International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

A Proposed DNA Postfix Hiding Method

Ghaith DehiaaKahdum, SaharAdillAl-Bawee, Mahdi NsaifJasim

Abstract*: Process of sending data through the Internet is vulnerable to tampering and theft, so protection has become a necessity. Many solutions have proposed to solve this problem in order to protect the sending data and conceal them in a way that cannot be penetrated or proven by nature. The paper proposes a method to hide data in a Deoxy ribonucleic acid (DNA) using Postfix conversion according to the embedded bits. The method has proved its security in concealing the information, not predicting its nature, and preserving the biological structure of the DNA sequence.

Keywords:DNA Sequence, Codon postfix, Mutation, Security, information hiding.

I. INTRODUCTION

Deoxyribonucleic acid (DNA) has special characteristics that qualify it to be the best medium for hiding data, these characteristics are:DNA has a high capacity to store a huge amount of data in it firstly. Secondly, The DNA sequence can be formed with a different length depending on what is working on it [1]. Due to these properties, several techniques have recently been proposed using DNA to hide data to take advantage of its high randomization and the high degree of hardness of detection[2, 3, 4]. This paper, proposes a method for hiding data in a DNA in an efficient and secure manner using codon-encoding Postfix. The proposed method issue is to convert the postfix codon in a manner preserves the corresponding amino acids with enduringness against the mutation of extracted bits. The paper layout is: related work presented in 2, proposed method in 4. Section 9 shows the experimental results. Finally, the conclusion is presented in 10.

II. RELATED WORK

In [5], A new DNA sequence-based data-hiding scheme was presented, using an injective mapping between one complementary rule and two secret bits in a message, the mapping scheme can effectively hide two secret bits in a message by replacing one character. The proposed scheme proved an efficient embedding capacity with a low modification rate.

Manuscript published on November 30, 2019.
Correspondence Author
GhaithDehiaaKahdum*, computer science, Collage Science for Women;University of Babylon, Babylon, Iraq.
SaharAdillAl-Bawee, computer science, Collage Science for Women; University of Babylon, Babylon, Iraq.
Mahdi NsaifJasim, department of Managementinformation systems, college of Business Informatics,university of information technology and communications,Baghdad, Iraq,

In [6], proposed a DNA watermarking technique. This proposal based on allocate the codons to a random circular angle using a random mapping table with a number of selected codons for embedding the watermark message into random circular angles of codons. This table is used as the watermark key and can be applied to any codon sequence regardless of sequence length. Without knowledge of this table, it is very difficult to detect the length and location of sequences for extracting the watermark.The method in [7], coupled three techniques (arithmetic coding, the advantage of codon redundancy, and public key cryptography), the secret message is converted to decimal number to generates a sorted list of all the synonymy codons, then the embedding process takes place in a DNA region.In [8], Frank Carter and Catherine Cleland proposed a method for concealing coded messages in DNA. The method of concealing the DNA is to encode the message within a genomic DNA sample followed by a further concealment of the DNA sample to a microdot. Exploit the advantage of the complexity of the genome of an organism to hide a Secret message in the genomic DNA.Shiu[9] suggested three methods (insertion method, substitution method, complementary method) and compare between these methods in each method author represent the capacity of carrying the information and hardly for attacker for extract information inside cover. F. P. Petitcolas., et al. [15] suggests, “Arithmetic encoding “for hidden information into DNA chain. The thought of the algorithm was based on the feature of codon redundancy. Numerous codons were interpreting to the same amino acids of the central dogma. In this way, it begins by changeover message that to be covered up in double arrangement into a decimal number between and 1. Arithmetic encoding is utilized to parse through the different codon tables. The length of the resultant stego-DNA depends on the accuracy of the embedded division that clearly influences the precision of the blind recovery handle. In (Mona Sabry, Mohamed Hashem, Taymoor Nazmy, Mohamed Essam Khalifa, 2010)proposed a critical alteration to the ancient Play fair cipher by presenting DNA-based and amino acids-based structure to the center of the ciphering prepare. In this think about, a parallel frame of information, such as plaintext messages, or pictures are changed into groupings of DNA nucleotides. Hence, these nucleotides pass through a Play fair encryption prepare based on amino-acids structure.

III. DEOXYRIBONUCLEIC ACID (DNA

DNA is a recent carrier has been used in data hiding area. In this chapter, we focus on data hiding in DNA. In biology, a Deoxyribonucleic Acid (DNA) is the leading molecular structure for encoding the data required to create and direct all chemical elements in the human body. For this reason, DNA has been put forward as a viable option for use in computational applications [16].

