IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 8, August. 2019, ||Series -II || PP 38-42

Analysis of Data Hiding for DNA Based Cryptography

Shradhanjali Singh[1], Shalu[2]
[1](*Department of ComputerScience Rajasthan TechnicalUniversity Kota, India)*
[2](*Department of ComputerScience Rajasthan TechnicalUniversity Kota, India)*
Corresponding Author: Shradhanjali Singh
Received 15 August 2019; Accepted 30 August 2019

Abstract: In today's world, securing data is becoming one of the main issues, the elaboration of the fusion of cryptography and steganography are contemplating as the sphere of on-going research. This can be gain by cryptography, steganography, and fusion of these two, where message firstly encoding using any cryptography techniques and then conceal into any cover medium using steganography techniques. Biological structure of DNA is used as the cover medium due to high storage capacity, simple encoding method, massive parallelism and randomness DNA cryptography can be used in identification card and tickets. Currently work in this field is still in the developmental stage and a lot of investigation is required to reach a fully-fledged stage. In this paper, we used modify DNA based Playfair as cryptography method. This paper provides analysis of capacity, payload, and bpn of proposed method used to hide DNA based Playfair cipher into DNA sequence.

Keywords: DNA , DNA Cryptography, Playfair, and substitution method.

I. INTRODUCTION

Nowadays, Security of information is the importance part of any type of communication. Due to the development of huge technology in the world, every organization depends on its information systems. Information security protects the information from uncertified access and use, disruption of data, the disclosure of data, modification or destruction of data in order to provide the confidentiality, authentication, integrity, and availability of data. Data security can be accomplished using cryptography and steganography.Cryptography is a process of encryption and decryption. In cryptography data is changing from a meaningful form to an unreadable form using a key is called encryption and changing data from unreadable to original data using the same key or different key is known as decryption. While steganography is the process of hiding message into media like video, images, audio and biological structure of DNA both cryptography and steganography are the independent method but to accomplish high secure environment, they can be combined. When we combine the cryptography and steganography, the first original data is encrypted using cryptography encryption method and then encrypted data is hiding into media using steganography method. In this paper hiding media is DNA sequence. DNA steganography is a process in which data is converted into DNA format in order to hide into a DNA pattern. DNA based cryptography is the act of using DNA steganography along with DNA encryption. DNA steganography is becoming very beneficial because of high storage capacity, massive parallelism, and randomness. These features make it better than any other steganography [3]. In [1] presented the two DNA-based steganography method based on DNA dummy binary stands. In the first approach, the plaintext is encoded into binary then mixed with dummy strands in an equimolar manner. The second method was based on steganography but used for implementing the cryptography. The data is encrypted with a group of dummy strands having the same key pool sequence.In [2] proposed encryption method using DNA hybridization. Plaintext in ASCII is encoded into binary. The binary bit is encoded into DNA format. Length of OTP is equal to ten times the length of plaintext in DNA format. then proposed encryption algorithm is used. Encrypted data is placed between two primers and hidden into microdots.In [6] submit a DNA- based cryptographic method in which first data is encoded into ASCII forms and put in (4*4) matrix. In this paper the author used the concept of mathematical calculation and encoding of data is performed in cycles to make message unreadable. A secret key is XOR with an output from mathematical manipulation of (4*4) matrix. The important feature of this algorithm is that it always calculates a different unreadable message from the same message and key.In [7] used properties of DNA bases and amino acids for execution of Playfair cipher. In this paper, the message is converted into binary. Then convert into DNA sequence .after that maps the codons of DNA into an amino acid. Used the concept of Playfair, and the secret key is converted into the amino acid into cipher text.In [8] presented three data conceal methods based upon properties of DNA computing. Names of three proposed algorithm are the Insertion Method, the Complementary pair Method, and the Substitution Method.In [9]proposed DNA-based steganography. In this paper, the author merges DNA base steganography with DNA based cryptography. Proposed two- step for more secure exchange of data: First, encrypt the plaintext using amino-acid and DNA

