International Journal of Engineering &Technology, 7 (4) (2018) 6211-6215

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.23127
Research paper

Security for personal health records occupying terabytes using cryptography and encoder decoder

Sathya Sankaran [1]*, Rajkumar Rajasekaran[1]

[1]School of Computer Science and EngineeringVellore Institute of Technology, Vellore-632014
**Corresponding author E-mail:litcomsci@gmail.com*

Abstract

Personal Health Records may contain personal data along with medical data that is available in Electronic Health Records. This research proposes an encryption decryption method with encoder decoder for records in well established hospitals that store medical records in terms of terabytes. For encryption, the medical information including images is converted into binary. This is sent to an encoder and the information is stored in chunks to form a matrix. Then it undergoes matrix transformation and split into sub matrices. This is sent to the destination. The reverse is done in the decryption module. Cryptanalysis for the algorithm used and brute force attack is almost impossible or takes a number of years. A comparative study of the proposed algorithm is done with DES, AES, IDEA and BLOWFISH and it is realized the proposed algorithm is immune to cryptanalysis methods to which the compared algorithms are not. The personal health records are located in the cloud. The encryption decryption provides data security which is used to provide cloud security.

Keywords:*Encoder; Cloud Security; Cryptology; Decoder; Personal Health Records.*

1. Introduction

The cloud offers services such as Infrastructure as a Service, Platform as a Service, Software as a Service and Network as a Service. Infrastructure offered includes hardware, software, storage, servers etc. Platform as a Service offers a platform where it is possible to develop, run and manage applications without the user having to possess the infrastructure required. Software as a Service offers software to customers. Network as a service is one in which a network operator offers a network and its services to third parties. It also extends a private network over a public network like the Internet. It also offers bandwidth between two nodes or users.

In spite of all the above services provided at a low cost, the cloud is not relied upon to a great extent. This is because of the lack of security for the data or information stored or used in the cloud. Hence it is required to come up with security solutions to rectify the same.

This research discusses the encryption and decryption of personal health records to store health information about the patients. Data stored would include personal details of patients like patient's name, age, address,weight, blood pressure, pulse rate, sugar level, albumin, hemoglobin; medication prescribed including diagnosis and treatment information. The aim is to provide an encryption decryption algorithm to protect medical data. It will give data security as part of cloud security. The data to be stored in encrypted form is subjected to encryption and then stored on the cloud. For decryption the data is downloaded from the cloud and then decrypted. This method of achieving security is better than implementing security directly on the cloud, which may not be fully secure. Encryption and Decryption is implemented as shown in Figure 1. The original data which is the plaintext is converted into scrambled data by encryption with an encryption key and the same encryption key is used for decryption to obtain the original text at the destination.

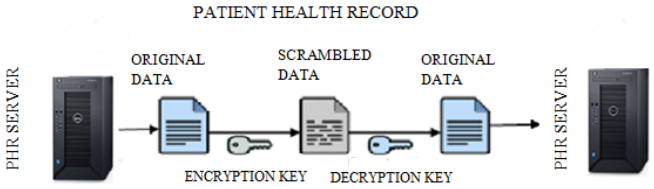

Fig. 1: Encryption-Decryption [1 - 3].

In this research work, the medical data of personal health records is converted into binary. The binary data is split into groups of size 2^{40}. Each bit is fed to an encoder. Depending on the position of the bit, a corresponding value of 40 bits is obtained at the output of the encoder. Chunks of 40 x 2^{40} bits obtained as above are stored as a matrix. This matrix is subjected to matrix transformation. Following this the matrix chunks of 40 x 2^{40} bits are split into sub matrices of various dimensions. The key is the indices to the sub matrices split and sent across to the destination. The destination does the opposite of the encryption in reverse order of encryption.

2. Literature review

Sharing important images on the network may lead to loss of the image Dawahdeh, Z. E. et al. [3]. Hence Hill Cipher method may be used to secure such images, but it has weak security. A combination of Elliptic Curve Cryptosystem and Hill Cipher is done to strengthen the Hill Cipher by making it asymmetric from symmetric. It makes Hill Cipher efficient and resistant to hackers.

