**Pak. J. Biotechnol. Vol. 13 special issue** II **(International Conference on Engineering and Technology Systems (ICET'16) Pp. 109 - 112 (2016)**

# CONJUNCTIVE KEYWORD SEARCH ON E-HEALTH RECORDS BASED 0N K-ANONY-MIZATION TECHNIQUE

S. Sneha and P.Asha

Computer Science and Engineering, Sathyabama University, Chennai, Tamil Nadu- 600 119, India.
1snehasagayaraj@gmail.com,

School of Computing, Sathyabama University,Chennai, Tamil Nadu-600 119, India.
ashapandian225@gmail.com

***Abstract***

An electronic health care system greatly enhances the patient healthcare records which are stored in the cloud server. Searchable encryption scheme is used which enhances the search mechanism. Conjunctive keyword search helps the authorized users to access the records by giving multiple keywords, so that it becomes difficult for the attackers to guess the keyword and retrieve the records. Re-encryption scheme provides more security to the records by re-encrypting the encrypted index before uploading them into cloud server. Since the patient's healthcare records consist of sensitive information, it may be inconvenient for the patient when his records are accessed by everyone. To overcome the problem in our proposed work we introduce the concept called K-Anonymity which is used so that it gives only a partial access to the authorized users by using two methodologies suppression and generalization. This has

been very efficient in the standard model.

*Keywords—K-Anonymity, Re-encryption, authorized*

## I.INTRODUCTION

In order to prevent the medical errors the Electronic Health Records makes the medical records to be computerized, by storing them in cloud. When the healthcare record of a patient is created in one hospital which will be centralized and stored in a cloud server so that when the patient moves to another hospital it will help him to manage and share information with others also. Electronic Health Records has lot of privacy issues. The patient's healthcare records may be vulnerable to attacks. Even though they promise to keep the data's safe, if the server is intruded or the misbehavior of even a single staff member who manages the records will cause the patients sensitive healthcare information to be leaked. So it is essential to keep track of the privacy of the records.

Re-encryption technique is used where the encrypted data's are re-encrypted, which will enhance the security. If the patient wants to move to an another hospital and he does not want his records to be accessed by the users from the previous hospital anymore, then he can use a new key to encrypt the records, will is more expensive. We have time-based proxy re-encryption scheme and we have a time limit which will be set for the authenticated users by mentioning the beginning and the closing time, such that the users have to access the records within that time limit or else he/she cannot access, the records will be deleted automatically. If the time period is one year then the users can access the records within that particular year, and after which they will lose their access rights.

Public Key Encryption Scheme with keyword search is used, which enables the user to search on the encrypted records without decrypting it. If the patient is the data owner he may give access right to the person's he wished to, by giving his private key to the trusted users. With the help of the private key, the users may search the encrypted records. If the user queries the private key, and if it matches with the record, then the record will be retrieved. One time password will be provided if the user request for the record. PKES is more efficient and secure scheme, which makes the hacker difficult to guess the keyword.

In the existing work conjunctive keyword search scheme with designated tester and timing enable proxy re-encryption function (Re-dtPECK) is used. Where the data owner, data users, time server, proxy server are provided with public and private key pairs. Designated tester is that only the designated tester will carry the test algorithm, mostly the server. And proxy re-encryption with conjunctive keyword search allows the server to re-encrypt the key, so that the records which have been encrypted by the data owner with his public key can be decrypted by the user with his private key. This scheme is called proxy re-encryption with keyword search (RE-PEKS). Since this scheme allows only a single keyword search

so, conjunctive keyword search has been proposed which is (RE-PECK) allows multiple keyword search over encrypted records. Designated tester allows only the server to carry out the test algorithm. And the timing enabled function allows the user to search the records within that time period mentioned in the algorithm. AES algorithm is used in order to encrypt the public key and also the documents.

The patient who is the data owner may also doesn't want the trusted users to view his full disease details. And also the timing enabled technique may not be appropriate for some users. In order to overcome the problem, we propose an approach called K-Anonymity which provides only a partial access to the users, by using two techniques namely, suppression and generalization. In suppression certain values of the attributes are replaced by an asterisk '*'. All or some values of a column may be replaced by '*'. In

generalization individual values of attributes are replaced by with a broader category.

