Mansoura Journal of Computers and Information Science

# A DNA Based Image Data Hiding Using 3D Logistic Chaotic Map

Zeinab Awad
Faculty of computers and information systems, C.S dep.
Mansoura University, Egypt
Zeinabaaa@yahoo.com

Magdy Zakria
Faculty of computers and information systems, C.S dep.
Mansoura University, Egypt.
Magdi_z2011@yahoo.com

Rasha Hassan
Faculty of computers and information systems, C.S dep.
Mansoura University, Egypt.
Yssf_hamid@yahoo.com

## ABSTRACT

Since steganography and cryptography are information security concerns, they gain wide importance in the researcher's studies. DNA based steganography is based on exploiting DNA nucleotides as confidential information carriers which enhance the data hiding process massively. A special kind of hiding mechanism is adopted in this work by using the DNA sequences that depend on iterating a 3D logistic chaotic map to generate three pseudo-random sequences. The proposed approach encodes the plain message bits together with two of the random keys into DNA sequences. It ensures extra DNA encryption, by carrying out some DNA sequence operations. Besides, it creates a fake sequence by embedding the generated sequence into another public reference sequence. The fake sequence is then hidden back into the cover image to achieve a double layer of data hiding. Experiments verifies that our approach satisfies the robustness of data hiding under reasonable average embedding capacity of 0.83 ,blindness, zero-payload, low modification rate of 5.16% (on average), zero-expansion rate, and high probability cracking.it is also considered as a double layer of security , dual hiding , not pure data hiding system, with adequate visual imperceptibility.

## General Terms

Image Steganography, Cryptography, Security, DNA, Chaotic maps.

## Keywords

DNA, Security, Image Steganography, Cryptography, Chaotic Maps, Cracking probability, Embedding Algorithm, Extraction Algorithm.

## 1. INTRODUCTION

Steganography is the art of hiding the secret information within the cover media in an undetectable manner. The steganography main goal is to conceal every presence of the secret data within cover media [1]. Various kinds of covers such as audio, videos, and images are utilized depending on numerous factors such as capacity, security level, robustness, and the encoding and decoding time. Recently, several sorts of modern steganography techniques and concealer media are investigated by scientists to improve system security.

Using image steganography, the confidential information could be concealed within the cover image by many algorithms that greatly change and enhance the system security. Due to low image embedding capacity, a large amount of information could not be hidden within it. DNA steganography is introduced to compensate for its capacity shortage.

Steganography based on DNA is considered as one of the innovative techniques in the area of image steganography. The distinctive features that are allowed by DNA molecules such as superb energy efficiency, exceptional information density, and massive parallelism could be exploited for cryptographic purposes. Due to its low visibility property, it is very difficult for an intruder to recognize the real DNA from the fake one.

In 1994 Adleman [2] initiated the work in DNA computing by finding a solution to a trivial example of a Hamiltonian path problem. Clelland et al[3] present a mechanism for concealing messages containing DNA sample included in microdot within the sequence of a normal letter. In 2000 Leier and et al[4] mapped DNA nucleotides into binary bits. In [5] more powerful data hiding methods that deploying DNA sequences were suggested.

Even though the work which utilizing natural DNAs is very valuable and brings up a new paradigm of hiding techniques, they may have downsides in terms of the errors that may emerge according to the mutation and hard of DNA system implementation[6]. Using the abstract DNA sequences model and utilizing its natural characteristics to reinforce the current

steganography techniques is a possible solution to this problem.

This paper proposed an algorithm that depends on generating three different 3D-chaotic map keys and utilizing them for embedding the confidential text message bits within another DNA carrier sequence available publicly from DNA databases to produce the faked one. The faked sequence is then hidden back within the original image to produce the output stego image. to recover the original text message the steps in the embedding algorithm are reversed.

In section 2 the most common basics of biomolecular technology background have been presented, whereas most recent literature studies have been discussed in section 3. Section 4 illustrates the proposed approach including the explanation charts and the extra details. Sections 5 consists of the evaluation of the designed experiment, analysis the results and discuss the results in terms of comparisons verse recent approaches. Finally, the conclusion of our work and future recommendations are listed.

## 2. BIOMOLECULAR TECHNOLOGY BACKGROUND

For better conception and understanding of the proposed algorithm we introduce a short description of the DNA associated terms and technology:

### 2.1 DNA

Genetic information is saved in chains of DNAs in living organisms cells. A single DNA molecule has four nucleotides: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). Watson-Crick base pairing is a standard complementarily rule that[6] makes a hydrogen bond between A and T; or C and G nucleotides.

The DNA is the Deoxyribonucleic acid and each of its nucleotides combines six tiny molecules inside it, a phosphate molecule, a sugar molecule called deoxyribose, and four distinct nitrogenous bases (Adenine, Thymine, Cytosine, and Guanine) that form a strand sequence. Because of the huge size of the DNA sequences they provide a large embedding capacity to conceal a big data inside of it. Techniques like DNA substitution, DNA insertion, and DNA complementary pair are utilized to hide the confidential information inside the DNA sequence.