Bensikaddour, E. H. et al. [4] gives method to secure multispectral satellite images. Chaos based confidentiality has been used for the past two decades. In this paper Fridrich scheme has been used to

ensure security by using confusion and diffusion. The experiment proves that the method used is efficient, with low power and hardware complexity. It gives a throughput of 120 Mbits/sec.

Karmel.A. et al. [5] suggest the following. Health care has a lot of information gathering frameworks, such as medical records, health reviews, managerial records and also wearable gadgets. This information is accumulated by various entities. The information contains the medical history of the patient and personal information. De-identification is the process employed to prevent a person's identity from being connected with information. For this cryptographic techniques are used.

Sherry-Ann Brown et al., [6] suggest that including genetic risk information in electronic health records (EHRs) will facilitate implementation of genomic medicine in clinics. This work studies patient's attitudes towards inclusion of genetic risk information as a part of personal health information in EHRs. The study resulted that the patients who participated showed positive attitudes towards incorporating genetic risk information in EHR.

Clemens Scott Kruse et al., [7] discuss the following. Privacy and security of patients' information is a big barrier when trying to store electronic health records. This work takes into account legal regulations, and discusses adoption of prominent security techniques for ensuring a secure electronic health record system. A study was done on Pub Med (MEDLINE), CINAHL, and ProQuest Nursing and Allied Health Source. 25 journals were collected and analyzed. The security measures mentioned were categorized into three themes viz., administrative, physical and technical. Advanced security measures are taken for sensitive nature of information.

Daniel M Walker, et.al. [8] propose that while electronic health records (EHRs) become a practice in health care, privacy breaches are increasing and being made public. These breaches may consumers withdraw from technology, thereby being a hindrance to its potential to improve care coordination and research. On study, no difference was found on the effect of privacy and security concerns on withholding behaviours between 2011 and 2014. In 2011 and 2014, withholding behavior was found to have an impact on high quality of care, but not between 2011 and 2014.

Roger Clarke, [9] in this work gives a definition of Privacy, Information Privacy and Dataveillance. Privacy is the interest that individuals have in keeping a 'personal space' free from interference by other people and organizations. Information Privacy is the interest a person has in controlling, or at least significantly influencing, the handling of data about themselves. Data Surveillance or Dataveillance is the systematic use of personal data systems in the investigation or monitoring of the actions or communications of one or more persons.

Vincent Lozupone [10] discusses the systematic procedures that should be available to recover data in case of disasters such as fires, hurricanes, other natural disasters, sabotages or security incidents. This paper discusses disaster recovery and recovery of data in medical record company.

Sujin Kim et al., [11] discuss the certification that should be given by the patients for their medical records when they are shared by medical personnel. An infringement of medical records could become a big issue. In the background of the above, FIDO based authentication system such as UAF and U2F are applied to the authority and work scope of medical personnel and medical support assistants.

Kotz, D. et al., [12] discuss a wearable master electronic device for secure control of physiological sensors and medical devices along with a secure storage of medical records viz., Amulet is discussed. This ensures safety of patients by the use of physiological sensors. Also, the medical records are stored securely.

Yüksel, B. et al., [13] present the following. With the current information and communication technology, electronic health services are used to store personal medical details among doctors, staff and other medical personnel. This helps in reducing the cost while providing efficient healthcare. However, this leads to security, integrity and privacy issues. This paper is a study of works that deal with solutions given for such security, privacy and integrity issues.

Walters, B. H. [14] narrates the growing health care delivery in Netherlands by having a networked electronic medical record. Sharing care for chronic complaints takes place through this network among general practitioners, dieticians and internists. Securing the medical research was done by watching the records by project leaders, clinicians and also patients. The watching was in the form of coveillance, self-surveillance, dataveillance etc.

3. MEDI-SENSI algorithm for personal health records

MEDI-SENSI is an encryption decryption method with encoder decoder.

The following is the encryption decryption algorithm to protect data in the personal health records. The personal health records in various formats such as integers, strings, images, etc. are converted into binary values and the stream of binary values are encrypted at the source and decrypted at the destination as explained below.Encryption is illustrated in Figure 2 and Decryption is illustrated in Figure 3.

