Proceedings of the Sixth International Conference on Inventive Computation Technologies [ICICT 2021]
IEEE Xplore Part Number: CFP21F70-ART; ISBN: 978-1-7281-8501-9

A Study of Securing Healthcare Big Data using DNA Encoding based ECC

Lipsa Nayak
Research Scholar, Dept.of computer Science
Vels Institute of Science, Technology and Advanced Studies (VISTAS)
Chennai, India
Email: info.lipsa@gmail.com

V.Jayalakshmi
School of Computing Sciences
Vels Institute of Science, Technology and Advanced Studies (VISTAS)
Chennai, India
Email:jayasekar1996@yahoo.co.in

2021 6th International Conference on Inventive Computation Technologies (ICICT) | 978-1-7281-8501-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICICT50816.2021.9358546

***Abstract*—IT world is migrating towards utilizing cloud computing as an essential data storing and exchanging platform. With the amelioration of technology, a colossal amount of data is generating with time. Cloud computing provides an enormous data storage capacity with the flexibility of accessing it without the time and place restrictions with virtualized resources. Healthcare industries spawn intense amounts of data from various medical instruments and digital records of patients. To access data remotely from any geographical location, the healthcare industry is moving towards cloud computing. EHR and PHR are patient's digital records, which include sensitive information of patients. Apart from all the proficient service provided by cloud computing, security is a primary concern for various organizations. To address the security issue, several cryptographic techniques implemented by researchers worldwide. In this paper, a vigorous cryptographic method discussed which is implemented by combining DNA cryptography and Elliptic Curve Cryptography to protect sensitive data in the cloud.**

Keywords: Healthcare, Cloud, Security, Medical big data, DNA cryptography, Elliptic Curve Cryptography

I. INTRODUCTION

The Healthcare industry is widening its services with the help of advanced information technology. A lot of time and money can be saved by modernizing our traditional healthcare system. As information about any patient is a vital thing in the Healthcare industry, storing and managing these data is also an essential part of it. Electronic health records (EHR), holds all vital information about a patient [1]. It is accepted that EHR can reduce medical costs by circumventing the repetition of expensive diagnosis, drug prescription. This vast data with variety is known as medical Big data. Cloud computing has become a boon for the healthcare industry as it can store and manage this medical big data without limitation, and data can be accessed from any geographic location and at any time[2]. Medical knowledge can be shared in any critical situation of a patient across the world by cloud computing. Tele-medicine, Digital pathology, Tele-dermatology, etc. are few applications of a healthcare cloud which provides cost-effective and streamlined services[3]. All administrative exertion like Registration, Billing, Appointments, Insurance related work of a healthcare organization can be done effortlessly in cloud computing ambiance[4]. Medical big data is transmitting through a broad network of the remote server, which are incorporated and viewed as one unified environment from the various location, so it is sensitive to intrusion or breach. As medical data is highly vulnerable and confidential, storing these in a third party server inevitably increases the risk of security. To fortify healthcare data in a cloud environment, various cryptographic techniques are there. Cryptography is a complete process of encoding and decoding of data. Plain text is transmuted in ciphertext in the process of encryption so that unauthorized access can not be possible. Here, a fused technique is proposed to secure the healthcare cloud by combining Elliptic Curve Cryptography (ECC) with DNA encoding. On the basis of time, efficiency and massive parallelism, DNA cryptography is chosen to protect the healthcare bigdata. The DNA sequences are used to encrypt the data as it holds some appealing properties. ECC is chosen to protect cloud as it gives the same security level as other public-key cryptography are giving but with a smaller key size. And by fusing DNA based encoding with Elliptic curve

cryptography, it became a little more complex to solve, which makes the data more secure. A robust, dynamic fused cryptographic technique with two levels of security is presented. Section two describes some concepts of the fused technique. Next part of the paper presented proposed algorithm, and last part gives the conclusion of this work.

