European Journal of Applied Sciences 8 (3): 164-169, 2016
ISSN 2079-2077

DOI: 10.5829/idosi.ejas.2016.8.3.22994

K-Nearest Neighbor Categorization on Secure Data Access in Cloud

[1]V. Balamurugan and [1]B. Muthu Kumar

[1]Final year MCA Student, Sathyabama University,
Chennai-600119, Tamilnadu, India
[2]Department of MCA, Faculty of Computing,
Sathyabama University, Chennai-600119, Tamilnadu, India

Abstract: For the last few years, a extensive research has been going on query processing of relation data and more practical and theoretical solution have been suggested to query processing under different scenarios. Now days cloud computing technology is increasing rapidly, so users now have the chance to store their data in remote location. However, different privacy issues are raised on cloud computing, important data needs to be encrypted before store the data on cloud storage. In extra, query processing methods have to be supported by cloud storage; otherwise, there is a no chance to store data on remote location of cloud storage. To perform the operation by queries on encrypted data without the decrypting by cloud is an important challenging issue. In our proposed system we take focus for resolving the k-nearest neighbor (kNN) query issues over the encrypted outsourced data on cloud storage: user issues of encrypted query information to cloud storage and return the k closest information to user by cloud. We propose k-nearest neighbor protocol that protects the input query of user, confidentiality of data and access pattern of data. Also we examine our protocol efficiency by different experiments. However, as stated above Privacy Preserve k-nearest neighbor (PPkNN) is composite issues and it cannot be achieved straightly by method of the existing k-nearest neighbor techniques on encypted data. We improve our proposed system and produce new solution for Privacy Preserve k-nearest neighbor (PPkNN) classifier issues on encrypted data.

Key words: PPkNN · k-NN Classifier · Encryption · Privacy preserving · Security · Outsourced Databases

INTRODUCTION

As an increasing paradigm of computing, cloud computing attracts more users and many organization to examine utilizing the advantages of cloud because of the flexibility, administrative offload overhead and cost-efficiency. In model of cloud computing, data owner stores his/her to database T and the functionalities of DBMS to remote location of cloud that structure to keep outsourced databases and produces querying acquire mechanisms and controls provided database. In addition, by outsourcing the data, the data owner acquires advantages of minimizing the costs for management of data and increase the quality of service. On addition, holding and data query processing out of the data owner control provides chances for security issues such as protecting data confidentiality and privacy of query. One simple way to keep protects the outsourced data from unauthorized user and maintain the confidentiality of data is achieved by encrypt process on data by data owner before the data outsourcing. By using this process, data owner can protect the privacy of their data. In extra, to protect the query privacy, authorized users needs to be encrypting all queries before transferring to cloud storage for estimation. Moreover, through the time of query processing, the cloud storage can also obtain sensitive and valuable information about real data records by perceive the access pattern of data even if input query and data are encrypted. Hence, from the above conversation, secure way query processing needs to promise confidentiality of encrypted data user's query information confidentiality hiding the access patterns of data. For achieving the data confidentiality on data by encryption may reason for another problem during the

Corresponding Author: V. Balamurugan, Final year MCA Student, Sathyabama University, Chennai-600119, Tamilnadu, India.

step of query processing on cloud storage. In common, it is critical to doing the process on encrypted data without decrypting it. Here the question arises is how cloud storage can run the queries on encrypted data when it is encrypting format for all the time. In some literature survey different techniques have been proposed related to query processing on encrypting data including aggregate and range queries. However these above technique either not inefficient or not applicable to resolve advanced queries like kNN. Given user query Q1, the purpose of the SkNN issue is to recognize the k-closest data tuples securely to Q1 by encrypted database in cloud of T1, without permitting the cloud to absorb anything concerning the real contents of database T1 and record of query Q1. More particularly, when storing the encrypted data to cloud, we detect that a powerful SkNN protocol have to fullfil the following criteria.

- Protect the Q1 confidentiality and T1 at all times.
- Keep hiding data access patterns from cloud.
- Exactly calculate the query Q of k-NN.
- Suffer overhead of low computation on end-user.