base Playfair method. Then applies complementary substitution method to conceal the DNA based unreadable data to some reference DNA sequence. In [10] proposed DNA-based cryptographic approaches for data conceal in DNA medium. In this paper author first, convert the data into DNA format then encrypt the data using modify Playfair cipher algorithm. After this encrypted data is concealed into the reference DNA sequence. This algorithm provides a significantly higher hiding capacity and security. In our paper algorithm have three-phase. The first is the preprocessing phase, which converts the plaintext into DNA format. The second phase is New DNA based Playfair encryption method and the third phase is the hiding method used to hide cipher data after encryption into the DNA sequence. Then we analysis the different parameters like capacity, payload and bpn of algorithm. The rest of this paper is organized as follows: II Section Provide background knowledge of the biological structure of DNA, DNA steganography, and cryptography. III Section contains our proposed method for DNA based cryptography and Substitution algorithm. IV section provides experimental results .V section provide the performance comparison with existing method. In the VI section at last conclusion is drawn.

II. BACKGROUND

This section explains the biological background on DNA, the overview of the DNA steganography and DNA cryptography

1. Biological background of DNA

Deoxyribonucleic acid or DNA is a biological molecule that contains genetic information of living organisms.it contains the repeating cell known as nucleotides. each nucleotide is made of sugar and a phosphate macromolecule, which act as a backbone of DNA and one of four organic bases. These bases are thymine(T), guanine(G), adenine(A) and cytosine(C). Combination of these bases forms DNA sequences. Normally A pair with T and C pair with G.A-T and C-G are base pair. Any three pairs of nucleotides are known as a codon.

2. DNA Steganography

DNA steganography is a process in which data is embedded into the biological structure of DNA. first data must be converted into DNA pattern and then merge with taken DNA pattern, so that resultant fake DNA sequence looks like actually existing DNA pattern. Different method is used for encoding the information into DNA pattern. for example, the nucleotides of DNA sequence are mapped with digital coding according to TABLE I. Usually audio, video, image, and game are used as coveringmedia. however, due to DNA high storage capability and randomness, DNA is getting beneficial as an embedding medium Different hiding method are insertion method , complementary pair rule techniques and substitution method[8].

TABLE I MAPPING OF DNA SEQUENCE INTO ABINARY

Bases of DNA	Decimal	Digital coding
A	0	00
C	1	01
G	2	10
T	3	11

3. DNA Cryptography

Cryptography is a process which converts the electronic data (plaintext) into unreadable data (ciphertext) form using secrets key. Encryption method and decryption method are the main component of cryptography. There is basically two encryption method:

a. Conventional method.

b. Public-Key encryption

In conventional encryption, both the parties' sender and receiver receive the same key. Electronic data is converted into unreadable data using an encryption method and secret key, and unreadable data is converted back into a human-readable form using decryption algorithm and secret key. In asymmetric sender and receiver have two related key.one of the related key is used for encryption algorithm and the other related key used for decryption [4].DNA cryptography makes use of DNA technologies for encryption, decryption and key generation as given in table 2.Polymerase chain reaction (PCR), DNA decoding and DNA chip are the most well-known DNA technology.

TABLE II DNA COMPUTINGTECHNOLOGY

DNA computing Technology	Key points
Polymerase chain reaction (PCR)	In this method, it contains two primer- oligonucleotides to increase the DNA pattern in every cycle This method is used in key generation, encryption, and decryption.
One time Pad (OTP)	OTP is randomly chosen secret DNA sequence but occurs only once. This method is used for encryption, key generation, and decryption.
DNA decoding	In this method, the plaintext is first converted into ASCII code, and then into binary code. After this decoded into DNA pattern This method is used in a preprocessing step.
DNA chip	Biochip has many sports conceal on a solid surface. These sports contain DNA pattern which is used to find expression of thegene.

III. DATASET

Dataset is taken from national center for biotechnology information (NCBI). More than 1.635×10^{10} dataset are presented in the NCBI[5].

TABLE III. DATASET

Refseq locus	Number of bases
AC153526 (Musmusculus)	200117
AC167221 (Bos Taurus clone)	204841
AC168901 (Bos Taurus clone)	191456
AC168907 (Bos Taurus clone)	194226
AC16908(Bos Taurus clone)	218028

IV. THE PROPOSED METHOD

Our proposed method contain three phase. First one is preprocessing step. Next one is modify DNA based Playfair algorithm and last one is hiding method to conceal the encrypted data into cover DNA sequence. We used DNA based Playfair of 4 x 4 matrix. Playfair rule are apply on 4 x 4 DNA based Playfair matrix as shown in TABLE IV. The encryption method at sender side is done as follow:

TABLE IV. 4 X 4 DNA BASED PLAYFAIR

AA	TC	CG	TG
GC	TT	TA	GT
GG	AT	CT	CC
CA	AC	AG	GA

1. Preprocessing step

a. Convert the plaintext(msg) into ASCII value(A).

b. Transform (A) into its binary format (BIN).

c. Then convert (BIN) into DNA format(BD) as shown in TABLE II.

d. The key is randomly used to shuffle the 4 X 4 DNA based Playfair grid.