3.1 DNA Structure

DNA defined as the genetic drawing of every living creature. Each body cell has a unique complete set of DNAs; a polymer comprised of monomers referred to as deoxyribose nucleotides, which made up of three components as shown in Figure (1).

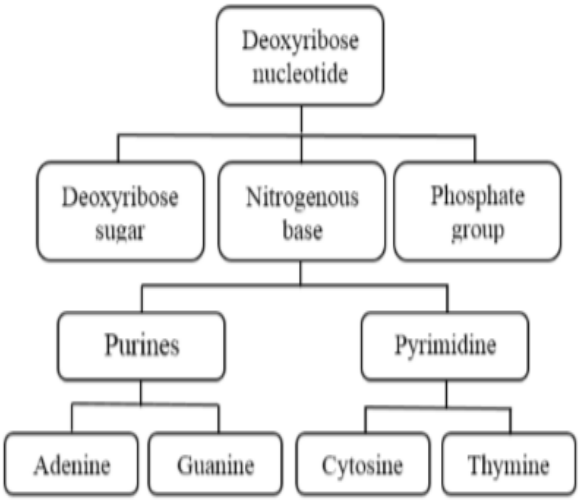

Figure (1): The structure of Deoxyribose Nucleic DNA [16].

The human body consists of trillions of cells each of which serves different functions. Each cell contains a nucleus that has a number of chromosomes as illustrated in Figure (2). Most of the DNA contents are found in a nucleus called nuclear DNA, and the rest of its contents are found in mitochondria which are called mitochondria DNA (mtDNA). DNA controls the function of each cell. Each DNA's chromosome consists of a DNA molecule, which holds genes. The gene is the entire genetic makeup,

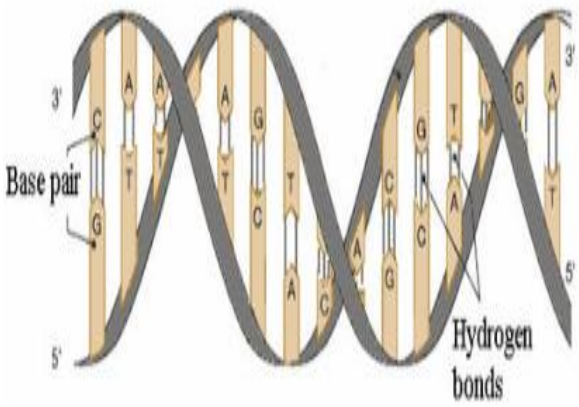

Figure (2): Helical Structure of DNA subjects [18].

which essentially contains information from all the chromosomes.Given that each nucleotide could have any of the four chemical bases and each codon is comprised of three nucleotides, then there is a sum of 4^3 = 64 different possible combinations. These combinations determine the amino acids to beused by living organisms, whose arrangement determines the structure and function of the resultant protein. Translation is the procedure through which RNA, a mediator duplicate of the directions contained in DNA, is made. The RNA is additionally comprised of four bases: adenine (A), cytosine (C), uracil (U) and guanine (G) [17].

IV. MOTIVATIONS OF THE PROPOSED METHOD.

The nature of the codon structures and the high randomness make it very attractive to use as rich hiding media to for messages with high security degrees, because the experts to analyses the DNA sequences are not widely available and the sequences don't attract more attention for raw users.

V. THE CONTRIBUTION

This research presents new technique to hide text messages in the sequences of strings representing the biological DNA data.

VI. THE METHODOLOGY

The research methodology is based on converting both the cover and message to binary representation, then using very complex method to choose the hiding positions of message data in the cover data. The positions are generated randomly using magic cube with changeable dimension to get maximum randomness. High dimension magic cubes can generate keys very hard to guess. The proposed algorithms are discussed in section 7.

