ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

AES Based Classification over Semantically Secured Encrypted Relational Data in Cloud

Madhuvandhi B R[1] , Aravind T[2]

PG Student, Dept. of C.S.E, Muthayammal Engineering College, Rasipuram, Tamilnadu, India

Assistant Professor, Dept. of C.S.E, Muthayammal Engineering College, Rasipuram, Tamilnadu, India

ABSTRACTData mining concepts are used in the effective discovery of correlations among the underlying data in large databases. Classification is one of the usually used tasks in data mining applications. This can be achieved by using either K-NN classifier or FW classifier. Compared to K-NN classifier, FW classifier can provide more security. The data on the cloud is in encrypted form, privacy preserving catalogue techniques are not applicable. K-NN classifier does not support centralized data sharing mechanisms.This work describes a FW classifier based on AES algorithm. The popularity of cloud computing, to expand a secure FW based classifier over encrypted data. Then improve the data sharing, security mechanisms in multiple clouds. The security system is dynamic key generator.

KEYWORDS: Security, FW Classifier, Encryption, Cryptography, AES.

I. INTRODUCTION

Data mining is the process of evaluating data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs. Data mining software is one of a number of analytical tools for evaluating data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Although data mining is a relatively new term, the technology is not. Companies have used powerful computers to shift through volumes of supermarket scanner data and evaluate market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.

Data mining is the practice of automatically probing large stores of data to discover patterns and trends that go beyond simple analysis. Data mining uses cultured mathematical algorithms to section the data and evaluate the probability of future events. Data mining is also known as Knowledge Discovery in Data (KDD).

The cloud computing paradigm is revolutionizing the organizations way of operating their data particularly in the way they store, access and process data. As an emerging computing paradigm, cloud computing attracts many organizations to consider seriously regarding cloud potential in terms of its cost-efficiency, flexibility, and offload of administrative overhead. Most often, organizations delegate their computational operations in addition to their data to the cloud. Despite tremendous advantages that the cloud offers, privacy and security issues in the cloud are preventing companies to utilize those advantages. When data are highly sensitive, the data need to be encrypted before outsourcing to the cloud.When data are encrypted, irrespective of the underlying encryption scheme, performing any data mining tasks becomes very challenging without ever decrypting the data. There are other privacy concerns, demonstrated by the following example. Suppose an insurance company outsourced its encrypted customers database and relevant data mining tasks to a cloud. When an agent from the company wants to determine the risk level of a potential new customer, the agent can use a classification method to determine the risk level of the customer. First, the agent needs to generate a data record q for the customer containing certain personal information of the customer, e.g., credit score, age, marital status, etc. Then this record can be sent to the cloud, and the cloud will compute the class label for q. Nevertheless, since q contains sensitive information, to protect the customer’s privacy, q should be encrypted before sending it to the cloud.

Data Mining over Encrypted Data (DMED) on a cloud also needs to protect a user’s record when the record is a part of a data mining process. Moreover, cloud can also derive useful and sensitive information about the actual data items by observing the data access patterns even if the data are encrypted. Therefore, the privacy/security requirements of the

DOI: 10.15680/IJIRCCE.2016. 0402183

DMED problem on a cloud are threefold: (1) confidentiality of the encrypted data, (2) confidentiality of a user's query record, and (3) hiding data access patterns.

Existing work on 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. Proposed novel methods to effectively solve the DMED problem assuming that the encrypted data are outsourced to a cloud. Specifically 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, concentrates on executing the k-nearest neighbour classification method over encrypted data in the cloud computing environment.

It is possible to use the existing secret sharing techniques in SMC, such as Shamir's scheme, to develop a PPkNN protocol. However, our work is different from the secret sharing based solution in the following aspect. Solutions based on the secret sharing schemes require at least three parties whereas our work require only two parties. For example, the constructions based on Share mind, a well-known SMC framework which is based on the secret sharing scheme, assumes that the number of participating parties is three. Thus, our work is orthogonal to Share mind and other secret sharing based schemes.

