K-Nearest Neighbor Categorization Problem over Semantically Protected Encrypted Relational Database

Ms. Radhika S. Morey
Computer Science & Engineering
Prof. Ram Meghe Institute of Technology & Research
Badnera.

Prof. R.R.Tuteja
Computer Science & Engineering
Prof.Ram Meghe Institute of Technology & Research
Badnera

Abstract— Data Mining has so many uses in various fields as banking, medicine, scientific research and among government sector. Classification is a method used in data mining applications. For the last few years, because of increase in various privacy issues, different type of solutions to the classification problem have been proposed under different security models. Since the data on the cloud is in encrypted form, existing privacy preserving. Classification techniques are not applicable. We focus on solving the classification problem over encrypted data. In this, we suggest a secure k-NN classifier over encrypted data in the cloud. The proposed protocol protects the security of data, privacy of user input data, and hides the data access patterns.

***Index Terms*- Security, k-NN classifier, Outsourced databases, Encryption, Relational database, Data mining, Classification**

I. INTRODUCTION

For the last few years, because of increase in various privacy issues, different type of solutions to the classification problem have been given under various security models. In this paper, we focus on solving the classification problem over encrypt data. In this, we suggest a secure classification of data in encrypted relational data in database. Data mining used in encrypted data (denoted by DMED) needs to protect a user's record when the record field used in a data mining process. As well as, cloud can be derive useful and sensitive information about the actual data by seeing the data entry, if the data are encrypted [2], [3]. Hence, the security requirements for DMED problem in data mining are given below:

(1) Confidentiality of the encrypted data,

(2) Confidentiality of a user's database,

(3) Hiding data entry sample.

In this paper, we propose a novel PPkNN protocol, a privacy preserving k-NN classifier over semantically secure encrypted data.

II. RELATED WORK

Due to space limitations, here we briefly review the existing related work and provide some definitions as a background. At first, it seems fully homomorphic crypto systems can solve the DMED problem since it allows a third party (that hosts the encrypted data) to execute arbitrary functions over encrypted data without ever decrypting them. However, we stress that such techniques are very expensive and their usage practical applications have yet to be explored.

Shamir's scheme [6], to develop a PPkNN protocol. For example, the constructions based on Sharemind [7], a well-known SMC framework which is based on the secret sharing scheme, assumes that the number of participating parties is three.

- Privacy-Preserving Data Mining
- Query Processing over Encrypted Data

Agrawal and Srikant [2], Lindell and Pinkas [3] 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:

1) Data perturbation and

2) Data distribution.

Agrawal and Srikant [10] proposed the first data perturbation technique to build a decision-tree classifier, and many other methods were proposed later (e.g., [4], [5], [8]).

However, as mentioned earlier, data perturbation techniques cannot be applicable for semantically secure encrypted data. Also, they do not produce accurate data mining results due to the addition of statistical noises to the data. On the other hand, Lindell and Pinkas [3] proposed the first decision tree classifier under the two-party setting assuming the data were distributed between them. We claim that the PPkNN problem cannot be solved using the data distribution techniques.

Various techniques related to query processing over encrypted data have been proposed, e.g., [12], [13], [14].However, we observe that PPkNN is a more complex problem than the execution of simple kNN queries over encrypted data [9].

In our most recent work, 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. However, as mentioned above, PPkNN is a more complex problem and it cannot be solved directly using the existing secure k-nearest neighbor techniques over encrypted data.

III. Proposed Work

A. Motivation

Most often, organizations delegate their computational operations in addition to their relational data. When data are highly sensitive, the data need to be encrypted before outsourcing. Whenever, data are encrypted, under some encryption scheme, performing any data mining tasks becomes very challenging without ever decrypting the relational data. Also we were classifies our relational data base on their type and stored the data in the system.

