International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 20 Issue 3 – MARCH 2016.

EFFECTIVE DATA SEARCH FOR ENCRYPTED RELATIONAL DATA IN CLOUD USING K-NEAREST NEIGHBOR ALGORITHM

Abirami Swetha.L[#1],Gokila.R[*2],Mr.Vijaya Ragavan.P[$3]

[#]*B.E, Computer Science and EngineeringDepartment, Dhanalakshmi College OfEngineering, Chennai, India.*
[*]*B.E, Computer Science and EngineeringDepartment, Dhanalakshmi College OfEngineering, Chennai, India.*
[$]*B.E., M.E., (Ph.D.), Associate Professor, Computer Science and EngineeringDepartment, Dhanalakshmi College Of Engineering, Chennai, India.*

***Abstract*—In Data mining, classification is one of the commonly used tasks. Due to the rise of privacy issues many theoretical and practical solutions to the classification problems have been proposed under different security models. In cloud the data of the stored information is made secure by encrypting the data. Since the data is in encrypted form, the existing privacy preserving classification techniques are not applicable. The proposed protocol protects the confidentiality of data, privacy of user's input query, and hides the data access patterns. The proposed protocol is to develop a privacy preserving k-NN classifier over encrypted data under the semi –honest model. Also, we empirically analyze the efficiency of the proposed protocol using a real world data set under different parameter settings. 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, this protocol concentrates on executing the k-nearest neighbor classification method over encrypted data in a cloud.**

Index Terms -PPkNN classifier, outsourced database, encryption.

I. INTRODUTION

Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet. Users need not have knowledge of, expertise in, or control over the technology infrastructure in the "cloud" that supports them. 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. WhenData 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.Data Mining is defined as extracting information from huge sets of data. In other words, we can say that data mining is the procedure of mining knowledge from data. There is a huge amount of data available in the Information Industry. This data is of no use until it is converted into useful information. It is necessary to analyze this huge amount of data and extract useful information from it.

Cloud needs to protect a user's record when the record is a part of a data mining process. Moreover, cloud can also derive useful and sensitive information about the actual data items by observing the data access patterns even if the data are encrypted [1], [2]. Therefore, the security requirements of the DMED problem on a cloud are threefold: (1) confidentiality of the encrypted data, (2) confidentiality of a user's query record, and (3) hiding data access patterns.

1.1. PROBLEM DEFINITION

Let us consider a database D of n records t_1, t_n and m + 1 attributes which is owned by person A. Let $t_{i,j}$ denote the jth attribute value of record t_i. Initially, person A encrypts his database attribute wise, that is, he computes E_{pk} ($t_{i,j}$), for 1 <= i <= n and 1 <=j <= m + 1, where column (m + 1) contains the class labels. Let person B be an authorized user who wants to classify his input record q= (q_1; . . . ; q_m) by applying the k-NN classification method based on D_0. We refer to such a process as privacy-preserving k-NN (PPkNN) classification over encrypted data in cloud.

PPkNN (D^1, q) ->c_q,

Wherec_q denotes the class label for q after applying kNN.

1.2. OUR CONTRIBUTIONS

In present era the data access patterns, such as the records corresponding to the k-nearest neighbors of q, should not be revealed to person B and the cloud .So we propose aPPkNN protocol, a secure k-NN classifier over semantically secure encrypted data. Thus, which data records correspond to the k-nearest neighbors in cloud storage and the output class label is not known the cloud. In addition, after sending his encrypted query record to the cloud, person Bdoesnot involve in any computations. Hence data access patterns are further protected from Bob. Here we discuss the general problem of secure computation on an encrypted database and propose a SCONEDB (Secure Computation ON an Encrypted Database) model, which captures the execution and security requirements.

II. RELATED WORK

In recent years privacy preserving techniques has arisen in various fields in cloud. Hence the execution of PPkNN is a more complex problem than the execution of

simplekNN queries over encrypted data. Hence we propose a novel secure k-nearest neighbor query protocol over encrypted data that protects data confidentiality, user's query privacy, and hides data access patterns.

Secure nearest neighbor revisited [3] investigates the secure nearest neighbor (SNN) problem, in which a client issues an encrypted query point E(q) to a cloud service provider and asks for an encrypted data point in E(D) (the encrypted database) that is closest to the query point, without allowing the server to learn the plaintexts of the data or the query (and its result).

Sharemind: a framework for fast privacy-preserving [4] computes a provably secure and efficient general-purpose computation system to address the problem of privacy preserving .The solution SHAREMIND is a virtual machine for privacy-preserving data processing that relies on share computing techniques. This is a standard way for securely evaluating functions in a multi-party computation environment. The novelty of the solution is in the choice of the secret sharing scheme and the design of the protocol suite.

Processing Private queries over untrusted data cloud through privacy homomorphism [5], the query processing preserves both the data privacy of the owner and the query privacy of the client is a new research problemdatasets by leveraging an index-based approach. Based on this framework, a secure protocol for processing typical queries such as k-nearest-neighbor queries (kNN) on R-tree index.

