Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 6 (2015), pp. 4541-4546

http://www.ripublication.com

Cryptographic Mehod For Hiding Data In Genomic DNA Using Calculus

[1]Rama Devi. K and [2]Dr. Prabakaran. S

[1] Research Scholar, Department Of Computer Science & Engineering, SRM University, Chennai, Tamil Nadu, India
[2] Professor, Department Of Computer Science & Engineering, SRM University, Chennai, Tamil Nadu, India
E-mail: [1] ramadevi.sarav@gmail.com, [2] prabakaran.mani@gmail.com

ABSTRACT:

The biological research in the field of cryptography paves the exploitation of storage capabilities, parallelism and security for data transmission. Nowadays people have been transmitting more information using minimum space and time but with maximum security. New generation computers and high speed internet facilities are available nowadays. It is a remarkable achievement to transmit the data in a secured manner. To address the security attacks researchers are still working on cryptographic algorithms. The DNA cryptography is a novel and challenging area to gain higher information security. In this paper, DNA sequence is generated for the private key that has to be sent. The polynomial equation is created, sender uses differential calculus and receiver uses integration calculus.

Keywords DNA Cryptography, private key, polynomial, calculus.

I. INTRODUCTION

Cryptography is the art and science of achieving information security by encoding a simple message into a unreadable. Basically we have Symmetric and asymmetric Cryptographic algorithms. In Symmetric Cryptography single private key is used for both encryption and decryption. In asymmetric cryptographic algorithm couple of keys are used for encryption and decryption process. Cryptography rely on the degree of randomness and uncertainty in the generation of cipher text from plain text for which many phenomenon have been introduced for instance elliptic curve cryptography, quantum cryptography, visual cryptography and DNA cryptography.

DNA Cryptography, a new branch of cryptography exploits DNA as an information carrier with the help of molecular techniques. DNA cryptography gains interest due to the huge storage capacity of DNA which is a basic computational tool of this field. One gram of DNA can store 10^8 tera-bytes. DNA cryptography provides intense complication and unpredictability in the DNA structure for encoding and decoding.

II. RELATED WORKS

Zoheng et al., [1] presented data hiding in pdf texts for secret communication. In this paper the author point out secret channel in a kind of PDF English texts, that are generated from documents that occupy full line width and positions each character individually. Later stegnographic system pdfstegno is discussed. Lee et al., [2] proposed a novel approach by entrenching secret message in PDF file. A message is string of bits or characters encoded using ASCII code. The encoding result is encoded in between one-word or in between one-character in pdf file.

Francis M.Grosges.T [3]proposed an encryption algorithm in iterative process based on chaotic function and xor operator. Main advantage of this algorithm is it produces very lengthy key so it resists brute force attack and assures confusion and diffusion properties in cipher.

Gehani.A LaBean TH, Reif JH [4]proposed DNA based cryptography that is based on one time pad method that is unbreakable. Zhang Q.Guo L, Wei XP, designed an image encryption using DNA addition combining with chaotic maps[ ]. In this paper DNA sequence matrix is acquired by encoding the actual image, then divide the DNA sequence into equal blocks and use DNA sequence addition operation to add these blocks. Next complement the result by using logistics maps. Finally decode the matrix to get encrypted image.

In 2010, Lai Xuejia et al., [5] introduced DNA based asymmetric encryption algorithm and signature method. DNA (PKC) algorithm possesses both security and authentication. Private key and cipher-text is a genetic molecule in DNA (PKC). In this algorithm key and cipher-text are transmitted physically and it's hard to replicate. This public key encryption algorithm is based on DNA microarray chip. It is constructed with examinations for encryption and decryption. Existing probes are used as a key. If the intensity is greater then some fixed value then the probe is denoted as one otherwise the probe value is 0. For encryption process couple of keys is used PKs and PKr. First plaintext is converted into its ASCII code and then it converted into binary code, binary code is arranged in the form of matrix. Then the matrix is encrypted using a biological molecule that is referred as key.

In 2011, Bibhash Roy et al., [6] proposed an improved symmetric key cryptography with DNA based strong cipher. In this paper the author focused on DNA computing logic, used for encryption, storage and transmission of the data. This paper proposed the exclusive cipher-text procedure and key generation procedure.

