A Novel Secure KNN Classifier with a Light Weight highly Secure Encryption Scheme over Encrypted Data in the Cloud

[1]P. Bharathi, [2]N. Anand Reddy
[1]PG Scholar, Dept. of CSE, Siddhartha Educational Academy Group of Institutions, Tirupati
[2]Asst. Prof. & HOD, Dept. of CSE, Siddhartha Educational Academy Group of Institutions, Tirupati

Abstract: Data Mining has wide applications in many areas such as banking, medicine, scientific research and among government agencies. Classification is one of the commonly used tasks in data mining applications. For the past decade, due to the rise of various privacy issues, many theoretical and practical solutions to the classification problem have been proposed under different security models. However, 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-NN classifier with light-weight highly secure encryption scheme 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 secure k-NN classifier over encrypted data under the semi-honest model. Also, we empirically analyze the efficiency of our proposed protocol using a real-world dataset under different parameter settings**.**

Index Terms **- Classifiers, cloud computing, data mining, encryption.**

I. INTRODUCTION

Recently, the cloud computing paradigm is revolutionizing the organizations' way of information processing, storage, and delivery in which highly centralized physical resources are furnished to remote clients on demand .Rather than purchasing actual physical devices—servers, storage, and networking equipment clients lease these resources from a cloud provider as a outsourced service that abstracts away physical devices [1][2]. 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. Cloud computing is flexible and portable in that it can be accessed anytime from anywhere. By using redundant sites and backup storage, cloud providers can also provide greater reliability than local computing systems. 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 [3]. 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 [4][5].

II. EARLIER WORK

In this section, we first formally dene the privacy preserving k-NN problem, then briefly review the privacy preserving protocol, upon which our work is built.

2.1 Privacy preserving Data mining

In privacy-preserving Data mining where private data and private knowledge must be protected, two main approaches to preserve privacy are data medication approach and cryptographic approach. Our paper takes the latter approach where secure multiparty computation (SMC) plays a critical role. In this setting, data is distributed across multiple parties either horizontally (the records are distributed among the parties) or vertically (columns of a record are distributed among the parties). The parties want to conduct a computation based on their private inputs such that they learn the outputs without revealing their own inputs or outputs to others, and no party can infer anything other than what can be learned from its own input and output.

2.2 Query Processing over Encrypted Data

For the past decade, query processing on relational data has been studied extensively, and many theoretical and practical solutions to query processing have been proposed under various scenarios. With the recent popularity of cloud computing, users now have the opportunity to outsource their data as well as the data management tasks to the cloud. However, due to the rise of various privacy issues, sensitive data (e.g., medical records) need to be encrypted before outsourcing to the cloud. In addition, query processing tasks should be handled by the cloud; otherwise, there would be no point to outsource the data at the first place. To process queries over encrypted data without the cloud ever decrypting the data is a very challenging task. In this paper, we focus on solving the k-nearest neighbor (kNN) query problem over encrypted database outsourced to a cloud: a user issues an encrypted query record to the cloud, and the cloud returns the k closest records to the user.

We first present a basic scheme and demonstrate that such a naive solution is not secure. To provide better security, we propose a secure kNN protocol that protects the confidentiality of the data, user's input query, and data access patterns. Also, we empirically analyze the efficiency of our protocols through various experiments. These results indicate that our secure protocol is very efficient on the user end, and this lightweight scheme allows a user to perform the kNN query.

2.3 Paillier Cryptosystem

The Paillier Cryptosystem named after and invented by French researcher Pascal Paillier in 1999 is an algorithm for public key cryptography. The distinguishing technique used in public key cryptography is the use of asymmetric key algorithms, where the key used to encrypt a message is not the same as the key used to decrypt it. Each user has a pair of cryptographic keys — a public key and a private key. The private key is kept secret, whilst the public key may be widely distributed. Messages are encrypted with the recipient's public key and can only be decrypted with the corresponding private key. The keys are related mathematically, but the private key cannot be feasibly (i.e, in actual or projected practice) derived from the public key.

III. PRIVACY PRESERVING DATA MINING

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 [6][7].

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. Earlier the organizations outsource their data to cloud and the users working for organization access the data. Even though, each user may have different user id's and password there may be chances to hack the data.