3.1.Encryption of medical data (operations to be done at the source)

1) Medical data such as numbers, strings and images are converted into binary data.
2) Divide the data to be sent from source to destination into groups of 2^{40} bits.
3) Input each bit in its position as 0 or 1 and the rest of the bits as 0 from above to a 2^{40} x 40 encoder.
4) Construct a 40 x terabit matrix of bit sequences obtained as above in each row of the matrix.
5) In the next step XOR the bits in the matrix with all 1s.
6) Divide the matrix into sub matrices of dimension 1024x1024. In each such sub matrix
    a) Exchange row1 and row2.
    b) Take a transpose.
    c) Repeat steps a and b for all pairs of rows.
7)

a) Split the matrix into smaller matrices of varying dimensions and add some false additional bits to the data such as integers, strings, images etc. These false bits add to the actual bits on transit.
b) The keys are generated as the index of the sub matrices split. Generating the keys takes time depending on how the matrix is split and is assumed to done on a desktop with speed of 1GHz. Any one combination is used for encryption which may take minutes/hours/days. The maximum time taken for the minutest split of matrix, i.e., splitting the 40 terabit matrix into single dimension matrices takes 27.7 days. Generating all possible combinations of the keys (index of split matrices) takes several years.For encryption generating any one combination is sufficient.The keys generated are to be remembered.
c) After generating the keys, the split sub matrices are sent out of order randomly. This sequence is to be remembered.

3.2. Decryption (operations to be done at the destination)

1) Remove the additional bits that were added and join the sub matrices that were split during encryption using the order in which they were sent and the position/index of the smaller split matrix which was remembered.
2) The result is a 40 terabit matrix.
3) Divide the matrix into sub matrices of dimension 1024x1024. In each such submatrix
    a) Exchange col1 and col2
    b) Transpose the matrix

c) Repeat steps a and b for all pairs of columns such col3 and col4 and so on till the last two columns.

4) XOR the bits of the resultant matrix with all 1s.

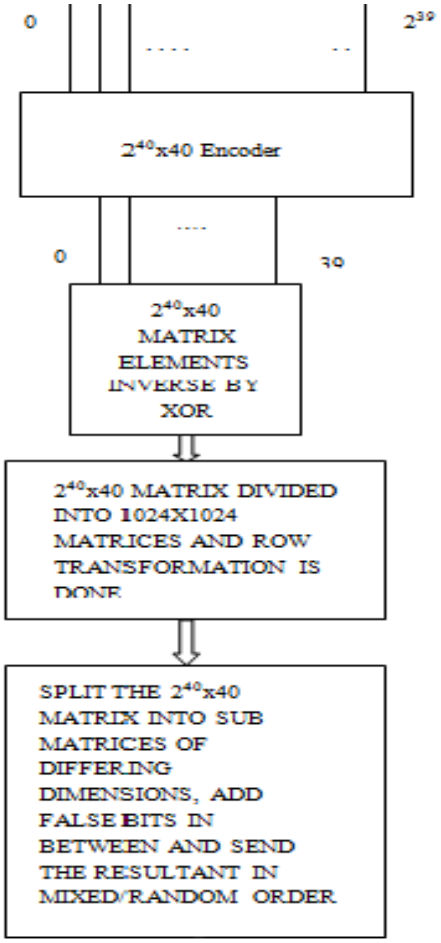

Fig. 2: Encryption.

5) Using a decoder, decode the bits which is 40 bits to 240 bits.

6) The decoder produces binary data that can be structured into strings, numbers and images.