In 2012, Yunpeng Zhang et al., [7] projected a DNA cryptography based on DNA fragment assembly. In this paper the author points out the features of DNA molecular, key generation technologies, DNA digital coding and related software. Using the DNA digital coding and DNA fragmentation author designed symmetric encryption

algorithm, it converted plaintext into binary ASCII code and then into DNA sequences. Pramanik Sabari, and Sanjit Kumar Setua, "DNA cryptography, "[8] proposed a symmetric key algorithm implemented in VB.net. DNA hybridization and molecular structure is used. This algorithm enhances parallel technique and reduces time complexity.

III. DNA Digital coding and Calculus:

In information technology the most basic encoding method is binary encoding method that is anything can be encoded in the form of 0 or 1 or combination of 0 and 1. Nowadays, in Bio-technology we have four different nucleotide bases which are Adenine (A), Cytosine(C), Thymine (T) and guanine(G) in DNA sequence. This simplest coding pattern is used to encode the nucleotide base A, C, G, T by means of digits 00-A, 01-T, 11-C, G-10. Evidently, there are 4! =24 possible coding patterns. This pattern could reflect the biological characteristics of 4 nucleotide bases and certain biological significance. DNA coding has following advantages: 1) DNA coding decreases the redundancy of the information coding. 2) The digital coding of DNA can be easily preprocessed for encryption or decryption. 3) This coding technique facilitates direct conversion from biological information into digital information.

In this paper we generate an elementary function in the form of polynomial function P(x).

$P(x)= a_4x^4 + a_3x^3 + a_2x^2 + a_1x^1 + a_0$.

For this polynomial function P(x) we determine the derivate with respect to x using differential calculus

$\frac{d}{dx}(x^n)=nx^{n-1}$

In receiver side, the integration of a received polynomial P'(x} is determined using integral calculus.

$\int x^n \, dx = \frac{x^{n+1}}{n+1}$

IV.PROPOSED ALGORITHM

Encryption Algorithm:

Step 1: The message is scanned for the occurrence of continuously two same characters, if so a special character is padded in between the two characters. For instance "Hello", x is padded in between two 'l', so the message becomes "Helxlo".

Step 2: Now the text is encoded by ASCII value and the decimal value is converted into binary value. Now left padding with 0 is made to equalize the binary value of length 8 bits.

Step 3: The binary string of length 8 bits is now encoded by DNA codon using the DNA codon table (A-00, T-01, C-11, G-10). We get 4 - codon for each byte in the

combination of ACGT. (For Hello – DNA string is TAGATATTTACATTGATACATACC)

Step 4: Generate a polynomial equation with exponent values T-4, G-3, C-2, A-1. For coefficient values, count the number of occurrences of A, C, G and T. In our example the coefficients are 9, 2, 4, 9 since number of T in the DNA sequence is 9 and similarly for other codes.

Step 5: Now the polynomial equation $P(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x^1 + a_0$. ($a_4$=9, a_3=2, a_2=4, a_1=9). So the equation is $P(x)=9x^4 + 2x^3 + 4x^2 + 9x$.

Step 6: Differentiate P(x) with respect to x. Now $P'(x)= a_3'x^3 + a_2'x^2 + a_1'x + a_0$ is determined where a_3'=36, a_2'= 6, a_1'=8, a_0'=9. So the equation $P'(x)= 36x^3 + 6x^2 + 8x + 9$.

Step 7 : Now DNA sequence is generated from P'(x) which encompasses a_3' number of 'T', a_2' number of 'G', a_1' number of 'C' and a_0' number of 'A'. DNA sequence with 36-'T', 6-'G', 8-'C' and 9-'A' is generated. This DNA string is sent as file1 in the form of PDF to the receiver.

Step 8 : The DNA string generated in step 3 is mixed with other DNA strands and sent as file2 to the receiver.

Decryption Algorithm:

Step 1: Receiver received a text document from file1 that encompasses only DNA codons. From the lengthy DNA sequence the number of occurrence of A, C, G, T is decoded as a_3'=36, a_2'= 6, a_1'=8, a_0'=9

Step 2: From this information a polynomial P'(x) is devised with coefficients a_3', a_2', a_1' and a_0' and exponents 3, 2, 1 and 0. The equation is $P'(x)= 36x^3 + 6x^2 + 8x + 9$ is generated.