Secure multidimensional range queries over outsourced data[6],the problem of supporting multidimensional range queries on encrypted data. The problem is motivated by secure data outsourcing applications where a client may store his/her data on a remote server in encrypted. The solution approach is to compute a secure indexing tag of the data by applying bucketization (a generic form of data partitioning) which prevents the server from learning exact values but still allows it to check if a record satisfies the query predicate.

Secure knn computation on encrypted databases [7], service providers like Google and Amazon are moving into the SaaS (Software as a Service) business. They turn their huge infrastructure into a cloud-computing environment and aggressively recruit businesses to run applications on their platforms. The problem of k-nearest neighbor (kNN) computation on an encrypted database. So a new asymmetric scalar-product-preserving encryption (ASPE) that preserves a special type of scalar product is developed.

III. SYSTEM DESIGN AND ARCHITECTURE

This study describes the secure Knn classification of encrypted data in cloud. The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. Efficient and intelligent output design improves the system's relationship to help user decision-making.

3.1. PRIVACY-PRESERVING DATA MINING

The privacy preserving data mining proposes a number of techniques to perform the data mining tasks in a privacy-preserving way. These techniques generally fall into the following categories: data modification techniques, cryptographic methods and protocols for data sharing, statistical techniques for disclosure and inference control, query auditing methods, randomization and perturbation-based techniques. The existing PPDM techniquescan broadly be classified into two categories: (i)data perturbation and (ii) data distribution. The first data perturbation technique was proposed to build a decision-tree classifier, and many othermethods were proposed later (e.g., [8], [9], [10]).

3.2.QUERY PROCESSING OVER ENCRYPTED DATA

PPkNN is a more complex problem than the execution of simple kNN queries over encrypted data. There are various techniques related to query processing overencrypted data have been proposed, e.g., [6], [11], [12].For instance, the intermediate k-nearest neighbors for the classification process should not be disclosed to the cloud or any users. The recent method [3]reveals the k-nearest neighbors to the user. A novel secure k-nearest neighbor query protocol over encrypted data [3][7]that protects data confidentiality, user's query privacy, and hides data access patterns .However, PPkNN is a more complex problem and it cannot be solved directly using the existing secure k-nearest neighbor techniques over encrypted data.

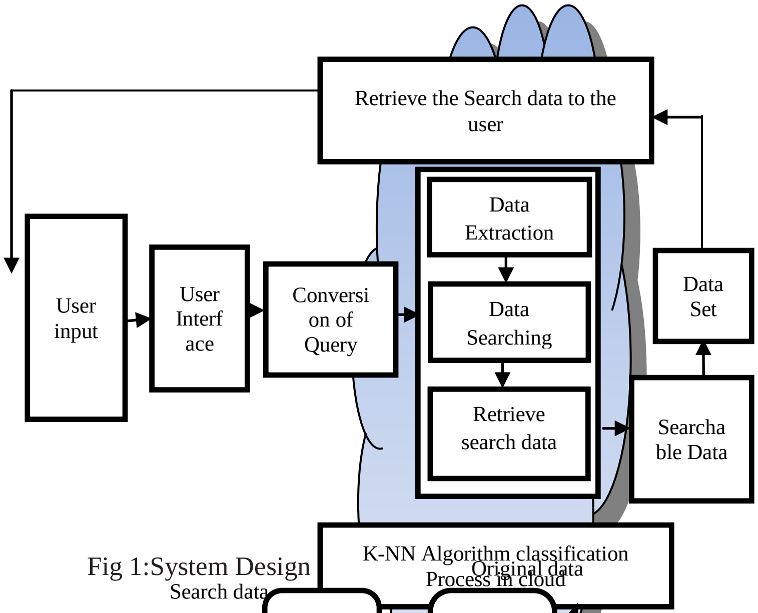

Fig 1:System Design

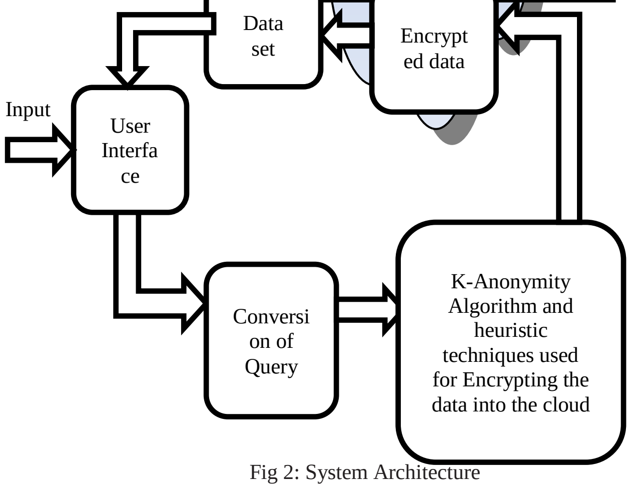
Search data
Original data
Data set
Encrypt ed data
Input
User Interfa ce
Conversi on of Query
K-Anonymity Algorithm and heuristic techniques used for Encrypting the data into the cloud

Fig 2: System Architecture

IV. PRIVACY-PRESERVING PRIMITIVES

The protocols specified below are considered under two-party semi-honest setting.