II. RELATED WORK

The cloud computing paradigm is revolutionizing the organizations way of operating their data particularly in the way they store, access and process data. As an emerging computing paradigm, cloud computing attracts many organizations to consider seriously regarding cloud potential in terms of its cost-efficiency, flexibility, and offload of administrative overhead. Most often, organizations delegate their computational operations in addition to their data to the cloud. Despite tremendous advantages that the cloud offers, privacy and security issues in the cloud are preventing companies to utilize those advantages. When data are highly sensitive, the data need to be encrypted before outsourcing to the cloud. However, when data are encrypted, irrespective of the underlying encryption scheme, performing any data mining tasks becomes very challenging without ever decrypting the data. There are other privacy concerns, demonstrated by the following example [1].

First, the agent needs to generate a data record q for the customer containing certain personal information of the customer, e.g., credit score, age, marital status, etc. Data Mining over Encrypted Data (DMED) on a cloud also needs to protect a user's record when the record is a part of a data mining process. Moreover, cloud can also derive useful and sensitive information about the actual data items by observing the data access patterns even if the data are encrypted. Therefore, the privacy/security requirements of the DMED problem on a cloud are threefold: (1) confidentiality of the encrypted data, (2) confidentiality of a user's query record, and (3) hiding data access patterns. Existing work on Privacy-Preserving Data Mining (PPDM) cannot solve the DMED problem. Perturbed data do not possess semantic security, so data perturbation techniques cannot be used to encrypt highly sensitive data [2].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. To effectively solve the DMED problem assuming that the encrypted data are outsourced to a cloud. The classification problem since it is one of the most common data mining tasks..Arrayal and Srikant, Lindell and Pinkas were the first to introduce the notion of privacy-preserving under data mining applications. The existing PPDM techniques can broadly be classified into two categories: (i) data perturbation and (ii) data distribution. Agrawal and Srikant proposed the first data perturbation technique to build a decision-tree classifier, and many other methods were proposed later. Data perturbation techniques cannot be applicable for semantically secure encrypted data. Do not produce accurate data mining results due to the addition of statistical noises to the data [3]. The Various techniques related to query processing over encrypted data have been proposed, however observe that PPkNN is a more complex problem than the execution of simple kNN queries over encrypted data. For one, the intermediate k-nearest neighbors in the classification process, should not be disclosed to the cloud or any users. The recent method in reveals the k-nearest neighbors to the user. Second, even if the k-nearest neighbors, it is still very difficult to find the majority class label among these neighbors since they are encrypted at the first place to prevent the cloud from learning sensitive information. Third, the existing work did no addressed the access pattern issue which is a crucial privacy requirement from the user's perspective [4]. In general, if the data involved in clustering belongs to a single entity (hereafter referred to as a user), then it can be done in a trivial fashion. However, in some cases, multiple users, such as companies, governmental agencies, and health care organizations, each holding a dataset, may want to collaboratively perform clustering task on their combined data and share the clustering results. Due to privacy concerns, users may not be willing to share their data with the other users and thus the distributed clustering task1 should be done in a privacy-preserving manner [5]. Privacy-Preserving Distributed Clustering (PPDC), can be best explained by the following example: Consider two health agencies (e.g., the U.S. CDC and the public health agency of Canada) each holding a dataset containing the disease patterns and clinical outcomes of their patients Therefore, they have to collaboratively perform the clustering task on their joint datasets in a privacy-preserving manner. Once the clustering process is done, they can exchange necessary information (after proper sanitization) if needed [6]. Cloud adoption in the NIST U.S. Government Cloud Computing Technology Roadmap. It developing privacy preserving solutions under federated cloud environment will become increasingly important in the near future.The Extensible Markup Language (XML) has been widely adopted in many financial institutions in their daily transactions; this adoption was due to the flexible nature of XML providing a common syntax for systems messaging in general and in financial messaging in specific [7].

