Available online at sjuoz.uoz.edu.krd

journals.uoz.edu.krd

Science Journal of University of Zakho

Vol. 7, No. 4, pp. 184–189, December.-2019

p-ISSN: 2663-628X
e-ISSN: 2663-6298

NEW DATA HIDING APPROACH BASED ON BIOLOGICAL FUNCTIONALITY OF DNA SEQUENCE

Abdullah A. Abdullah [a] *, Adel S. Eesa [a], Ahmed M. Abdo [a]

[a] Dept. Computer Science, Faculty of Science, University of Zakho, Kurdistan Region, Iraq (abdullah.abdullah; adel.eesa ; ahmed.abdo)@uoz.edu.krd,

Received: Sep., 2019 / Accepted: Nov., 2019 / Published: Dec.,2019 ***https://doi.org/10.25271/sjuoz.2019.7.3.612***

ABSTRACT:

Data hiding or steganography has been used ever since a secret message was needed to be transferred. Data hiding methods need a medium to be cover for secret message that is to be sent. Different mediums are used such as image, video, audio, and last decade the deoxyribose nucleic acid (DNA). In this paper, a new data hiding approach based on the DNA sequence is proposed. Unlike many existing methods, the proposed method does not change the biological functionality of the DNA reference sequence when the sequence is translated into amino acids. The proposed method is consisting of two steps: the first step is encrypting the message using the Toffoli quantum gate. The second step is embedding the encrypted message into DNA sequence by taking one codon at a time and considering amino acids' biological functionality during the embedding process. Experimental results show that the proposed method outperforms the existing schemes preserving biological functionality in terms of cracking probability, and hiding capacity for bit per nucleotide.

KEYWORDS: Steganography, DNA sequence, Toffoli quantum gate, Information hiding, information Security.

1. INTRODUCTION

The need for information transfer through networks rapidly increased in the last two decades since it is a faster and cheaper way to transfer data between two users. Although it is a convenient way to send data it has a major drawback which is a secure transmission of data (Dodis, Mironov, & Stephens-Davidowitz, 2016). Due to the expansion of computer networks, the number of hacking and intrusion incidents is increasing every year(Eesa, Orman, & Brifcani, 2015). For that, two different strategies have come to widely used to make secure data transfer which includes Cryptography and Steganography. Both cryptography and steganography are independent of each other but they can be used together side by side to the higher secure atmosphere for data transferring. Cryptography (Abdo, Sabry, & A., 2018) is a technique to encrypt data from the sender by converting a message to noisy data that is meaningless for the reader yet it can be recovered again to the original message. On the other hand, Steganography (Mstafa & Elleithy, 2017) is a technique to hide a message in a medium without changing much of the medium to prevent suspiciousness from it. Different mediums have been used for hiding purpose and they mostly include video (Ramadhan J Mstafa & Elleithy, 2017), audio (Kakde, Gonnade, & Dahiwale, 2015), and pictures (Jiang, Zhao, & Wang, 2016). However, such medium may alter or distort the medium to some level and thus may give a hint to an attacker that the medium is suspicious. In search of better and different mediums; research has been done on Deoxyribonucleic Acid (DNA) as a data hiding arrangement (Hafeez, Khan, & Qadir, 2014; Huang, Chang, & Wu, 2014; Leier, Richter, Banzhaf, & Rauhe, 2000; Shiu, Ng, Fang, Lee, & Huang, 2010). Some of the researches were based on chemical features biological DNA (Peterson, 2001)while others depend on the DNA sequence characteristics and structure(Huang et al., 2014; Shiu et al., 2010).

Quantum computing is a theoretical field of computer science that is based on quantum mechanics. Unlike traditional computing model, the quantum computing is based on the probability of state of an object and not the binary model which makes it exponentially faster and can factor probabilities in fraction of time that needed for classical computer. The qubits are used instead of classical bit to indicate the state of the used object(Gyongyosi & Imre, 2019). Toffoli gate (Toffoli, 1980) which is considered a quantum operator is a reversible gate that means it can contribute to any logical circuit to make a reversible gate and it has three inputs and outputs.