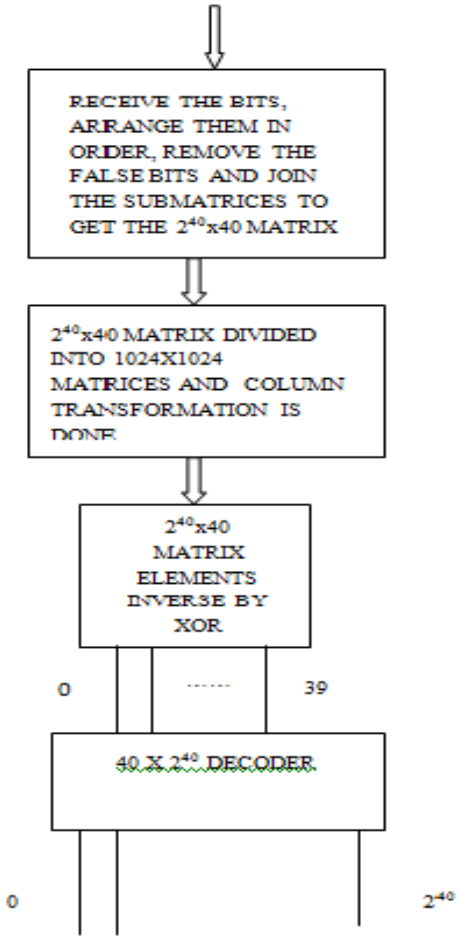

Fig. 3: Decryption.

4. Cryptanalysis of medi-sensi

Brute force attack will try to generate all possible keys. But this takes several years.And MEDI-SENSI algorithm adds false bits along with the sub-matrices that are split from the original matrix. False bits help in concealing the actual data. The false bits should mislead the data so that they give misleading data that avoid obtaining the real data. Also there is an adoption of order in which the sub matrices are sent. (For example, where n is 40 terabit, all combinations of n-1 2 bit sub matrices, n-4 3 bit sub matrices, n-8 4 bit sub matrices and n-9 1 bit sub matrices takes111.11 days and all combinations of n-1 2 bit sub matrices, n-4 3 bit sub matrices, n-9 4 bit sub matrices n-14 5 bit sub matrices and n-15 1 bit sub matrices take 138.88 days. So generating all types of combinations of keys would take several years.)

Chosen plaintext attack involves giving access to the encryption module to the adversary temporarily. By giving several plaintext to the encryption module, the corresponding ciphertext is obtained. One can obtain some k number of <p, c> pairs. Using this one has to try to get the key or the current plaintext that is sent from the sender to the receiver.It is neither possible to get the key nor the current plaintext, as the encryption scheme suggested, is to split matrices organized from plaintext of varying sizes and they are sent in a random order. It is not possible to predict this order and size of the split matrix or the key because of such an encryption scheme.

Chosenciphertext attack involves giving access to the decryption module to the adversary temporarily. By giving several ciphertext to the decryption module, the corresponding plaintext is obtained. From this some <p,c> pairs are obtained. Using this, it is required to find the key or the current plaintext transmitted to the receiver. The decryption module works by using the key to get the order in which the split matrices are to be read. Neither the key nor the current plaintext can be obtained from this.

Adaptive chosen plaintext attack involves using the encryption module temporarily to get some plaintext ciphertext pairs and then choosing a plaintext based on the already given plaintext and using this to guess the key or the current plaintext transmitted to the receiver. In the suggested encryption module, it is neither possible to predict the key nor the current plaintext as it not possible to work out a relation between the plaintext nor the cipher text.

Adaptive chosen cipher text attack involves using the decryption module temporarily to getsome plaintext ciphertext pairs. Every next cipher text is chosen such that it has a relation with the previous ciphertext used.From these pairs of plaintext ciphertext pairs predicting the key or the currentplaintext transmitted to the receiver is done.In the proposed encryption decryption module, it is not possible to find a relation between one ciphertext and another.

Differential attack works by taking difference of any arbitrary plaintext pairs, and obtaining the corresponding difference between any cipher text pairs, followed by working out the key for a round. Here, the keys are the indices of the sub matrices and are distributed over the matrix. So, keys are obtained for each sub matrix. So differential attack is difficult and takes hundreds of days or several years.

