Solving Classification Issues Over Encrypted Data

Shubham Bhaskar*, Arindam Das*, Swarnika Shubham*, Mrs. Sridevi G.M**
*Department of ISE, SJB Institute of Technology, Kengeri, Bengaluru- 560060
**Asst Professor, Department of ISE, SJB Institute of Technology, Kengeri, Bengaluru- 560060

***Abstract*—*Data Mining and Cloud computing are the two very technologies that makes the classification and storage of the data so simple. The classification tasks of data mining is very useful, however it leads to certain privacy issues. Hence, several solutions have been suggested over the years. With the invention of cloud computing users can now outsource their data over to the cloud along with several data mining tasks that can be performed there. Even if the cloud provides so splendid features, it however gives rise to certain security issues which makes data storage difficult. So, the data is uploaded over it in 'encrypted form' to ensure encryption of data. The classification technique, however, does not apply over the 'encrypted data'. Hence, we aim to solve the classification (k-NN Classifier) problem over the encrypted data.*

Keywords*—*Security, k-NN Classifier, Encryption

I. INTRODUCTON

Cloud computing means storing and accessing data and programs over the Internet instead of your computer's hard drive. As per National Institute of Standards and Technology's (NIST) [1] :-

"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

It comprises of of five characteristics, three service models, and four deployment models. Cloud computing is a computing paradigm, where in a large pool of systems are connected in private or public networks, to provide infrastructure for application, data and file storage. With its arrival, the cost of computation, application hosting, content storage and delivery is reduced drastically.

Cloud computing refers to the delivery of computing resources over the Internet. Instead of keeping data on your own hard drive or updating applications for your needs, you use a service over the Internet, at another location, to store your information or use its applications. Doing so may give rise to certain privacy issues.

Example: When you store your photos online instead of on your home computer, or use webmail.

With the advent of cloud computing, a lot of data has been uploaded over the cloud. Also, the data mining tasks are being assigned to the cloud to classify the data into the "labels" .Now since the cloud is insecure we cannot just upload the data over the cloud.

The data needs to be encrypted and then uploaded over the cloud. Once, the data are encrypted and uploaded, we can perform various data-mining operations on it and hence classify the data with the labels. This is only called as Data mining over encrypted data (DMED) where in the privacy of user record must be protected if it is a part of the data-mining process.

However, when data are encrypted, irrespective of the underlying encryption scheme, performing any data mining tasks becomes very challenging without ever decrypting the data

Fig: Cloud Computing

II. BACKGROUND

At first C. Gentry proposed [2]: ,"Fully homomorphic encryption using ideal lattices", that executes arbitrary

functions over encrypted data without ever decrypting them, can solve the DMED problem but they are quite expensive as well as quite Impractical.

Afterwards, A. Shamir's [3], "How to share a secret,". ACM, vol. 22, proposed that we can use Secret sharing scheme.

However, solutions based on the this scheme requires at least three parties but our work require only two parties. So since these works seem to be quite reluctant to our terms and conditions, hence we propose our K-NN classifier over the encrypted data.

III. LAYERS OF CLOUD COMPUTING

Cloud Providers offer services that can be grouped into three categories.

1. *Software as a Service (SaaS):*
Here, a complete application is offered to the customer, as a service when demanded. One instance of the service runs over the cloud and many end users are serviced. Today it is offered by Google, Microsoft etc.

2. *Platform as a Service (Paas):* In PaaS, an operating system, hardware, and network are provided, and the customer installs or develops its own software and applications, Google's App Engine is it's example.

3. *Infrastructure as a Service (Iaas):* IaaS model provides just the hardware and network; the customer installs or develops its own operating systems, software and applications. Some common examples are Amazon, GoGrid, 3 Tera, etc.

IV. EXISTING SYSTEM

The PPDM technique (that includes perturbation technique or secure multi-party computation based approach) cannot solve the DMED problem. Perturbed data don't possess semantic security at all hence, it cannot be used to encrypt the highly sensitive data.

