ENCRYPTED DATA SEARCHING USING KNN ALGORITHM

Remya R S [#1] , Aswathy G S [*2]

[#]*Master of Engineering , Dept of CSE , Sivaji College of Engineering and Technology*

[*]*Assistant Professor, Dept of CSE , Sivaji College of Engineering and Technology*

Abstract--In medical cloud computing, a patient can store her medical data to the cloud server. By storing the medical data then only the authorized doctor can access the data by the appropriate patient. Before storing the data they are encrypted .thus the patient only needs to send the encrypted key to the authorized doctor. Using this encrypted data it is very difficult to search the data .thus there proposed a Secure and Efficient Dynamic Searchable Symmetric Encryption (SEDSSE) techniques. In this technique that leverages two schemes the secure K-Nearest Neighbor(KNN) and to find the keyword thus leverage a technique of Attribute Based Encryption(ABE).also achieve two security features of Backward and Forward privacy. That can solve the key sharing problem .further can solve the duplication of files that are uploaded can be solved by Mergle hash function.

Index Terms-Medical data, searching encrypted data, KNN algorithm, duplication in updating

I. INTRODUCTION

Recently, the cloud computing is the way of operating their data in the way that they can store and process data. it also consider the terms of cost efficiency, offload of administrative Overhead ,flexibility. Thus it can also derive the useful information and sensitive information of the actual data item that the data are encrypted the cloud that act as third party. That the cloud environment presents the opportunity for the data owner to outsource their database, and database management functionality and data miming tasks to access mechanism for querying and managing the hosted database.

Encryption technique will be a direct way to protect the confidentially of the particular data from the cloud as well as from the unauthorized user by encrypting the data before outsourcing it. Thus the data owner can protect the privacy of his own data. The authorized user required to encrypt there data before sending them to the cloud. For this a scare and efficient dynamic searchable symmetric encryption techniques for the security.

Computing paradigm that enables resource allocation, measurements of service, self demands. such, a patient store her data on the cloud server ,namely data out souring, and then open her cloud data to the doctor it includes

private information such as medical cases and diagnosis' reports. Other challenging factors such as forward and backward privacy for the encryption. For this the real health care service enivornment the health records are owned by the multiple data owners such as unauthorized users. For this case the health related medical data will be stored in the cloud server as by using knn algorithm to encrypt the details and stored. By this method only the authorized doctor will sent the corresponding encrypted key to the patient as the request and that will be verified only after that the patient send the permission and then the authorized doctor will view the document by the keyword.

The main contribution of this paper is that the many of the files are uploaded such a case there occurs the duplication. this duplication that can be existed by the mergle hash function. By using this function the details that are uploaded by the patient side will be duplicated thus it can be rectified by the hash function. Thus there avoids the d-duplication of the data that are stored in the cloud server. Such a case only one file will uploaded onces if there occur more then there will show the alternative message that the file is uploaded earilier thus the patient cannot upload the file also they know that the already the file is uploaded.

By these techniques thus there occurs the d-duplication scheme. and also avoid the unauthorized access of the data that are stored in the cloud. thus the authorized user can only access the data that are encrypted by sending the corresponding encrypted key.

II. RELATED WORK

The encrypted data cannot provide good usability due to the difficulty of searching over encrypted data. For this issue, searchable Symmetric Encryption (SSE) technology has been proposed in the fundamental approach for the keyword search over the encrypted cloud data. In system they will implement mutikey word searching; fuzzy keyword searching technique in this method the data owners cannot rechange there updated documents. Also they implement the searching method in a plain text so the security for their schema is very low compare to ours. in existing system there must be a key problem occur that in this there is only one data owner and he uploaded his file using a single key this may lose our security if any one access the key and they can download every single file in the system.

The concept of Static SSE that enables SPE was first proposed by Boneh et al. which supports single keyword search on encrypted data but the computation overhead is heavy. Curtmola et al. refined the definition of SSE later. After this work, Boneh et al. proposed conjunctive, subset, and range queries on encrypted data. Recently in

static searchable symmetric encryption, Wang et al. have developed the ranked keyword search scheme in and proposed a novel scheme supporting similarity search in. However, these schemes cannot efficiently support multi-keyword search. To overcome this problem, Sun et al. Proposed a multi-keyword scheme which also considers the relevance scores of keywords, and it can achieve efficient query by utilizing the multidimensional tree technique. In, Yu et al. proposed a multi-keyword took retrieval scheme with fully homomorphism encryption, which can return ranked results and achieve high security. Cao et al. proposed a multi-keyword ranked search scheme, which can return ranked results of searching according to the number of matching keywords and its extended versions achieve higher efficiency. As mentioned by Ren et al., there still exist many security challenges for public clouds.