In this paper, a new robust hiding scheme is proposed. The proposed method does not change the biological functionality of DNA sequence and hence it has a low modification rate which makes the DNA sequence less suspicious by attacker. The cracking probability for the proposed method is high which makes it hard to detect the original message.

The rest of this paper is organized as follows: in section 2 and an overview of the DNA sequence presented. Section 3 briefly overviews the related work. Section 4 is the presentation of the proposed scheme for hiding a message. Section 5 illustrates there cover phase. Section 6 discusses the analysis of the performance of the presented scheme. In the last section, section 7 has the conclusion of this paper.

2. DNA SEQUENCE

DNA sequence is made from four labeled nucleotides which are: Adenine (A), Thymine (T), Cytosine (C), and Guanine (G), and when the nucleotide is not equal to each one of them they referred to as non-labeled nucleotide and DNA sequence may have few non-labelled nucleotides denoted by (N) as well (Huang et al.,

* Corresponding author

2014). The complementary rule for DNA sequence to make RNA (ribonucleic acid) and then protein based on its chemical structure and bonds are defined as A-T, T-A, C-G, G-C (Adams et al., 1991).

The four bases are important in the bio-system of living organisms. The organization of nucleotide defines the type of protein, and protein is responsible for all activities in living cells and different protein has different functionality (Adleman, 1994). DNA converts to RNA which is an intermediate process towards protein creation; the process is called transcription. The process of converting RNA to amino acid called translation which then creates proteins (Hartwell, Hopfield, Leibler, & Murray, 1999). In the translation process; the combination of three nucleotides is taken from RNA which called codon and each codon denotes an amino acid. There are twenty amino acids that derived from different codons and most of them came from multiple codons as it is shown in Table 1. In addition to amino acids, there are three STOP codons indicating the end of protein sequence (Hartwell et al., 1999).

Table 1. Codons list with their represented Amino acid

Codon	Amino acid	Codon	Amino Acid
AAA	Lys	CAA	Gln
AAC	Asn	CAC	His
AAG	Lys	CAG	Gln
AAT	Asn	CAT	His
ACA	Thr	CCA	Pro
ACC	Thr	CCC	Pro
ACG	Thr	CCG	Pro
ACT	Thr	CCT	Pro
AGA	Arg	CGA	Arg
AGC	Ser	CGC	Arg
AGG	Arg	CGG	Arg
AGT	Ser	CGT	Arg
ATA	Ile	CTA	Leu
ATC	Ile	CTC	Leu
ATG	Met	CTG	Leu
ATT	Ile	CTT	Leu
GAA	Glu	TAA	STOP
GAC	Asp	TAC	Tyr
GAG	Glu	TAG	STOP
GAT	Asp	TAT	Tyr
GCA	Ala	TCA	Ser
GCC	Ala	TCC	Ser
GCG	Ala	TCG	Ser
GCT	Ala	TCT	Ser
GGA	Gly	TGA	STOP
GGC	Gly	TGC	Cys
GGG	Gly	TGG	Trp
GGT	Gly	TGT	Cys
GTA	Val	TTA	Leu
GTC	Val	TTC	Phe
GTG	Val	TTG	Leu
GTT	Val	TTT	Phe

3. RELATED WORK

DNA is a good and resourceful medium for data security and transferring; it has great storage capacity and can outperform magnetic and optical disc for storage (Adleman, 1994; Bancroft, 2001). For that, many researchers have used DNA for the purpose of information hiding (Abdo et al., 2018; Artz, 2001; Peterson, 2001; Shimanovsky, Feng, & Potkonjak, 2003). DNA based information hiding is still attracting many research until it accomplishes these three criteria: 1- Invisibility, which hide information from attackers, 2- Capacity of hiding, which must have satisfactory size for message to hide, 3- Consistency, in which the medium must not be disturbed by hiding message (Hafeez et al., 2014). The part of hiding information in the DNA sequence is nucleotide. The hidden process can be done by manipulating nucleotides arrangement or altering. Although DNA nucleotide represents in English letters (European Nucleotide Archive), but converting it to binary makes the process of hiding data much easier. Leier (Leier et al., 2000) was the first to use binary representation of DNA. Peterson in 2001 proposed a scheme for encrypting data in DNA sequence by replacing a codon with a character or letter and then converts it to proper DNA sequence and hence he could use up to 64 characters. However, this scheme can easily be cracked due to the frequency of some English letters that are represented as one codon. Other researchers (Dagher et al., 2019) have been working on coding area of DNA sequence. However; such as non-labelled nucleotide; coding area is also having its drawback for information hiding because DNA sequence is not a random sequence and research discovered that less than 1% of DNA sequence in human body is different (European Nucleotide Archive). So, by altering too much of DNA sequence, attacker would notice that the DNA sequence is suspicious and does not match real biological DNA sequence.