Also the perturbed data do not give precise data mining results. Secure multi-party computation (SMC) approach assumes data are distributed and not encrypted. We solve the classification problem over encrypted data. In other words, we propose a secure k-NN classifier over encrypted data over the cloud.

We make sure to solve the DMED problem by outsourcing encrypted data to cloud. We implement the k- nearest neighbor classification technique over the encrypted data in the cloud computing domain.

V. SYSTEM ARCHITECTURE

`` The various steps of system architecture are:

1. ENCRYPTION

The data cannot be just send over the cloud because of certain privacy issues. Hence, they are 'Encrypted' and then sent over the cloud.

2. DATABASE

Once the data from the database is encrypted it is sent over into the database where in all the data (of the similar kind) is collected and subsequently stored for the further use.

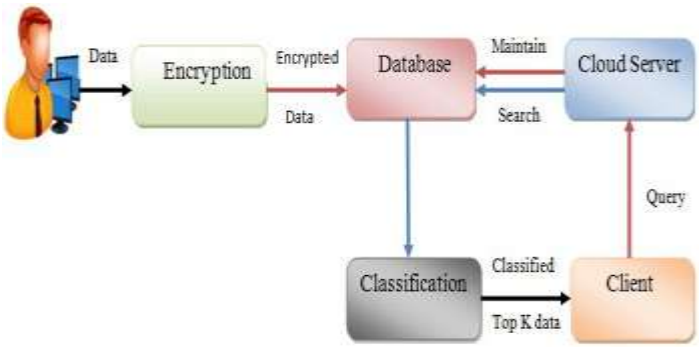

Fig: System Architecture

3. CLIENT

The client sends the query over to the cloud, to get it classified by the classification algorithm.

The query is always in the encrypted format and hence is not readable for normal users.

4.CLOUD SERVER

The cloud server maintains this database and also when it has to search something for reference it refers the database only.

5. CLASSIFICATION

Now, the classification method classifies the entire encrypted data over the database.

Lastly, once the data are classified i.e the various 'Class -Labels' are formed, the top K data is retrieved by entering its value.

VI. PROPOSED SYSTEM

We try to propose a Privacy preserving k-NN classifier algorithm. Here, the data is semantically secured and encrypted as well.

Once, the data is encrypted and uploaded over the cloud the owner of the database does not involve in any computations. So no information is revealed to the database owner. Also, content of database D and the query 'q' is not revealed to cloud.

Lastly, the 'Class Label : Cq' is revealed only to the authorized user and no one else. In addition, after sending his encrypted query record to the cloud, user does not involve in any computations. Hence, data access patterns are further protected [4].These are the ways that protects the confidentiality and integrity of the data or information of the user.

$$PPkNN(D',q) \rightarrow cq$$

PPKNN stands for privacy preserving K-NN protocol [5].
This protocol ensures data confidentiality, user's input query's confidentiality, and also hides the access patterns of data.

VII. MODULE DESCRIPTION

There are three basic modules of the system:

1. *Data Owner*
The owner once registers itself as owner provides login details to approve their authentication. Afterwards, they can carry out various operations as per their consent like uploading the data. It performs these operations:

- ✓ Register
- ✓ Login
- ✓ Upload Data

2. *Admin*
It acts as an admin for both the other modules and hence acts as a mediator for both. It basically acts as a server that provides services to both the modules when requested with a service

It basically act as a service provider to the Owner as well as User. Both of them carry out sign-up, login and carry out various operations like upload the data, classification.
It performs these operations:

- ✓ Register
- ✓ Login
- ✓ Send Request.

3. *Data User*

It signs up or registers at first and then logins, afterwards it sends a request to the admin .This request is nothing but the query to be processed.

