ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Efficient Privacy Preservation Techniques for Maintaining HealthCare Records Using Big Data

S.Jegadeesan, S.Pooja, Vidhya.T

Assistant Professor/IT, Dept of Information Technology, Velammal Institute of Technology, Panchetti, Tamilnadu, India

B.Tech/IT, Dept of Information Technology, Velammal Institute of Technology, Panchetti, Tamilnadu, India

B.Tech/IT, Dept of Information Technology, Velammal Institute of Technology, Panchetti, Tamilnadu, India

ABSTRACT: Efficient privacy preservation technique for health records is have been directed to big data processing for its high volume ,velocity challenges.Two level encryption is performed where cloud provider maintains first level encryption and second level encryption is controlled by patient server.we are using doubleencryption level to maintain the privacy of the patient in which the records are maintained in two tables it makes the privacy even more enhanced. We are using two algorithms which are linear congruential generator and Triple DES Algorithm. Linear congruential generator is done in the first level encryption which provides the encryption key. For index level table we use same key encryption which makes searching very easy. The index table consist of the basic information about the patient such as name, ssn no, age (i.e it consist the user searchable keywords).Big Data are now rapidly expanding in all science and engineering domains, including biological and biomedical sciences. In this re-identification is performed in which if the information is same, it will be stored in the buffer and the re-identification page will be opened in which the unique identification will be verified. If it is not the same it will be showed after the authentication is performed. The authentication will be performed based on some criteria. Once the authentication is matched the, a privacy key will be generated only then it will be able to taken from the database and it will bedecrypted and will be shown to the requestor.

KEYWORDS: Linear Congruential Generator, Triple DES Algorithm, Two Level Encryption, Authentication, Security.

I.INTRODUCTION

The main aim of the system is to maintain the privacy of the patients records, since most of the records are being stolen by the third party or the unknown person. So it become a privacy issue. In this paper we are storing the documents of the patient the database by creating a client side web page.in such way the data are being stored in the database. While searching the data are stored in two level: index level and the privacy level. In the index level it contains the basic information about the patient. In the index level, we use same key for both encryption and decryption. It makes the patients record more secured. we have overcome the disadvantages of the de-identification by using the re-identification concept. It checks the unique identification of each patient. When the searching, an authentication is checked and it is based on certain criteria if it is matched it forwards the request to the patient server and it generates a private key and using that private key the records are decrypted and it is displayed to the insurance company or the hospital.[1] Efficient privacy preservation technique for health records is have been directed to big data processing for its high volume ,velocity challenges [2]HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective.[6]The concept of smart grid has emerged as a convergence of traditional power system engineering and information and communication technology. It is vital to the success of next generation of power grid, which is expected to be featuring reliable, efficient, flexible, clean, friendly and secure characteristics. In the existing systemCosine Similarity Check protocol is used for searching over the encrypted datas, where all the stored documents have to be decrypted before the start of the comparison process . All stored patient information must be encrypted using same public key. De-identification is a

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

process which maintained the secrecy of the data or records. De-identification is a process which displays the records of the user when the patients have the same name.

II.PROPOSED SYSTEM

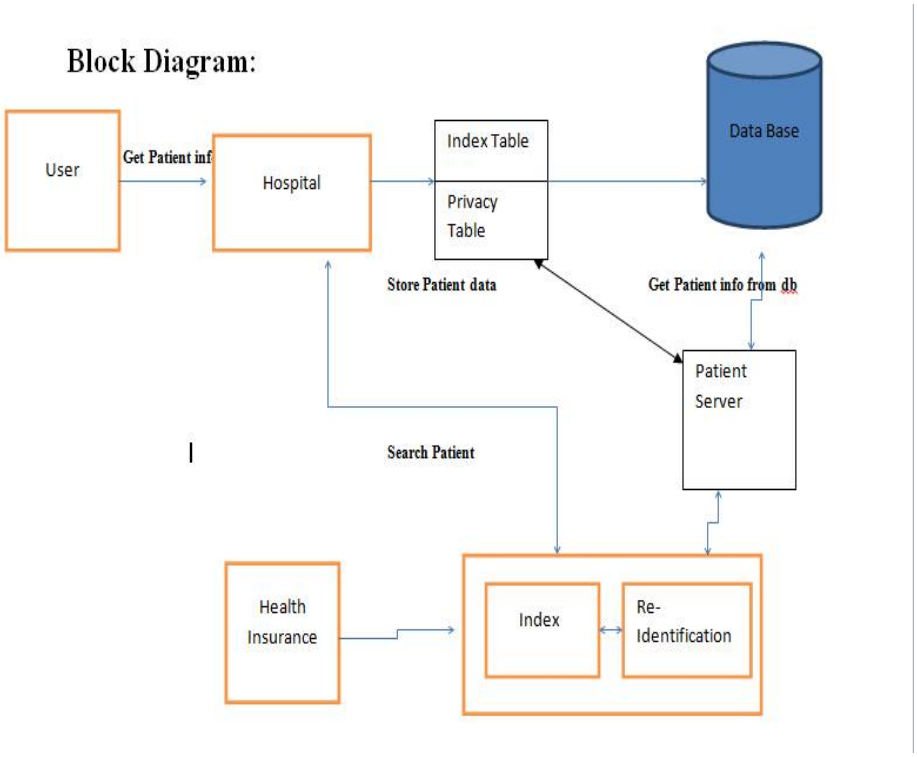
This paper addresses these issues by proposing privacy preserving using two level encryption. The data are stored in the data base, when searched it will be stored in two tables: index table and privacy table. Index table consist of user searchable keywords and the privacy table contains the most private information about the patient.We are using two algorithms which are linear congruential generator and Triple DES Algorithm. Linear congruential generator is done in the first level encryption which provides the encryption key. For index level table we use same key encryption which makes searching very easy. The index table consist of the basic information about the patient such as name, ssn no, age (i.e it consist the user searchable keywords).Big Data are now rapidly expanding in all science and engineering domains, including biological and biomedical sciences. In this re-identification is performed in which if the information is same, it will be stored in the buffer and the re-identification page will be opened in which the unique identification will be verified. If it is not the same it will be showed after the authentication is performed. The authentication will be performed based on some criteria.

