Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 13, Number 7 (2017), pp. 3367-3380

http://www.ripublication.com

Privacy preserving on E-Health records based on Anonymization technique

S.Sneha

Student of (M.E.) Computer Science and Engineering,
Sathyabama University, Chennai, Tamil Nadu-600 119, India.

P.Asha

Assistant Professor, School of Computing,
Sathyabama University, Chennai, Tamil Nadu-600 119, India.

Abstract

Patient's health care records contain a lot of sensitive information which are vulnerable to attacks. The healthcare records of the patients are stored in the hospital server. Even if a single member in the hospital misbehaves there are chances of leakage of sensitive information. Searchable encryption technique is used which enhances the search mechanism and also quick retrieval of the records. Keyword search scheme is used which helps only the authorized users to access the records by giving keywords, so that it becomes difficult for the intruders to guess the keyword and retrieve the records and makes it convenient for the authorized users to access the details. Re-encryption technique is used which improves security to the medical records by re-encrypting the encrypted index before uploading them in cloud server. Since the patient's healthcare records contains sensitive information, it may be inconvenient for the patient when his records are accessed by everyone. To overcome the problem, the proposed work implements the K-Anonymity technique, which is used to hide the originality of the dataset, so the only the admin can view all the details of the patient where other members of the hospital can view only the details applicable to them. This can be achieved by using two methodologies suppression and generalization.

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 the cloud server. 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 single member in the hospital 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.

The patient who is the data owner may also doesn't want the trusted users to view his full disease details. 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 users in the ACL, so that only the users in the ACL can access this group of files. 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. K-anonymity has been proposed, 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 are nurse, doctors, pharmacists, patients who play different roles in the hospital. 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 Health care Records contains the most sensitive information, the patient does not want his disease details to be leaked. To encrypt the HER AES algorithm is used in the proposed work and 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.

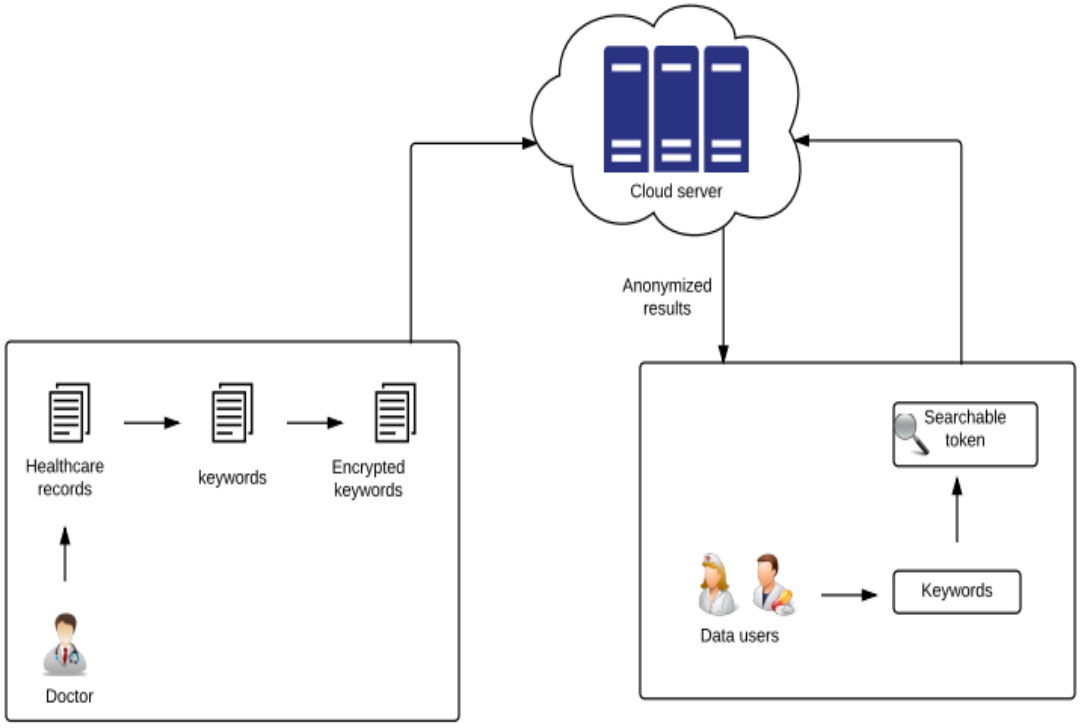

Fig.4.1. Architecture diagram

A. ADVANTAGES

There advantages of the proposed work includes

- Security is more
- Retrieval is easy
- Timing oriented
- Prevents keyword guessing attacks

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 full details of patient and the nurse can access the symptoms and the medicine. The chemist can access only the medicine details.