Once the query is processed, the classified data or data with the 'Class Labels' are provided to the user who is authorized and no one else.
This figure will make it clear :

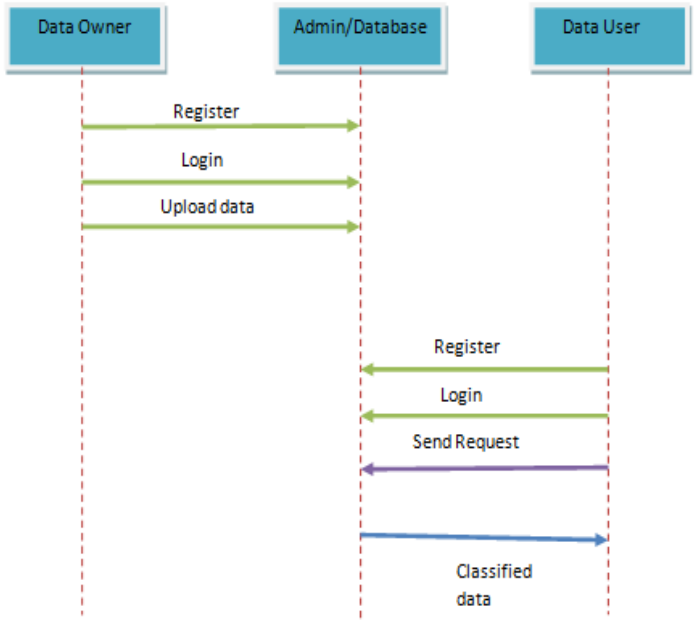

Fig : Sequence Diagram

All the functionality of the three modules comprised helps formulate the working of the system.

Lastly, the Admin returns the classified data or the Class label to the authorized users and that is the end result.

VIII. CONCLUSIONS

This paper can easily sort out the problem of the classification methods over the 'Encrypted data',as the usual classification technique cannot be applied to the encrypted data.This protocol solves the huge problem of classifying the encrypted data and hence is similar to a revolution in making : data storage,data mining operations ,infrastructure etc

It proposed a novel privacy-preserving kNN classification protocol over encrypted data on the cloud. It also ensures data confidentiality, user's as well as input query's confidentiality and also hides the access patterns of data.These are the norms where in the confidentiality and the integrity of the data is preserved till the last point.

Now the user can not only enjoy the benefits of the secure cloud but also can perform various operations of data mining on them.Hence, create class label as the end producct that is provided to the end user that has a proper authorization over the sytem,

This protocol can also be modified in future for some better classification technique over the encrypted data.

IX.REFERENCES

[1] National Institute of Standards and Technology's (NIST) "definition of cloud".
[2] C. Gentry, "Fully homomorphic encryption using ideal lattices," in ACM STOC, pp. 169–178, 2009.
[3] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, pp. 612–613, Nov. 1979.
[4] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, "k-nearest neighbor classification over semantically secure encrypted relational data." print arXiv:1403.5001, 2014.
[5] R. Agrawal and R. Srikant, "Privacy- preserving data mining," in ACM Sigmod Record, vol. 29, pp. 439–450, ACM, 2000.
[6] Y. Lindell and B. Pinkas, "Privacy preserving data mining," in Advances in Cryptology (CRYPTO), pp. 36–54, Springer, 2000.
[17] L. Xiong, S. Chitti, and L. Liu, "K nearest neighbor classification across multiple private databases," in CIKM, pp. 840–841, ACM, 2006.
[18] Y. Qi and M. J. Atallah, "Efficient privacy-preserving k-nearest neighbor search," in IEEE n ICDCS, pp. 311–319, 2008.
[19] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in ACM SIGMOD, pp. 563– 574, 2004.
[20] H. Hacig¨um¨us¸, B. Iyer, C. Li, and S. Mehrotra, "Executing sql over encrypted data in the database-service-provider model," in ACM SIGMOD, pp. 216–227, 2002.
[21] B. Hore, S. Mehrotra, M. Canim, and M. Kantarcioglu, "Secure multidimensional range queries over outsourced data," The VLDB Journal, vol. 21, no. 3, pp. 333–358, 2012.
[22] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in ACM SIGMOD, pp. 139–152, 2009.