International Journal of Advanced Research Methodology in Engineering & Technology,
Special Issue, March 2017, ISBN 978-1-63535-324-2

Enhancing Data Security with System Performance and Fault Tolerant in Big Data

Aishwarya J [1], Ashwini N [2]
Student Final Year[+]
[+]Department of Computer Science Engineering,
[+]Jeppiaar SRR Engineering College,
Chennai, India.
aishwaryajags@gmail.com [1], ashwiniachu0108@gmail.com [2]

Rajasekaran G [3]
Assistant Professor[+]
[+]Department of Computer Science Engineering,
[+]Jeppiaar SRR Engineering College,
Chennai, India.
call2rajasekar@gmail.com [3]

***Abstract*---Big Data consists of extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions. The new technologies and their data generation at substantial rate gave birth to the Big Data and a robust platform is required to capture, retrieve, store, and process the data. . Data generated by Health care services and applications such as sensors, human centric applications, social networks, and smart-phones need to be collected and processed to provide in-depth knowledge for advancement in the field. In this paper we proposed secure re-encryption technique , it perform two level encryption before actual data store in remote server , for using this technique need two server one local server another remote server. Here we are using big data concept, so that, all patient data must be split into multipart then encrypt each part under privacy key. Main goal of the project is fast and securely store and retrieve data from big data for that we used storage level cache, that store uploaded data temporarily until move to actual location, from this way we can avoid data loss and also obtain fast response. We proposed two Encryption key index level, privacy level. Here we are also implementing distributed servers to avoid fault tolerance.**

***Keywords*—Big data, two level encryption, Cache servers, Distributed servers.**

I. INTRODUCTION

Big Data is a cluster of large datasets that cannot be processed by using traditional computing techniques. Big data is not merely a data; rather it has become a complete subject, which involves various tools, techniques and frameworks. Big data involves the data produced by different devices and applications. Big data technologies are important in providing more accurate analysis, which may lead to more concrete decision-making resulting in greater operational efficiencies, cost reductions, and reduced risks for the business. Challenges include capture, storage, analysis, data accuracy, search, transfer, visualization, querying updating and information privacy. The term "big data" often refers simply to the use of predictive analytics, user behavior analytics or certain other advanced data analytics methods that extract value from data and seldom to a particular size of dataset.

II. EXISTING SYSTEM

In existing system they used big data for maintaining data's but they have some drawbacks such as data losses or low of system performance may occur in case of multiple request hitting simultaneously. Most of the industrial data's are stored in cloud, but cannot predict all stored data's are secured. Even, more encryption algorithms are invented, sensitive information can be leaked if that one key by which the data is encrypted is hacked, so less secure.

III. ISSUES IN EXISTING SYSTEMS

1. Only Single level encryption is used.
2. Most of the encryption keys are managed by cloud providers, so providers may break all information**.**
3. No storage level caches are used to avoid data losses.
4. There are no distributer servers to avoid fault tolerance.

IV. PROPOSED SYSTEM

In Proposed System, "Two level Encryption" concept is introduced. Patient information is divided into two category, Index information and Privacy information. Unique secret key is maintained for each patient.
Storage level caches are used to obtain fast response. Distributed servers are introduced to avoid fault tolerance.

V. SYSTEM DESIGN

In System Design , the architecture explains about the novel concept "Two level Encryption" , all stored patient information have two category, one for search index another privacy table. Search index contain only searchable keywords, so that encryption keys are common to all patient .Privacy table are maintained by network admin that contain unique encryption keys for all patient. Those keys only provide authorized request, that means patient can set instruction for access , instruction can be of any type such as IP address or Unique id. Main goal of the project is fast and securely store and retrieve data from big data for that we used storage level cache, that store uploaded data temporarily until move to actual location, from this way we can avoid data loss and also obtain fast response.