VII. THE PROPOSED DNA POSTFIX METHOD

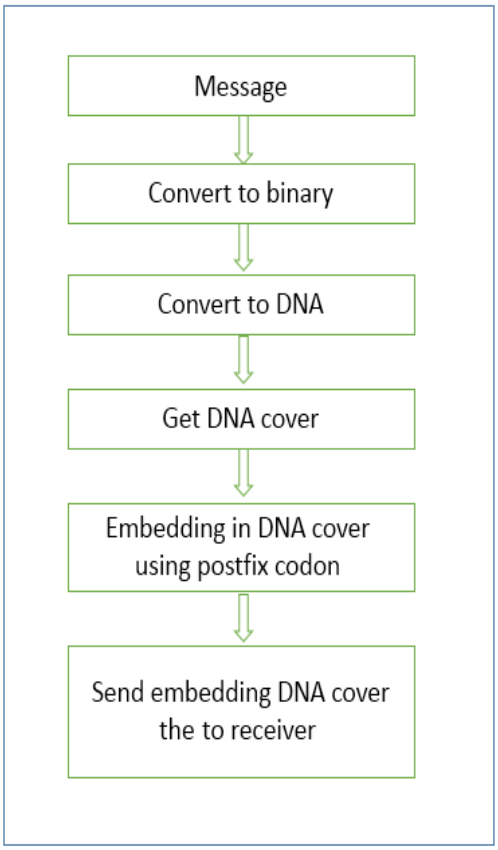

Figure (3)IllustratesDNA Postfixmethod structure:
Fig (3): DNA PostfixMethod

Retrieval Number: C5321098319/2019©BEIESP
DOI:10.35940/ijrte.C5321.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VIII. GENERAL STRUCTURE

The proposed method is based on: first, convert the sending message to a binary based on DNA digital coding rule as shown in a table (1) to form a binary message[9] (BinMsg), the BinMsg then converted to DNA bases according to a DNA conversion mapping rules. These bases will be divided into codons using a standard genetic code rules, in turn each amino acid is pointed mapped to one or more of DNA nucleotides (codons). Second, the codon consist of three bases the third base is called Postfix and the embedding process in the amino acid codons are aggregated by codon postfix as shown in a table (2), because of a codon postfix aggregation, the amino acid codons are classified to two groups of codon postfix: {A or C} and {G or T} each with the corresponding first two DNA bases.

Table (1): DNA digital coding

DNA nucleotide	*Decimal*	*Binary*
A	0	00
C	1	01
G	2	10
T	3	11

TABLE (2): CODON POSTFIX

Amino acid	Codon Postfix
Ala/A	GC{A,C,G,T}
Arg/R	AG{A,G} + CG{A,C,G,T}
Asn/N	AA{C,T}
Asp/D	GA{C,T}
Cys/C	TG{C,T}
Gln/Q	CA{A,G}
Glu/E	GA{A,G}
Gly/G	GG{A,C,G,T}
His/H	CA{C,T}
Ile/I	AT{A,C,T}
Leu/L	CT{A,C,G,T} + TT{A,G}
Lys/K	AA{A,G}
Met/M	AT{G}
Phe/F	TT{C,T}
Pro/P	CC{A,C,G,T}
Ser/S	AG{C,T} + TC{A,C,G,T}
Thr/T	AC{A,C,G,T}
Trp/W	TG{G}
Tyr/Y	TA{C,T}
Val/V	GT{A,C,G,T}
Stop:	TA{A,G} + TG{A}

Algorithm (1) describe the general steps of proposal DNA Postfix Embed Method, while Algorithm (2) describe the message retrieval steps. Fig (4) An example of a DNAPostfix method.

Algorithm (1): DNA Postfix Method

Input: message (M), DNA cover (S)
Output: DNA cover (S')

1: read the message (M)
2: convert a message to a binary (BinMsg)
3: convert BinMsg to a DNA codon (C)
4: read DNA sequence (S)
5: embedding process
  5.1. ignore (ATG, TGG, TGA) codons
  5.2. case (M)
    (M)= [A]
      If codon postfix = (A) or (G)
        (c') = [A]
      Else
    (M)= [G]
      If codon postfix = (G) or (A)
        (c') = [G]

    End
    (M) = [T]
    If codon postfix = (C) or (T)
      (c') =[T]
else
    (M) = [C]
    If codon postfix = (C) or (T)
(c') =[C]
end
end embed process
6: get next codon
7: $S' = C'1C'2C'3C'4 \cdots\cdots C'p$

Algorithm (2): S' Retrieval Method

input: DNA sequence (S')
Output: Message (M)

1 Read S'
2 retrieving process
  2.1 for each codon in the sequence S'
  2.2 ignore (ATG, TGG, TGA) codons
  2.3 If current codon postfix = (A) or (G)
  M = [A] or [G]
  else
  M = [C] or [T]
  2.4 Next codon
3 Apply the steps in algorithm (1) in inverse order to retrieve M
4 Return M