Advantages:

1. Two level encryption
2. One encryption key is managed by top cloud server ,while another key managed by regionalserver
3. High secure than existing system

III.ARCHITECTURE DIAGRAM

The architecture diagram of the proposed system that described above is as follows:



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

V.SUBSYSTEMS

1.Client server Authentication
2. Key Generation
3.DatabaseEncryptionand improve de-identification

1. Client Server Authentication

A web page is created to actively interact with the user and to collect the details of the user. The user's details are entered in the webpage and it is stored into the database. While searching the details are stored in two levels: index level and privacy level. The index table consist of the primary or basic details about the patient (user searchable keywords), the privacy table consist of the private information of the patient.Once Server Authentication process completed, we design patient info web port to get all necessary information about patient health report. Each patient uploaded document must be stored in Own Encryption Key,which is known only to the user.

2. Key Generation

A unique encryption key is generated once the authentication is matched. The key is generated by the linear congruential generator which provides a random values for encryption. once it is matched**,** the request is forwarded to the patient server**.** The patient server is accessed by both the database and the insurance. It has an authentication check criteria which is the request will be accepted only based on some constraints(particular IP, hospital).

3.Database Encryption and improve De-identification

After the authentication is matched a privacy key will be generated. Only using this privacy key the database can be accessed else it cannot be accessed, so the database is completely authenticated. It decrypts the documents of the patient from the database. The documents will be shown only visible who is requestingthe information (i.e) either the hospital or insurance.This system we improve de-identification, when there are two or more patients have the same name then a re-identification page is opened where a unique identification key has to be entered. Then a private key will be generated when the authentication is matched. The key will be sent to the database using that key the documents will be decrypted and it will be downloaded and shown to the one who is requesting the document. Only if the details are correct the insurance company will pay the amount to the hospital where the patient has admitted.

- Therefore, to mitigate the threats from re-identification, the concepts of *k*-anonymity, *l*-diversity, and *t*-closeness have been introduced to enhance traditional privacy-preserving data mining.
- Obviously, de-identification is a crucial tool in privacy protection, and can be migrated to privacy- preserving big data analytics.
- However, as an attacker can possibly get more external information assistance for de-identification in the big data era.
- we have to be aware that big data can also increase the risk of re-identification. As a result, de-identification is not sufficient for protecting big data privacy.

End-user and regional server configuration:

In this module we develop end-user interface page (client side web page) and host regional server, And also establish network connection between them.

Regional server means it is intermediate server between end-user and top level cloud-server, which is used for implement first and second level encryption Eg: end-user submitted data first process by regional server for implement first and second level encryption, then post to top level cloud server implement second level encryption , eventually that data store cloud repository.