IV. IMPLEMENTATION

The hardware requirements of the proposed work includes

- Processor: i5
- Clock speed: 550MHz
- Hard Disk: 500 GB
- RAM: 4 GB
- Cache Memory: 512KB
- Operating System : Windows 7

The doctor logins to the hospital sever and once he logs in he will get OTP to his mail id. If the OTP is correct he will login in to the hospital server and upload the details of the patient in the mango lab, which is a cloud database. And he will extract the keywords separately and encrypt the keywords which are also stored in mango lab. Later when the users want to search for the records, they will enter the keyword which has been encrypted and anonymized results will be displayed to the users based on their roles. The cloud server will have the database where the entire login details of the patient, their records and encrypted keywords.

A. MODULES REQUIRED

The following are the modules required for the implementation of the proposed work. There are five modules used which includes

1) Cloud server deployment

In this module the cloud server is deployed. Where all the details of the patient will be stored in the cloud server. In our proposed work three databases are maintained, where the first database contains the login details of the patient, the second database contains records of the patient, and the third database contains the encrypted keywords.

2) User registration

The different roles in the hospital include doctor, patient, pharmacists, and nurse. Based on the roles each user will login with their user id and password, and if a new user wants an access they have to register based on the roles that they play. And only the members of the hospital can login and access the details.

3) Data Anonymization

K-Anonymity 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 '*'. In generalization individual values of attributes are replaced by with a broader category.

4) Mongo lab – user data separation

Mango lab is a cloud database which stores all the details of the patients and the login details of the patient are also stored. Three databases are maintained which includes the login details, record details, and encrypted keywords.

5) Dynamic data transfer

If the authorized users wants to access the details of the patient then he have to access with his registered mail id and password and an OTP will be sent to his registered mail id if the OTP is correct then he can retrieve the records.