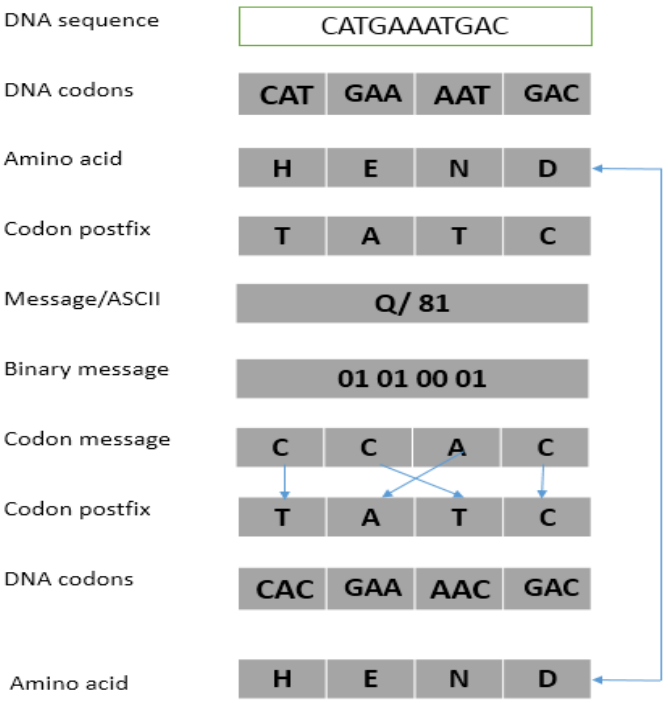

FIG (4): A DNA POSTFIX EXAMPLE

IX. EXPERIMENTAL RESULTS

The experimental results is illustrated as follows.

9.1 Capacity

The capacity of a media is measured by the maximum amount of information that can be hidden in it. In our method the media is DNA and measured in bits per codon (bits/codon) or bits per nucleotide (bpn)and computed using equation (1).

$$capacity = \frac{Maxsizeof \text{ DNA Cover } bits}{DNAsequancelengthin\ bases} \quad \ldots\ldots \quad (1)$$

$$= \frac{(1/3) \times |n|}{|n|} = \frac{1}{3}\ [bpn]$$

$$= 1\ [bit/codon]$$

|n| is a length of the DNA cover, i.e. the number of nucleotides consisting the DNA sequence**.**

Table (3) illustrate capacity comparison between our method and ([11], [12], [13] [14]) methods. Our method recorded high capacity.

Table (3):Comparative results with the related works

Title	capacity (bit/codon)
M.Hassan [10]	%74.32
Adnan Gutub [11]	%33.68
F.AL-Azawi [12]	%18.019
Adnan Abdul-Aziz [13]	%10.25
The Proposed	%90.25

9.2 Mutation

Mutation relates to (Mutation/Base and Bit Error Rates (BER)) of the extracted DNA Sequence, and computed as in an equation (2):

$$\frac{Extractednumberbitsfrom Mutated\boldsymbol{S'}}{Extractednumberbitsfromoriginal\boldsymbol{S}} \quad \ldots\ldots (2)$$

S represent the original DNA sequence, S' represent the mutated DNA sequence.

Table (4) shows different Mutation/Base and BER of the extracted DNA Sequences.

Table (4): Mutation/Base and BER of DNA Sequences

Sequence ID	Sequence Length	Message Length	Mutation/Base	Base change rate (%)
BC094877	1304	200	50	3%
NM000520	2751	600	180	6%
NM001179490	1584	460	200	12%
AC167229	197711	60000	20000	10%
AC166259	198972	60000	30000	15%
JQ439993	256	70	33	12%

X. CONCLUSIONS

The new proposed method for hide the sending data in a DNA sequence using Postfix codon property has been confirmed according to the obtained results, The proposed method of data security and amino acid preservation and attain has achieved high capacity in data hiding.