Step 3: Integration is applied to the polynomial equation P'(x) to get P(x).

$P(x) = \int P'(x) = \int (a_3'x^3 + a_2'x^2 + a_1'x + a_0) = a_3'\int x^3 + a_2' \int x^2 + a_1'\int x + a_0$

In our example $P(x)= 9x^4 + 2x^3 + 4x^2 + 9x$ is obtained.

Step 4: Now the receiver traps the number of A, C, G and T as (T-9, G-2, C-4, A-9). Now the total length of the DNA string is considered as N. (N=24)

Step 5: Now the DNA strands in file2 are partitioned by a length of N (i.e 24 codons). Now check the each block for the occurrence of coefficients a_3 number of 'T', a_2 number of 'G', a_1 number of 'C' and a_0 number of 'A'.

Step 6: The matched blocks are extracted separately and decoded into binary and its respective Character. Among the codons that produce meaningful English text is found to be a plain text sent by the sender.

V.SECURITY

The Security of this encryption algorithm comes under two levels: The first level is biological security and the complexity lie under biological problems. It is extremely difficult to amplify the message that is coded in DNA strands. The second level is the mathematical strength that increases the security level. This algorithm resists brute force attack because in any way DNA strand that acts as a key cannot be predicted.

This algorithm is also free from other security attacks and hence the complexity increases as the size of the polynomial coefficient increases.

VI. EXPERIMENTAL RESULT

Tests are undertaken on 2.0GHz Intel CPU employing 3GB RAM. We evaluated the execution time according to the size of the key embedded in file1 and the size of the plain text embedded in file2. Files used in the test are in PDF format.

Length of file1 (Key)in bytes	Length of file2 (message) in bytes	Computation Time	
		Encryption Time (ms)	Decryption Time (ms)
69	10	7.8	3.8
555	100	8.53	6
5500	1000	15	11.5
55000	10000	70.34	68.9
553000	100000	530.23	520.34

V.CONCLUSION AND FUTURE WORK

In this paper, we designed an encryption and decryption algorithm with the use of DNA codons. In encryption side differential calculus is used and in decryption process integral calculus is used to generate and regenerate a polynomial equation. Bothe the files file1 and file2 that holds cipher text and key are in the form of DNA code that are sent in PDF format. The Key is generated for each message so the key cannot be used for trapping other messages. It improves the confidentiality resists security attack. It can be concluded that the proposed algorithm can enhances the secure data transmission through internet. Further this algorithm can be improved by providing authentication and error detection and correction mechanisms like hamming code can be incorporated.

REFERENCE:

[1] Shangping Zhong, Xueqi Cheng, Tierui Chen “Data Hiding in a kind of PDF texts for Secret Communicaton “ International Journal of Network Security, Vol 4 No.1, PP 17-26 Jan 2007

[2] Lee IS, Tsai WH “ A new approach to covert communication via PDF files” Journal Signal Processing Volume 90 Issue 2 February 2010 Pages 557-565

[3] M.Francois, T.Grosges, D.Barchiesi and R.Erra, ”Image Encryption Algorithm Based on a Chaotic Iterative process” Applied Mathematics, Vol 3 No 12, 2012

[4] Gehani.A., LaBean TH, Reif JH “ DNA based cryptography” aspects of molecular computing Lecture notes in computer Science Volume 2950.

[5] Lai, XueJia, "Asymmetric encryption and signature method with DNA technology, " Science China Information Sciences 53.3, page 506-514, (2010).

[6] Bibhash Roy, Pratim singha, “An improved symmetric key cryptography with DNA based strong cipher”. Devices and Communications (ICDeCom), IEEE, 2011 International Conference on 24-25 Feb. (2011).

[7] Yunpeng Zhang, Bochen Fu, and Xianwei Zhang, "DNA cryptography based on DNA Fragment assembly, " In Information Science and Digital Content Technology (ICIDT), IEEE International Conference on, vol. 1, pp. 179-182, (2012).

[8] Pramanik Sabari, and Sanjit Kumar Setua, "DNA cryptography, 7th IEEE International Conference on, pp. 551-554, (2012).