Excessive use of XML in financial transactions messaging created an aligned interest in security protocols integrated into XML solutions in order to protect exchanged XML messages in an efficient yet powerful mechanism. There are several approaches proposed by researchers to secure XML messages.. This can involve fully encrypting the entire message, partially encrypting it by selecting parts of each message, or even encrypting external elements attached to the message itself. Although this model is able to secure XML messages, some issues arose concerning performance and inefficient memory usage, leaving room for more improvements and enhancements. However, financial institutions (i.e. banks) perform large volume of transactions on daily basis which require XML encryption on large scale [8]. Fuzzy Logic (FL) approach can be used here to distinguish sensitive parts within each XML document. FL provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information.The FL model is empirically-based, relying on an operator's experience rather than their technical understanding of the system. Their approach is able to encrypt the whole message, full nodes, and sub-trees; however, it is not able to encrypt an element while keeping the descendants of the same node unchanged, and also it cannot handle attribute encryption [9]. A solution has been proposed to handle this limitation. Ed Simon proposed changing the attribute so that it is encrypted with the Encrypted Data Manifest attribute and including any other details inside the element. Another solution proposed was to use XSLT for attribute transformation into elements to perform the encryption process. However, this suggested solution did not face success, as the decrypted parts need to be transformed back to the original attributes for message validation against the corresponding XML schema. A system has been proposed by for pool encryption, which has the capability of removing sensitive information from the output file.These nodes are removed from their original position in the XML message into a pool which contains all other encrypted nodes [10].

III. PROPOSED ALGORITHM

Suggest a secure FW based classification technique to improve the data sharing in multiple clouds classifier over encrypted data in the cloud. The proposed protocol protects the confidentiality of data, privacy of user’s input query, and hides the data access patterns. To the best of our knowledge, our work is the first to develop a FW based classification technique over encrypted data under the semi-honest model. Also, we empirically analyse the efficiency of our proposed protocol using a real-world dataset under Different parameter settings.

- Centralized security control in cloud server
- AES 256bit Security Mechanisms for Cloud Computing
- Automate deployment in data sharing in cloud server
- OTP based random key process
- Time limit for data sharing in cloud server
- Unauthorized user not hacking the data in cloud server

Algorithm Description

AES is based on a design principle known as a substitution-permutation network, combination of both substitution and permutation, and is fast in both software and hardware. Unlike its predecessor DES, AES does not use a Feistel network. AES is a variant of Rijndael which has a fixed block size of 128 bits, and a key size of 128, 192, or 256 bits. By contrast, the Rijndael specification per *se* is specified with block and key sizes that may be any multiple of 32 bits,

both with a minimum of 128 and a maximum of 256 bits. The key size used for an AES cipher specifies the number of repetitions of transformation rounds that convert the input, called the plaintext, into the final output, called the cipher text. The number of cycles of repetition are as follows
Each round consists of several processing steps, each containing four similar but different stages, including one that depends on the encryption key itself. A set of reverse rounds are applied to transform cipher text back into the original plaintext using the same encryption key.

SYSTEM ARCHITECTURE

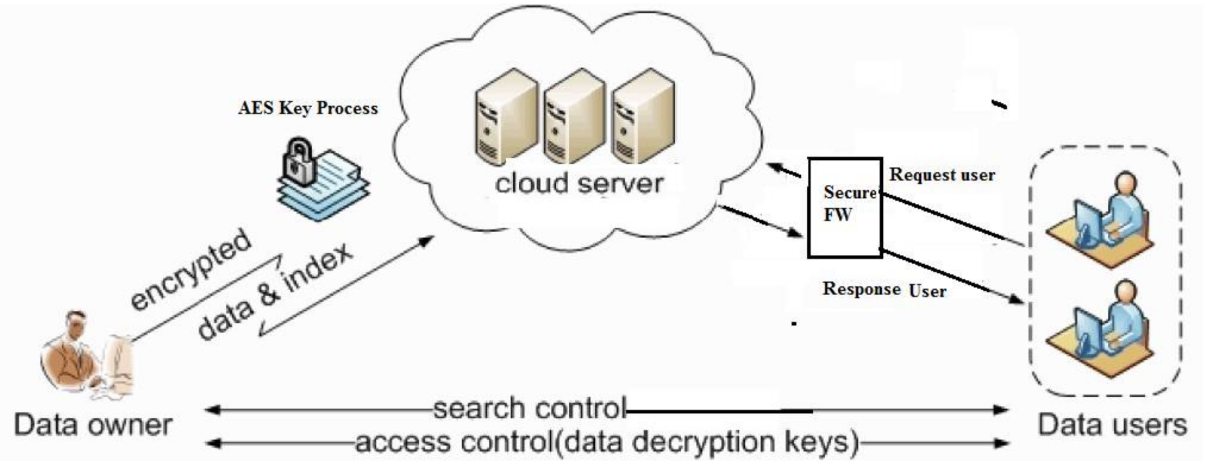


Fig. System Architecture

IV. SYSTEM ORGANIZATION

MODULES DESCRIPTION

User Login

In this module using user and data owner login process maintaining here. The data owner login data owner only login, data owner can using all the process of the network in cloud data sharing processes. User login limited processes only allowed here.