The concept of Dynamic SSE (DSSE) can be referred to Song et al., which explicitly considers the problem of searchable encryption and can support insertions/deletions of documents in a straightforward way. However, the straightforward way causes heavy overhead of updating. Kamara et al. Proposed a dynamic scheme which achieves security against adaptive chosen-keyword attacks and can add and delete documents efficiently. Two schemes with sub linear search and updating time were developed in and such schemes have better security property, i.e., forward privacy.

Besides, some ORAM schemes seem to be the most secure way to query encrypted data, which can achieve both forward and backward privacy but they are of high updating complexity which limits their application in practice. Recently, Yuan et al. have made several significant contributions in the area of image-centric social discovery e-health monitoring system with minimum service delay and privacy preservation by exploiting geo-distributed clouds. In the system, the resource allocation scheme enables the distributed cloud servers to cooperatively assign the servers to the requested users under the load balance condition. Thus, the service delay for users is minimized.

III.PROPOSED SYSTEM

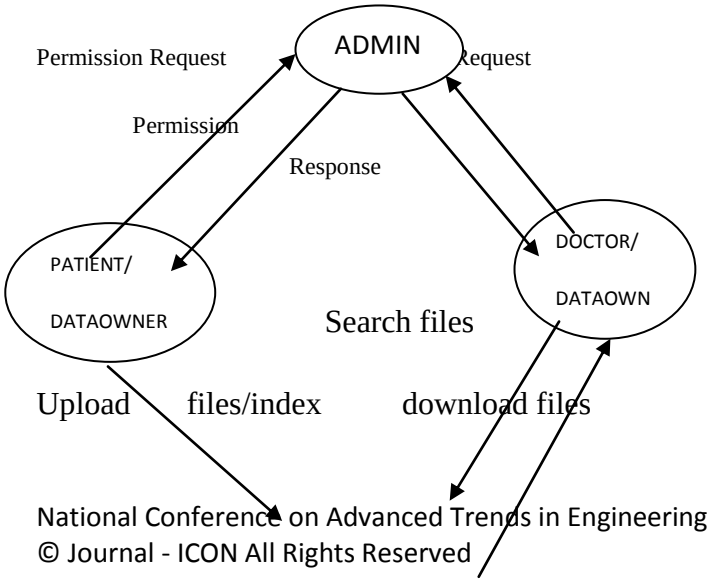
A Secure and Efficient Dynamic Searchable Symmetric Encryption (SEDSSE) scheme over medical cloud data. That extends and improves our previous research. This paper addresses two new issues: the collusion between the cloud server and search users as well as different secret key distribution among search users. In addition, we apply the new design to the health care system, the security and performance are analyzed.

Firstly, we combine the k-Nearest Neighbor (KNN) and Attribute-Based Encryption (ABE)

techniques to propose a Secure and Efficient Dynamic Searchable Symmetric Encryption scheme, named SEPSSE I. The proposed scheme can achieve forward privacy, backward privacy, and collusion resistance between the cloud server and search users.

Secondly, based on the scheme, we further propose an enhanced scheme, named SEPSSE II to solve the key sharing problem which widely exists in the KNN based searchable encryption schemes. Compared with the existing DSSE schemes, our proposed schemes are have less storage costs, search and updating complexity. to patient and will give permission then only the data file can be taken. And then can be uploaded in the cloud. Thus there occurs the security.

Each of this case the doctor and patient sent the corresponding details to the trusted authority that is admin. The admin play a great role of the SE DSSE.



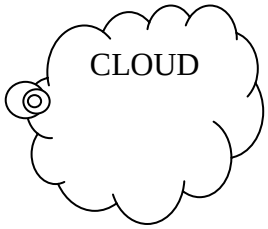


Fig 1.Architecture of proposed system

First, the admain set the site of the medical data to be stored in the cloud. The cloud is the third party access. Thus there does not have any security thus there implement a security scheme of Secure and Efficiency of Dynamic Searchable Symmetric Encryption. If the patient upload the medical records. That will be encrypted by the admin using KNNalgorithm.it will convert the plain text data into cipher text thus it will be in the form of unreadable format. Thus the patient create a keyword for each of the document. That will be set into the cloud by the Attribute based encryption technique. Thus there set a vector value for the keyword. After that only searching the keyword of particular document will be identify the document related to the keyword.