Boomerang attack works by dividing the cipher E to component ciphers E0 and E1. E0 stands for the cipher of the first two stages and E1 stands for the cipher of the third stage. The attack is a dif-

ferential attack on the third stage, which takes hundreds of days or several years.
In Integral cryptanalysis, set or multiset of plaintext that have an XOR sum of 0 is chosen. Even if such a plaintext is chosen, it is not possible to find the corresponding ciphertext as the ciphertext is split over different submatrices.
Linear cryptanalysis works by obtaining linear equations of plaintext, keys and cipher text. It is a known plaintext attack.It works by finding the bias. Even if linear relationships exist among the plaintext and ciphertext, solving these keys would lead to taking time equivalent to hundreds of days or years.
Related Key attack works by finding relationship between keys that are assumed and observing the cipher, where some mathematical relation connecting the keys are known to the attacker. Such mathematical relation between the keys does not exist.
XSL cannot be used to cryptanalyse MEDI SENSI as it is not possible to deduce quadratic relations or monomials which is required to implement XSL.
Key recovery attack looks for all the keys, and this is similar to brute force attack explained above.
Side channel attack can be misled by including false bits that creates signals, along with the split matrices.
Meet-in-the-middle attack is not possible as decryption of one data is not the encryption of the other one. Even if working out the decryption of one stage becomes the encryption of the other stage, the split up of matrix ultimately into sub matrices cannot be obtained as part of cryptanalysis.
Davies' attack is a dedicated attack for DES.
The known-key distinguishing attack works by using a key to cryptanalyse. In this model one has to work all possible combinations of the sub matrices to get the keys. Usually the key is not known to the attacker.
A chosen key distinguishing attack works in the same way except that the key is chosen.
Birthday attack aims at finding two functions which give the same value for two different inputs. Such a function cannot be found in the model suggested.
Differential Linear attack works by applying differential attack to a few of the stages followed by linear cryptanalysis. Whatever may be the order and choice of the stages for differential and linear cryptanalysis, the last stage poses a challenge for the attack on it.

5. Comparative study of encryption-decryption algorithms

Table1: Comparison of Encryption Decryption Algorithms

Algorithm → Parameter	DES	AES	BLOWFISH	IDEA	MEDI-SENSI
Cryptanalysis	Susceptible to differential cryptanalysis, linear cryptanalysis and Davies' attack theoretical methods. Attack by brute force is the most basic method possible. Susceptible of Boomerang attack and Differential linear attack.	Susceptible to known-key distinguishing attack, key recovery attacks, side channel attacks. Brute force attack is possible.	Affected by birthday attack. Vulnerable to known-plaintexts on weak keys.	Broken by meet-in-the-middle attack. Weak keys with large number of 0s produce weak encryption.	Immune to Brute force attack, Chosen plaintext attack, Chosen ciphertext attack, Adaptive chosen plaintext attack, Adaptive chosen ciphertext attack, Differential attack, Boomerang attack, Integral attack, Linear attack, Related key attack, XSL attack, Key recovery attack, Side channel attack, Meet in the middle attack, Davies attack, known key distinguishing attack , chosen key distinguishing attack, birthday attack, differential linear attack.
Structure of the algorithm	Feistel structure	Iterative rather than feistel structure	Similar to DES	Uses rounds with subkeys	Sequential Structure
Type of operation	Uses permu-tation, subs-titution and key mixing	Uses substitution and permutation	Almost similar to DES	Modular addition, multi-plication and XOR operations	Uses encoding, decoding, XOR and matrix transformation .
Algorithm → Parameter |V	DES	AES	BLOWFISH	IDEA	MEDI-SENSI
No .of bits in plaintext, ciphertext	64 bit plaintext64 bit cipher text	128 bit plaintext, 128 bit cipher text	64 bit plain text, 64 bit ciphertext	64 bit plaintext, 64 bit ciphertext	2^{40} bit groups plaintertext, 40 terabit bits + false bits in ciphertext
Key size and rounds	56 bit key, 16 rounds	128 bit key for 10 rounds,192 for 12 rounds, 256 for 14 rounds	Key size between 32 and 448, 16 rounds	128 bits key size, 8.5 rounds	Keys depending on the number of split matrices, lack of concept called round
Works on no. of bits	Each round works on split up 32 bits	128 bits as-16 bytes as- 4x4 matrix	Works on data having size multiple of 8, if not padded	64 bits are divided into 4 parts of 16 bits each.	Works on different number of bits at different levels
Operations done	At each stage one half goes to the next stage while the other goes thro' permutation, substitution and key mixing	Does Sub bytes, Shift Rows, Mix Columns, Add Round Key	64 bits divided to 32, one half XORed with P array, transformed with F function and XORed with other half.	Modular, addition, multiplication, XOR applied to 16 bit sub blocks. Each round has diff keys. Totally 52 keys are generated	Includes encoding, XOR, matrix transformations including exchange of rows, transpose and exchange of columns.