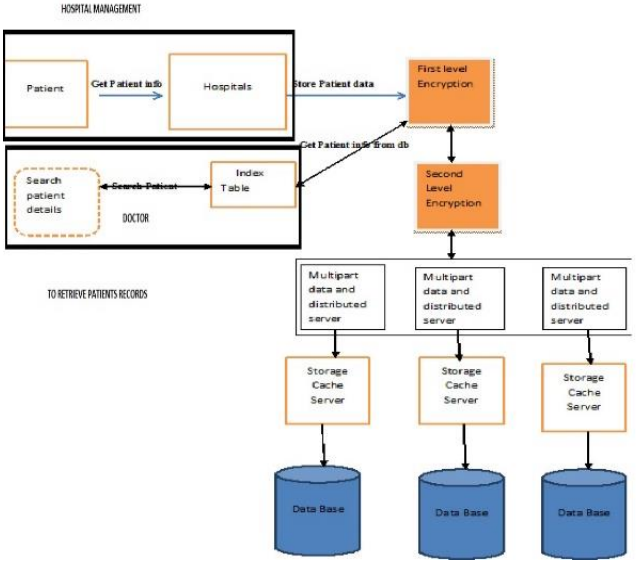

Fig1. Architecture diagram

V. ADVANTAGES OF PROPOSED SYSTEM

1. Two level encryption
2. Second Key are maintained by local server.
3. Highly secured than existing system
4. Fault tolerant.

VI. MODULES

A. *DATA CLASSIFICATION:*

In this module we develop interface page : In this project wet have two types of client, Hospital and patient to access patient data's. For apply security service we implement two level encryption concept. Patient data's are classified into two types' searchable keyword and privacy information. This system maintains two type of patient information; in-patient and out-patient.All Request must be Authenticated before go to patient database.

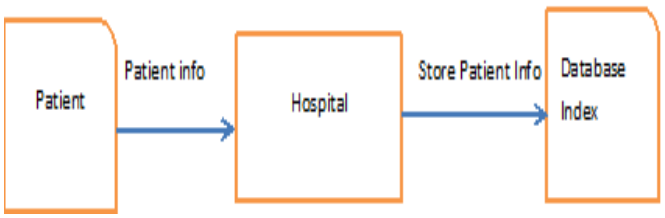

Fig2. Data classification

B. *ENCRYPTION USING LINEAR CONGRUENTIAL GENERATOR:*

In this module we implement first level encryption, once the patient registration process is completed, that data's are classified into two type: searchable keyword's and Privacy data's. Both these data's are encrypted using an algorithm named ,**linear congruential generator.** First level encryption key are maintained by cloud service provider for searching specific user.

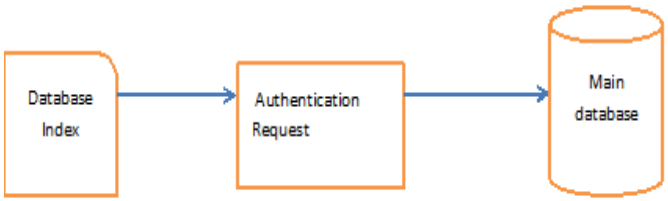

Fig3. Encryption using LCG

C. *ENCRYPTION USING DATA ENCRYPTION STANDARD:*

Second level encryption keys are generated when user submit their registration form, that keys are maintained securely in network admin system. That system have one frontend layer (security layer) .using this layer user can set criteria for accessing key. Uploaded data split into multipart data then encrypt each part using privacy key , finally store in Database

server. For second level encryption we are using an algorithm named , Data Encryption Standard.

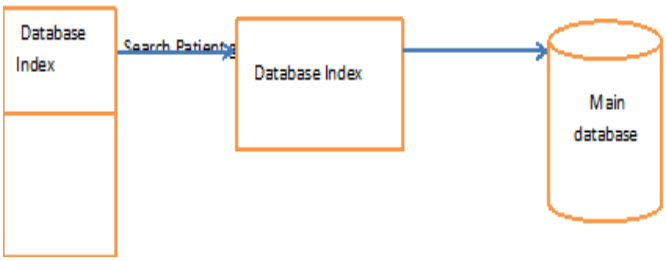

Fig4. Encryption using DES

D. CACHE AND DISTRIBUTED SERVERS:

In this module we implement two level of servers, cache server and distributer servers, cache servers are intermediate server placed between master and slave , job of this servers are storing incoming data's until receive from slave server.
Distributer servers are master servers which are located on different places; through these servers we avoid fault tolerance.