2. Encryption by 4 x 4 DNA based Playfair grid.

a. BD is input for 4 x 4 DNA based Playfair grid

b. apply the three basic rule of Playfair on BD using shuffle 4 x 4 DNA based Playfair grid to get the final encrypted data in DNA format(CD).

3. Hiding Method

Cipher text(CD) in DNA format is conceal in reference DNA sequence as shown in TABLE III. This paper used hiding method [20] for concealing message in DNA sequence to achieve Fake DNA sequence.

4. Decryption Method

Information of 4 x 4 DNA based Playfair and key used to shuffle the DNA based Playfair is shared between sender and receiver side. Receiver first recovery cipher text (CD) from DNA sequence as given in [11]. deciphering is givenas follow:

a. apply the inverse of 4 x 4 DNA based Playfair and result is (BD)
b. Map (BD) into binary (BIN) using TABLE II.
c. Map 8 bit of (BIN) into ASCII code(A)
d. Finally we get msg from (A).

V. EXPERIMENTAL RESULT

Analysis of different techniques depends on capacity, Payload and bpn. Capacity means the total length of fake sequence after encryption. Payload is defined as remaining DNA bases after hidding encrypted DNA bases. bpn read as bit per nucleotide used for calculation of hiding capacity. We randomly take plaintext (msg) of 20000 byte. The result for 20000 bytes plaintext is given in TABLE V

TABLE V. 4 X 4 DNA BASED PLAYFAIR

Refseq	Number of bases	DNA based Playfair sequence	capacity	payload	Bpn
AC153526	200117	80000	150599	49518	53.31
AC167221	204841	80000	156742	48099	51.04
AC168901	191456	80000	147894	43894	54.13
AC168907	194226	80000	146581	47645	54.57
AC16908	218028	80000	174896	43132	54.57

VI. PERFORMANCE ANALYSIS

We have done some changes on the simple DNA based Playfair method, to provide better hiding capacity and high security level. We used 4 x4 matrixes rather than 5x5 matrixes. 4 x4 matrixes remove ambiguity bit problem in the previous method. This removal provides high hiding capacity.

1. Security Analysis

It provides two level of security. Only sender and receiver knew the key and DNA sequence. There are more than 1.63 million real DNA sequence and we used 4 x 4 matrixes. Hiding method is based on the complementary rule.so overall cracking probability (cp) is given below.

$$p(cp) = \frac{1}{4 \times 3 \times 2 \times 1 \times 1.6 \times 10^{10}\ 16 \times 6 \times 24}$$

2. Comparative Analysis

In this section we compare the most recent method with our proposed method. Hiding capacity of our work and [18] is given in TABLE VI. Our proposed method provides good hiding capacity and security. Comparison graph is shown in fig1.

TABLE VI. COMPARATION OF HIDING CAPACITY

Refseq	Proposed work	S.Marwan
AC153526	53.31	48.85
AC167221	51.03	50.01
AC168901	54.13	46.8
AC168907	54.57	47.41
AC16809	54.57	53.32