## II. RELATED WORK

For searching the documents single keyword search was used which will take a long time for searching, and retrieve many documents that contain the keyword, so searching technique becomes inefficient. In order to overcome the above problem, we move on to conjunctive keyword search, which is not the multiple execution of single keyword search instead it enhances the search technique by enabling the users to query multiple keywords in turn they can retrieve the required data's or document. And it becomes more efficient since it extracts the exact result. It also enhances the privacy by making the users to know which documents are extracted by the users **[1].** Secure Channel Free-Public Encryption with Keyword Search (SCF-PKES), allows the server to have its own public/private key pairs which is (kp,ks)where k is the input and p is the public key and s is the private key. Where the user inputs the server's public key in the algorithm, it will be executed only when the public key matches with the private key. Since PKES has a drawback of using only one keyword for searching over encrypted data we propose a method called as PECKS Public Encryption with Conjunctive Keyword Search, where a secure channel is set between the sender and the receiver, Where the public key and the document is given as input in the algorithm and the cipher texted conjunctive keyword is the output. Similarly with the private key and the query as the input the trapdoor is generated as output. When the algorithm is run if the cipher texted conjunctive keyword matches with the trapdoor the result is returned or incorrect message is displayed **[2]**. Before outsourcing, the data owner will prepare an access control list (ACL), which is the list of users who can access the data's, and they will be grouped together under one file group. And each file group will be encrypted by using one symmetric key, and this key will be distributed to the users under each group. And with the help of the key the users can decrypt and retrieve their documents. First classify data with similar access control lists (ACLs) into a file group, and then encrypt each file group with a unique symmetric key. The symmetric key will be distributed to the data users in the list, so that only the users who are in the ACL can access the file. The main drawback of this approach is that the key size managed by the owner grows along with the number of file groups **[3].** In the existing work PEKS is proposed, which allows the users to search over the encrypted documents, and is based on the traditional encryption scheme along with the keyword search, where the owner has to prepare the keywords from the input k, and then encrypt these keywords, and these keywords will be indexed and outsourced to the cloud server, and later when the users wants to retrieve the documents he will query the chipper texted keyword to the server, and the document which matches with the keyword will be retrieved **[4].** Time Released Encryption is based on the time dependent type of encryption. And the decryption can also be controlled based on the time. The particular group of recipients will be given a time limit to access the records, and the can decrypt the records within that time limit after which they will lose their access rights and they will not be able to access the records. Time Released Encryption (TRE) along with Proxy Re-Encryption is used, which is found to be more effective. Proxy re-encryption technique enables the encrypted records to be re-encrypted **[5].** Searchable symmetric encryption technique is used, helps the user to retrieve the documents by using his private key, this scheme helps the user to query the keyword in such a way even the owner does not know what was the query, but the owner still has to authenticate the query. So the owner can authenticate the query without learning the policies **[6].** This scheme removes the secure channel which is used for security purpose. Instead Key Policy-Attribute Based Keyword Search is used (KP-ABKS), which is based on Key Policy-Attribute Based Encryption (KP-ABE). The user quires the attribute to retrieve the document, this scheme allows multiple users to carry out search mechanism which has been proved to be more flexible **[7].** The work is based on Searchable symmetric encryption (SSE), which supports both conjunctive search and also Boolean queries over the encrypted documents. It is suitable for very large databases and focuses mainly on the single keyword search **[8].**

### *A. Drawbacks*

In the existing system there are some drawbacks, storing the data in cloud has many security issues, so many of the companies do not prefer cloud. In order to enhance the security we prefer proxy re-encryption technique (PRE). Using Public Encryption Keyword Search allows only single keyword search, so conjunctive keyword search is preferred. When the patient moves to an another hospital he does not want his record to be accessed by his previous physicians anymore, so the earlier methodologies use a new key for encrypting the records, Which consumes lot of time and the cost is also high. In order to overcome the above issue timing enabled re-encryption was proposed, where the records are deleted automatically when particular time period is reached, and the time limit is set so that only the authorized users can access the records within that time period. This again brings incompatibility to many users but still has proved to be more secure. We present a concept called k-anonymity in our proposed work, which not only enhances the security but also displays only the necessary details to the authorized users.