Three data hiding scheme is proposed by Shiu (Shiu et al., 2010) based on DNA sequences. Three schemes were: insertion method, the complementary pair method, the Substitution Method. The insertion method inserts message bits to the sequence bits and then rearranges DNA sequence but it expands based on the size of the message. The complementary pair method takes complementary nucleotides based on the message bits; here the bit per nucleotide is low. The substitution method is done by comparing a message with a DNA sequence and taking their complementary based on the message; this method has a high modification rate. When modification rates increase; the sequence gets more suspicious by attackers. To overcome this problem; Huang (Huang et al., 2014) proposed a scheme with a low modification rate. Yet, the scheme had drawbacks such as low hiding capacity and the sequence cannot preserve the biological function. Hafeez (Hafeez et al., 2014) proposed a method to hide data in DNA sequence without changing gene biological function. Malathi (P, M, R, Raghavan, & R.E., 2017) proposed a method with high capacity however the distortion in the DNA sequence is also very high which make the sequence very suspicious.

Properties of DNA sequence such translation and transcription process that are early stage of protein making are used for cryptography (Kalsi, Kaur, & Chang, 2018; Patnala & Kiran Kumar, 2019) which depends on codons and their corresponding amino acid. The charastestics also been used on steganography based on amino acids (R.B., M.V., G.R., Johar, & G.S., 2019, p. 3).

4. PROPOSED METHOD

The proposed method is consisting of two steps: in the first step, a reversible Quantum gate is used to encrypt the message. In the second step, the encrypted message is embedded inside a DNA sequence. The details of these steps are described as follow:

Step1: Quantum gates are reversible gates which means inputs can be recovered based on outputs. In logic gate only NOT gate is reversible when output is zero the input must be one and vice versa. However, if the output of AND gate is zero the two inputs can be (00, 01, or 10) which means inputs cannot be detected based on outputs only. On the other hand, all quantum gates are reversible for instance controlled-NOT (CNOT) has two inputs (A, B) and two outputs (X, Y), the value of ‘X’ is equal to ‘A’ and the value of Y is equal to ‘A⊕B’ or ‘A’ XOR ‘B’. The proposed scheme used Toffoli gate (Toffoli, 1980), sometimes referred to as Controlled Controlled Not (CCNOT), Figure 1 describes three bits Toffoli gate which works as follows: when the first and second bit values equal one then the third one is changed. In the proposed scheme, the first two input bits (A, B) are outputs (P, Q) for the same gate. Only for the first gate inputs are taken directly from the binary message.

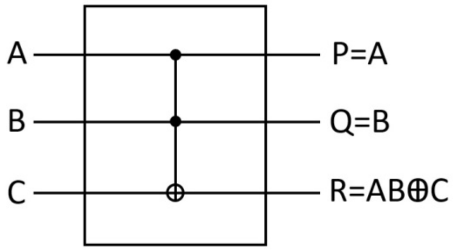

Figure 1. Toffoli quantum gate

An example of the proposed method works as follows: If the sender wants to send message M to the receiver, the sender first converts the message to binary representation. For example, the binary message BM is equal to 11001100. At the sender side, the BM is converted to BM' based on the Toffoli quantum gate. The first two bits remain as is; so, BM'=11. The third one alters if the previous two outputs bit from MB' are ones so the zero changes to 1 now BM'= 111. The fourth bits also changes because second and third bits on BM' are ones, now BM'= 1111. The fifth bits also changes because third and fourth bits are ones, so now BM'= 11110. The sixth bits remain unchanged because fourth and fifth bits of BM' are not both ones, now BM'= 111101. The same process is used for remaining bits the final BM'=11110100. The procedures work according to algorithm 1.

```
Algorithm 1:
Input:
        M: Original message
Output:
        Ciphered binary message (BM')
Method:
        1. Convert the message (M) into binary form
           (BM).
        2. Convert the binary message (BM) to a new
           ciphered binary message (BM') based on
           Toffoli Quantum interconnected gates.
```