### 2.2 DNA Encoding scheme

It is a mapping rule that is used to map the four DNA nucleotides mentioned above into a binary form to facilitate the computation with DNA nucleotides. Accordingly, C could be represented by 0(00), similarly G by 1(01), A by 2(10), and T= 3(11)[7]. So we have 4! Potential coding rules using this encoding configuration. Table (1) lists one of the possible encoding rules.

### 2.3 DNA arithmetic and logic operations

We can perform traditional arithmetic addition and subtraction operations on DNA sequences the same as in Z2 (i.e. mod 2)[8]. For example, 11 + 10 =01, so if (A, C, T, G) = (0, 1, 2, 3), we have G + T = C and C-T=G. Similarly, bitwise logic operations like XORing, ORing and ANDing can be carried out on the DNA sequences.

| Binary Value | DNA base |
|---|---|
| 00 | C |
| 01 | G |
| 10 | T |
| 11 | A |

**Table 1.Mapping of DNA bases into binary form.**

### 2.4 DNA complementary rules

In complementary pair rule, every nucleotide base is assigned to a unique equivalent pair. accordingly, there are six valid complementary pairs available as shown in figure 1[9].

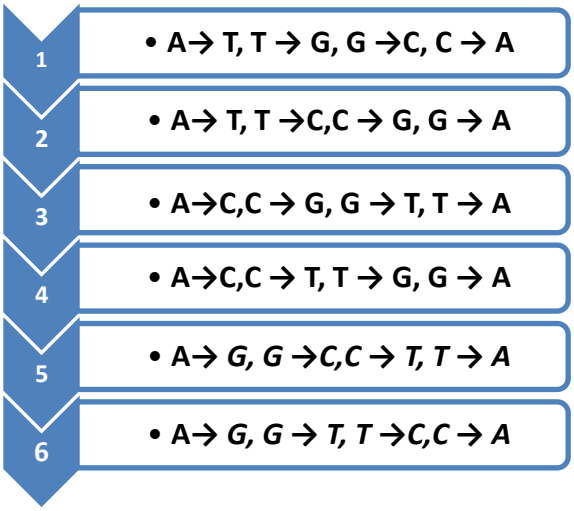


**Figure 1.The six legal complementary rules**

The complementary rule should satisfy that, for each nucleotide xi in the nucleotide string,
$x_i \neq B(x_i) \neq B(B(x_i)) \neq B(B(B(x_i)))$

$x_i = B(B(B(B(x_i))))$

where $B(x_i)$ is the base pair of $x_i$, which can ensure the complementary rule of injective mapping.[10]

## 3. RELATED WORK

Steganography based on DNA sequences obtained an increased interest among researchers for more than a decade. This is accomplished by using the mathematical and theoretical DNA prototype and exploiting the multiple distinct features of real DNA, like its inherent low visibility, massive storage capacity, and its unique ability to be synthesized in any length to provide a more efficient and faster information hiding algorithm. Below, we are going to highlight some of them:

Peterson et al [11] suggested a method to secrete the confidential data into DNA sequence by consecutive DNA bases substituting. 64 symbols were utilized for encodings such as D=CAG and S=AGC, and so on. It is not efficient to

implement since the attacker will concentrate on the frequently appearing letters to crack the encoded message.

Three DNA based hiding methods were introduced by Shiu et al [5] which are substitution, insertion and complementary methods. In the insertion method, the binary form of the secret data is embedded at the beginning of the selected DNA sequence binary equivalent, and then the substitution mechanism is applied with the complementary rule to substitute the secret message by the sequence bases. The complementary rule is given by ((AC) (CG) (GT) (TA)) is used to embed and hide the secret data in the DNA sequence. The performance evaluation of the above three methods is depending on measuring the payload, capacity and BPN (Bits Per Nucleotide) values. To check security obtained by the above algorithms the cracking probability of each of them is estimated. It is found that the insertion method has the minimum cracking probability among the other methods.
Mohammad et al[12] suggested mechanism for sharing the resources in the cloud environment and ensuring that the publicly available data remains secured by converting it to binary format and using the complementary algorithm for encoding it to DNA by complementing the carrier sequence bases with the DNA converted message bases to produce the faked sequence which is uploaded into the cloud. To recover the data at the client side the inverse of the complementary algorithm is adopted.

. Amal et al in [13]proposed a DNA-based steganography approach incorporated with a DNA based cryptography method to interchange the secret data through DNA carriers securely. The plain text message is encrypted by DNA based play fair cipher and Amino acid & then a unique substitution complementary algorithm is applied to conceal the DNA based cipher within another reference sequence.

Wang et al[14] presented a data hiding technique that depends on DNA sequences, in their work he encrypts the secret data using the vigenere cipher to enhance its confidentiality. The generating cipher is then transformed into a binary representation and then concealed into a DNA sequence by some DNA encoding method.
Samir et al in [1] presented an approach of image steganography that depends on concealing the secret image within the cover image by replacing its pixels values with different 256 DNA base combinations created by 4 nucleotides.

