. Clifton, "Privately computing a distributed k-nn classifier," in Proc. 8th Eur. Conf. Principles Practice Knowl. Discovery Databases, 2004, pp. 279–290.
[17] L. Xiong, S. Chitti, and L. Liu, "K nearest neighbor classification across multiple private databases," in Proc. 15th ACM Int. Conf.Inform. Knowl. Manage., 2006, pp. 840–841.
[18] Y. Qi and M. J. Atallah, "Efficient privacy-preserving k-nearest neighbor search," in Proc. IEEE 28th Int. Conf. Distrib. Comput.Syst., 2008, pp. 311–319.
[19] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2004, pp. 563–574.
[20] B. Hore, S. Mehrotra, M. Canim, and M. Kantarcioglu, "Secure multidimensional range queries over outsourced data," VLDB J.,vol. 21, no. 3, pp. 333–358, 2012.