ALGORITHM DESCRIPTION:

Linear congruential generator:

An LCG is essentially a formula of the following form:

number = (a * number + c) mod m

In other words, we begin with some start or "seed" number which ideally is "genuinely unpredictable", and which in practice is "unpredictable enough". For example, the number of milliseconds— or even nanoseconds— since the computer was switched on is available on most systems. Then, each time we want a random number, we multiply the current seed by some fixed number, *a*, add another fixed number, *c*, then take the result modulo another fixed number m. The number *a* is generally large.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

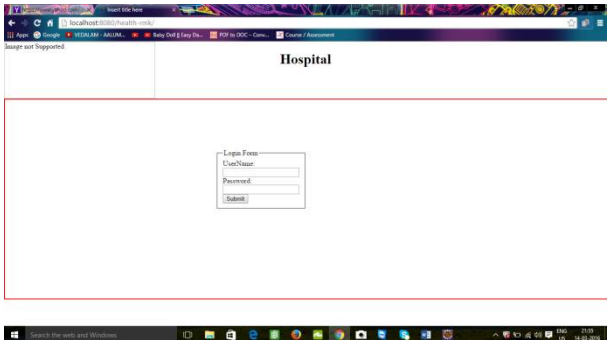
Triple DES Algorithm :
Triple DES was designed to replace the original Data Encryption Standard (DES) algorithm, which hackers eventually learned to defeat with relative ease. At one time, Triple DES was the recommended standard and the most widely used symmetric algorithm in the industry.Triple DES uses three individual keys with 56 bits each. The total key length adds up to 168 bits, but experts would argue that 112-bits in key strength is more like itDespite slowly being phased out, Triple DES still manages to make a dependable hardware encryption solution for financial services and other industries.

Advantages:

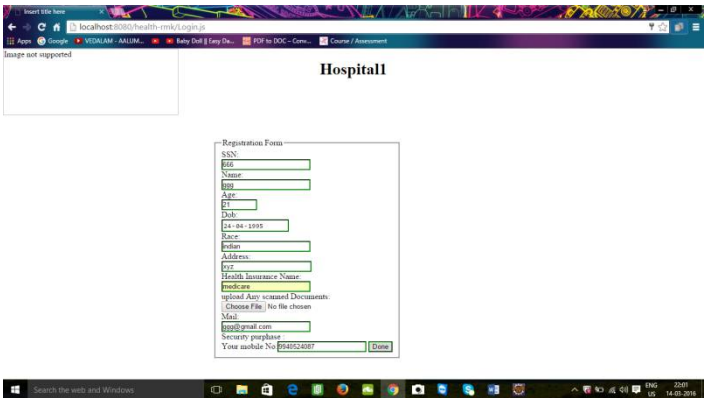
- **Re-identification provides more security to the database**
- **Only when the authentication is matched the access to the database will be provided.**
- **Searching is made more easy.**

VI.SIMULATION AND RESULT

1. **In this system a login in page created where the user(i.e the patient)will be logged in and the details will be collected from the user**



2. **In this system once the user(i.e the patient) logs in then the details will be collected and will be stored into the database**



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

3. Once the details are entered and it will be stored into the database in an encrypted format.

Encrypted Data are Stored Successfully

VII.CONCLUSION

In this article, we have investigated the privacy challenges in thebig data era by first identifying big data privacy requirements andthen discussing whether existing privacy-preserving techniquesare sufficient for big data processing. We have also introducedan efficient and privacy-preserving cosine similarity computingprotocol in response to the efficiency and privacy requirements of data mining in the big data era. Although we have analyzedthe privacy and efficiency challenges in general big data analyticsto shed light on the privacy research in big data, significantresearch efforts should be further put into addressing unique privacyissues in some specific big data analytics.

REFERENCES

[1] IBM, "Big Data at the Speed of Business," http://www-01.ibm.com/software/data/bigdata/, 2012.
[2] X. Wu et al. , "Data Mining with Big Data," IEEE Trans. Knowledge Data Eng. , vol. 26, no. 1, 2014, pp. 97–107.
[3] S. Liu, "Exploring the Future of Computing," IT Professional , vol. 15, no. 1, 2013, pp. 2–3.
[4] Oracle, "Oracle Big Data for the Enterprise," http://www.oracle.com/caen/technologies/big-data, 2012.
[5]"Big Data at CSAIL," http://bigdata.csail.mit.edu/.
[6] R. Lu et al. , "EPPA: An Efficient and Privacy-Preserving Aggregation Scheme for Secure Smart Grid Communications," IEEE Trans. Parallel Distributed. Sys. , vol. 23, no. 9, 2012, pp. 1621–31.
[7] P. Pailier, "Public-Key Cryptosystems based on Degree Residousity Classes," EUROCRYPT, 1999, pp. 223–38.
[8] M. Li et al. , "Toward Privacy-Assured and Searchable Cloud Data Storage Services," IEEE Network , vol. 27, no. 4, 2013, pp. 1–10.
[9] A. Cavoukian and J. Jonas, "Privacy by Design in the Age of Big Data," Office of the Information and Privacy Commissioner, 2012.
[10] R. Lu, X. Lin, and X. Shen, "SPOC: A Secure and Privacy-Preserving Opportunistic Computing Framework for Mobile-Healthcare Emergency," IEEE Trans. Parallel Distributed. Sys. , vol. 24, no. 3, 2013, pp. 614–24.