If the attacker hacks the database that contains users' password and logins data then secrecy of data will be lost or if the remote servers on the cloud compromises with the attacker then he can access the main database where security becomes a challenging issue. Later, with the advent of cloud computing organizations have started outsourcing data in encrypted form to the cloud [8]. One straightforward way to protect the confidentiality of the outsourced data from the cloud as well as from unauthorized users is to encrypt data by data owner can protect the privacy of her own data. Database owner encrypts the data in database attribute-wise using secret key (s_k) before outsourcing to cloud and does not involve in any computations on database.

In Multi-party computation, two parties are involved encrypted data (D') is stored in one party (P_1) and P_2 holds shared key [9]. While processing the query some intermediate computations will be processed on decrypted data. Therefore the data access patterns may get revealed to the hacker who has been listening throughout the query processing [10][11].

2.1 Limitations

- Data do not possess semantic security.
- Data mining results are not accurate.
- Existing approach assumes data are distributed and not encrypted.
- Reveals the intermediate neighborhood information, thus reducing the potential information leakage.
- Cloud infrastructure costs high.

IV. NOVEL SECURE KNN CLASSIFIER

A novel secure knn classifier was proposed in this paper with a light weight highly secure encryption scheme over encrypted data in the cloud to reduce the computational complexity [12][13]. The proposed model is depicted in Fig. 1. When data are encrypted irrespective of the encryption scheme, performing any data mining tasks becomes very challenging without ever decrypting the data. Therefore, the privacy/security requirements of the DMED problem on a cloud are threefold:

- Confidentiality of the encrypted data,
- Confidentiality of a user's query record, and
- Hiding data access patterns.

And also we propose a novel secure single party computation technique, to reduce cloud infrastructure cost instead of multi-party computation. To reduce the complexity of algorithms we use a light-weight encryption scheme called ROT (i) [14][15].

4.1 Advantages

Achieves Confidentiality of data, User's input query privacy and hiding data access patterns. Solves the classification problem over encrypted data. Increases the performance of proposed ppknn protocol. Reduces the cost of infrastructure on cloud. Reduces the complexity of algorithm by using light-weight encryption scheme.

4.2 Modules

Data Owner

This module helps the owner to register those details and also include login details. This module helps the owner to upload his file with encryption using RSA algorithm. This ensures the files to be protected from unauthorized user [16][17]. This module allows the Owner to view the uploaded files and downloaded files

Cloud Server

A cloud server is a logical server that is built, hosted and delivered through a cloud computing platform over the Internet. Cloud servers possess and exhibit similar capabilities and functionality to a typical server but are accessed remotely from a cloud service provider [18]. A cloud server may also be called a virtual server or virtual private sever.

Authorized User

This module includes the user registration, login details. This module allows the user to download the file using his secret key to decrypt the downloaded data [19].

Admin

This module assigns roles to the users. This module has authorities to add or remove users.

Crypt Util

Stage 1 - Secure Retrieval of k-Nearest Neighbors (SRkNN): In this stage, Bob initially sends his query q (in encrypted form) to P(secure single party server). After this, P involve in a set of sub-protocols to securely retrieve (in encrypted form) the class labels corresponding to the k-nearest neighbors of the input query q. At the end of this step, encrypted class labels of k-nearest neighbors are known only to C.

Stage 2 - Secure Computation of Majority Class (SCMCk): Following from Stage 1, P compute the class label with a majority voting among the k-nearest neighbors of q. At the end of this step, only Bob knows the class label corresponding to his input query record q.

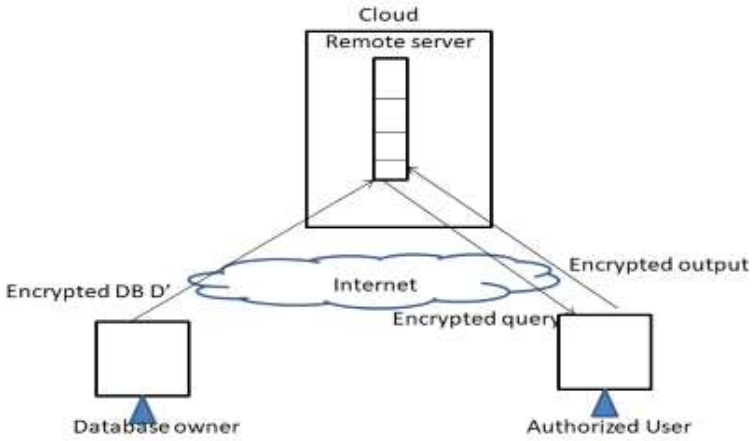

Fig. 1 Model of the proposed work

4.3 Algorithm

Algorithm: Novel PPkNN
R – Record set
R' – Encrypted record set
CS – Cloud Server
O – Owner
A – Authorized User
S_k – Secret key
Q – Query set
CL – Class Labels