In [15]Secure Blind Data Hiding algorithm was proposed .the secret message is encrypted before hiding using the Playfair cipher and the secret data concealed using a Generic Complementary Substitution algorithm. It has high embedding capacity, support for high robustness method, and high cracking probability but it has the disadvantage of high modification rate and redundancy and the payload does not equal to zero.

in [16] a new data hiding algorithm that allows for high embedding capacity was suggested .initially it Encrypts the secret data by vigener or Playfair cipher. After hiding DNA reference will be sent in a microdot in a Paper before sending it to the receiver. If the paper is contaminated The algorithm regenerate another key and DNA sequence.it provides a double amount of data hiding with high security. as well as Preserve the functionality of DNA sequence and avoid any mutations.

But it has some weak points of the Unblind algorithm. The need to Send multiple data to the receiver for the extraction process. and High modification rate in the non-coding region.
Malathi et al[17]., implement an Improved, not pure DNA Based Steganography scheme that allows for High security., Double random key generator. And very high Probability cracking, but it has the drawbacks of the Unblind algorithm., Does not preserve the functionality of DNA and the Payload not equal to zero.
Gururaj Maddodi1 et al[18] implement a new combined neural network and chaos-based pseudorandom sequence generator and a DNA-rules based chaotic encryption algorithm for secure transmission and storage of images. The proposed scheme uses a new heterogeneous chaotic neural network generator controlling the operations of the encryption algorithm pixel position permutation, DNA-based bit substitution and a new proposed DNA-based bit permutation method. The randomness of the generated chaotic sequence is improved by dynamically updating the control parameters as well as the number of iterations of the chaotic functions in the neural network.

## 4. The Proposed Data Hiding Algorithm

Initially, the algorithm accepts the secret message then converts it into equivalent ASCII values and then into a binary form $M_{bin}$ then encodes it into a DNA ($M_{DNA}$) sequence strand according to specific DNA encoding rule as in (A=00, C=01, G=10, T=11). The proposed algorithm uses a 3D logistic chaotic map to generate three different random chaos key sequences of the same length as the secret message. Two of them are encoded into two different DNA sequences using a similar encoding rule. The first DNA sequence is xored with DNA encoded message ($M_{DNA}$) and the resultant sequence is then added with the second chaos DNA sequence. The generated DNA sequence Message is embedded in another DNA carrier sequence selected from public several GENE BANK databases web sites such as EBI(the European Bioinformatics Institute) database or NCBI(National Center for Biotechnology Information 2014). The embedding data hiding process depends primarily on substituting specific DNA bases of the reference sequence with another base from the generated DNA sequence encoded message according to certain complementary rule. The substitution positions are selected randomly based on the third generated chaotic sequence that provides a random permutation order with the same length as the sequence encoding a message to produce the falsify fabricated DNA. The fabricated DNA is then hidden back randomly inappropriate cover image pixels to maintain a double hiding layer for the secret data. The procedures to retrieve the hidden secret text from the carrier cover image are inverse of the steps performed in the embedding algorithm.

The entire process consists of the following steps:

### 4.1 3D Chaos Generation

The integration of steganography with chaotic theory plays a vital role in information security. This is due to chaos some distinctive unique properties like sensitivity to initial control parameters, high randomness, and non-periodicity. The simplest process of chaos production or initiation is the logistic map and is given by an equation:

$X_{n+1}= \mu x_n (1-x_n)$ (1)

To make this equation chaotic, the condition $0 < x_n <1$ and $\mu \epsilon [0,4]$ should be satisfied. To reinforce and improve security, Hongjuan Liu. et al [10]introduced the 2D logistic map depending on the quadratic coupling. In its extended 3D version is suggested in [19] by using the next formula:

$x_{n+1}=\gamma x_n (1-x_n) +\beta y_n^2 x_n+\alpha z_n^3$ (2)

$y_{n+1}=\gamma y_n (1-y_n)+\beta z_n^2 y_n+\alpha x_n^3$ (3)

$z_{n+1}=\gamma z_n (1-z_n)+\beta x_n^2 z_n+\alpha y_n^2$ (4)

When $3.53<\gamma<3.81$, $0<\beta<0.022$, $0<\alpha<0.015$ and the values of x, y, z initially lies between 0 and 1, the previous equations show the chaotic behavior. The existence of quadratic coupling, cubic in addition to three constant terms ensures that the 3D logistic map even more robust and intricate. Figure 2. (a-1), (b-1), and (c-1) indicate the histograms of x, y, z chaos sequences values generated by the above equations with values of x(1)=0.2350; y(1)=0.3500; z(1)=0.7350; $\alpha$=0.0125; $\beta$=0.0157; $\gamma$=3.7700 respectively.