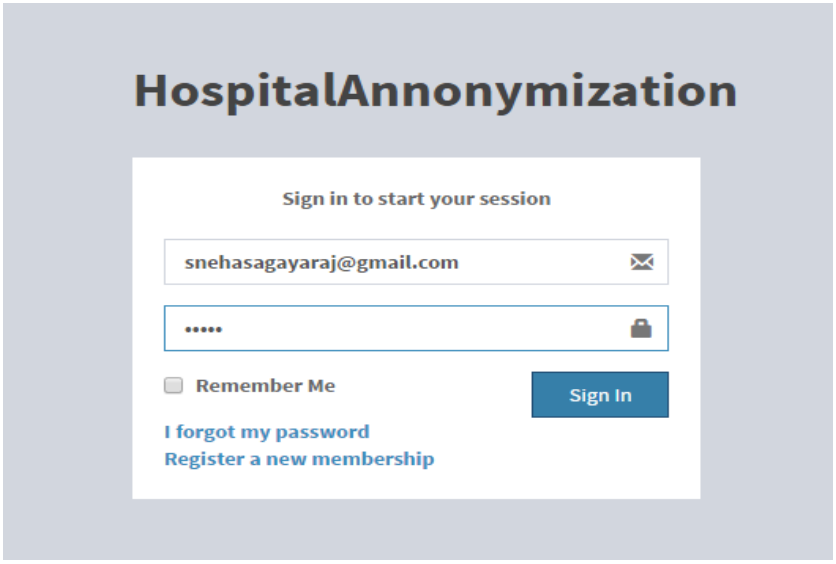

Fig.4.1. Login

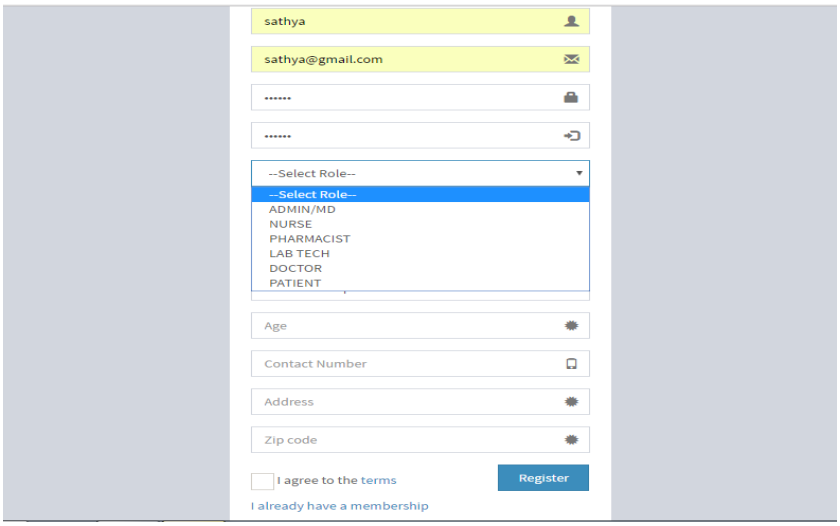

Fig 4.2. New user registration

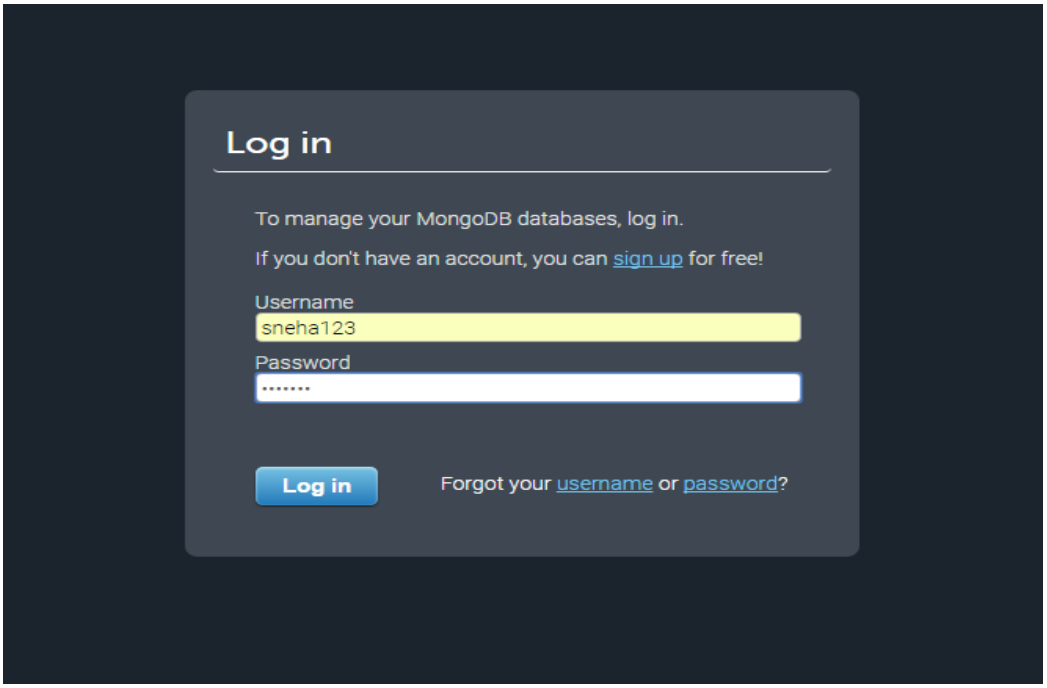

Fig 4.3. Mango lab login

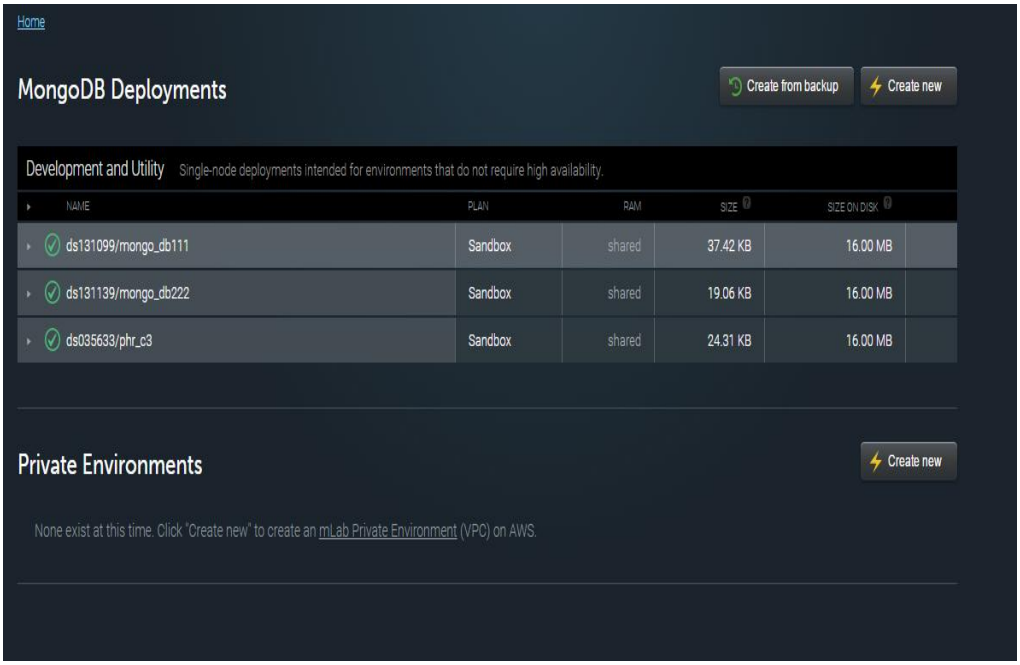

Fig.4.4. Patient dataset

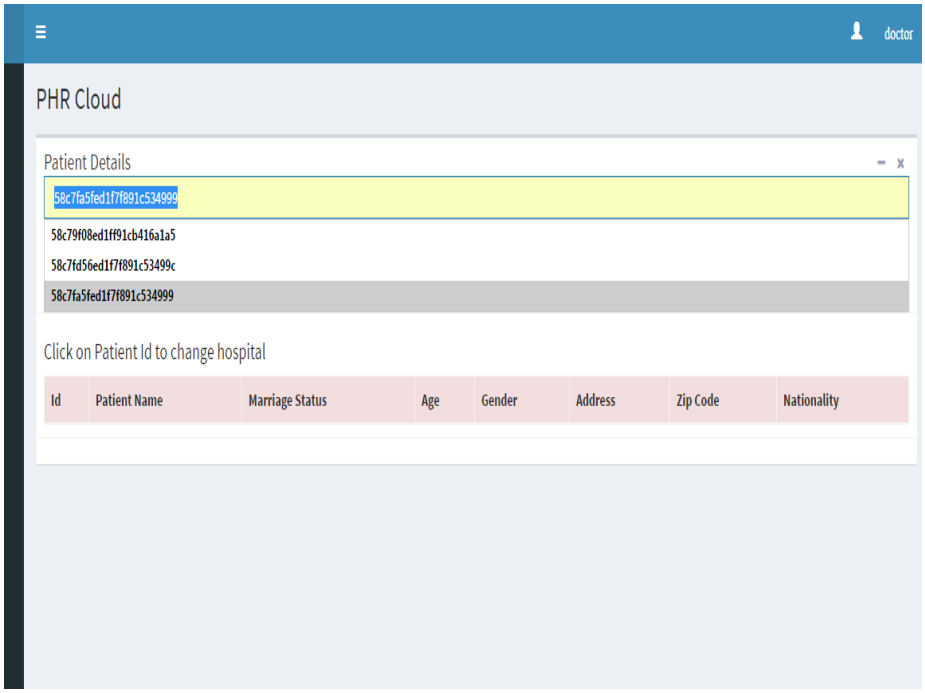

Fig 4.6. Search results

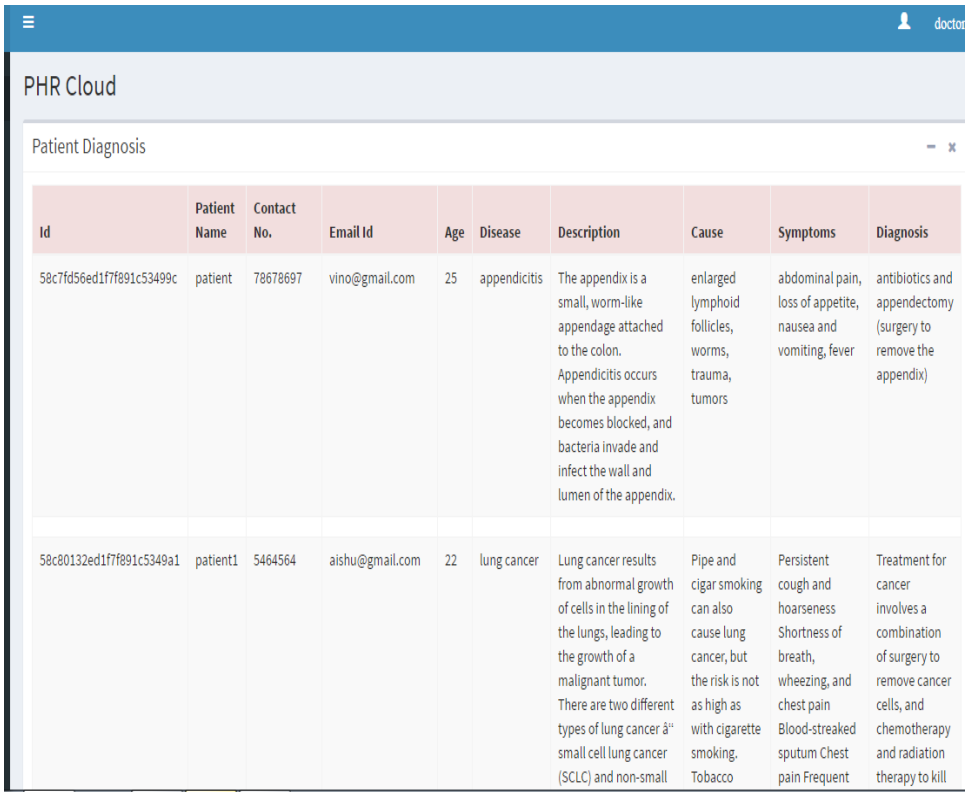