Secure squared Euclidean distance (SSED): In this protocol, P1 with input (E_{pk}(X), E_{pk}(Y))and P_2 with sk securely compute the encryption of squared Euclidean distance between vectors X and Y. Here X And Y are m dimensional vectors where$E_{pk}(X)=(E_{pk}(x_1),\ldots,E_{pk}(x_m))andE_{pk}(Y)=(E_{pk}(y_1),\ldots,E_{pk}(y_m))$.The output $E_{pk}(|X-Y|^2)$ will be known only to P_1.

Secure bit-decomposition (SBD): Here P1 with input E_{pk} (z) and P_2 securely compute the encryptions of the individual bits of z, where 0 <=z < 2^l. The output

[z]= (E_{pk} (z_1),….,E_{pk} (Zl) is known only to P1. Herez_1 and z_l are the most and least significant bits of integer z, respectively.

4.1. K-NN CLASSIFICATION ALGORITHM

K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms. The neighbors are taken from a set of objects for which the class (for *k*-NN classification) or the object property value (for *k*-NN regression) is known.

STEP 1: BEGIN
STEP 2: Input: D = {(x_1, c_1), . . . , (x_N , c_N)}
STEP 3: x = (x_1. . . x_n) new instance to be classified
STEP 4: FOR each labelled instance (x_i, c_i) calculate d (xi, x)
STEP 5: Order d (x_i , x) from lowest to highest, (i = 1. . . N)
STEP 6: Select the K nearest instances to x: D^k_x
STEP 7: Assign to x the most frequent class in D^k_x
STEP 8: END.

V. FEATURES

The k-nearest neighbor is the simplest form of all machine learning algorithms. It is based on the principle that the samples are similar, generally lies in closed vicinity. K-nearest neighbor is instance based learning method. The computation time is less for KNN algorithm when compared to Naive Bayes algorithm and neural networks.

KNN has high accuracy results when compared to Naive Bayes algorithm.

VI. EXISTING SYSTEM

Existing work on Privacy-Preserving Data Mining (either perturbation or secure multi-party computation 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 (SMC) based approach assumes data are distributed and not encrypted at each participating party.

Drawbacks of existing system:

The rise of various privacy issues. Privacy-preserving classification techniques are not applicable for Encrypted form in Cloud.

VII. PROPOSED SYSTEM

In our proposed system we focus on solving the classification problem over encrypted data. In particular, we propose a secure k-NN classifier over encrypted data in the Cloud. We focus on solving the classification problem over encrypted data. In particular, we propose a secure k-NN classifier over encrypted data in the Cloud. To the best of our knowledge, our work is the first to develop a secure k-NN classifier over encrypted data.

Benefits of proposed system:

It solves the classification problem over encrypted data in relational database. It protects the confidentiality of data, privacy of user’s input query, and hides the data access patterns. Data are highly sensitive.

VIII. CONCLUSIONS AND FUTURE WORK

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. This paper proposed a novelPrivacy-preserving k-NN classification protocol over encrypted data in the cloud. Our protocol protects the confidentiality

Of the data, user’s input query, and hides the data access patterns. We plan to investigate alternative and more efficient solutions to the SMIN problem in our future work. Also, we will investigate and extend our research to other classification algorithms.

ACKNOWLEDGEMENTS

I would like to thank my Project Guide, Mr.Vijayaragavan P, for his guidance, and constructive suggestions that helped me in coming so far in the project. I also thank our Head of Department, Dr.Sivasubramanian S, for his support and valuable suggestions.

REFERENCE

[1] S. De Capitani di Vimercati, S. Foresti, and P. Samarati,“Managing and accessing data in the cloud: Privacy risks and approaches,” in Proc. 7th Int. Conf. Risk Security Internet Syst., 2012,
pp. 1–9.

[2] P. Williams, R. Sion, and B. Carbunar, “Building castles out of mud: Practical access pattern privacy and correctness on untrusted storage,” in Proc. 15th ACM Conf. Comput. Commun. Security, 2008, pp. 139–148.

[3] X. Xiao, F. Li, and B. Yao, “Secure nearest neighbor revisited,” in Proc. IEEE Int. Conf. D.ata Eng., 2013, pp. 733–744.

[4].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.

[5]. H. Hu, J. Xu, C. Ren, and B. Choi, “Processing private queries over untrusted data cloud through privacy homomorphism,” in Proc. IEEE 27th Int. Conf. Data Eng., 2011, pp. 601–612.

[6] 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.

[7] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, “Secure kNN computation on encrypted databases,” inProc. ACM SIGMOD Int. Conf. Manage. Data, 2009, pp. 139–152.

[8] 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.

[9] 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.

[10] R. J. Bayardo and R. Agrawal, “Data privacy through optimal k anonymization,”
in Proc. IEEE 21st Int. Conf. Data Eng., 2005,pp. 217–228.

[11] 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.

[12] H. Hacigeumeu¸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.