II. PRELIMINARIES

A. DNA computing

In 1994 DNA computing was introduced by American scientist Richard Adlenman[5]. DNA cryptography is an emerging research area. It is a leading edges from which DNA cryptography is extracted. Due to its complex double helix structure, it is in the surveillance of data security researchers. Two strands of DNA that are coiled to each other and made up of millions of connected nucleotides[6]. Each of these nucleotides, Adenine (A), Thymine (T), Guanine (G), and Cytosine (C), is a nitrogen-containing nucleobase.Deoxyribonucleic Acid or DNA is the carrier of genetic information in living organisms(FIG 1).

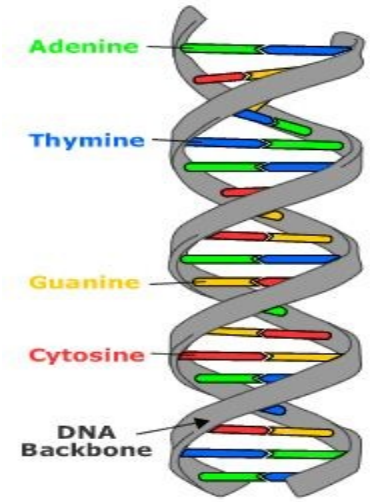

Figure 1: Structure of DNA

B. DNA Cryptography for Healthcare big data

A robust encryption design can be obtained by using DNA cryptography for its randomness and uniqueness. As in DNA Cryptography, data can be encrypted in A,C,G,T forms, any combination can be chosen to improve security[7]. These A, C, G ,T can be presented as a combination of 0s and1s. A can be referred to as 00, T as 01, C as 10, G as 11 (Table.1). As the computer understands only 0s and 1s, mathematically, data can be encoded using these A, T, C, G sequences. The data can be encrypted via three ways in DNA Cryptography, they are 1) Insertion Technique, 2) Complementary pair Technique, and 3) Substitution Technique. In each of these techniques, part of a DNA sequence from any publicly available DNA sequences would be selected secretly by the sender and receiver only. Then the sender converts the selected DNA sequence by adding a hidden message H to it. Sender sent this transformed DNA sequence to the receiver by inserting some other DNA sequences from which the receiver extracts the real course of DNA and hidden message H.

Table .1 Conversion table of DNA nucleotide base to Binary sequence

DNA nucleotide	A	C	G	T
Binary sequence	00	01	10	11

DNA Encryption and Decryption Algorithm for Healthcare Bigdata:

Input:DNA sequence D and hidden text H

Output:

Step1: From global database, DNA sequence(D) is selected. Both sender and receiver are aware of the DNA sequence.

Step2: Convert selected DNA sequence into Binary sequence by using binary coding scheme.

Step3: Segment the DNA sequence into equal parts of each 3 bits.

Step4: Convert the hidden text into its ASCII value and convert it into corresponding binary value.

Step5: Insert each bit of hidden text(H) before every even positioned segments of binary DNA sequence.

Step6: Concatenate all segments of binary DNA sequence.

Step6: Again, the binary sequence can be converted into DNA sequence by referring DNA encoding table. Hence, the encrypted DNA sequence is obtained.

Step7: After getting the encrypted message, receiver starts decrypting the message by converting it again to binary form.
Step8: After that receiver remove the bits of hidden text from the original sequence and by converting it again into DNA sequence, receiver will get the original DNA sequence.

Example of encoding plain text by using Dynamic DNA cryptography
Let Hidden text (H): "v."
ASCII value of text:118
The binary value of the text (H'): 01110110
Let the DNA sequence is (D): ATGGTTCCAATGC
Binary DNA sequence (D1): 001110101111010100000111001.
Segmented Binary DNA sequence (where x=3): 001 110 101 111 010 100 001 110 01.
Then, each bit of H is added to each even number segment of D1.
001, 0110, 101, 1111, 010,1100 ,001, 1110, 01
After adding all the segments together, the below mentioned sequence has been obtained:
00101101011111010110000111100 1
By converting above binary sequence to DNA sequence, below sequence is obtained and this is not real:
00-10-11-01-01-11-11-01-01-10-00-01-11-10-01
D'=AGTCCTTCCGACTGC

C. Elliptic curve cryptography:.

In year 1985 Victor Miller and Neal Koblitz developed Elliptic Curve Cryptography or ECC (Figure 2). which is a topical research area in the current era[8]. In our paper, we choose ECC because by using a lesser key size, it gives high-level security. Equation of ECC is in the form of

$$y^2 = x^3 + cx + d$$

where c and d are the constant with

$$4c^3 + 27d^2 \neq 0$$

For every c and d value, different curve will be there[10]. G is the generator point in the elliptic curve. By multiplying G with the private key, we can get the public key[11].