Step2: In this part, the values of the amino acid are considered and the hiding method should not change the biological functionality of the DNA sequence. For hiding, the message bit is not directly injected to the DNA sequence, but rather it depends on the codon value itself. For embedding; codon value is checked if codon value is equal to (ATG) or (TGG) which represents Met, and Trp amino acid respectively, the value remains unchanged based on Table 3 because these two amino acids have just one codon to represent them and any change in codon will be ruled to another amino acid. Otherwise; the codon is converted to binary representation 6-bit based on Table 2 values for nucleotides. For example, when codon value equal to (CAT) the C will be substituted by 01, A will be substituted by 00, and T will be substituted by 11. Thus, the codon CAT will have a binary representation of 010011. Then the value of 6-bits binary is changed based on the number of even/odd bits of ones on codons and the change is based pairs exchange given in Table 3 base on the following rules when the message bit is equal to one: a) If the number of bits represented by 1 in 6-bits is an even number and the message bit equal to 1 then the 6-bits is changed based on complementary pair value. b) If the number of bits represented by 1 in 6-bits is an odd number and the message bit is 1 then the 6-bits remains unchanged. In contrast, the rules will work differently when the message bit is equal to zero as follow: a) If the number of bits represented by 1 in 6-bits is an even number and the message bit equal to zero then the 6-bits remains unchanged. b) If the number of bits represented by 1 in 6-bits is an odd number and the message bit is zero then the 6-bits are changed based on complementary pair value. The pairing exchange for nucleotides that represents the same amino acids is not surprisingly ending with (A, G) or (C, T) except for two codons (ATA) and (TGA) because the other pairs that end with flipped letters have different amino acids that can be represented by one codon only. Table 3 explains the pair exchange values in detail.

Table 2. The binary representation of nucleotides

Nucleotide	Binary representation
A	00
C	01
G	10
T	11

Table 3. Amino acids codon's exchange dictionary

Amino acids exchangeable codon pairs	
Amino acid	Codon pairs
Lys	(AAA, AAG)
Asn	(AAC, AAT)
Thr	(ACA, ACG) , (ACC, ACT)
Arg	(AGA, AGG), (CGA, CGG), (CGC, CGT)
Ser	(AGC, AGT), (TCA, TCG), (TCC, TCT)
Gln	(CAA, CAG)
His	(CAC, CAT)
Pro	(CCA, CCG), (CCC, CCT)
Leu	(CTA, CTG), (CTC, CTT), (TTA, TTG)
Glu	(GAA, GAG)
Asp	(GAC, GAT)
Ala	(GCA, GCG), (GCC, GCT)
Gly	(GGA, GGG), (GGC, GGT)
Val	(GTA, GTG), (GTC, GTT)
Tyr	(TAC, TAT)
Cys	(TGC, TGT)
Phe	(TTC, TTT)
Ile	(ATC, ATT)
STOP	(TAA, TAG)
Amino acids with one-way exchange codon pair	
Ile	ATA ➔ ATC
STOP	TGA ➔ TAA
Amino acids with a unique representation of codon	
Met	ATG
Trp	TGG