## III. PROPOSED SYSTEM

There are various advantages in cloud computing, since cloud provides a large storage space and also we can access our files from anywhere and anytime we want, our goal is to move the patients' health care records into the cloud server.

This will help to prevent the errors in the medical records. The users for the EHR can be nurse, doctors etc. Since the patients records are centralized, the details for the patient can be accessed from anywhere

and it can be shared between the members of the hospital. Even if the patient moves from one hospital to another, since the records are stored in the cloud server, they can be easily accessed by the authenticated members of another hospital. Since the Electronic Healthcare Records contains the most sensitive information, the patient does not want his disease details to be leaked. To encrypt the EHR we use AES algorithm. In the proposed work we have used conjunctive keyword search scheme with designated tester and in the existing work timing enable proxy re-encryption function (Re-dtPECK) is used which makes the search technique more effective. But the main drawback is that if the user is given a particular time limit it becomes difficult for him to access the records whenever he wishes to, and at the same time the patients details should be kept safe and secure. In order to overcome the above drawback we move on to a technique called k-anonymity, which can be achieved using two techniques namely suppression and generalization.

K-anonymity is the technique where the original dataset will be transformed so that it will be difficult for intruder to determine the identity of individual data. Two methods used in K-anonymization are generalization and suppression. Suppression is the process where the individual attribute will be replaced by asterisk. For e.g. if Joan has heart disease then heart disease will be replaced by asterisk.

And generalization is the process of replacing the values of attribute with a border category .e.g. if Joan age is 19 instead of mentioning as 19, it can be replaced by 20<age<=30, which is the border category.

And the main advantage is that only the certain users will be able to access certain details, like the doctors can access the full details of patient and the nurse can access the symptoms and the medicine. The chemist can access only the medicine details. In our proposed system if the user requests for the records he gives the query along with his private key to the sever, if the key matches with the severs public key, then an One Time Password will be sent to the users mobile number by the server. If the user enters the correct OTP he will be allowed to search the records. This method is introduced to enhance the security to our proposed system.

*A. Advantages*

- No time limit
- Highly efficient
- More secure
- Prevents offline keyword guessing attacks

*B. Algorithm*

AES algorithm is used to encrypt the key and the electronic health care records. And k-anonymization is used for removing the personally identifiable information from the datasets.

*1). AES algorithm*

AES algorithm uses the concept of both substitution and permutation. It uses the block size of 128 bit. And the key sizes of 128,192,256 bits. For 128 bit key size we have 10 rounds of repetition, for 192 bit key size we have 12 rounds of repetition and for 256 bit key size we have 14 rounds of repetition. Each round consists of 4 steps based on key size.

*a). AES encryption*

For encryption four steps are follows,
(i) Substitute bytes, (ii) Shift rows, (iii) Mix columns, (iv)Add round key.

*b). Step 1: Substitution of bytes*

The 16-byte inputs are substituted in order to form a resultant matrix of four rows and four columns.

*c). Step 2: Shift rows*

Shifting the rows consists of 4 steps,
(i) Not shifting the first row, (ii) Circular shift of second row, (iii) Circular shift of third row with two bytes to the left, (iv) Circular shift of fourth row with three bytes to the left.

*d). Step 3: Mix columns*

Invertible linear transformation is used to combine four bytes in a column. A set of completely new 16-byte input is formed.

*e). Step 4: Add round key*

In this step the 16-byte input is transformed into 128 bit and then they are XORed with a round key of 128-byte. And the output produced is a cipher text and similarly the rounds are repeated based on the key size.