The protocols created in this system are secured under model of semi-honest. But, they can extend simply to secure protocols below other model of opponent, such as covert and malicious, using zero-knowledge proofs and threshold based cryptosystem.

Related Work: In this part, we have examined existing research papers associated to the PPkNN of encrypted data. The more evaluated previous related work as follows:

The researches C. Gentry [1] suggests completely Homomorphic cryptosystems for achieve the issues of DMED. It permits third party to run random function on encrypted data without decrypting that data. Problem of this method is, that techniques are more expensive and they are not.

Explored nearly yet. For instance, S. Halevi and C. Gentry [1] shows that it takes more time on even for inadequate security parameters on high performance machine.

A. Shamir [2] proposed a confidential sharing scheme to create a PPkNN protocol in SMC (secure multiparty computation). SMC related approach conclude data are splitted and it encrypted at every joined party, in-between computation are achieved on normal data not in encrypted data. In our suggested work has some different from this secret sharing solution. In the secret sharing related methods requires at least three parties in system whereas two parties are used in our system. So, in proposed there is minimum chances of duplicity of data security.

R. Srikant and R. Agarwal [3], B. Pinkas and Y. Lindell [4] initiate the idea of privacy preserving below application of data mining. The conventional techniques of PPDM split into two classification: (i) distribution of data and (ii) perturbation of data. Srikant and Agarwal [3] suggested the data perturbation method first to construct classifier of decision-tree and then more methods were suggested. However, data perturbation techniques cannot be relevant for semantically encrypted data. Because of the some extra statistical noises to data, the techniques of data perturbation do not provides exact mining results.

On the additional, Pinkas and Lindell [4] proposed first classifier of decision tree underneath the two party setting. Normally users do not have true keywords to define the queries concluding the data were dispensed between them. Since more works have been published using techniques of SMC. It is calculated that the issue of PPkNN cannot be achieved using the techniques of data distribution as data in our system is encrypted and not dispensed in plaintext between several parties.

Proposed Word

Overview: In existing system it cannot solve problem of Data mining on Encrypted data (DMED). Normally perturbed data do not have semantic security; one technique is obtainable like perturbation technique. But the perturbation technique cannot support the more sensitive data. Also this method does not produce exact data mining result. In this proposed system we create a secure kNN query protocol over on the encrypted data in semi-honest model. It helps to protects user query privacy, data confidentiality and hides pattern of data access. The Privcacy Perceiving k-Nearest Neighbor is more critical issue and it cannot be answered straightly using previous technique of k-NN on encrypted data. We increase our existing work and give new infusion to problem of Privacy Perceiving k-Nearest Neighbor classifier on encrypted data. For this we propose protocol of novel PPkNN over encrypted data. Here our proposed uses group of common protocols for building the suggested k-NN classifier.

Query Processing: There are more techniques have been proposed for query processing on encrypted data. In our proposed system we propose novel secure kNN query protocol on encrypted data. The suggested protocol

secures the data confidentiality, privacy of user query and hides access pattern of data. As noted above, PPkNN is critical issue and solved straightly by existing techniques. Therefore, our system helps to solve the problem of PPkNN over on encypted data [5-10].

KNN Query Process: The privacy preserving kNN classification uses the semantically secure scheme of Homomorphic encryption. The technique of kNN classification can arrange more amounts of to secured fashion. Normally kNN query not perform the process over encrypted data. For that we develop a kNN query procedure related on range queries. As result, index use in processing of range query also permit fast PPkNN queries processing.

The algorithm of PPkNN contains two rounds interaction operating between the server and client. The client will forward the range of start upper-bound, which holds more than k points and range of start lower bound, which holds less than k points, to server. Inner range is identified by server and send to client. Then outer range will be computed depend on inside range and send return to server. The server discovers information in the range of outer and sends return to client. In final, client decrypts the information and discover the data as final result which one having top range.