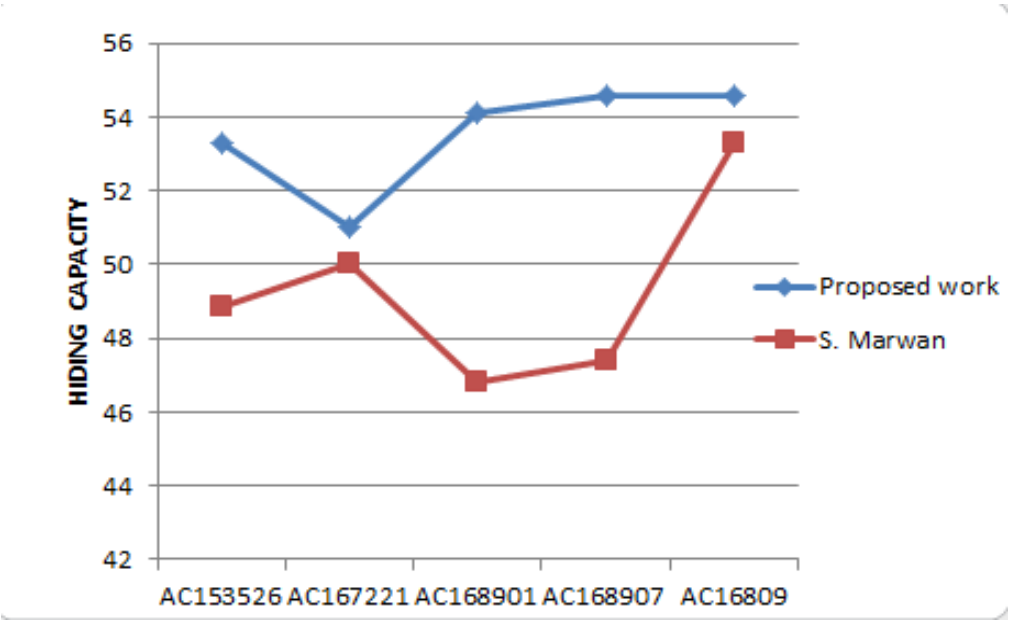

Fig. 1. Hiding capacity comparison.

Fig. 2.

VII. CONCLUSION

In this paper, we provide a better algorithm to communicate message securely. Algorithm had two stages. First encrypted the data by using 4 x 4 DNA based Playfair. After that encrypted data is hiding into DNA sequence by using substitution method. . We used 4 x4 matrixes rather than 5x5 matrixes. 4 x4 matrixes remove ambiguity bit problem in the previous method. This removal provides high hiding capacity. DNA based cryptography is one of the new emerging method in the cryptographic field, which used the concept of DNA computing. Due to high speed, minimal storage requirement, minimal power requirement, and parallel computability, it exploited for encryption technology. This paper discusses various parameters like capacity, payload, security and bnp to estimate the performance of algorithm. For the future work, we can used the DNA based public –private key method in place of modify DNA based Playfair.

REFERENCE

[1]. Leier, C. Richter, W. B., and H. Rauhe , Cryptography with DNA binary strands, *BioSystems 57*,pp.13-22,2000.

[2]. M. Borda, and O.Tornea, DNA secret writing Techniques,*8th IEEE International Conference on communicaton*, pp. 451-456,2010.

[3]. L. M. Adleman, Molecular computation of solutions to combinatorial problems,*Science, vol. 266*, pp. 1021-1025,1994.

[4]. Stallings, William, Cryptography and Network Security, Principles and Practice,*6th edition*,2014

[5]. National center for biotechnology informationhttp://www.ncbi.nlm.nih.gov

[6]. T. Mandge, and V. Choudhary, A DNA encryption technique based on matrix manipulation and secure key generation scheme, *International Conference on Information Communication and Embedded Systems*, pp.47-52 Feb 2013.

[7]. M. Sabry, M. Hashem, T.Nazmy, and M. E. Khalifa, A DNA and aminoacids-based implementation of playfair cipher,*international journal of computer science and information security, vol.8,* pp.129-137 ,2010

[8]. H. I. Shiu, K. L. Ng, J.F. Fang, R.C.T Lee, and C.H.Huang, Data hiding methods based upon DNA sequences,*information science , vol.180,* pp. 2196–2208, June2010

[9]. A. Khalifa, and A. Atito, High-capacity DNA-based steganography,*In the 8th International Conference on INFOrmatics and Systems,* pp. bio-76-bio-80 May2012

[10]. S. Marwan, A Shawish and Nagaty,DNA based cryptographic method for data hiding in DNA media, *Biosphere 150*,pp110-118, September 2016.

[11]. Cheng Guo, Chin-Chen Chang and Zhi-Hui Wang, A New Data Scheme Based on DNA Sequence, *ICIC International Innovation* .pp139-149, 2012.

Shradhanjali Singh. "Analysis of Data Hiding for DNA Based Cryptography." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 08, 2019, pp. 38-42.