B. System Architecture

To protect user privacy, various privacy-preserving classification techniques have been proposed over the past decade. This paper proposed a novel privacy-preserving encryption over encrypted data. The technique we used in this paper that protects the confidentiality of the data, user's input query, and hide the data entry. We also examine the result which can be produce by the technique under different conditions that improve the efficiency and performance of our application; we plan to investigate alternative and more efficient solutions to the problem in our future work. We are going to implement the system in which any user can upload his data, files to the server and that will stored in relational database in encrypted format. We used secrete key which can be occurred in algorithm for maintaining privacy. The detail description of modules that we are going to create in this system is described as follows:

Modules Description

- **Proprietor**

Data Owner as proprietor

1. View Profile
2. Update Profile
3. Upload Files
4. View Files
5. View Enjoyer Request
6. Send Key to Enjoyer

You have to register to the system, and then the request sends to the Administrator.

This user can only login when admin give access to this ssuser.

- **Enjoyer**

Data Miner as Enjoyer

1. Search File
2. Request File
3. Download File Using Secrete Key

This user must register first for getting login to system. After login the user can search file, send request for downloading the file.If user want to download any file then first of all request sent to the proprietor of that file with its secret key which will generated by system.

Now the proprietor has to login in the system, he/she can see the request of for that specific file. If this user accepts the request then secrete key will sent on the registered email of the enjoyer. Now enjoyer can use that key for downloading the file.