For Example: Suppose Alice maintains database D with n records r1….rn and n + 1 attributes. Let rl, k indicate kth record attribute value rl. Initially, Alice start encryption the database by attribute wise, that means she calculate Epk(rl, k), for 1= l = n and 1= k =n+1, where column (n+1) holds the labels of class. We presume that scheme of underlying encryption is secure semantically. Let encrypted database be indicated by D'. We presume that outsources D' of Alice as well as the process of future classification to cloud. Let Bob be authorized user, he wants to classify his input information q=hq1…..qsi by using the k-NN classification process based on D'. We mention to such process as PPkNN classification on encrypted data in cloud. Formally, we describe the PPkNN protocol as: PPkNN (D', q) → cq, where cq indicate class label for h after using k-NN classification process on D' and h.

In this protocol, once data is encrypted and stored to cloud, Alice does not engage in any computation.

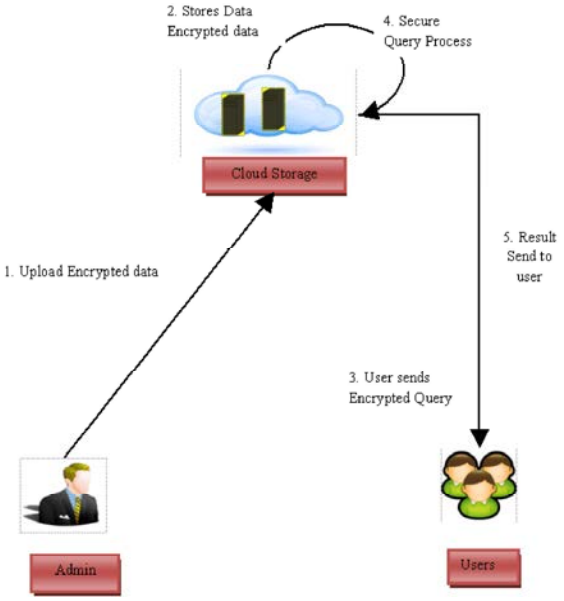
By this, no data information is disclose with Alice. In specially, our proposed protocol encounters the following computation.

- Records of D of any results have not been disclosed to cloud.
- Bob's query has not be disclosed to cloud.
- cl should be disclosed only with Bob. In extra no information should be disclosed other than cq.
- Access pattern of data, such as information similar to k-NN of h, should not be disclosed with cloud and Bob.

K-NN Classification Algorithm:

- Construct the normal training data set D;
- for every process Y in the test data do
- If Y has an undisclosed system call then
- Y is abnormal;
- else then
- for every process Dj in training data do
- compute sim (Y, Dj);
- if sim(X, Dj) equals 1.0 then
- Y is normal; exit;
- Identify k higher scores of sim(Y, D);
- Compute sim_avg for k-nearest neighbore;
- If sim_avg is grater than threshold then
- Y is normal;
- else then
- Y is abnormal;

Architecture



RESULT AND DISCUSSION

In this system we suggested novel privacy-preserving k -NN categorization technique on encrypted