## 4.2 Chaos Histogram Equalization

To ensure a higher security level and to improve the system robustness we require to equalize the x, y, z values histogram respectively. Thus for a color image of size M×N such that M is the rows number and N is the Columns number the following formulas are used to equalize x, y and z values:

$X = \text{int}(XN_1) \bmod N$ (5)

$Y = \text{int}(YN_2) \bmod M$ (6)

$Z = \text{int}(ZN_3) \bmod 512$ (7)

To simplify consider $N_1$, $N_2$, $N_3$ are equal large random values mostly above 10000. Figure 2. (a-2), (b-2) and (c-2) indicate the equalized x, y, z histograms at $N_1=N_2=N_3=$ 100000, M=512, N=512.

## 4.3 Embedding Algorithm

The algorithm allows us to secretly hiding the message bits inside the selected DNA reference sequence available from public GENEBANK databases web sites such as EBI (the European Bioinformatics Institute) database or NCBI (National Center for Biotechnology Information 2014).the generated fake sequence is then concealed within the carrier image. before the embedding operation, the secret text data bits are encrypted into a DNA sequence giving another layer of security. The embedding technique uses the six initial condition parameters for a 3D chaotic map along with the number of iterations K as a key.

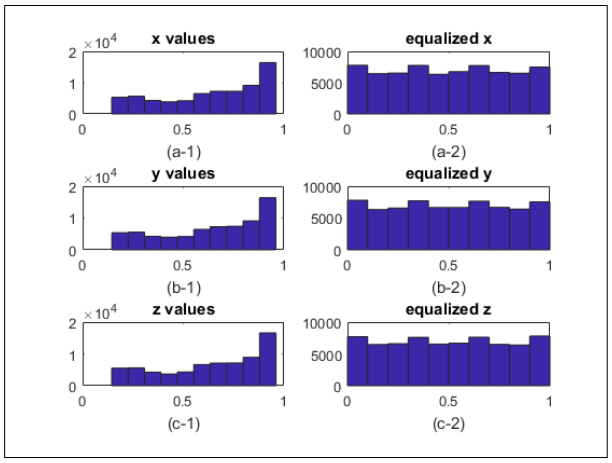


Figure 2. The equalized x, y, z histograms at N1=N2=N3= 100000, M=512, N=512.

Mainly the entire embedding algorithm includes the following steps:

Input: The original cover image, plain text message in binary form $M_{bin}$ of length m , $R_{DNA}$ which is DNA reference sequence of length n where (m <= n) , the DNA encoding rule ,complementary rule , initial chaotic parameters , random number generator seed SD, the number of iterations K,the offsets $O_x$,$O_y$ .

Output: Faked DNA sequence with the concealed secret text message FM, stego image.

Step1:Encode $M_{bin}$ into a DNA sequence $M_{dna}\{m_1,m_2,m_3,\ldots m_{m/2}\}$ of length |m/2| using the identified DNA encoding rule given, Get the length n of $R_{DNA}\{s_1, s_2, s_3, \ldots, s_n\}$.

Step 2: Iterate the 3d logistic chaotic map K times to generate three distinct chaotic sequences X, Y, Z.

Step 3: To improve the security level equalize the generated chaotic sequences X, Y, Z and select distinct integers from X, Y of the same length as $M_{dna}$ starting from the offset $O_x$, $O_y$ encode them into two DNA sequences $X_{dna}$, $Y_{dna}$ using the same DNA encoding rule.

Step4: Xor $M_{dna}$ sequence with the $X_{dna}$ sequence to produce the $Mx_{dna}$ sequence.

Step5: Perform DNA sequence addition between the output $Mx_{dna}$ and $Y_{dna}$ to produce the sequence $Mxy_{dna}$.

Step 6: Select a random unique permutation order $PO_z$ from Z with the same length as $Mxy_{dna}$ to identify the substitution positions in the reference sequence $R_{DNA}$.

Step 7: Define j, and i initially to be 1. Define FM to be an empty string to hold the output fake sequence to be hidden in the cover image.

Step 8: For each i in $R_{DNA}$, do the following:

if (i is equal to $PO_z(j)$ and $Mxy_{dna}(j)$ = 'A') such that ($1 \le j \le |m|/2$)) then fm(i) = Comp(Comp($Mxy_{dna}$( j )); // Apply the complementary rule

else if (i is equal to $PO_z(j)$ and $Mxy_{dna}(j)$= 'T')

then fm(i) = Comp(Comp($Mxy_{dna}(j)$) ;

else if (i is equal to $PO_z$(j )and $Mxy_{dna}(j)$ = 'C')

then fm(i) = Comp(Comp($Mxy_{dna}$(j )) ;

else if (i is equal to $PO_z(j)$and $Mxy_{dna}(j)$= 'G')

then fm(i )= Comp(Comp($Mxy_{dna}(j)$) ;

FM = FM + fm(i)

Step 9: Decode resulted fake sequence FM from the above loop with a DNA decoding rule and use the steganography encoding algorithm to hide them randomly into the cover image pixels using the random number generator with seed SD to produce the stego image .

