INTERNATIONAL JOURNAL OF RESEARCH
In Advanced Engineering Technologies

Volume 6, Issue 2 JULY 2017

Implementing an Efficient Data Classification Protocol for Protecting Data Access Patterns

[1]K. PRATHYUSHA,[2] J. A PAULSON

[1]M. Tech Student, Department of CSE, Nalanda Institute of Engineering and Technology, Village Kantepudi, Mandal Sattenapalli, District Guntur, Andhra Pradesh, India

[2]Associate Professor, Department of CSE, Nalanda Institute of Engineering and Technology, Village Kantepudi, Mandal Sattenapalli, District Guntur, Andhra Pradesh, India

ABSTRACT– ***Data mining emphases on producing beneficial information from the data sources rather than a simple data mining technology. Among several forms of data mining tasks, data classification is a key task for a user who wants to group a record at hand based on the database existing in the cloud. With the recent improvement of cloud computing users accepting the opportunity to outsource their information as well as the information management tasks to the cloud. The rise of different privacy problems, sensitive information (e.g., medical records) would like to be encrypted before outsourcing to the cloud. In addition, query process tasks should be handled by the cloud otherwise; there would be no purpose to source the information at the primary place. If user submits an encrypted query file to the cloud, and the cloud returns the k closest files to the user, such a simple solution is not secure. To produce higher security, a secure kNN protocol that assures the understanding of the information, user's input query, and information access patterns. These results specify that secure protocol is able on the user end, and this lightweight system allows user to benefit any mobile machine to perform the kNN query.***

Keywords: kNN Classification, Data Mining

1. INTRODUCTION

As of late, the cloud computing view is altering the associations' method for working there in format ion especially in the way they store, get to and process information. As a developing computing world view, cloud computing pulls in numerous associations to consider truly with respect to cloud potential regarding its expense productivity, adaptability, and offload of authoritative overhead. Regularly, associations appoint their computational operations notwithstanding their information to the cloud. In spite of colossal points of interest that the cloud offers, protection and security issues in the cloud are forestalling organizations to use those preferences. At the point when information is profoundly touchy, the information should be encoded before outsourcing to the cloud. Be that as it may, when in format ion are scrambled, independent of the basic encryption plan, performing any information mining errands turns out to be exceptionally testing while never decoding the

information. There are other protection concerns, showed by the accompanying illustration. As developing processing world view, distributed computing draws in numerous associations to consider genuinely with respect to database potential as far as its cost efficiency adaptability and offload of regulatory overhead. Regularly, associations designate their computational operations despite there in format ion to the cloud. Notwithstanding enormous preferences that the cloud offers, protection and security issues in the database are counteracting organizations to use those points of interest. At the point when information is exceedingly delicate, the information should be encoded before outsourcing to the database. Nonetheless, when information is scrambled, independent of the fundamental encryption plan, performing any information mining undertakings turns out to be extremely testing while never decoding the information. Assume an insurance agency outsourced its scrambled clients database and applicable in formation mining undertakings to a database. At the point when an operators from the organization needs to focus the danger level of a potential new client, the specialists can utilize an order strategy to focus the danger level of the client. Initially, the operators needs to create an information record q for the client containing certain individual data of the client, e .g., FICO assessment, age, conjugal status, and so on. At that point this record can be sent to the database, and the database will figure the class mark for q. All things considered, since q contains delicate data, to secure the client's protection, q ought to be encoded before sending it to the database. The above case demonstrates that information mining over encrypted information (indicated by DMED) on a database additionally needs to ensure a client's record when the record is a piece of an information mining procedure. In addition, database can likewise determine helpful and sensitive data about the genuine information things by watching the information access examples regardless of the fact that the information is encoded.

2. RELATED WORK

Alka Gangrade et.Al. Provide a unique answer for NaïveBayes classification over vertically partitioned records. Instead of using facts transformation, we outline a secure multiplication protocol to transform the version parameters even as retaining information values comfy. Our class system is quite efficient and speedy due to the fact the walking time of our classifier is much less than presents NaïveBayes classifier. It is also a lot much less than ID3 andC4.5 selection tree classifier because Bayesian classifier simplest needs to undergo the complete training records once. They are also space green due to the fact they increase a frequency desk in length of the product of the wide variety of attributes, number of sophistication values, and the range of values in step with characteristic no longer the actual fee of the characteristic.

Monika D.Rokade, Mr.S.A.Kahate look at is performed on processing troubles of private queries on listed facts in a cloud. A comfy traversal framework in listed environment is given to relaxed protocols for such traditional queries. The assumptions and approached stated on this paper are very well beneficial, efficient to perform and efficaciously used below settings of different parameters. It has been summarized that the process cited right here, on privacy homomorphism, is used to guard processing queries on cloud is high scalable.

Li Xiong, Subramanyam Chitti, Ling Liu supplied a trendy version for kNN class throughout more than one personal databases and multi-spherical algorithms for knowing the version. Their analysis and experiments showed the feasibility of the approach and its capability to acquire a balance between 3 crucial overall performance metrics: relative accuracy, performance, and privacy.

Mark Shaneck, Yongdae Kim and Vipin Kumar proven a protocol for privately computing the knearest neighbors of factors in a horizon-tally partitioned statistics set. We defined this set of rules inside the two birthday celebration case and proved security for each of the elements ofthe set of rules. In addition, we confirmed how this algorithm may be used to compute LOF outlier scores. For future work, we purpose to enhance the set of rules to no longer monitor the intermediate community records, accordingly decreasing the potential statistics leakage. Also, as this work is centered on horizontally partitioned information, another area of destiny paintings would be extending it to vertically partitioned statistics.