VII. ALGORITHM

Linear congruential generator (**LCG**) is an algorithm that extracts a sequence of pseudo-randomized numbers calculated with discontinuous piecewise linear equations. The method represents one of the best-known pseudo random no generator algorithms. Relatively easy to understand, the theory behind them and they're easily implemented and fast, especially on hardware which can provide modulo arithmetic by storage-bit truncation.

$$R_{n+1} = (aR_n+b) \bmod n$$

Where n can be in multiples of 2 and if a 32-bit random number is required, it can be n can be 2^32.
R_n and R_{n+1} are the current and next pseudo random numbers in the pseudo random number sequence.
A is the multiplier and *b* is the increment.
For generation of proper random number sequence, the *n* and *b* should be relatively prime. Further, *a-1* should be divisible by all prime factors of *n*. *a-1* should be divisible by 4 if *m* is divisible by 4.
Data Encryption Standard (DES) is a block-cipher, means cryptographic key &algorithm are applied to a block simultaneously rather than one bit of data at a time. Encryption of a plaintext, DES groups the plaintext into 64bit blocks. Each block is encrypted using the secret key into 64-bit encrypted text by means of permutation & substitution. This process involves 16 rounds and can run in 4 different modes, encrypt the blocks individually or making each encrypted block dependent on all the previous blocks. Decryption is the reverse of encryption, following the same ,reversing the order to which the keys are applied. For any encrypt, the most basic method of attack is brute-force, which means trying each key upto you find the right key. The length of the right key ascertain the no of possible keys -- and so the feasibility -- of this attack. DES use 64-bit key, but 8 of these bits are used for parity checks, actually limiting the key to 56-bits. It would take a maximum of 2 power 56, or 72,057,594,037,927,936, attempts to identify the correct key.

VIII. FUTURE IMPLEMENTATION

The project could be undertaken and can be implemented into future use.
Where we can use finger print sensors, so that the patient's finger print is taken as the searchable keyword to retrieve the privacy information from the database.

IX. CONCLUSION

Above article, we have concluded the challenges in Cloud computing by first identifying data privacy requirements and then discussing whether existing privacy-preserving techniques are sufficient for data processing. We have also introduced an efficient and privacy-preserving Two-level encryption in response to the efficiency and privacy requirements of data mining in the cloud.

REFERENCES

[1] D McGraw ,"Why the HIPAA privacy rules would not adequately protect personal health records: Center for Democracy and Technology (CDT) brief," 2008. [Online]. Available: http://www.cdt.org/brief/why-hipaa-privacyrules would-not-adequately-protect-personal-health-records.[Accessed: 20-Sep-2015].

[2] F. Chen, N. Mohammed, S. Wang, W. He, S.Cheng, and X. Jiang, "Cloud-Assisted Distributed Private Data Sharing," in The ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, 2015.

[3] K. Benitez and B. Malin, "Evaluating re-identification risks with respect to the HIPAA privacy rule," J. Am. Med.Informatics Assoc., vol. 17, no. 2, pp. 169–177, 2010.

[4] S. J. Nass, L. A. Levit, and L. O. Gostin, Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health ThroughResearch. The National Academies Press, 2009.

[5] F. McSherry and K. Talwar, “Mechanism Design via Differential Privacy,” in 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS’07), 2007, pp. 94–103.

[6] C. Dwork , “Differential Privacy” , Int. Colloq. Autom. Lang. Program. , vol.4052 , no. d , pp. 1-12 , 2006.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith “Calibrating noise to sensitivity in private data analysis,” Theory Cryptogr., vol. 3876, no. 1, pp. 265–284, 2006.