BEGIN
S_k ← Compute(pwd)
For each r in R
    r' ← ROT_i (r, S_k)
    CS ← r'
NEXT
For each q in Q
q' ← ROT_i (q, S_k)
CS ← q'
Exe(q')
CL ← PPkNN ← SMIN
CL' ← Majority (CL);
Return CL'
NEXT
END

V. CONCLUSIONS

To ensure client protection, different security safeguarding order procedures have been proposed over the previous decade. The current procedures are not appropriate to outsourced database situations where the information lives in scrambled structure on an outsider server. This paper proposed a novel security saving k-NN arrangement convention over encoded information in the cloud. Our convention secures the classification of the information, client's info inquiry, and shrouds the information access designs. We likewise assessed the execution of our convention under various parameter settings. Since enhancing the proficiency of SMINn is an essential initial step for enhancing the execution of our PPkNN convention, we plan to research option and more productive answers for the SMINn issue in our future work. Moreover, we will explore and extend our examination to other characterization calculations.

REFERENCES

[1] A Mell P, Grance T (2009) A NIST definition of cloud computing. National Institute of Standards and Technology. NIST SP 800-145. http://www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf
[2] IDC (2009) Enterprise Panel, September. http://www.slideshare.net/JorFigOr/cloud-computing-2010-an-idcupdate
[3] Cloud Industry Forum (2011) Cloud UK: Adoption and Trends 2011.
[4] Cloud Security Alliance (2010) Top Threats to Cloud Computing. v1.0, March.
[5] Horrigan JB (2008) Use of cloud computing applications and services. Pew Internet & American Life project memo, Sept.
[6] Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act (USA PATRIOT ACT) (2001) Title V, s 505.
[7] ENISA (2009) Cloud Computing: Benefits, risks and recommendations for information security. Daniele Catteddu and Giles Hogben (eds), November.
[8] Marchini R (2010) Cloud Computing: A Practical Introduction to the Legal Issues. London: BSI.
[9] McKinley PK, Samimi FA, Shapiro JK, Chiping T (2006). Service Clouds: A Distributed Infrastructure for Constructing Autonomic Communication Services. In: Dependable, Autonomic and Secure Computing, IEEE, 341-348.
[10] Warren S, Brandeis L (1890) The Right to Privacy. 4 Harvard Law Review 193.
[11] Westin A (1967) Privacy and Freedom. New York, USA, Atheneum.
[12] American Institute of Certified Public Accountants (AICPA) and CICA (2009) Generally Accepted Privacy Principles. August. http://www.aicpa.org/interestareas/informationtechnology/resources/privacy/generallyacceptedprivacyprinciples/downloadabledocuments/gapp _prac_%200909.pdf
[13] Solove DJ (2006) A Taxonomy of Privacy. University of Pennyslavania Law Review, 154(3):477, January.

[14] Nissenbaum H (2004) Privacy as Contextual Integrity. Washington Law Review, 101-139. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=667622
[15] Nissenbaum H (2009) Privacy in Context: Technology, Policy and the Integrity of Social Life. Stanford University Press.
[16] Swire PP, Bermann S (2007) Information Privacy. Official Reference for the Certified Information Privacy Professional, CIPP.
[17] European Commission (1995) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. http://ec.europa.eu/justice_home/fsi/privacy/docs/wpdocs/2007/wp136_ en.pdf
[18] Jaya Krishna Sunkara, E Navaneethasagari, D Pradeep, E Naga Chaithanya, D Pavani, D V Sai Sudheer, "A New Video Compression Method using DCT/DWT and SPIHT based on Accordion Representation", I.J. Image, Graphics and Signal Processing (e-ISSN: 2074-9082, p-ISSN: 2074-9074, IF: 0.11), pp. 28-34, May 2012.
[19] Privacy Protection Study Commission (1977) Personal Privacy in an Information Society, United Statues Privacy Protection Study Commission Fair Information Practices. http://epic.org/privacy/ppsc1977report/