The patient first set the details and get registered. There will be a id and password. by using the id and password the patient can get login. the doctor can also set there details to register there details. After the successful registration the doctor also set the id and password. Also the doctor can view there profile. And if there occurs any changes then can be

changed. Also can apply the photos and days of appointment and time of appointment. After that the admin login can view how many of the patient and doctor are registered to the site. After that the patient can upload the data in the uploading side. And can set their key word for the related files.

Thus the doctor will login and can search the files by sending the appropriate keyword to the searching side. Thus the related files will be occurs due to the priority based and choosing the file the doctor can send the request to the patient. After that the patient give the conformation then only the doctor can get the files in the form of decrypted. thus the patient send the request to the admin. And admin send the response then only the patient can upload the files in to the cloud. Thus in the cloud they are in the form of encrypted. And doctor search the files. They are in the form of encrypted. Thus the doctor send request to the admin and admin will give response by the checking the keyword. And give response and the doctor can download the files.

The privacy. In this section there occurs the knn algorithm and attribute based encryption technique that to protect the data from the unauthorized access. Each patient's medical data are highly sensitive. So they are encrypted and stored as the unreadable formate thus there occurs the privacy of the data. In this experiment there occurs the index creation and searching of keyword technique. thus there occurs the main method to protect the data on the particular cloud server. By the keyword only the data can be reterived thus there occurs the basic method. Also the future there occur the duplication thus can be rectified by the technique. Thus there one user will upload each file only once. Thus there duplication can be avoid.

IV.ALGORITHM

In this section, we overview the concept that from a foundation of our research. These include KNN algorithm, ABE Encryption.

A.KNN ALGORITHM

A k-nearest-neighbor algorithm, often abbreviated knn, is an approach to data classification that estimates how likely a data point is to be a member of one group or the other depending on what group the data points nearest to it are in.The k-nearest-neighbor is an example of a "lazy learner" algorithm, meaning that it does not build a model using the training set until a query of the data set is performed.

KNN algorithm is particularly sensitive to outliers and noise contained in the training data set. In this paper, we use the reverse cloud algorithm to map the training samples into clouds. Each attribute is mapped to a cloud vector. .

It stores all a valuable cases and classifies new cases based on a similarity measured.it is distance function.also used for the statistical estimation and pattern recognition already.

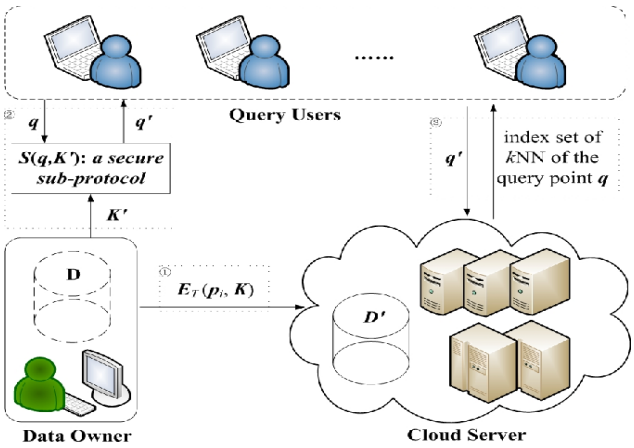


Fig 2.KNN Algorithm

By comparing the similarity of clouds in the cloud vector, we can calculate the attributes weights. For those attributes with a low weight of properties, we find out merger them to a new attribute which can generate more significant attribute weight than original ones. We present a new KNN algorithm based on Cloud Model and compare our algorithm with classic KNN algorithms and other well-known improved KNN algorithms using 10 data sets. Experiments show that our approach could achieve a better or at least a comparable classification accuracy with other algorithms.

B.ABE ENCRYPTION

Attribute-based encryption is a type of public-key encryption in which the secret key of a user and the cipher text are dependent upon attributes (e.g. the country in which he lives, or the kind of subscription he has). In such a system, the decryption of a cipher text is possible only if the set of attributes of the user key matches the attributes of the ciphertext

.A crucial security aspect of attribute-based encryption is collusion-resistance: An adversary that holds multiple keys should only be able to access data if at least one

individual key grants access.

Although ABE concept is very powerful and a promising mechanism, ABE systems suffer mainly from two drawbacks: non-efficiency and non-existence of attribute revocation mechanism. Other main challenges are:

- Key coordination
- Key escrow
- Key revocation

Revocation of users in cryptosystems is a well-studied but nontrivial problem. Revocation is even more challenging in attribute-based systems, given that each attribute possibly belongs to multiple different users, whereas in traditional PKI systems public/private key pairs are uniquely associated with a single user.

In principle, in an ABE system, attributes, not users or keys, are revoked. The following paragraph now discusses how the revocation feature can be incorporated.