## 4.4 Extraction algorithm

The data retrieval process is based on the following steps:

Input: The stego image, the reference(carrier) DNA sequence $R_{DNA}$ with length n, DNA encoding rule and complementary pair rule used in the embedding process, two chaotic DNA encoded sequence $X_{dna}$, $Y_{dna}$, chaotic-based permutation order $PO_z$ with length| m/2 |, random number generator seed SD.

Output: Secret plain message M.

Step 1: Use the steganography decoding algorithm to retrieve the hidden fake sequence from the stego image based on the given seed SD.

Step 2: Encode the extracted sequence into a faked DNA sequence FM of length (n) containing the hidden secret Message.

Step 3: Define i and j with an initial value of 1.

Step 4: For i from 1 to n do the following:

if (i is equal to $PO_z(j)$ and Comp(Comp(fm(i) ))= 'A' such that ($1 \le j \le |m|/2$))

then set $Mxy_{dna}$(j ) to 'A' , and increment j

else if (i is equal to $PO_z(j)$ and Comp(Comp(fm(i) ) = 'T')

then set $Mxy_{dna}$(j )to 'T' , and increment j

else if (i is equal to $PO_z(j)$ and Comp(Comp(fm(i))) = 'C'

then set $Mxy_{dna}(j)$ to 'C' , and increment j

else if (i is equal to $PO_z(j)$ and Comp(Comp(fm(i))) = 'G'

then set $Mxy_{dna}$(j )to 'G' , and increment j

Step 5: obtain The resulted $Mxy_{dna}$ sequence by joining all $Mxy_{dna}$ps where ($1 \le p \le j - 1$).

Step 6: perform DNA sequence subtraction between the output $Mxy_{dna}$ and $Y_{dna}$ to produce the sequence $Mx_{dna}$.

Step 7: Xor the $Mx_{dna}$ sequence with the $X_{dna}$ sequence to produce the $M_{dna}$ sequence.

Step 8: Decode the generated sequence $M_{dna}$ into the message $M_{bin}$ of length m using the given identified DNA decoding rule.

Step 9: Convert $M_{bin}$ into its equivalent ASCII form to get the original message M.

## 5. Performance analysis

The performance is measured in terms of certain evaluation measures such as payload, capacity, and BPN (bit per nucleotide) for verifying the exactness and confidentiality of concealed secret data within the DNA carrier sequence.

## 5.1 Capacity

It is determined by expanded DNA sequence overall size after the embedding of the confidential hidden information [7].here the carrier sequence capacity is given by:

Capacity = |S| (1)

## 5.2 Payload

It is the new sequence length after extracting out the DNA reference sequence from the fake reference sequence.

Payload = 0 (2)

Which means that the length of the fake reference sequence does not expand anymore, And hence the expansion rate(ER) of the actual reference sequence is equal to zero, which is very important in an minimizing the distortion of the cover media in order not to draw the intruder attention to detect the secret plain message . The modification rate is the second measure used to evaluate the quality of the fake DNA sequence. The modification rate (MR) could be described as follows:

$$MR = \frac{r_i \oplus f_j}{R_l}$$

Where $r_i$ is the i-th binary bit of the binary sequence of the reference DNA sequence, fj is the j-th binary bit of the binary sequence of the fake DNA sequence, and $R_l$ is the reference sequence length. Table (3) shows a Comparison of the modification rate (MR) of DNA sequence among Shiu et al.'s two methods [5] and our method when embedding 2500 bytes in the reference sequence.

## 5.3 BPN

The total number of bits Concealed inside each DNA nucleotide (bit per nucleotide).

BPN = |M| / |S| (3)

Here,

M refers to a secret plain text message length.

S identifies the faked DNA Sequence size.

## 5.4 Cracking Probability

The total probability needed by an intruder to anticipate the secrecy information stashed within the reference DNA sequence is called the cracking probability. Hence, to break or interpret the hidden secret plain message the intruder needs to obtain the following information:

Since there are about 163 million DNA reference sequences publicly obtainable so first information required by an attacker to breach the confidential message is the DNA reference sequence size, and the possibility to anticipate it is:

1/ 1.63×108 (1)

With the four DNA bases A, C, G, T We have 24 distinct binary encoding rules that could be constructed using various Permutations of two bits, and hence the probability to predict the encoding rule is given by:

1 /4! (2)

The intruder needs to find the secret plain text length in addition to the carrier DNA sequence length to decipher the secret text, so it is possible to predict the plain message and reference DNA sequence length with probability:

1/ (n-1) (3)

Since there are six legal complimentary pair rules, so the probability for an attacker to recover the plain message when an algorithm conceals a secret data by selecting a specific complementary rule is given by:

1/6 (4)

In the introduced approach, we build a kind of an injective mapping to map the six possible complementary pair rules with the two probable bases bits of the binary encoding rules. Thus the number of the injective mappings is:

4 * 3 * 2 *1 = 4! = 24. (5)

The six 3D chaotic logistic map initial condition parameters and the number of iterations K should be known to the attacker to produce the chaotic sequences X, Y, Z. Thus, the number of guesses needed for finding those initial conditions along with the iteration no is:

