RESEARCH ARTICLE OPEN ACCESS

Privacy Preserving Data Analysis in Mental Health Research Using Cloud Computation

Dr. B. Anni Princy [1], G.Tamilselvi [2], S.Shruthakeerthi [3], B.Sowmya [4]
Professor [1], UG student [2], [3] & [4]
Department of Information Technology
Panimalar Engineering College
Chennai

ABSTRACT

In health care related research studies, there exists a need for retrieving patient's health record from multiple sites. So here comes the digitization of health records, which leads to wide range of access to various users such as doctor, patient, psychiatrist and pharmacist. The sensitive nature of individual health care data poses a security issue. Moreover the increased access of health information by the users threatens the privacy and confidentiality of the stored data. Notwithstanding the existing privacy protection approaches used for mental health records, we suggest a privacy preserving data analysis methodology enabling protection of health records, once user access to records are granted. This paper mainly focuses on utilizing the data analysis approach in preserving privacy of personal health records to overcome the drawbacks of existing approaches.

Keywords:-Health record digitization, data analysis approach, security issue, privacy, confidentiality.

I. INTRODUCTION

Health care research often involves data analysis of huge amount of data on medical records which are more sensitive in nature. The sensitive health information needs privacy preserving

which is a major concern in health domain. Protecting the privacy of Individual information is more important when such information is used for health services related research. For

preserving the privacy of records, we need to ensure that the query result do not reveal any identifiable information to the querying party, privilege for each user will differ from one other. For instance

Case 1: The pharmacist should not be able to learn the patient's personal detail, who have been tested positive for HIV or which patients are receiving treatment for severe psychiatric disorder.

Case 2: The doctor must know all the medical information such as blood pressure, sugar level of the patient, personal information such as age, name but not the contact information of the patient in treatment.

This paper focuses on the challenges of research on patient data, and assures that all data sources have the same, shared, complete identifier, like social security number. The proposed approach provides security by using Advanced Encryption Standard (AES) for the encryption of all patient detail. Every time the updated information is notified to the users to reduce data duplication. This is the major concern when there is huge amount of data in personal health records.

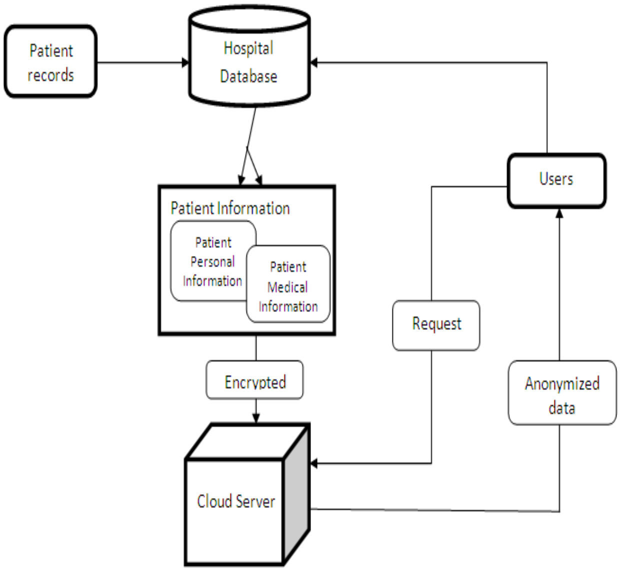
II. EXISTING SYSTEM

The privacy and security of the sensitive personal information are the main need of the users, which could hinder further development and widely adoption of the systems. Data duplication occurs when there is any update to the patient records.