After the encryption process, the produced ciphered binary message is then embedded inside the selected DNA sequence. The main steps for the hidden process are described in algorithm 2.

```
Algorithm 2:
Input:
        M: Ciphered binary message (BM')
        F: DNA file
Output:
        Covered DNA file (DNA')
Method:
        1. Initialize the pairing dictionary based on Table 3.
        2. Find non-labeled nucleotides on DNA sequence to be
           excluded in the embedding process.
        3. Convert DNA sequence to codon array "three
           nucleotides each".
        4. If codon is equal to (ATA) or (TGA); go to the next
           codon array and repeat this step, otherwise go to next
           step.
        5. Convert the codon to binary 6-bits string, each letter
           represented by 2-bits based on table 2 and count
           number of ones on the binary number string.
        6. Take a bit from BM' and compare:
```

```
    • If BM' bit is zero and the codon in the array has an even
      number of ones, the codon stays the same. Otherwise,
      change the codon value based on the amino acid pair
      dictionary.
    • If BM' bit is one and the codon in the array has odd
      numbers of ones, the codon stays the same. Else,
      change the codon values the amino acid pair dictionary.
7. If BM' is not empty, take next bit in BM' and next value in codon array then go to step 3. Otherwise, go to the next step.
8. Convert the codon array to DNA' sequence.
9. Restore the non-labeled nucleotide to the DNA' sequence.
```

An example of the message embedding process works as follows. Suppose the DNA sequence for the embedding process is (TAC CCT CGA GAT GCA ATG ATA TTA ATC). To hide a bit, one codon (three nucleotides) is taken then convert codon to binary representation for the first codon binary value is equal to 110001 number of (1)'s in sequence is odd and the message bit is equal to (1) so the codon remains unchanged. The second codon is (CCT) the binary representation of it is 010111 the number of (1)'s in the sequence is even, for that the codon is changed based on the pairing dictionary in Table 3 and second codon is converted to (CCC) which has binary representation of 010101 and thus gets odd number of (1)'s. the same process goes for the remaining message bits and codons. Note, when codon number six (ATG) of DNA sequence is reached, it is discarded in the embedding process because it is not paired in table 3 of pairing amino acid, for that next codon is taken for embedding process.

5. MESSAGE RECOVERY PHASE

In this phase, the receiver has the DNA sequence that contains the hidden message as well as the DNA itself which is able to preserve its biological functionality. The recovery process works as follow:

```
Algorithm 2:
Input:
    Covered DNA file (DNA')
Output:
    M: Original message
    F: DNA file holds its biological functionality
Method:
    1. Find non-labeled nucleotide from DNA' sequence and exclude them from the recovery process.
    2. Convert DNA' sequence to codon array "three nucleotides each".
    3. Define empty string BM' to store binary data.
    4. If codon array is equal to 'ATA' or 'TGA'; go to the next codon array and repeat this step, otherwise go to next step.
    5. Find the number of zeros and ones in codon array,
        • If codon has even number of zeros and ones; insert '0' to BM' string.
        • If codon had odd numbers of zeros and ones; insert '1' to BM' string.
    6. If codon array is not null or the stop condition has not met, take the next codon array value and go to step 4.
    7. Define binary string BM to store original binary data.
    8. Take 3 bits from MB' from the end of string to the beginning. And inject them to the Toffoli gate to get the original value of the third bit and insert it to MB string. Repeat this step until reaching the beginning of the string.
    9. Convert the BM string to the original message M.
```

The recovery process for the given example in the proposed method is started by checking the codon and calculating the number of (1)'s in the DNA sequence. Firstly, codons that equal (ATG, TGG) discarded for the recovering process. Then if codon has even number of (1)'s in its binary representation then the message bit is zero; otherwise the message bit is one. After completing the process, the BM' can be found which in our case is equal to 11110100. After this process, the BM' is converted to BM by taking three bits in reverse order, in given example 100 if the first two bits are one then alter the third bit otherwise third bit remains unchanged. In the given example, it remains unchanged. Then the same process is done for the rest of the bits until the beginning of the binary message is reached by then the receiver can get the original binary message. The last step receiver simply converts the binary message to the original message.