7! (6)

To encode the confidential plain text inside the carrier sequence, DNA XOR operations are applied with a combination of:

$$\frac{1}{2^{8m}} \quad (7)$$

Similarly, The DNA addition operations are applied also to encode the confidential text within the reference DNA, with a combination of:

$$\frac{1}{2^{8m}} \quad (8)$$

So, the cumulative probability to discover the secret message inside the DNA reference sequence (cracking probability) based on the suggested approach is:

$$CP = \frac{1}{1.63\times10^{8}\times24\times7!\times24\times6\ \times(n-1)\times2^{\,8m}\ \times2^{\,8m}} \quad (9)$$

Where,

n: The reference DNA sequence bits number which is equal to the length counterfeit DNA sequence.

m: Is the number of the secret message bits ( plain message length ).

# 6. EXPERIMENTAL RESULTS

This section depicts the multiple experiments performed to evaluate the effectiveness of the proposed data hiding algorithm. It was examined on the MATLAB R2017a platform on Intel(R) Core (TM) i7- 4600(U) CPU @ 2.70 GHz personal computer with 8GB RAM. The data set of the real sequences were obtained from the NCBI database. The plain text M size is 1.5 KB and the 3D chaotic map initial state parameters for generating the chaotic sequences X,Y,Z respectively are of x(1)=0.2450; y(1)=0.3510; z(1)=0.7360; α=0.0135; β=0.0167; γ=3.7800. The random seed value is 80 and 2 bits binary coding rule is utilized in the simulation. Table-2 shows the resulted capacity, payload and BPN (bit per nucleotide), when 20100 bytes of data bits are concealed inside eight reference sequences. It is clear from the obtained results in Table-2 that the secret message nucleotides are substituted entirely in the reference sequence so the expansion rate (ER) of the real original sequence is equal to zero.

As shown in Table-2, since the data embedding scheme depends upon substituting the bases, its payload is zero, so it indicates that the fabricated sequence size does not expand any more after the embedding process and hence it prevents

| | Number of DNA bases | Payload | Capacity | BPN=\|M\|/C | Sender side Time(sec) | Receiver side Time(sec) |
|---|---|---|---|---|---|---|
| AC153526 | 200117 | 0 | 200,117 | 0.8035 | 2.49 | 3.518 |
| AC166252 | 149884 | 0 | 149884 | 1.0728 | 1.98 | 2.408 |
| AC167221 | 204841 | 0 | 204841 | 0.7850 | 2.616 | 3.385 |
| AC168874 | 206488 | 0 | 206488 | 0.7787 | 2.59 | 3.397 |
| AC168897 | 200203 | 0 | 200203 | 0.8032 | 2.411 | 3.175 |
| AC168901 | 191456 | 0 | 191456 | 0.8399 | 2.2380 | 3.101 |
| AC168907 | 194226 | 0 | 194226 | 0.8279 | 2.5200 | 3.310 |
| AC168908 | 218028 | 0 | 218028 | 0.7375 | 2.650 | 3.665 |

**Table 2.The proposed algorithm results obtained when hiding 20100 bytes within the selected DNA sequences**

attracting the third parity attention. The BPN is within the range [1.0728, 0.7375] which gives an acceptable embedding capacity (in bits) between 160,826 and 160,796. The proposed algorithm has a relatively high stable capacity that does not expand the reference sequence length. Finally, the time needed to conceal all the secret message bits at the sender side in a sec is in the range [1.98, 2.65] whereas at the receiver side is in [2.4, 3.6]. It is obvious from Table-2 that the execution time at both sides together with the capacity is based on reference sequence size utilized which means that the total execution time is directly related to the sequence size.

## 6.1 Comparisons

Table-3 differentiates the basic characteristics of the presented method, with the basic substitution and insertion methods mentioned in [5]. Based on the evaluation criteria showed in Table-3.we could verify that our proposed scheme overcomes the drawbacks that were found in the other schemes.

| Sequence-Id | Number of DNA bases | Insertion Method In ref[5] | Substitution Method In ref[5] | Proposed method |
|---|---|---|---|---|
| AC153526 | 200117 | 14.28% | 95.00% | 4.99% |
| AC166252 | 149884 | 18.76% | 93.33% | 6.66% |
| AC167221 | 204841 | 13.96% | 95.12% | 4.87% |
| AC168874 | 206488 | 13.86% | 95.16% | 4.84% |
| AC168897 | 200203 | 14.27% | 95.01% | 4.99% |
| AC168901 | 191456 | 14.89% | 94.78% | 5.21% |
| AC168907 | 194226 | 14.69% | 94.85% | 5.14% |
| AC168908 | 218028 | 13.16% | 95.41% | 4.58% |

**Table 3. Modification rate (MR) comparison among methods in ref[5] and the proposed method**

## 6.2 Histogram Analysis

Histogram exhibits exactly the distribution of each image pixel. The higher similarity between the carrier image histogram and the produced stego image histogram indicates little distortion happened after hiding the confidential data bits into the carrier image [20]. Figure 3(a) represents the cover image histogram, and Figure 3(b) represents the histogram of the stego image after hiding the secret message into the carrier image.