III. PROPOSED SYSTEM

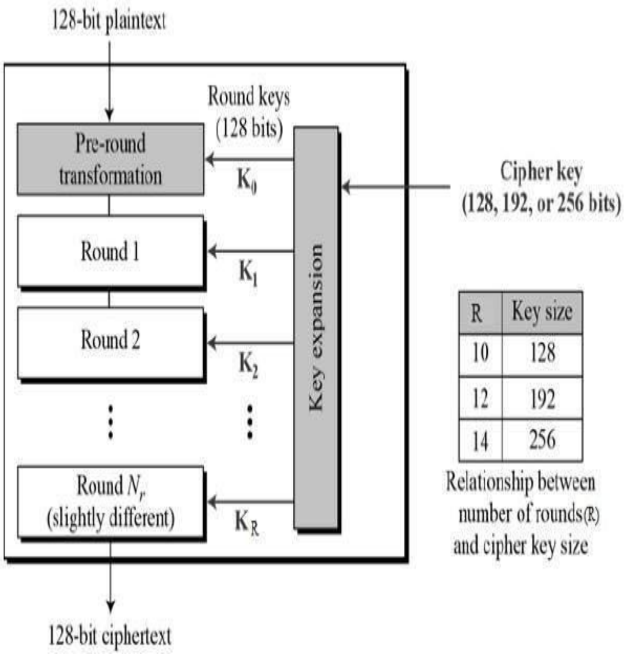
In the proposed system, we aim to apply AES encryption model to electronic health records that are stored in centralized cloud server. Attribute based encryption ensures the respective privileges of users. Anonymization of data is done to hide the sensitive information regarding patients, particularly their personal details. The main concern is about duplication of data which is overcome by sending a mail alert to all the users, when an updation is made.

IV. ARCHITECTURE



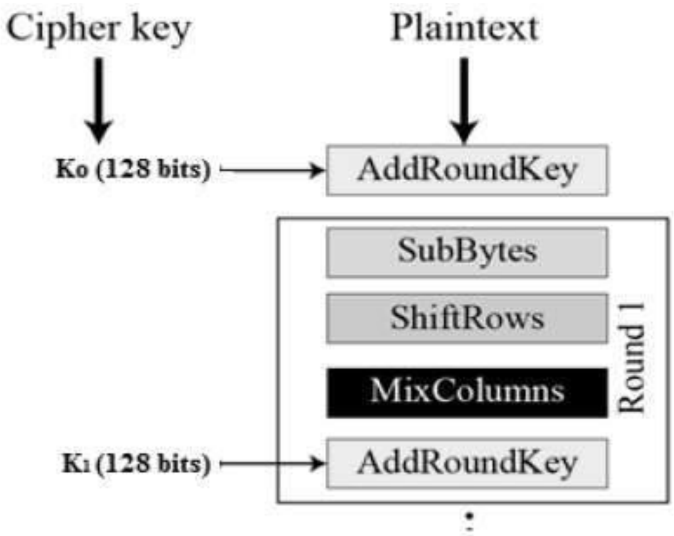
V. ALGORITHM

AES is a symmetric block cipher. That is same key for both encryption and decryption. The number of rounds in AES depends on the key length. For 10 rounds it uses 128bits whereas 12 and 14 rounds for 192 and 256 bits. In this proposed system we are going to use 128bits i.e. 10 rounds. AES Encryption algorithm is used because it resists against all known attacks and design simplicity. The algorithm begins with an Add round key stage which means 9 rounds of four stages and a tenth round of three stages. The four stages are Substitute bytes, Shift rows, Mix Columns, Add Round Key. The tenth round simply leaves out the Mix Columns stage.



Encryption Process

Each round consists of four sub-processes. The rounds are



A. Sub Bytes

The 16 input bytes are substituted in a fixed table i.e. S-box. The result will be in a matrix of four rows and four columns.

B. Shift rows

Each rows of the matrix is shifted to the left. Any entries that fall off are reinserted on the right side of row. First row is kept as same it is not shifted. Then second row is shifted one position to the left and then third row is shifted two positions to the left. Finally, fourth row is shifted three positions to the left; the result is a new matrix consists of 16 bytes.

C. Mix Columns

In this step each column of four bytes is transformed using a mathematical function. This operation takes input as four bytes of one column and outputs four new bytes, which replace the original column. The result is new matrix consisting of 16 new bytes. This step is not performed in the last round.

D. Add round key

The 16 bytes of the matrix are now considered as 128 bits and XOR operation is done to the 128 bits of key. If this is the last round then the output is the cipher text. Otherwise, the 128 bits are used as 16 bytes and we begin another similar round.

Decryption Process

The process of decryption in AES cipher text is similar to the encryption process in the reverse order. Each round consists of the four processes Add round key, Mix columns, Shift rows, Byte substitution.