Fig 4.6. Doctor login

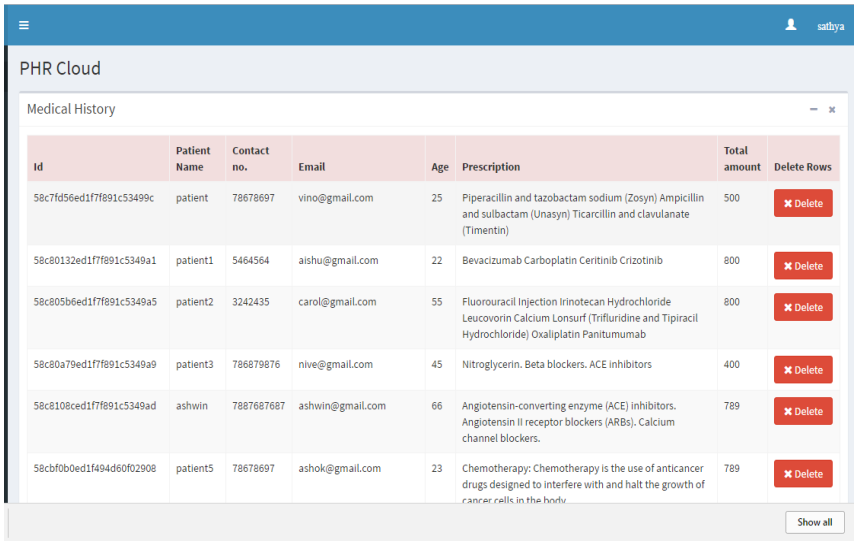

Fig 4.6. Anonymized results

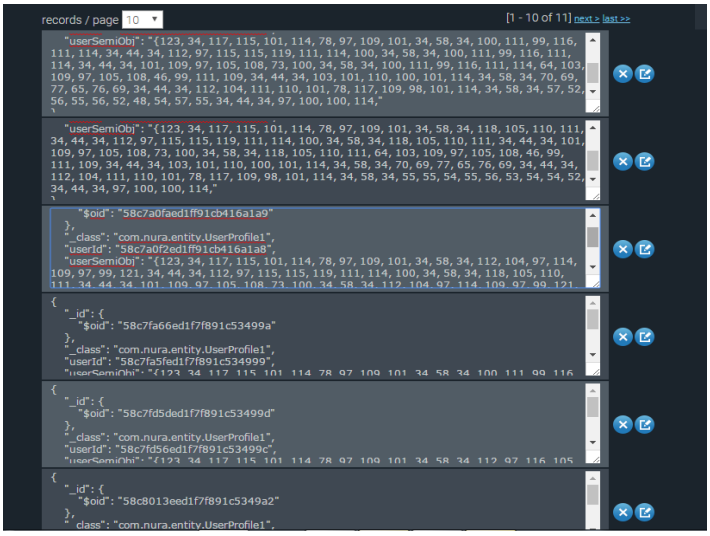

Fig 4.7. Encrypted results

V. 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.

f). 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

3). 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 by

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 any 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 technique which are listed as follows

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

V. PERFORMANCE ANALYSIS

The performance analysis for the total number of records retrieved and the comparison with the existing work are illustrated below