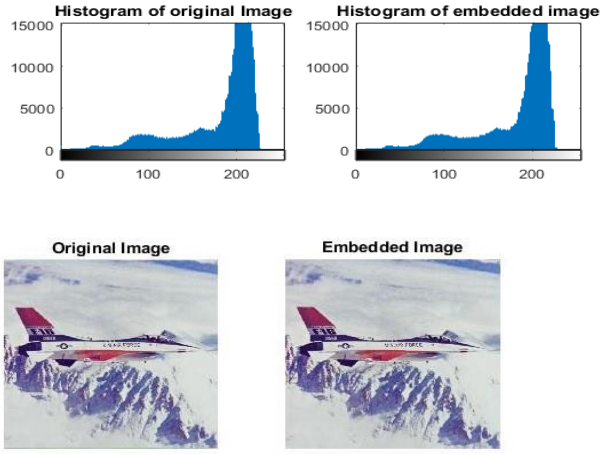


Figure (3-a) Figure (3-b)

Figure3. Histogram analysis (a) cover image (b) Stego image

| Criteria | The proposed scheme | The main substitution scheme [5] | The insertion scheme [5] |
|---|---|---|---|
| Capacity | \|S\| | \|S\| | \|S\|+\|M\|/2 |
| Preserve DNA biological function | No | No | No |
| Blindness(yes/No) | Exists | Not Exist | Exists |
| Payload | Zero | Zero | \|M\|/2 |
| Double layer of security (DNA encryption) | Yes | No | No |
| Cracking probability | $1/\ 1.63 \times 10^8 \times 24 \times 6! \times 24 \times 6 \times 1/(n-1) \times 2^{8m} \times 2^{8m}$ | $1/\ 1.63 \times 10^8 \times 6$ | $1/\ 1.63 * 10^8$ X 1/ 24 X 1/ (n − 1) X 1/ $(2^m - 1)$ X 1/ $2^{s-1}$ |
| Pure data hiding | No | Yes | No |

**Table 4. A comparison among the suggested scheme, the main insertion, and the main substitution schemes**

## 6.3 Visual Inspection Analysis

There is no distinction between the original cover image and the stego image that appeared through the visual inspection. Visually, they are the same without any clue about the secret message existence.

## 6.4 Peak Signal-to-Noise Ratio Analysis

The possible maximum signal power to the power of noise the signal includes ratio is called Peak Signal-to-Noise Ratio (PSNR)[20]. And it is defined as:

$$PSNR = 10log_{10}\frac{(2^d-1)^2}{MSE}\ \ (dB)$$

Such that d is the pixel bit depth and MSE defines the mean square error between the original cover and stego images. MSE and given by:

$$MSE = \frac{1}{nn}\sum_{i=1}^{n}\sum_{j=1}^{n}(p[i,j]-p'[i,j])^2$$

Here p[i, j], p'[i,j] are the pixel's ith-row, jth-column of the original image and the stego image respectively.

The steganography encoding algorithm is used to hide the decoded faked sequence FM randomly into the cover image pixels using the random number generator with seed SD to produce the stego image.

The steganography encoding algorithm could be explained as follows:

Input: the cover (carrier image), the decoded DNA faked sequence(FM), the random number generator seed(SD).

Output: the stego image with the secret message hidden inside.

Step 1: produce a random permutation set by intializing a random number generator with a seed SD.

Step 2: Rearrange the random permutation set into groups such that each group refers to the positions of three random distinct pixels in the cover image.

Step3: for each group of three pixels
- isolate each pixel in the group
- apply the RGB encoding for the first pixel
- apply the BGR encoding for the second pixel
- apply RG encoding for the third pixel.
- repeat

In Table-4, we have compared the PSNR value of our embedding scheme with similar values of LSB steganography schemes in references,[21],[22] and [23]. We observe that our technique has the highest PSNR values, and hence it proofs that the suggested DNA based image steganography technique has an acceptable security level.

| .Stego image | REF [21] | REF [22] | REF [23] | proposed |
|---|---|---|---|---|
| Airplane (F16) | 51.14 | 44.24 | 44.24 | 54.37 |