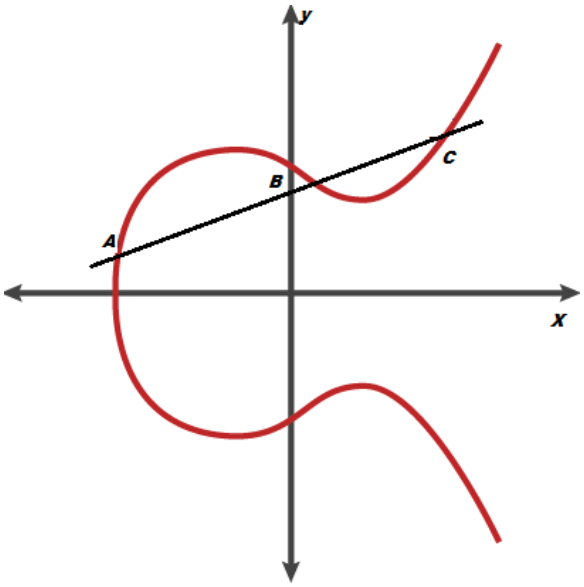

Figure 2 : Elliptic Curve

III. PROPOSED DNA ENCODING BASED ELLIPTIC CURVE CRYPTOGRAPHY

To achieve a neoteric security system for Healthcare big data, Here we combined DNA Cryptography with Elliptic Curve Cryptography[12]. These two are the most recent and powerful techniques to secure data in the cloud[13][14]. In the first phase of our proposed technique, plain text is encoded by using DNA Cryptography by using a DNA encryption algorithm. In DNA Cryptography, the encoding of data happened in three ways from which to choose the insertion technique to encode plain text. First plain text is transformed into corresponding ASCII value then converted into equivalent binary series. From publicly available DNA nucleotide sequences, a known sequence should be chosen, which is converted into binary value using table-1. A known DNA nucleotide series for sender and receiver should be chosen. The binary value of the nucleotide sequence is divided into segments of size greater than or equal to 2. After this step we have to add each bit of H, to each even number segment of DNA nucleotides and a new binary sequence can be obtained. After that, the resultant binary sequence again converted into DNA nucleotide. In the second

phase of our proposed technique the resultant sequence converted into some decimal values. After that, these decimal numbers encoded into an Elliptic Curve point by using Koblitz method [15][16].

A. ECC Encryption

- A data breach is the intentional or unintentional Let M - be the message.
- Let G - be the generator point.
- Let K - is a random positive integer.
- Let P_B - public key for receiver
- Encode the message M into an Elliptic Curve Point P_M by using Koblitz's method.
- Then encrypt that Elliptic curve point into a cipher point C_M.

$$C_M = \{ kG,\ P_M + kP_B \}$$

B. ECC Decryption

- When the attackers or hackers try to prevent valid For decryption, first, we have to multiply the encrypted point of Elliptic curve with receiver B's private key $(n_B) = kG * n_B$
- Then subtract it from second coordinate of the encrypted Elliptic curve point

$$(P_M + kP_B)P_M + kP_B - (kG * n_B) = P_M + kP_B - kP_B = P_M \quad \text{(equation 2)}$$
as $(P_B = n_B * G)$

By using the ECC decryption equation(2), the cipher points are deciphered., then these points turned numerals by using koblitz's method[16]. These numerals decoded into the DNA nucleotide sequence, which seems unidentified but the receiver decodes the original text by using the DNA decryption algorithm mentioned above.

