IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727
PP 45-48
www.iosrjournals.org

k-Nearest Neighbor Classification Over Encrypted Cloud Data

Pooja Bajare [1] , Monika Bhoyate [2] ,Yogita Bhujbal [3] , Erandole Monika [4] ,Vaishali Shinde [#5]

[1]*(Department of Computer Engineering, SavitribaiPhule Pune University, Pune, India)*
[2]*(Department of Computer Engineering, SavitribaiPhule Pune University, Pune, India)*
[3]*(Department of Computer Engineering, SavitribaiPhule Pune University, Pune, India)*
[4]*(Department of Computer Engineering, SavitribaiPhule Pune University, Pune, India)*
[5]*(Department of Computer Engineering, SavitribaiPhule Pune University, Pune, India)*

Abstract: *Today data mining is used in many applications areas medical, scientific research, banking and many more. From last decade, Internet has given rise to many privacy issues. To solve these issues many theoretical and practical solutions to the classification problem have been proposed using different security models. However, cloud computing allow users to outsource their data to cloud. User prefers to encrypt the data before storing it on cloud, but performing any classification on encrypted data is main issue. Today's privacy-preserving classification techniques are not useful for encrypted data, so here we uses k-NN classifier over encrypted data in the cloud. The proposed technique protects the security of data, privacy of user's input query, and hides the access patterns. Our aim is to develop a secure k-NN classifier on encrypted data using the semi-honest model. Also, efficiency of K nearest neighbor classification is analyzed using real world data set under different parameters conditions.*

Keywords: *cloud****,*** *encryption, k-NN classifier, Security, outsourced databases*

I. Introduction

Recently, the cloud computing paradigm[1] is revolutionizing the organizations in way of operating their data particularly in the way they store, access and process data [5]. As an emerging computing paradigm, in cloud computing, it attracts many organizations to related cloud potential in terms of its cost-efficiency, flexibility, and offload of administrative overhead. cloud can also derive useful and sensitive information about the actual data items by observing the different data access patterns even if there is a data is encrypted [2], [3]. Most often, organizations delegate their computational operations in addition to their data to the cloud. The advantage of cloud is that the privacy and security issues in the cloud which are prevents the companies to utilize those advantages. The data need to be encrypted before outsourcing to the cloud when data is highly sensitive. However, when data are encrypted, irrespective of the underlying encryption scheme, it is very challenging to performing any data mining tasks ever decrypting the data

II. System Architecture

Description:

In this architecture system,

- First, data owner registers and get login details after that data owner upload the documents on cloud.
- To prevent uploaded data from accessing unauthorized user, we are applying encryption algorithm Elliptic Curve Cryptography (ECC).
- When registered users tries to retrieve data from cloud, here we are using k-nn classification technique
- Then for decryption of uploaded data, Secret key is provided to authorized user when she/he downloads file from cloud

III. Data Model And Description:

- User Module
- Data Owner Module
- Verifier Module
- Cloud module

1.Data Owner Module:

Input : data files

Output: upload data files to cloud

2.Data User Module:
Input : Send file request,
Output: Get file, download using secret key.

3.Verifier :
Input: Check users identity, generate key
Output : Issues secret key to registered user

4.Cloud server:
Input: Stores encrypted data
Output: View Properties, send cipher text when they get permission

IV. Technical Study

1)Classification: Classification technique is one of the techniques used in data mining. Classification means assigning class label to unknown data. Classification is analysis to extract data models. there are many techniques are available. Popular from them are decision trees, naïve bays, K nearest neighbor classification. Many classification methods are developed by researchers in machine learning, pattern recognition, and statistics. Classification has many applications including, Marketing, manufacturing ,and medical etc. Data classification is in two step process, one is construction and model used to prediction class label

2)Cryptography:
The concept of the cryptography includes
Plaintext : It is the original message before it encryption
Encryption Algorithm: The algorithm which gives the encrypted plain text called cipher text.
Decryption Algorithm: The algorithm which gives the decrypted cipher text and which is plain text.
Encryption or Decryption Key: The key which used in encryption or decryption algorithm. It may be of various size, numeric or string.The modern field of the Cryptography can be divided into several areas of study. The Chief ones are mentioned here.
a. Symmetric or Secret key cryptography
b. Asymmetric or public key Cryptography.

HARDWARE RESOURCES REQUIRED

System	:	Pentium IV 2.4 GHz.
Hard Disk	:	40 GB.
Monitor	:	15 VGA Colour.
Mouse	:	Logitech.
Ram	:	512 Mb

SOFTWARE RESOURCES REQUIRED

Operating system	:	Windows XP/7.
Coding Language	:	JAVA/J2EE
Database	:	MYSQL
Tools	:	Eclipse.