6. EXPERIMENTAL RESULTS

The proposed scheme is tested and evaluated using different DNA sequences from (European Nucleotide Archive) database; details of the used dataset are described in Table 4. The recovery phase used in this work is not changing the functionality of DNA sequence when it translates to amino acids to produce protein in later stages. The results are compared to other works that used the same datasets (Hafeez et al., 2014, 2014; Huang et al., 2014; Shiu et al., 2010). For the evaluation and comparison purpose, two different measurements are considered which are hiding capacity, and cracking probability

Table 4. Dataset of DNA sequences imported from EBI.edu

DNA sequence	Number of nucleotides	Number of non-labeled nucleotides	Sequence definition
AC153526	200,117	0	Mus musculus 10 BAC RP23-383C2
AC166252	149,884	0	Mus musculus 6 BAC RP23-100G10
AC167221	204,841	0	Mus musculus 10 BAC RP23-3P24
AC168874	206,488	1,300	Bos taurus clone CH240-209N9
AC168897	200,203	5,186	Bos taurus clone CH240-190B15
AC168901	191,456	250	Bos taurus clone CH240-18511
AC168907	194,226	809	Bos taurus clone CH240-19517
AC168908	218,028	918	Bos taurus clone CH240-195K23

In general, hiding capacity determines the maximum size of bits that can be inserted to the medium based on methods that been used, in case of DNA sequence it is measured by the number of bits to the number of nucleotides in sequence or (bpn). The following equation is used for determining hiding capacity:

$$Hiding\ Capacity = \left(\frac{number\ of\ bits\ in\ message}{DNA\ senquence\ size}\right) bpn$$

The hiding capacity is slightly different from one DNA sequence to another because of the nature of the sequence. In the case of hiding one bit per nucleotide; the capacity will reach one-third of the total sequence size. In the presented scheme; it is slightly different than one third because non-labeled nucleotides are excluded and to maintain the biological functionality of used sequence; another two codons are excluded (ATA, TGA) because these two codons are unique when representing amino acids of sequence. Table 5 shows the hiding capacity for each sequence that been tested.

Table 5. Hiding capacity per DNA sequence in bps

DNA sequence	capacity	bpn
AC153526	200,117	0.3293
AC166252	149,884	0.3294
AC167221	204,841	0.3292
AC168874	205,188	0.3273
AC168897	195,017	0.3201
AC168901	191,206	0.3284
AC168907	193,417	0.3276
AC168908	217,110	0.3277

The cracking probability is the probability of attackers making a correct guess to discover a hidden message. Even though the presented work presented is simple to achieve; the cracking probability in the proposed method is high. Furthermore, the DNA sequence does not change its amino acid which makes the DNA sequence less suspicious. To successfully guess the original message; attackers need four types of information as follow: (1) the original DNA sequence that used as reference for the hiding purpose; (2) number of codon pairing used in proposed scheme with their potential representation of bit; (3) the codons that are excluded for embedding process; (4) how the original message is cyphered.

First: the number of publicly available DNA sequences nowadays approximately 163 million sequences. The probability of attackers making successful guess out of this number is $\frac{1}{1.63\times10^{8}}$

Second: number of pairing codons used is 31 pair the probability making correct guess is $\frac{1}{31}$ and the probability of guessing whether the original bit is zero or one is $\frac{1}{2}$ so this step will have the probability of $\frac{1}{62}$

Third: the probability of making successful guess of codons that are excluded ('ATA', 'TGA') for hiding process is $\frac{1}{64}$ for the first one and $\frac{1}{63}$ for the second one, the general it is $\frac{1}{64\times63}$

Fourth: in order to successfully guess the original message from the crypto message the attacker must guess and try all possible values for the message which will equal to exponent value and that is equivalent to $\frac{1}{2^{\mathrm{M}}}$

In general, the cracking probability is equal to:

$$Cracking\ Probability = \frac{1}{1.63\ \times 10^{8}} \times \frac{1}{62} \times \frac{1}{64 \times 63} \times \frac{1}{2^{M}}$$

Table 6 illustrates the performance of the proposed method compared with some existing work. The obtained result shown in Table 6 indicates that in addition to maintaining the biological functionality of the DNA sequence, the performance of the proposed method is much better when compared to the other schemes in term of cracking probability which makes the proposed method more robust than the other methods. In term of hiding capacity suing bit per nucleotide (bpn), the proposed method is obtained around one-third of the sequence size which means one bit can be hiding in every three nucleotides (one codon). Thus, the proposed method outperforms the existing method in term of hiding capacity when biological functionality is preserved. However, some methods can have higher hiding capacity but cannot preserve the biological functionality of the sequence as indicated in Table 6.