C. Proposed DNA Encoded Elliptic Curve Cryptography Algorithm

Input:

R // any selected DNA sequence
M // Plain or original Text
b // The number of bits such that R is segmented as each segment contains b number of bits

Output:

E // Encrypted Text or Cipher Text Points

Segmented_DNA_ECC Algorithm:

Let M1 -> binary representation of the plain text M
Let R1 -> binary representation of the selected DNA sequence R
Split R1 into R1/b segments A, such that each segment Ai in A has b number of bits.
For each bit r in M1 loop
For each segment Ai in A where i=2,4,6,…, loop //only even segment
Ai = r U A|//Inserted r at the //beginning of Ai only
Loop
Loop
C={} // empty
For each segment Ai in A where i=1,2,3,…, loop //all segments
C = C U Ai// Join all segments into //single one C
End Loop
D -> DNA nucleotide conversion(C) //A=00,C=01,G=10,T=11
D1-> Decimal representation of D
Call koblitz(D1) to obtain encrypted text E and ECC point

End Segmented_DNA_Cryptography.

IV. CONCLUSION

Health care organizations are getting tremendous benefits by moving their Big data to the cloud. It is an indispensable and ever going challenge to secure healthcare big data in this distributed environment. In this pape, our research is based on a complex multi level security technique by using DNA encoding and Elliptic curve cryptography. Both these techniques give high level security itself. To provide multifaceted protection for the attacker, we combined some features of DNA cryptography with Elliptic curve cryptography. This hybrid cryptography offers better security in less time, memory, and smaller key size than traditional security techniques.

References

[1] Alex Mu-Hsing, "Opportunities and Challenges of Cloud Computing to ImproveHealth Care Services," Journal of Medical Internet Research, vol. 13, no. 3, 2011.

[2] KH Huang et al.," A Secure Electronic Medical Record Sharing Mechanism in the Cloud Computing Platform" 2011 IEEE 15th International Symposium on Consumer Electronics.

[3] Yang, J. et al., A 5G cognitive system for healthcare. Big Data and Cognitive Computing 1 (1) (2017) 1-2.

[4] S. Namasudra et al.,"Cloud computing: fundamentals and research issues," in Proceedings of the 2nd International Conference on Recent Trends and Challenges in Computational Models IEEE, 2017.

[5] Ganesh Chandra Deka et.al., "Advances of DNA computing in cryptography," Taylor & Francis, 2018.

[6] Jie Chen,"A DNA-based, biomoleccular cryptography design," in IEEE International Symposium on Circuits and System (ISCAS),2003,822-825

[7] Y.F. Wang et al.," An Encryption scheme using DNA Technology",2008 3rd International Conference on Bio-Inspired Computing (2008),37-42

[8] Victor S. Miller, Use of Elliptic Curves in Cryptography, Advances in Cryptology. Springer, vol. 218, pp. 417–426, (2000).

[9] Jansma, N. and Arrendondo, B. Performance comparison of elliptic CURV and RSA digital signatures. Univ. Michigan College Eng., MI, USA, TechRep., 2004.

[10] Lawrence C. Washington, Elliptic Curves Number Theory and Cryptography, Taylor & Francis Group, Second Edition (2008).

[11] Prokash Barman & Banani Saha, " E-Governance Security using Public Key Cryptography With special focus on ECC", International Journal of Engineering Science Invention, Vol-2, Issue 8, August 2013, PP 10-16.

[12] V. Jayalakshmi, Lipsa Nayak, "Protecting Medical Big Data in a Healthcare Cloud sing Elliptic Curve Cryptography" Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 11-Special Issue,2018

[13] P. Vijayakumar, V. Vijayalakshmi & G. Zayaraz, "DNA Computing based Elliptic Curve Cryptography", International Journal of Computer Applications (0975 – 8887), Volume 36– No.4, December 2011, pp 18-21.

[14] William Stallings, Cryptography and Network Security, Fifth Edition, Pearson. pp. 344.

[15] N. Koblitz, "Elliptic Curve Cryptosystem, Mathematics of Computation", Vol A8, 1987, PP 203-209

[16] Prokash Barman & Banani Saha, An Efficient Hybrid Elliptic Curve Cryptography System with DNA Encoding, International Research Journal of Computer Science(IRJCS), Vol 2, 33-39, (2015)