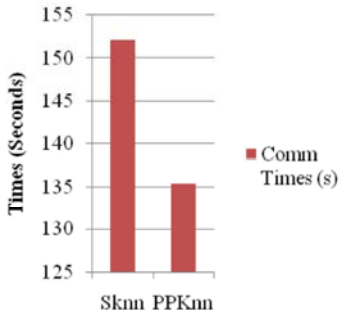


Fig. 2: Communication comparison of SKnn and PPKnn

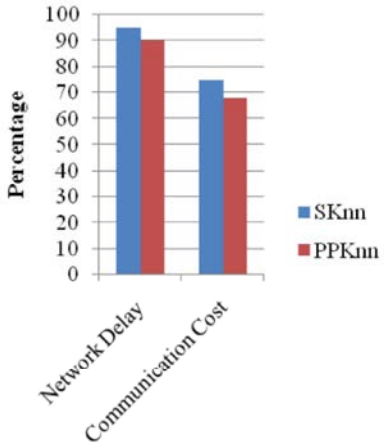


Fig. 3: Comparisons of SKnn and PPKnn

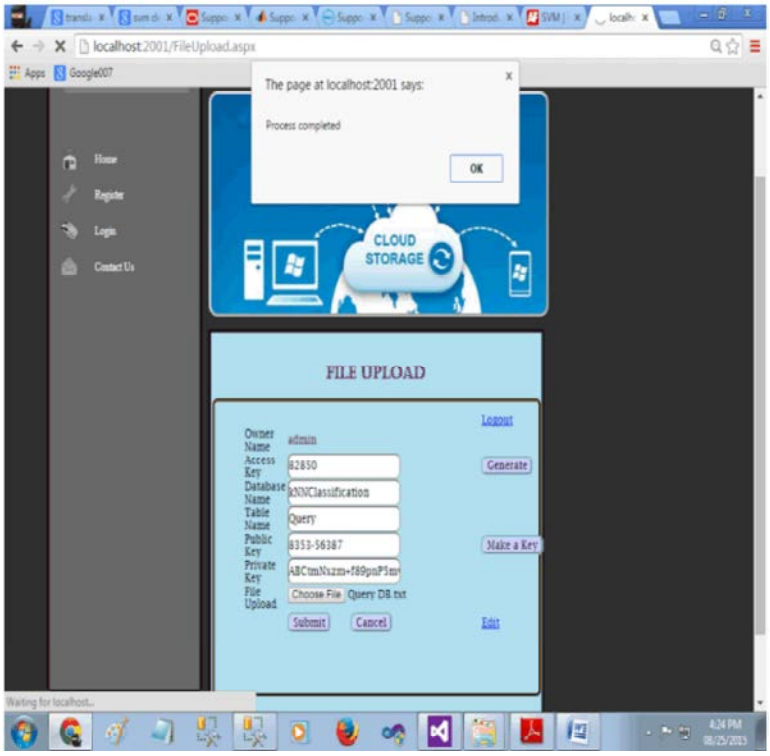


Fig. 4: File uploaded to Cloud Storage by Admin with Encrypted Format

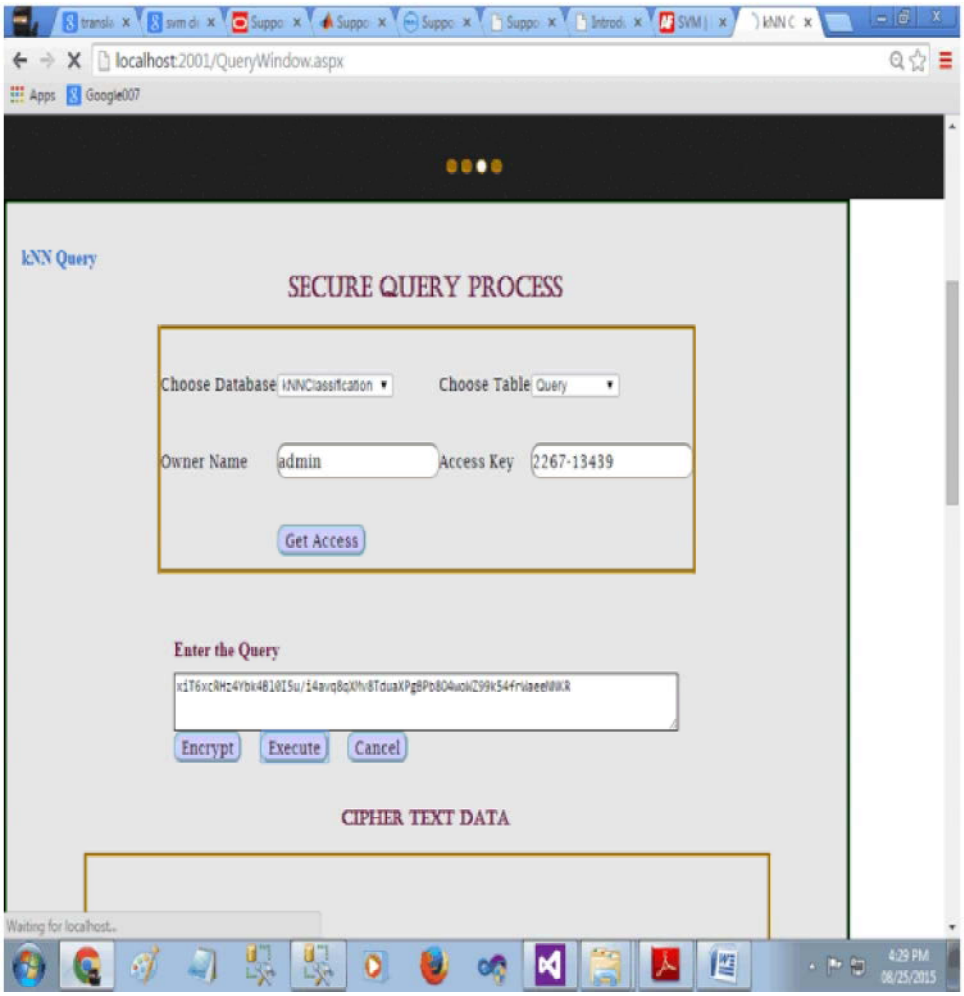


Fig. 5: Input the query in Encrypted Format

data from cloud. This technique protects confidentiality data, input query of user's and hides data access pat terns. Estimate presentation of methods under various parameter settings. Here we discussed some experiments representative performance of Privacy Preserving k-Nearest Neighbor (PPKNN) classification method with some parameter settings [11, 12].

The above figure explains the communication times of existing and proposed system protocols. Our proposed protocol PPKnn protocol utilize minimum time for identifying the records in the storage space based on the user input query.

The above figure shows the percentage results in cost and network delay with comparison of existing and proposed systems. If no of records increases the SKnn protocol takes more time to respond for results. But our proposed system takes minimum times if it records size is constantly increasing.

The above figure explains the file uploading process in admin side. For this admin have to generate access key for sending to user and secret key for encrypt the selected file. Once the file is encrypted then it will upload on cloud storage server.

The figure 5 shows the user process screen. In this process user have to choose the data base name and provide the access key that has been generated on the admin side for enter into to cloud server. Then user provides input query in encrypted format for do not reveal the query information to others even cloud server also. After the query processing on cloud server final result will be send to user in encrypted format and text format also.

CONCLUSION

This proposed system analysis the different methods of PPkNN (privacy preserving k-nearest neighbor) on