REFERENCES

1. Gehani A, LaBean T, Reif J: DNA-based cryptography. In: Aspects of Molecular Computing. Springer; 2003: 167-188.
2. Chang C-C, Lu T-C, Chang Y-F, Lee R: Reversible data hiding schemes for deoxyribonucleic acid (DNA) medium. International Journal of Innovative Computing, Information and Control 2007, 3(5):1145-1160.
3. Khalifa A, Elhadad A, Hamad S: Secure Blind Data Hiding into Pseudo DNA Sequences Using Playfair Ciphering and Generic Complementary Substitution. Appl Math 2016, 10(4):14831492
4. Shiu H, Ng K-L, Fang J-F, Lee RC, Huang C-H: Data hiding methods based upon DNA sequences. Information Sciences 2010, 180(11):2196-2208.
5. Guo1Cheng, Chang2Chin-Chen, and Wang3 Zhi-Hui; "A NEW DATA HIDING SCHEME BASED ON DNA SEQUENCE", ICIC International, ISSN 1349-4198, Vol 8, No. 1(A), January 2012
6. Lee S-H: DNA sequence watermarking based on random circular angle. Digital Signal Processing 2014, 25:173-189.
7. Shimanovsky B, Feng J, Potkonjak M: Hiding data in DNA. In: International Workshop on Information Hiding: 2002. Springer: 373-386.
8. Carter Frank, Clelland ,Catherine "DNA-BASED STEGANOGRAPHY", 2001.
9. H. J. Shiu, K. L. Ng, J. F. Fang, R. C. T. Lee, and C. H. Huang," Data hiding methods based upon DNA sequences," Inf. Sci., vol. 180, pp. 2196-2208, 2010.
10. Khalifa and A. Atito. High-capacity dna-based steganography. In Informatics and Systems (INFOS), 2012 8th International Conference on,, pages BIO–76– BIO– 80, May 2012.

11. M. Hassan Shirali-Shahreza, Mohammad Shirali- Shahreza, "A New Approach to Persian/Arabic Text Steganography,"5th IEEE/ACIS International Conference on Computer and Information Science (ICIS-COMSAR 06), July 2006.
12. Adnan Gutub ,ManalFattani, "A Novel Arabic Text Steganography Method Using Letter Points and Extensions", WASET International Conference on Computer, Information and Systems Science and Engineering (ICCISSE), Vienna, Austria, May ,2007.
13. F. Al-Azawi, MoayadFadhil, "Arabic Text Steganography using Kashidaa Extensions with Huffman code", Asian network for
14. Science Information, 2010.
15. Adnan Abdul-Aziz Gutub, Wael Al-Alwani, and Abdulelah Bin, Mahfoodh"Improved Method of Arabic Text Steganography Using the Extension Kashida Character",Bahria University Journal of Information & Communication Technology, December, 2010.
16. S. Mona, H. Mohamed, N. Taymoor, and K. Mohamed Essam," A DNA and Amino Acids-Based Implementation of Playfair Cipher," International Journal of Computer Science and Information Security, vol. 8, pp. 126-133, 2010.
17. D. Zebari, H. Haron, and S. R. M. Zeebaree, "Security Issues in DNA Based on Data Hiding : A Review Security Issues in DNA Based on Data Hiding : A Review,". International Journal of Applied Engineering Research ISSN.vol. 12, no. December, pp. 15363–15377, 2017.
18. M. Torkaman, N. Kazazi, and A. Rouddini, "Innovative approach to improve hybrid cryptography by using DNA steganography," Int. J. New …, vol. 2, no. 1, pp. 225–236, 2012.
19. M. Borda and O. Tornea, "DNA secret writing techniques," 2010 8th Int. Conf. Commun. COMM 2010, no. May, pp. 451–456, 2010.

AUTHORS PROFILE

Name: GhaithDaaiKadhum
University: Babylon
Collage: Science for Women
Department: Computer Science
e-mail: ghaith90daai@gmail.com
Certification: B.Sc. from Al-MustasiriyahUniversity, MSc in Computer Science from University of Babylon
Specialize : bioinformatics DNA computing Interest Fields: Security, steganography in DNA ,magic cube

Name: SaharAdillKadhum Al-Bawee
University: Babylon
Collage: Science for Women
Department: Computer Science
e-mail: salaa_38@yahoo.com
Certification: PHD in Computer Science from Higher Studies Institute for computer & Informatics
Interest Fields: Security, secure DNA in bioinformatics, network

Assit.Prof.Dr.Mahdi Nsaif Jasim, university of information technology and communications, college of Business Informatics Dept. of Management information systemsBorn in Babylon / Iraq, lives in Baghdad.
Interest:information systems , data and information security, mining in vector data, GIS, database systems....The researcher has interrest in SDN data acquisition and data processing. He also Supervised a number of PhD and MSc. Students in different Iraqi universities