- **Server (Administrator)**

1. Activate proprietors
2. View proprietor
3. View All Data in Classified Form

C. Proposed System Design

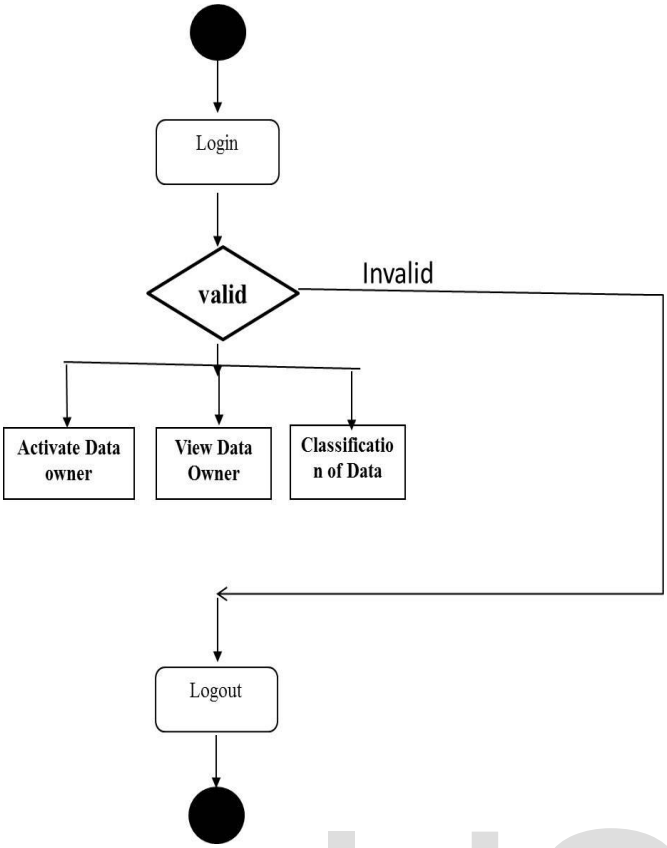

Fig 1. DFD for Administrator

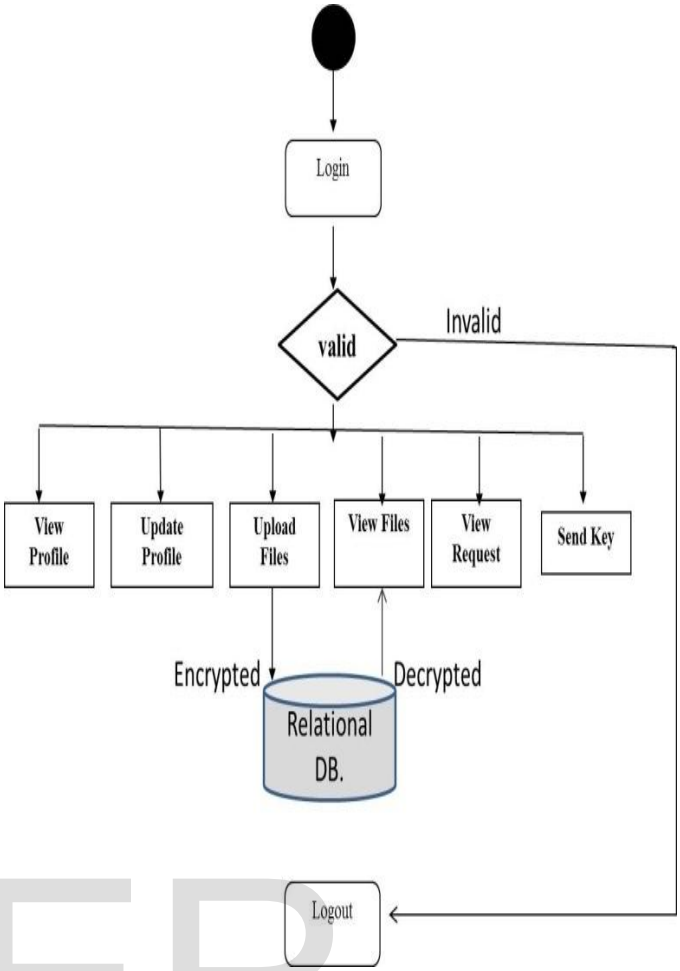

Fig 2. DFD for Data Owner

D. Proposed System Algorithm

AES Algorithm:-

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. When some round-based process for encryption can be occur, the input state array is XORed with the first four words of the key which can be schedule. The process which can be occurred in encryption same process performed during decryption except that we XOR the cipher text state array with the last four words of the key which can be schedule.

AES is the cryptographic algorithm for maintaining privacy policies for encryption and decryption. In AES algorithm a block cipher with a block length of 128 bits and the key length is 128,192,256 bits. For 10 rounds of processing for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for

256-bit keys.AES will become de facto standard encrypting all forms of electronic information.

IV. 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. We proposed a novel privacy-preserving k-NN classification protocol over encrypted data in the cloud. Our protocol gives protection to the data, and user input database, and hides the data access patterns. We also evaluated the achievement of protocol under different conditions.

Proposed a system in which user can upload, download the files. Whole data can be stored in encrypted form. For data storing relational database can be used. As per the security purpose privacy should be given to the data.

Acknowledgment

I wish to thank my guide prof. R.R.Tuteja madam for their valuable support, guidance and suggestions.

REFERENCE

[1] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, ―k-nearest neighbor classification over semantically secure encrypted relational data,"eprint arXiv:1403.5001, 2014.

[2] R. Agrawal and R. Srikant, ―Privacy-preserving data mining," ACM Sigmod Rec., vol. 29, pp. 439–450, 2000.

[3] Y. Lindell and B. Pinkas, ―Privacy preserving data mining," in Proc. 20th Annu. Int. Cryptol. Conf. Adv. Cryptol., 2000, pp. 36–54.

[4] 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.

[5] 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.

[6] A. Shamir, ―How to share a secret," Commun. ACM, vol. 22,pp. 612–613, 1979.

[7] D. Bogdanov, S. Laur, and J. Willemson, ―Sharemind: A frameworkfor fast privacy-preserving computations," in Proc. 13th Eur.Symp. Res. Comput. Security: Comput. Security, 2008, pp. 192–206.

[8] R. J. Bayardo and R. Agrawal, ―Data privacy through optimal k-anonymization," inProc. IEEE 21st Int. Conf. Data Eng, 2005, pp. 217–228.

[9] 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.

[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.

[12] 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.

[13] 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.

[14] H. Hacig €um€ų s , B. Iyer, C. Li, and S. Mehrotra, ―Executing sql over encrypted data in the database-service-provider model," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2002, pp. 216–227.