Data Owner

To completely outsource its data management and operational tasks to a cloud. For example, Alice may want to access her data and analytical results using a smart phone or any device with very limited computing capability. Suppose Bob wants to keep his input query and access patterns private from Alice.

File Upload

Files can be stored and uploaded to Cloud File Server using a variety of methods. Smaller files can be uploaded using the web browser interface. Then upload the file shared permission based folders around projects or departmental needs and allow both users and customers to securely access their files from anywhere.

Searching and Download the File

This modules are performance about searching the data anywhere of the place and taken here and download the file in anywhere of the place and eve time user available location.

Security Module

This modules are performance about the security mechanism in cloud storage. Cryptographic Security Mechanisms for Cloud support for secure key process and file encryption process has been used here.

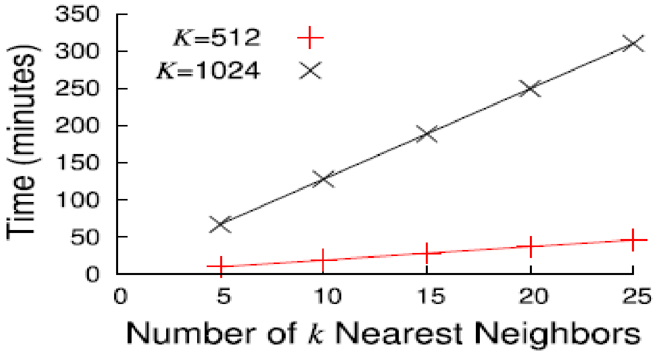
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

V. SIMULATION AND RESULTS



(a) Total cost of Stage 1

An execution image generally includes the messages exchanged and the information computed from these messages. To prove a protocol's security under the semi-honest model, the well-known security definitions from the literature of SMC. Even when an adversary modifies the intermediate computations he/she cannot gain any additionalInformation. The main goal of SMC is to ensure the honest parties to get the correct result and to protect theirPrivate input data from the malicious parties.

VI. CONCLUSION AND FUTURE WORK

The secure FW based classification techniques is commonly used queries in many data mining applications. Under an outsourced database environment, where encrypted data are stored in the cloud, secure query processing over encrypted data becomes challenging. The existing k-NN techniques over encrypted data are not secure.This proposed method over encrypted data in the cloud. In this method performing the task, which acts as a basic solution, leaks some information to the cloud and fully secure, that is, it protects the confidentiality of the data, user's input query, and also hides the data access patterns Also, evaluated the performance of our method under different parameter settings. As a future work, we will investigate and extend our research to other complex conjunctive queries over encrypted data.

As a future work, in order to reduce the human expert knowledge intervention and increase performance of the phishing detection system. This can be achieved by generating classification rules using well known classifiers; and improve to performing the security and user interaction of cloud storage.

REFERENCES

1. R. Agrawal and R. Srikant, "Privacy-preserving data mining," ACM Sigmod Rec.,vol.29, pp. 439–450, 2000
2. 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.
3. 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 Securit Internet Syst., 2012, pp. 1–9.
4. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving Mining of association rules," Inf. Syst., vol. 29, no. 4,pp. 343–364, 2004.
5. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. 41st Annu ACM Sympos. Theory Comput., 2009, pp. 169–178.
6. 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.
7. Y. Lindell and B. Pinkas, "Privacy preserving data mining," inProc. 20th Annu. Int. Cryptol. Conf. Adv. Cryptol., 2000, pp. 36–54.
8. P.Paillier, "Public key cryptosystems based on composite degree residuosityclasses,"in Proc. 17th Int. Conf. Theory Appl. Cryptographic Techn., 1999, pp. 223–238.
9. B.k.Samanthula, Y.Elmehdwi "k-nn classification over semantically secure Encrypted relational data", vol 27, no.5, 2015.
10. 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.