VI. WORKING

This paper builds on earlier work of privacy preserving; it uses commutative encryption which is less secure. To overcome this we are going to use AES (Advanced Encryption Standards). This ensures that only the recognized and privileged users can query on the records. Each user will be provided with their related privilege to access records and result will be provided according to the request. The privilege will differ from one user to other and it is given by Abstract Based Encryption. Information about patient is divided into two categories, such as Personal information and Health records. This is encrypted and stored in cloud server. The stored information is future retrieved by various users when required. We proposed an approach for querying and privacy preserving for sensitive information.

A. Cloud server deployment

In this module main cloud server is deployed. Within this all access are maintained and monitored, this is a Main server where all details about service provider and user's information are stored. If any New Request comes from the user then the server will collect all request and process that request. Based on the Request it will redirect it to that particular service providers. The cloud server deployment acts as a database for whole architecture.

B. Hospital server deployment

In this module, hospital server is deployed, which includes all the hospital information about doctors, patient will be stored. In this integrated hospital information system, this addresses all the major functional areas of hospitals. This server is temporary for storage and later it is moved to main server i.e., cloud server deployment.Any access need can be done directly to the cloud server.

C. Mongo lab – user data separation

Mongo DB is an open-source database, and leading none SQL database. Mongo DB concepts needed to create and deploy a high scalable and performance oriented database. Mongo DB is a cross-platform document-oriented database. This module stores the information about user profile like user name, password, Email id, phone number, address and medical reports. Mongo DB supports field, range queries, regular expression. Queries will return results in fields or information.

D. Double encryption & data anonymization

In this module, we designed and implemented double encryption for patient personal information like name, disease etc. The intention of data anonymization is privacy protection. It is the process of removing personal information from data sets, so that the people whom the data describe remain anonymous. Double Encryption is provided by AES algorithm to provide security to sensitive information. Anonymization is used to give privilege for various users according to their query request. For example, if nurse request for patient disease information then only patient name, disease name, medical report only is shown because it is not necessary for nurse to know about patient personal information like address, Email-id, phone no etc., which are considered as sensitive information and privilege is provided.

E. Dynamic data transfer

If user wants medical details then just request to cloud server. At the time generate one OTP and it is sent to your mail id for verification process. After verifying the OTP the data is transferred or retrieved from cloud server. If there is any update in patient information it is done by hospital management and information about update is shown to users by sending a mail. This step is done to solve the problem of data duplication.

VII. CONCLUSION

In this paper, we propose an approach that involves advance encryption standard (AES) encryption involving analysis of data, thereby providing more data security and privacy in the cloud. All the users are given different privileges by attribute based encryption methodology. An email alert is given to all users, when an updation process is made. This firmly avoids the duplication of data or existence of multiple sets of data. These encryption techniques in the clod data storage ensure a secured retrieval and manipulation of sensitive patient's health records.

REFERENCES

[1] Rindfleisch TC. Privacy, information, technology, and health care. Communications of the ACM. 1997:93–100.

[2] Health Services Research and the HIPAA Privacy Rule. http://privacyruleandresearch.nih.gov/healthservicesprivacy.asp

[3] Kelman C, Bass A, Holman C. Research use of linked health data – a best practice protocol. Australian and New Zealand J. of Public Health. 2002;26:251–255.

[4] Agarwal R, Evmfimievski A, Srikant R. Information sharing across private databases. ACM SIGMOD. 2003. pp. 86–97.

[5] O'Keefe CM, Yung M, Gu L, Baxter R. Privacy – preserving data linkage protocols. ACM WPES. 2004:94–102.

[6] Vaidya J, Clifton C. Secure set intersection cardinality with application to association rule mining. J of Computer Security. 2005;13:593–622.

Dr.B.Anni Princy M.E Computer Science and Engineering, Ph.D. CSE-Software Engineering, currently working as professor in Panimalar Engineering College, Chennai

Tamil Selvi.G currently persuing bachelor degree in information technology in panimalar engineering college,Chennai.

Shruthakeerthi.S currently persuing bachelor degree in information technology in panimalar engineering college,Chennai.

Sowmya.B currently persuing bachelor degree in information technology in panimalar engineering college,Chennai.