6. Conclusion

The MEDI-SENSI algorithm discussed above can be used to protect data in personal health records in the order of terabytes. Particularly, the brute force on the proposed algorithm attack takes several years. Cryptanalysis that works on other algorithms, i.e., DES, AES, IDEA, BLOWFISH, does not work on MEDI-SENSI algorithm. The MEDI-SENSI algorithm is a useful option for personal health records in hospitals that are well established and those that are developing. It helps when the personal health records are to be shared on the cloud from one hospital to another hospital, from one doctor to another doctor, from one state to another across the globe in the cloud.

References

[1] Data Server Implementation. https://www.indiamart.com/proddetail/data-server-implementation-8993920612.html. India Mart Member since 2013. Accessed Aug 2018.

[2] IT Engg Portal. Symmetric Key Encryption. http://www.itportal.in/2011/12/encryption-decryption-information.html. Copyright @2011. Accessed Aug 2018.

[3] Z.E. Dawahdeh, S.N.Yaakob, R.R. bin Othman, A new image encryption technique combining Elliptic Curve Cryptosystem with Hill Cipher, Journal of King Saud University-Computer and Information Sciences, 30(3) (2018) 349-355.https://doi.org/10.1016/j.jksuci.2017.06.004.

[4] E.H. Bensikaddour, Y. Bentoutou, N. Taleb, Embedded implementation of multispectral satellite image encryption using a chaos-based block cipher, Journal of King Saud University-Computer and Information Sciences. (2018)https://doi.org/10.1016/j.jksuci.2018.05.002.

[5] A. Karmel, A. Multi-Level Privacy Preservation and Transmission of Medical Records Using Cryptographic Technique, International Journal of Applied Engineering Research, 12(17) (2017), 6735-6740.

[6] S.A. Brown, H. Jouni, T.S. Marroush, I.J. Kullo, Disclosing Genetic Risk for Coronary Heart Disease: Attitudes Toward Personal Information in Health Records, American journal of preventive medicine, 52(4) (2017), 499-506.https://doi.org/10.1016/j.amepre.2016.11.005.

[7] C.S. Kruse, B. Smith, H. Vanderlinden, Nealand, A, Security Techniques for the Electronic Health Records, Journal of Medical Systems, 41(8) (2017), 127.https://doi.org/10.1007/s10916-017-0778-4.

[8] D.M. Walker, T. Johnson, E.W. Ford, T.R. Huerta, Trust Me, I'ma Doctor: Examining Changes in How Privacy Concerns Affect Patient Withholding Behavior, Journal of medical Internet research, 19(1) (2017).https://doi.org/10.2196/jmir.6296.

[9] R. Clarke, Introduction to dataveillance and information privacy, and definitions of terms. Roger Clarke's Dataveillance and Information Privacy Pages. (1999)

[10] A. Stubbs, Ö Uzuner, De-identification of medical records through annotation. In Handbook of Linguistic Annotation, Springer Netherlands (2017) 1433-1459.https://doi.org/10.1007/978-94-024-0881-2_55.

[11] S. Kim, J. Jung, J. Kim, A Study on FIDO Authentication System for Reinforcing the Security of Electronic Medical Records. In MATEC Web of Conferences, EDP Sciences (Vol. 108 (2017)., p. 09001).https://doi.org/10.1051/matecconf/201710809001.

[12] D. Kotz, R. Halter, C. Cornelius, et.al., U.S. Patent No. 9,595,187. Washington, DC: U.S. Patent and Trademark Office.

[13] B. Yüksel, A. Küpçü, Ö. Özkasap, Research issues for privacy and security of electronic health services. Future Generation Computer Systems, 68, (2017), 1-13.https://doi.org/10.1016/j.future.2016.08.011.

[14] B. H. Walters, Veillance and Electronic Medical Records in Disease Management Programs in the Netherlands. In *Under Observation: The Interplay Between eHealth and Surveillance*, Springer International Publishing (2017) 91-106.https://doi.org/10.1007/978-3-319-48342-9_6.