**Table 5.** The proposed method PSNR (DB) compared to other methods.

## 7. CONCLUSION & FUTURE WORK

In this work, a DNA based image steganography technique is suggested by consolidating means of steganography and cryptography in an attempt to obtain a double layer of security on the system as well as providing dual cover steganography to greatly enhance the security level. The first contribution is the use of 3D logistic chaotic map and exploits their inherent characteristics of several initial state parameters to enlarge the key size used which makes our proposed scheme more secured and further impenetrable by any third party and gives relatively higher cracking probability compared to other approaches. Furthermore using the chaotic system ensures the randomness and non-periodicity due to their extreme sensitivity of initial conditions in producing pseudorandom chaotic sequences that has the direct effect on encrypting the secret message and embedding it in the reference sequence so it massively supports the DNA cryptography sequence confusion. The second contribution is the blindness feature, in contrast to some other data hiding techniques we can extract the secret text without any knowledge about the original DNA carrier sequence. The third contribution is the improved efficient combined complementary pair and the substitution technique that result in zero-payload and relatively low modification rate algorithm which, evades attracting attention to the fake sequence and makes it unremarkable while improves the system's performance further as the generated fake sequence does not expand. Modifying the implemented system to maintain the biological functions of the output DNA sequence to achieve zero or low modification rate of DNA. Testing the system with several other parameters of imperceptibility, and developing a secure DNA key exchange and authentication protocol to transfer many keys between the sender and the receiver, since the entire process mainly depends on using many keys beginning from DNA encoding to data embedding, would be our main future research target.

## 8. REFERENCES


[1] Bandyopadhyay S K and Chakraborty S 2013 Image steganography using DNA sequence *ASIAN J. Comput. Sci. Inf. Technol.*

[2] Adleman L M 1994 Molecular computation of solutions to combinatorial problems *Science (80-. ).* **266** 1021–4

[3] Clelland C T, Risca V and Bancroft C 1999 Hiding messages in DNA microdots *Nature* **399** 533

[4] Leier A, Richter C, Banzhaf W and Rauhe H 2000 Cryptography with DNA binary strands *Biosystems* **57** 13–22

[5] Shiu H J, Ng K-L, Fang J-F, Lee R C T and Huang C-H 2010 Data hiding methods based upon DNA sequences *Inf. Sci. (Ny).* **180** 2196–208

[6] Najaf Torkaman M R, Nikfard P, Sadat Kazazi N, Abbasy M R and Tabatabaiee S F 2011 Improving Hybrid Cryptosystems with DNA Steganography BT - Digital Enterprise and Information Systems ed E Ariwa and E El-Qawasmeh (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 42–52

[7] Mousa H, Moustafa K, Abdel-Wahed W and Hadhoud M M 2011 Data hiding based on contrast mapping using DNA medium. *Int. Arab J. Inf. Technol.* **8** 147–54

[8] Wasiewicz P, Mulawka J J, Rudnicki W R and Lesyng B 2000 Adding numbers with DNA *S 2000 conference proceedings. 2000 IEEE international conference on systems, man and cybernetics. 'cybernetics evolving to systems, humans, organizations, and their complex interactions'(cat. no. 0* vol 1 (IEEE) pp 265–70

[9] Zebari D A, Haron H and Zeebaree S R M 2017 Security Issues in DNA Based on Data Hiding : A Review **12** 15363–77

[10] Liu H, Wang X and kadir A 2012 Image encryption using DNA complementary rule and chaotic maps *Appl. Soft Comput.* **12** 1457–66

[11] Peterson I 2001 Hiding in DNA *Proc. Muse* **22**

[12] Abbasy M R and Shanmugam B 2011 Enabling data hiding for resource sharing in cloud computing environments based on DNA sequences *2011 IEEE World Congress on Services* (IEEE) pp 385–90

[13] Khalifa A and Atito A 2012 High-capacity DNA-based steganography *Informatics and Systems (INFOS), 2012 8th International Conference on* (IEEE) p BIO-76-BIO-80

[14] Wang Z, Zhao X, Wang H and Cui G 2013 Information Hiding Based on DNA Steganography 946–9

[15] Khalifa A, Elhadad A, and Hamad S 2016 Secure blind data hiding into pseudo-DNA sequences using Playfair ciphering and generic complementary substitution *Appl. Math. Inf. Sci* **10** 1483–92

[16] Marwan S, Shawish A and Nagaty K 2017 Utilizing DNA Strands for Secured Data-Hiding with High Capacity. *Int. J. Interact. Mob. Technol.* **11**

[17] Malathi P, Manoaj M, Manoj R, Raghavan V and Vinodhini R E 2017 ScienceDirect Highly Improved DNA Based Steganography *Procedia Comput. Sci.* **115** 651–9

[18] Maddodi G, Awad A, Awad D, Awad M and Lee B 2018 A new image encryption algorithm based on heterogeneous chaotic neural network generator and DNA encoding

[19] Khade P N and Narnaware M 2012 3D chaotic functions for image encryption *Int. J. Comput. Sci. Issues* **9** 323

[20] Kordov K and Stoyanov B 2017 Least Significant Bit Steganography using Hitzl-Zele Chaotic Map *Int. J. Electron. Telecommun.* **63** 417–22

[21] Wang X, Zhao J and Liu H 2012 A new image encryption algorithm based on chaos *Opt. Commun.* **285** 562–6

[22] Amirtharajan R and Rayappan J B B 2012 An intelligent chaotic embedding approach to enhance stego-image quality *Inf. Sci. (Ny).* **193** 115–24

[23] Aziz M, Tayarani-N M H and Afsar M 2015 A cycling chaos-based cryptic-free algorithm for image steganography *Nonlinear Dyn.* **80** 1271–90