INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

PRIVACY PRESERVING TECHNIQUE USING K-NN CLASSIFICATION IN OUTSOURCES ENVIRONMENT

RASHMI SHESHRAO KODANE[1], PROF. KARUNA BAGDE[2]

1. Student of Master of Engineering (CS&IT), HVPM's college of Engineering and Technology Amravati, India.
2. Associate Professor (CSE), HVPM's College of Engineering and Technology Amravati, India.

Accepted Date: 15/03/2016; Published Date: 01/05/2016

Abstract*:* Data Mining has wide use in many fields such as banking, medicine, scientific research and among government agencies. Classification is one of the commonly used tasks in data mining applications. With the recent popularity of cloud computing, users now have the opportunity to outsource their data, in encrypted form, as well as the data mining tasks to the cloud. Since the data on the cloud is in encrypted form, existing privacy-preserving-classification techniques are not applicable. In this paper, we focus on solving the classification problem over encrypted data. In particular, we propose a secure k-Nearest Neighbour (k-NN) 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. Our work is the first to develop a secure k-NN classifier over encrypted data under the semi-honest model.

Keywords: Security, K-NN Classifier, Outsourced Databases, Encryption, Privacy Preserving.

Corresponding Author: MS. RASHMI SHESHRAO KODANE

Access Online On:

www.ijpret.com

How to Cite This Article:

Rashmi Sheshrao Kodane, IJPRET, 2016; Volume 4 (9): 938-943

PAPER-QR CODE

INTRODUCTION

Lately, the cloud computing model [1] is changing the landscape of the organizations' way of working their information especially in the way they save access and process data. As a growing processing model, cloud processing draws many organizations to think about seriously concerning cloud potential with regards to its cost-efficiency, versatility, and offload of management expense. 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. Data mining over encrypted data (denoted by 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[2],[3]. 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. 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.

I. GOALS AND OBJECTIVE

- Improving the efficiency of SMINn is an important first step for improving the performance of our PPkNN protocol.
- 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.

II. PRIVACY-PRESERVING DATA MINING

Agrawal and Srikant [6], Lindell and Pinkas [7] 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 [6] proposed the first data perturbation technique to build a decision-tree classifier, and many other methods were proposed later (e.g., [8], [9]).

In our most recent work [10], we proposed a novel secure k-nearest neighbor query protocol over encrypted data that protects data confidentiality, user's query privacy, and hides data access patterns. More specifically, this paper is different from our preliminary work [10] in the following four aspects. First, in this paper, we introduced new security primitives, namely secure minimum (SMIN), secure minimum out of n numbers (SMINn), secure frequency (SF), and proposed new solutions for them. Second, the work in [10] did not provide any formal security analysis of the underlying sub-protocols. On the other hand, this paper provides formal security proofs of the underlying sub-protocols as well as the PPkNN protocol under the semi-honest model. Additionally, we discuss various techniques through which the proposed PPkNN protocol can possibly be extended to a protocol that is secure under the malicious setting. Third, our preliminary work in [10] addresses only secure kNN query which is similar to Stage 1 of PPkNN. However, Stage 2 in PPkNN is entirely new. Finally, our empirical analyses in Section 6 are based on a real dataset whereas the results in [10] are based on a simulated dataset. Furthermore, new experimental results are included in this paper.

III. PROPOSED TECHNIQUE:

Cryptography based PPDM (Privacy Preserving Data Mining)

This technique includes secure multiparty computation where a computation is secure if at the completion of the computation, no one can know anything except its own input and the results. Cryptography based algorithms are considered for protective privacy in a distributed situation by using encryption techniques. Transformed data are exact and protected. Better privacy compare to randomized approach. It offers a well-defined model for privacy that includes methods for proving and quantifying it. Second a vast set of cryptographic algorithms and constructs to implement privacy preserving data mining algorithms are available in this domain.

Advantage of Cryptography based PPDM

1) Because the process is transparent, it is easy to implement and debug.

2) In situations where an explanation of the output of the classifier is useful Cryptography based PPDM can be very effective if an analysis of the neighbors is useful as explanation.

3) There are some noise reduction techniques that work only for Cryptography based PPDM that can be effective in improving the accuracy of the classifier.

AES and DES:

- DES is really old while AES is relatively new.
- DES is breakable while AES is still unbreakable.
- DES uses a much smaller key size compared to AES.
- DES uses a smaller block size compared to AES.

AES and 3DES:

- AES is more secure (it is less susceptible to cryptanalysis than 3DES).
- AES supports larger key sizes than 3DES's 112 or 168 bits.
- AES is faster in both hardware and software.
- AES's 128-bit block size makes it less open to attacks.
- AES is required by the latest U.S. and international standards.

IV. PROPOSED SYSTEM

We use this algorithm for maintaining privacy policies. The overall structure of AES encryption/decryption is shown below. The number of rounds is for the case when the encryption key is 128 bit long. (As mentioned earlier, the number of rounds is 12 when the key is 192 bits and 14 when the key is 256.)Before any round-based processing for encryption can begin, the input state array is XO Red with the first four words of the key schedule. The same thing happens during decryption except that now we XOR the cipher text state array with the last four words of the key schedule.

For encryption, each round consists of the following four steps:

1) Substitute bytes,

2) Shift rows,

3) Mix columns, and

4) Add round key.

The last step consists of XO Ring the output of the previous three steps with four words from the key schedule.

For decryption, each round consists of the following four steps:

1) Inverse shift rows,

2) Inverse substitute bytes,

3) Add round key, and

4) Inverse mix columns.

The third step consists of XO Ring the output of the previous two steps with four words from the key schedule. Note the differences between the order in which sub-situation and shifting operations are carried out in a decryption round vis-a-vis the order in which similar operations are carried out in an encryption round. The last round for encryption does not involve the “Mix columns” step. The last round for decryption does not involve the “Inverse mix columns” step.

V. CONCLUSION:

To protect user privacy, various privacy-preserving classification techniques have been proposed over the past decade. The k-nearest neighbors is one of the commonly used query 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. This paper proposed a novel privacy-preserving k-NN classification protocol over encrypted data in the cloud. 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. Since improving the efficiency of SMINn is an important first step for improving the performance of our PPkNN protocol, we plan to investigate alternative and more efficient solutions to the SMINn problem in our future work. Also, we will investigate and extend our research to other classification algorithms.

VI. 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. B. K. Samanthula, Y. Elmehdwi, and W. Jiang, "k-nearest neighbor classification over semantically secure encrypted relational data," eprint arXiv: 1403.5001, 2014.
5. R. Agrawal and R. Srikant, "Privacy-preserving data mining," ACM Sigmod Rec., vol. 29, pp. 439–450, 2000.
6. Y. Lindell and B. Pinkas, "Privacy preserving data mining," in Proc. 20th Annu. Int. Cryptol. Conf. Adv. Cryptol., 2000, pp. 36–54.
7. 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.
8. R. J. Bayardo and R. Agrawal, "Data privacy through optimal annonymization," in Proc. IEEE 21st Int. Conf. Data Eng., 2005, pp. 217–228.
9. Y. Elmehdwi, B. K. Samanthula, and W. Jiang, "Secure k-nearest neighbor query over encrypted data in outsourced environments," in Proc. IEEE 30th Int. Conf. Data Eng., 2014, pp. 664–675.
10. 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.
11. 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.