Table 6. Comparison of performance

Method		Hiding Capacity Average bpn	Cracking probability	Expansion	Biological functionality preserved
Shiu(Shiu et al., 2010)	Complementary pair	0.07	$\frac{1}{1.63\times10^{8}}\times\frac{1}{24^{2}}$	$\frac{M}{2}$	No
	Substitution	0.82	$\frac{1}{1.63\times10^{8}}\times\frac{1}{6}$	0	No
Huang(Huang et al., 2014)		0.0236	*Not given*	0	No
Hafeez(Hafeez et al., 2014)		0.268	$\frac{1}{1.63\times10^{8}}\times\frac{step\ size}{key\ length!\times X}\times\frac{1}{25}$	0	Yes
Malathi(P et al., 2017)		>1	$\frac{1}{1.63\ \times10^{8}}\times\frac{1}{24}\times\frac{1}{n-1}\times\frac{1}{2^{M}-1}\times\frac{1}{2^{s-1}}\times\frac{1}{2^{8m}}$	$\frac{M}{2}$	No
Proposed method		0.327	$\frac{1}{1.63\ \times10^{8}}\times\frac{1}{62}\times\frac{1}{64\times63}\times\frac{1}{2^{M}}$	0	Yes

7. CONCLUSION

This work used the combination of cryptography and hiding information techniques using DNA sequence as a medium to hide messages. The proposed scheme can hide data in reference sequence without modification on representative amino acids of the sequence nucleotides thus the biological functionality of the DNA sequence is preserved. In this work, no compression techniques are used and the technique is blind, thus, no extra information needs to be sent alongside the DNA sequence. In the proposed technique; the basic interconnected Toffoli quantum gates is used to secure the massage and make massage hard to be detected. For the embedding phase, the message bits did not inject to sequence directly, instead, it depended on the odd and even number of ones in codons based on the dictionary table to represent message bits. Experimental results indicate that the proposed method outperforms existing techniques that preserve the biological functionality of sequence and many other methods in terms of cracking probability, and hiding capacity.

REFERENCES

Abdo, A. M., Sabry, A., & A., A. (2018). A New Message Encryption Method based on Amino Acid Sequences and Genetic Codes. *International Journal of Advanced Computer Science and Applications*, *9*(8). https://doi.org/10.14569/IJACSA.2018.090872

Adams, M., Kelley, J., Gocayne, J., Dubnick, M., Polymeropoulos, M., Xiao, H., … et, a. (1991). Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science*, *252*(5013), 1651–1656. https://doi.org/10.1126/science.2047873

Adleman, L. M. A. (1994). Molecular Computation of Solutions to Combinatorial Problems. *Science*, *266*, 1021–2024.

Artz, D. (2001). Digital steganography: Hiding data within data. *IEEE Internet Computing*, *5*(3), 75–80. https://doi.org/10.1109/4236.935180

Bancroft, C. (2001). Long-Term Storage of Information in DNA. *Science*, *293*(5536), 1763c–11765. https://doi.org/10.1126/science.293.5536.1763c

Dagher, G. G., Machado, A. P., Davis, E. C., Green, T., Martin, J., & Ferguson, M. (2019). Data storage in cellular DNA: Contextualizing diverse encoding schemes. *Evolutionary Intelligence*. https://doi.org/10.1007/s12065-019-00202-z

Dodis, Y., Mironov, I., & Stephens-Davidowitz, N. (2016). Message Transmission with Reverse Firewalls—Secure Communication on Corrupted Machines. In M. Robshaw & J. Katz (Eds.), *Advances in Cryptology – CRYPTO 2016* (Vol. 9814, pp. 341–372). https://doi.org/10.1007/978-3-662-53018-4_13

Eesa, A. S., Orman, Z., & Brifcani, A. M. A. (2015). A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert Systems with Applications*, *42*(5), 2670–2679. https://doi.org/10.1016/j.eswa.2014.11.009

European Nucleotide Archive. (n.d.). Retrieved 6 1, 2019, from The European Bioinformatics Institute: https://www.ebi.ac.uk/ena