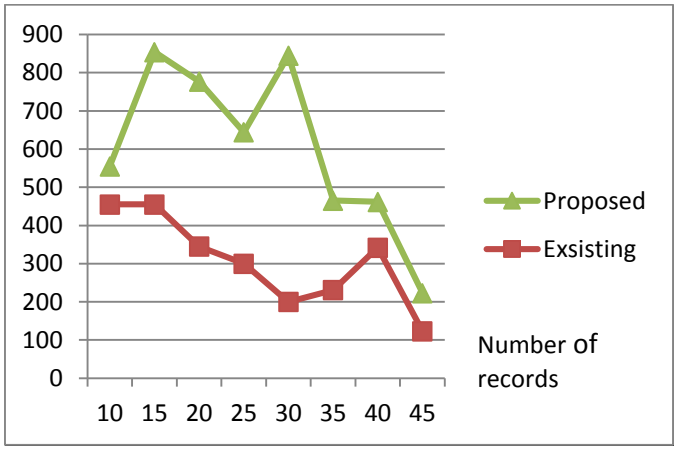

Fig: 5.1 Total number of records retrieved

The above graph fig 5.1 is showing the number of records retrieved within a particular time period. The result is showing that the proposed technique retrieves more records in a shorter period of time when compared to the existing work.

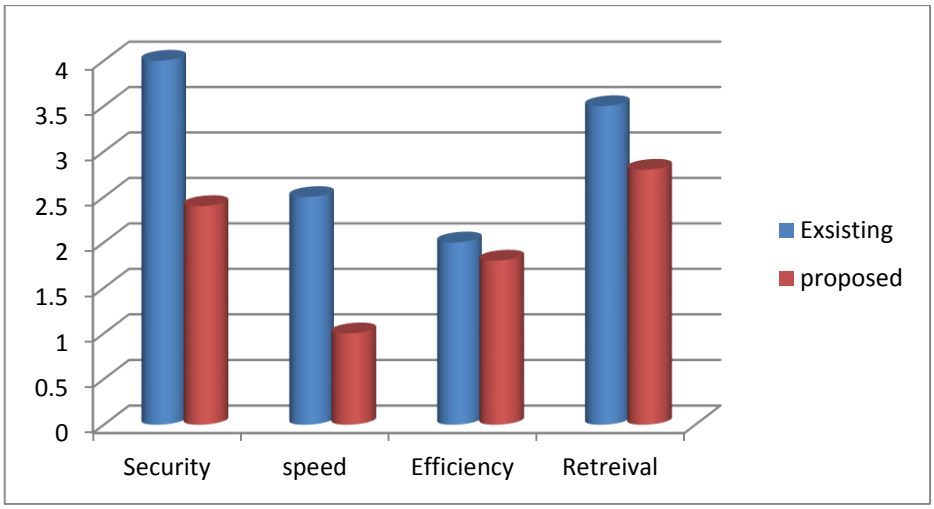

Fig: 5.2 Comparison with the existing work

The above graph Fig 5.2 is showing the security level of records is comparatively high when compared with the proposed system. Similarly the speed, efficiency, and retrieval of the records is comparatively high than the existing work.

CONCLUSION

In the proposed work the implementation of the protection and adding security to the health care records of a patient has been implemented. Since the records contain much sensitive information. Proxy Re-Encryption (PRE) along with K-Anonymity adds more security to our medical records. Since re-encryption technique allows the encrypted keywords 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 asecurity model of conjunctive keyword search over encrypted relational database,” J. Syst. Softw., vol. 84, no. 8, pp. 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, vol. 43, no. 3, pp. 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., vol. 258, pp. 355–370, Feb. 2014.

[4] D. Boneh, G. Di Crescenzo, R.Ostrovsky, and G.. Persiano, "Public key encryption with keyword search," in. Proc.EUROCRYPT, vol. 3027. Interlaken, Switzerland, May 2004, pp. 506–522.

[5] K. Emura, A. Miyaji, and K. Omote,"A timed-release proxy re-encryption scheme," IEICE Trans. Fundam.Electron., Commun.Comput. Sci., vol. 94, no. 8, pp. 1682–1695, 2011.

[6] S. Jarecki, C. Jutla, H. Krawczyk, M.Rosu, and M. Steiner, "Outsourced symmetric private information retrieval," in Proc. ACM SIGSAC Conf. Comput. Commun. Security, 2013, pp. 875–888.

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

[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, 2013, pp. 353–373.