3. FRAMEWORK

A. Overview of Proposed Protocol

The proposed system focus on solving the classification problem over encrypted data. In the proposed system, anew privacy preservation protocol based on KNN classification method is introduced to protect the confidentiality of data, privacy of user's input query and to hide the data access pattern. Diagram Using Homomorphic encryption allowscomplex mathematical operations to be performed on encrypted data without using the original data and provides the data security in cloud.

The proposed algorithm to preserve intermediate k nearest neighbor in the classification process should not reveal to cloud server or any other user. The proposed algorithm develops a solution for privacy–preserving k-nearest neighbor classification which is one of the commonly used data mining tasks. It determines which the closest results are by identifying the class of minimum distance using K nearest neighbors. The new privacy preservation protocol implementation for the input query record classification over the encrypted database in the cloud is carried by the steps:

1. Secure Data Upload
2. Query Processing
3. Secure KNN query process

The proposed PPKNN protocol mainly consists of two stages:

Stage 1: Secure retrieval of K-nearest Neighbors(SRKNN). In this stage, the authorized user sends a query (in encrypted form) to cloud. After this cloud involve in a set of sub-protocols to securely retrieve (in encrypted form) the class labels corresponding to the K-nearest neighbors are known only to cloud server.

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

B. PPkNN Algorithm

Inputs: encrypted dataset, Cloud1 (C1), Cloud2 (C2), User query

1. User send the query to C1 and that query passed to C1 in the form of encryption
2. C1 compute query to generate vectors by using SMINn protocol

   //here, in SMINn protocol we are generating random functions and encrypt those
3. After that C1 sends query to C2
4. C2 compute the nearest data to received query and again forward the resulted query results to C1
5. C1 send that query to in the form of bits that is based on Secure Bit-OR (SBOR) protocol.

   //SBOR converts the output as bits.
6. Finally data user gets the result from c1 as class labels

C. Secure Minimum Out of n Numbers (SMINn)

The main goal of the SMINn protocol is to compute [min $(d_1,\ldots,d_n)$]=[d_{min}] without revealing any information about d_i's to C1 and C2

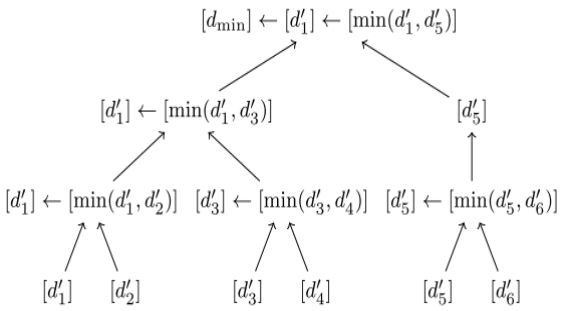
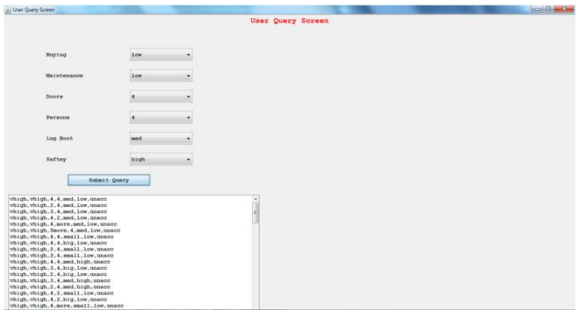


Fig1. Binary Execution Tree based on SMINn Protocol

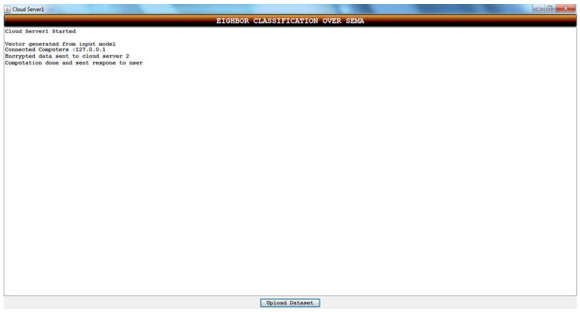
Suppose C1 holds {[d_1],….,[d_6]} (i.e., n = 6). For simplicity, here we are assuming that there are no secrets associated with d_i's. Then, based on the SMINn protocol, the binary execution tree (in a bottom-up approach) to compute [min $(d_1,\ldots,d_6)$ is shown in Fig. 1. Note that, initially [d_0]=[d_i].

4. EXPERIMENTAL RESULTS

Run the two cloud servcers named as cloud server1 and cloud server2. After, we need to upload the dataset from cloud server1 and thje keys will be generated by cloud server2.



Here, user can pass the query through the user interface.



After key computation the two servers send the query results to the user.

5. CONCLUSION

This paper proposed a new privacy preserving k-NN classification protocol over encrypted data. This protocol protects the confidentiality of the information, user's input query, and hides the data access patterns.

REFERENCES

[1] P. Williams, R. Sion, and B. Carbunar, "Building castles out ofmud: Practical access pattern privacy and correctness onuntrusted storage," in Proc. 15th ACM Conf. Comput. Commun.Security, 2008, pp. 139–148.

[2] P. Paillier, "Public key cryptosystems based on composite degree residuosity classes," in Proc. 17th Int. Conf. Theory Appl. Cryptographic Techn., 1999, pp. 223–238.

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

[4] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. 41st Annu. ACM Sympos. Theory Comput., 2009, pp. 169–178.

[5] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in Proc. 30th Annu. Int. Conf. TheoryAppl. Cryptographic Techn.: Adv. Cryptol., 2011, pp. 129–148.

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

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

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

[9] P. Zhang, Y. Tong, S. Tang, and D. Yang, "Privacy preservingNaive Bayes classification," in Proc. 1st Int. Conf. Adv. Data MiningAppl., 2005, pp. 744–752.