Gyongyosi, L., & Imre, S. (2019). A Survey on quantum computing technology. *Computer Science Review*, *31*, 51–71. https://doi.org/10.1016/j.cosrev.2018.11.002

Hafeez, I., Khan, A., & Qadir, A. (2014). DNA-LCEB: A high-capacity and mutation-resistant DNA data-hiding approach by employing encryption, error correcting codes, and hybrid twofold and fourfold codon-based strategy for synonymous substitution in amino acids. *Medical & Biological Engineering & Computing*, *52*(11), 945–961. https://doi.org/10.1007/s11517-014-1194-2

Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, *402*(S6761), C47–C52. https://doi.org/10.1038/35011540

Huang, Y.-H., Chang, C.-C., & Wu, C.-Y. (2014). A DNA-based data hiding technique with low modification rates. *Multimedia Tools and Applications*, *70*(3), 1439–1451. https://doi.org/10.1007/s11042-012-1176-z

Jiang, N., Zhao, N., & Wang, L. (2016). LSB Based Quantum Image Steganography Algorithm. *International Journal of Theoretical Physics*, *55*(1), 107–123. https://doi.org/10.1007/s10773-015-2640-0

Kakde, Y., Gonnade, P., & Dahiwale, P. (2015). Audio-video steganography. *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 1–6. https://doi.org/10.1109/ICIIECS.2015.7192885

Kalsi, S., Kaur, H., & Chang, V. (2018). DNA Cryptography and Deep Learning using Genetic Algorithm with NW algorithm for Key Generation. *Journal of Medical Systems*, *42*(1), 17. https://doi.org/10.1007/s10916-017-0851-z

Leier, A., Richter, C., Banzhaf, W., & Rauhe, H. (2000). Cryptography with DNA binary strands. *Biosystems*, *57*(1), 13–22. https://doi.org/10.1016/S0303-2647(00)00083-6

Mstafa, R. J., & Elleithy, K. M. (2017). Compressed and raw video steganography techniques: A comprehensive survey and analysis. *Multimedia Tools and Applications*, *76*(20), 21749–21786. https://doi.org/10.1007/s11042-016-4055-1

P, M., M, M., R, M., Raghavan, V., & R.E., V. (2017). Highly Improved DNA Based Steganography. *Procedia Computer Science*, *115*, 651–659. https://doi.org/10.1016/j.procs.2017.09.151

Patnala, B. D., & Kiran Kumar, R. (2019). A Novel Level-Based DNA Security Algorithm Using DNA Codons. In Ch. Satyanarayana, K. N. Rao, & R. G. Bush, *Computational Intelligence and Big Data Analytics* (pp. 1–13). https://doi.org/10.1007/978-981-13-0544-3_1

Peterson, I. (2001). Hiding in DNA. *Proceedings of Muse*, *22*.

Ramadhan J Mstafa, & Elleithy, K. M. (2017). *Efficient and Robust Video Steganography Algorithms for Secure Data Communication*. https://doi.org/10.13140/rg.2.2.33095.50083

R.B., S., M.V., N., G.R., M., Johar, S., & G.S., H. (2019). DNA based Steganography Using 2-3-3 Technique. *2019 International Conference on Data Science and Communication (IconDSC)*, 1–6. https://doi.org/10.1109/IconDSC.2019.8816945

Shimanovsky, B., Feng, J., & Potkonjak, M. (2003). Hiding Data in DNA. In F. A. P. Petitcolas (Ed.), *Information Hiding* (Vol. 2578, pp. 373–386). https://doi.org/10.1007/3-540-36415-3_24

Shiu, H. J., Ng, K. L., Fang, J. F., Lee, R. C. T., & Huang, C. H. (2010). Data hiding methods based upon DNA sequences. *Information Sciences*, *180*(11), 2196–2208. https://doi.org/10.1016/j.ins.2010.01.030

Toffoli, T. (1980). Reversible computing. In J. Bakker & J. Leeuwen (Eds.), *Automata, Languages and Programming* (Vol. 85, pp. 632–644). https://doi.org/10.1007/3-540-10003-2_104