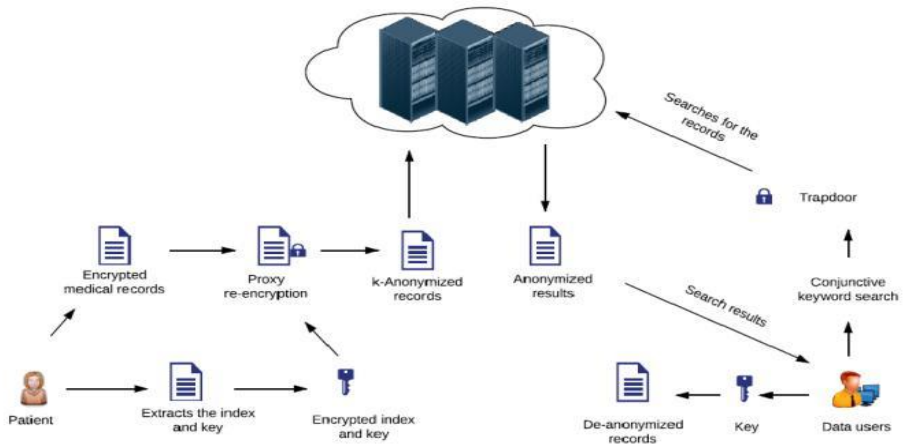


Fig.1.Architecture diagram

*B.AES decryption*

For decryption each round contains four processes which is carried in reverse order. Which includes, (i) Substitute bytes, (ii) Mix columns, (iii) Shift rows, (iv) Add round key. There are various advantages by using AES algorithm for encrypting the key and the records. This includes

- More security
- Faster
- Large key size
- Easy to implement

*C. K-Anonymity*

Anonymization technique is used for protecting the privacy of the health care records. In order to make the medical records anonymous the details of the patient like the disease, symptoms, age and other details could be hidden from the inappropriate users either by using boarder category or using asterisk symbols for some data's. And later de-anonymization could be applied to retrieve the original records.

*1). Suppression*

Suppression is one of the techniques used in k-anonymity, where some of the sensitive information like the disease and the symptom details of the patients could be replaced with asterisk, incase if an inappropriate users wants to access the records.

*2). Generalization*

Generalization is an another technique in k-anonymity which uses the broader categories of values .In case of the medical records where the age of the patient could be broadly categorized, if the age is 22 it will be categorized as it is above 21 and below 30. This makes it difficult for the intruders to guess the data.

And there are various advantages in using anonymization techniques which are listed as follows

- Enhances security
- Highly scalable
- Effective in larger datasets
- Less computing time

ACKNOWLEDGMENT

I express my heart full thanks Mrs. P. Asha school of computing for providing her endless support and courage for bringing up this paper.

CONCLUSION

The proposed work, discuss about protecting the health care records of a patient, since they contain many sensitive information's. Proxy Re-Encryption (PRE) along with K-Anonymity adds more security to our medical records. Since re-encryption technique allows the encrypted data to be re-encrypted and K-Anonymity gives only partial access to the users.

As a future enhancement, we can improve the anonymity techniques so that it will provide easy and secure access to all the records.

REFERENCES

**[1].** J. W. Byun and D. H. Lee, On a security model of conjunctive keyword search over encrypted rela-tional database. J. Syst. Softw. 84 (8): 1364–1372 (2011).

**[2].** M.-S. Hwang, S.-T. Hsu and C.-C.Lee, A new public key encryption with conjunctive field keyword search scheme. Inf. Technol. Control 43 (3): 277–288 (2014).

**[3].** Q. Liu, G. Wang and J. Wu, Time-based proxy re-encryption scheme for secure data sharing in a cloud environment. Inf. Sci. 258: 355–370 (2014).

[4]. D. Boneh, G. Di Crescenzo, R.Ostrovsky, and G. Persiano, Public key encryption with keyword search, Proc. EUROCRYPT Interlaken, Switzer-land 3027: 506–522 (2004).

**[5].** K. Emura, A. Miyaji and K. Omote, A timed-release proxy re-encryption scheme. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. 94(8): 1682–1695 (2011).

[**6].** S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu and M. Steiner, Outsourced symmetric private infor-mation retrieval, Proc. ACM SIGSAC Conf. Comput. Commun. Security Pp. 875–888 (2013).

**[7].** P. Liu, J. Wang, H. Ma and H. Nie, Efficient verifiable public key encryption with keyword search based on KP- ABE, Proc. IEEE 9th Int. Conf. Broadband Wireless Comput. Commun. Appl. (BWCCA) Pp. 584–589 (2014)

[8]. D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Ro ̧su and M. Steiner, Highly-scalable searchable symmetric encryption with support for Boolean queries, in Advances in Cryptology, Berlin, Germany, Springer Pp. 353–373 (2013).