V. Figure

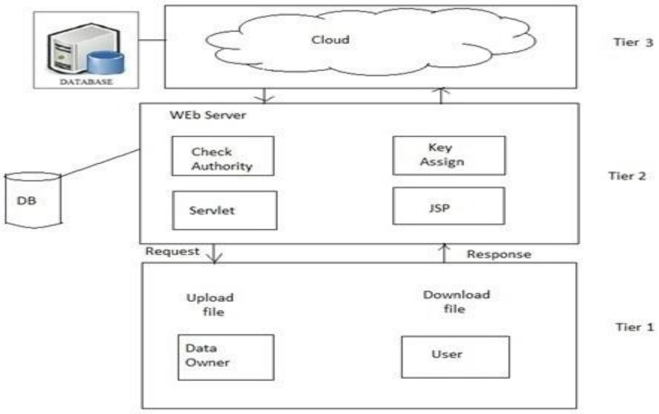

Figure 1: Proposed system

VI. Proposed System

Existing work on Privacy-Preserving Data Mining which cannot solve the DMED (Data Mining on Encrypted Data)problem. Perturbed data do not possess semantic security, There is one technique is available i.e. perturbation technique. But this technique cannot be used to encrypt the highly sensitive data. Also the perturbed data do not produce very accurate data mining results. (SMC) Secure multi-party computation is based approach assumes data are not encrypted at each participating party as well as it is distributed. In this, we are first develop a secure KNN classifier over encrypted data in the semi-honest model. Using a real-world dataset, we analyze the efficiency of our proposed technique using with different parameter settings.

We plan to find alternative and more efficient solutions to the SMINn problem. We propose secure privacy preserving k-NN classifier on semantically secure encrypted data. In our technique, we are outsource our encrypted data on cloud and retrieve with help of KNN classification technique. We are using elliptic curve cryptography (ECC) for encryption which is alternative to SMC
Usefulness of cloud:
1)Cloud needs to protect a user's record when the record is a part of a data mining process. Cloud can derive useful and sensitive information about the actual data items by observing the data access patterns even if the data are encrypted.
2)When data are highly sensitive, the data need to be encrypted before outsourcing to the
cloud. And secure retrieve information using a novel privacy-preserving k- NN classification
technique.
In this paper, if the encrypted data are outsourced to the cloud, then the Alice does not participate in any computations. So, there is no information is disclose to Alice side. In addition, our algorithm meets the following privacy requirements:

- Any intermediate results or Contents of
D should not be disclose to the cloud.
- Bob's query q should not be disclose to the cloud.
- Cq should be tell result only to Bob. Also, no other information should be disclose to Bob.

Data access patterns, like as the records corresponding to the k nearest neighbors of q should not be disclose to Bob and the cloud.In our technique are either newly generated randomized encryptions or random numbers we emphasize that the intermediate results seen by the cloud. Thus, which data records correspond to the k -nearest neighbors and the output class label are unknown to the cloud. In addition, after sending his encrypted query record to the cloud, Bob does not involve in any computations. So our data access patterns are further secured from Bob.

VII. Conclusion

To protect user privacy, various privacy- preserving classification techniques have been proposed over the past decade. The existing techniques are not useful to store database environment ,where the data resides in encrypted form on a third- party server. This paper proposed a novel privacy- preserving k-NN classification technique over encrypted data in the cloud. Our technique protects the confidentiality of the data, user's input query, and hides the data access pat terns. We also evaluated the performance of our technique under different parameter set tings. Since improving the efficiency of SMIN n is an import ant first step for improving the performance of our PPkNN technique, we plan to investigate alternative and more efficient solutions to the SMIN n problem in our future work. Also, we will investigate and extend our research to other classification algorithms.

Acknowledgements

We would like to acknowledge the guidance of Ms V.D.Shinde for her insightful support and inspiration throughout the various stages of this paper. We sincerely appreciate the help and advice given by her which went a long way in helping us understanding the key concept of this paper.

References

[1]. Mell and T. Grance, "The nist definitionof 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 CRiSIS, pp. 1 –9, 2012.
[3]. P. Williams, R. Sion, and B. Carbunar, "Building castles out of mud: practical access pattern privacy and correctness on untrusted storage," in ACM CCS , pp. 139–148, 2008.
[4]. P. Paillier, "Public key cryptosystems based on composite degree residuosity classes," in Eurocrypt , pp. 223–238, 1999.
[5]. B. K. Samanthula, Y. Elmehdwi, and W. Jiang, "k-nearest neighbor classification over semantically secure encrypted re-lational data." eprint arXiv:1403.5001, 2014.
[6]. C. Gentry, "Fully homomorphic encryption using ideal lattices," in ACM STOC , pp. 169–178, 2009.