encrypted data. To protect user privacy records, different techniques have been proposed related the privacy preserving classification over encrypted data. The obtainable methods are not relevant to outsourced environment of database wherever data stays encrypted form third party server. Our intended system introduces new PPkNN sorting protocol on encrypted data from cloud. The protocol achieves the data privacy, input user query and data secretes access pattern.

REFERENCES

1. Gentry, C. and S. Halevi, 2011. Implementing gentry's fully homomorphic encryption scheme, in EUROCRYPT, Springer, pp: 129-148.
2. Shamir, A., 1979. How to share a secret, Commun. ACM, 22: 612-613.
3. Agrawal, R. and R. Srikant, 2000. Privacy-preserving data mining, in ACM Sigmod Record, 29: 439-450.
4. Lindell, Y. and B. Pinkas, 2000.Privacy preserving data mining, in Advances in Cryptology (CRYPTO), pp: 36-54.
5. Mell, P. and T. Grance, 2011. The nist definition of cloud computing (draft), NIST Special Publication, 800: 145.
6. Hu, H., J. Xu, C. Ren and B. Choi, 2011. Processing private queries over untrusted data cloud through privacy homomorphism, in ICDE. IEEE, pp: 601-612.
7. Gentry, C., 2009. Fully homomorphic encryption using ideal lattice, in ACM STOC, pp: 169-178.
8. Williams, P., R. Sion and B. Carbunar, 2008. Building castles out of mud: practical access pattern privacy and correctness on untrusted storage, in CCS. ACM, pp: 139-148.
9. Hore, B., S. Mehrotra and G. Tsudik, 2004. A privacy-preserving index for range queries, in VLDB, pp: 720-731.
10. Hacıgumus¸, H., B. Iyer and S. Mehrotra, 2004. Efficient execution of aggregation queries over encrypted relational databases, in Database Systems for Advanced Applications. Springer, pp: 125-136.
11. Islam, M.S., M. Kuzu and M. Kantarcioglu, 2012. Access pattern disclosure on searchable encryption: Ramification, attack and mitigation, in NDSS.
12. Li, M., S. Yu, W. Lou and Y.T. Hou, 2012. Toward privacy-assured cloud data services with flexible search functionalities, in ICDCSW. IEEE, pp: 466-470.