It can be a type of public key encryption in which the secret key of the user and the cipher text are depend upon the attributes.instead of encrypting each part of a log with the keys of all recipients,it is possible to encrypt the log match attribute.

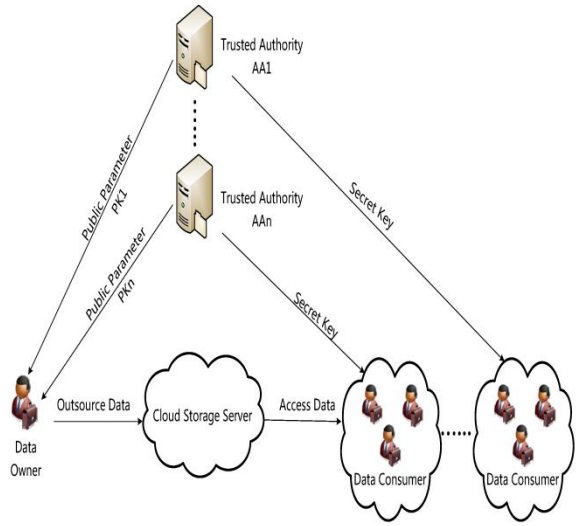


Fig 3.Attribute Based Encryption

A simple but constrained solution is to include a time attribute. This solution would require each message to be encrypted with a modified access tree T0, which is constructed by augmenting the original access tree T with an additional time attribute. The time attribute, ζ represents the current 'time period'. Formally, the new access structure T0 is as follows: T0 = (T AND ζ). For example, ζ can be the 'date' attribute whose value changes once every day.

It is assumed that each non-revoked user receives his fresh private keys corresponding to the 'date' attribute once each day directly from the mobile key server MKS (which is the central authority) or via the regional delegates. With a hierarchical access structure, the key delegation property of CP-ABE can be exploited to reduce the dependency on the central authority for issuing the new private keys to all users every time interval.

There are significant trade-offs between the extra load incurred by the authority for generating and communicating the new keys to the users and the amount of time that can elapse before a revoked user can be effectively purged.

This above solution has the following problems: Each user X needs to periodically receive from the central authority the fresh private key corresponding to the time attribute; otherwise X will not be able to decrypt any message. It is a lazy revocation technique the revoked user is not purged from the system until the current time period expires. The certain method that the attribute based encryption can be obtained.

V.DESIGN

The patient and doctor module are connected to the admin and the cloud. The patient send the request and accept then only the files that can be uploaded. And also the doctor also have to send the request and access the files to download.

They are connected to the visual studio. Thus the sql server also will access the cloud. The information those are encrypted and accessed. Thus they can be decrypted by the knn algorithm.

VI.CONCLUSION AND FUTURE WORK

In this paper, the K-Nearest Neighbor (knn) algorithm that will be used for the unauthorized access to the data. And the data in the medical cloud are safe. Due to the encryption of the data also occurs the key sharing problem propose two dynamic searchable encryption schemes with high security level. The first one can not only achieve collusion resistance between the cloud server and search users, but also can achieve both forward privacy and backward privacy. The second one further solves the key sharing problem which widely exists in the kNN based searchable encryption scheme. One of the major issues of achieving secure and efficient dynamic searchable symmetric encryption over medical cloud data that involves duplication. To avoid duplication over the medical cloud data that can be overcome by the mergle hash function. The mergle hash tree that can avoid the duplication over the cloud data that are uploaded by the patient. It can be represented by the binary tree method. This problem that cannot avoid duplication over the data. The tree structure of the hash function that can reduce the duplication.

REFERENCE

[1]Hongwei(2018),"achieving secure and efficient dynamic searchable symmetric over medical cloud data".

[2] Q. Shen, (2014)"Exploiting geodistributed clouds for e-health monitoring system with minimum service delay and privacy preservation".

[3]W.Sun(2014)"Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking".

[4]Y.Zhu(2013) "Towards secure multikey word top-k retrieval over encrypted cloud data".

[5] Y. Zheng(2013)"Scalable and secure sharing of personal health records in cloud computing using attribute- based encryption".

[6] H. Liang(2012)"An smdpbased service model for interdomain resource allocation in mobile cloud networks".

[7] C. Wang(2012) "Enabling secure and efficient ranked keyword search over outsourced cloud data".

[8] M.-H. Kuo, (2011)"Opportunities and challenges of cloud computing to improve health care services.

[9] K. Ren“Securing personal health records in cloud computing: Patient-centric and fine-grained data .

[10] L. M. Vaquero(2008) “A break in the clouds: towards a cloud definition.