A New Image Encryption Algorithm based on DNA Approach

Ritu Gupta
Dept. Of CSE
Inderprastha Engineering College,
Ghaziabad, up, India

Anchal Jain
Dept. Of CSE
Inderprastha Engineering College,
Ghaziabad, up, India

ABSTRACT
In recent years, various DNA based cryptographic algorithms have been suggested to develop secure image encryption techniques but still many of them have low computing security as they have low avalanche effect and require to send long key . In this regard, this paper proposes a new method of image encryption based on DNA computation technology. The original image is encrypted using DNA computation and DNA complementary rule. First, a secret key is generated using a DNA sequence and modular arithmetic operations. Then each pixel value of the image undergoes the encryption process using the key and DNA computation methods. The researcher prove the validity of the algorithm through simulation and the theoretical analysis on the parameters such as sensitivity to plaintext, key sensitivity, histogram analysis, correlation analysis including bio-security and math security. Further, the algorithm has huge key space generated using key expansion algorithm while keeping the original key sequence small. It is shown the algorithm has achieved the satisfactory computing security level in the encryption security estimating system.

Keywords
DNA sequence; key generation; encryption; DNA addition; DNA complementary rule

1. INTRODUCTION
With the ever increasing growth of multimedia applications, security has become an important issue on communication and storage of images. In recent years, plenty of image encryption approaches have been proposed. Traditional encryption algorithms, such as DES, IDEA and AES etc. are not suitable for image encryption. DNA cryptography is emerging as a new cryptographic field where DNA is used to carry the information. The ability for huge storage and parallelism are making it suitable for image encryption. Many DNA based encryption schemes are being proposed by researchers [1,2]. DNA encryption means combining DNA technique with cryptology and producing new cryptography to provide safe and efficient cipher services. The interesting feature about the structure of DNA is the complementary rule proposed by Watson and Crick. Many researchers have used this concept for proposing image encryption methods. Hongjun Liu et al. [3] proposed an encryption method using DNA complementary rule where piecewise linear chaotic map is used for permutation and then substitution is performed using complementary rule. Various operations performed on DNA include synthesis, insertion, truncation, deletion, transformation, ligation and polymerase chain reaction etc. Qiang Zhang et.al [4] proposed an encryption method that used DNA subsequence operations such as truncation, deletion and transformation operation combined with chaos system to scramble the location and value of image pixel. Encryption method proposed by Nirmalya Kar et al. [5] used DNA sequencing along with cryptographic algorithm to generate the cipher text similar with the biological structure of the DNA Strands. Another image encryption algorithm [6] used index based chaos for permutation and key generation and DNA addition operation for substitution. Qiang Zhang et al. [7] also used DNA addition operations along with chaotic map to encrypt image but the cryptanalysis of this algorithm by Houcemeddine Hermassi et al. [8] show that it is non invertible and also it is weak against chosen plain text attack. Recently indexed based symmetric DNA encryption schemes [9, 10] are proposed where DNA gene bank is used for key which avoids the sending of long key over the channel.

This paper also proposes a new method of image encryption in which there is no need to send the long key over the secure channel. First, a secret key sequence is generated using two prime numbers and a selected DNA sequence. Next, each gray scale pixel value of the original image is added with the key element using CBC-mode of encryption to generate intermediate values. These intermediate values are then encoded to four nucleotides by DNA coding and finally DNA complementary rule and DNA addition operation are used to generate the encrypted image.

2. BASIC THEORY OF PROPOSED ALGORITHM

2.1 DNA Coding and Complementary Rule
A DNA sequence consists of four nucleic acid bases A (adenine), C (cytosine), G (guanine), T (thymine), where A and T are complementary, and G and C are complementary. In this paper we use C, T, A and G to denote 00, 01, 10, 11 (the corresponding decimal digits are “0123”) [1, 2]. Using this encoding method each 8-bit pixel value of the gray scale image can be represented as a nucleotide string of length four. Inversely to decode the nucleotide string we can get a binary sequence easily.

In total 4! = 24 kinds of coding, there are only 8 of them can meet complementary rule , for example, the decimal digits “0123” (the corresponding binary number is “00011011”) can be encoded in to one of them, such as “CTAG”, “CATG”, “GATC”, “GTAC”, “TCGA”, “TGCA”, “ACGT” or “AGCT”. There are total six legal complementary rules [3] which are as follows:

(AT)(TC)(CG)(GA),(AT)(TG)(GC)(GA),
(AC)(CT)(TG)(GA),(AC)(CG)(GT)(TA),
(AG)(GT)(TC)(CA),(AG)(GC)(CT)(TA).

Any one of them for example, (AG)(GC)(CT)(TA) can be applied to proposed method.

2.2 DNA Addition and Subtraction Operation
With the rapid development in DNA computing some algebraic operations other than biological operations based on DNA sequence are also proposed by the researchers such as

addition operation. The DNA addition operation is similar to the binary addition operation for example in binary 10+11 = 01 similarly, A+G = T. in this paper, DNA addition operation is used during encryption and subtraction is used for decryption process. The addition and subtraction tables [7] are as follows:

Table 1. Addition Table

+	**C**	**T**	**A**	**G**
C	C	T	A	G
T	T	A	G	C
A	A	G	C	T
G	G	C	T	A

Table 2. Subtraction Table

+	**C**	**T**	**A**	**G**
C	C	G	A	T
T	T	C	G	A
A	A	T	C	G
G	G	A	T	C

3. PROPOSED ALGORITHM

3.1 Key Generation and Expansion

In this encryption process, the key sequence is generated using pseudo random sequence generation method with some modification on it.

3.1.1 Key Generation

The key generation process consists of following steps:

1. A special gene sequence [9] of length 4n is selected from the gene bank. The start and end points of the sequence are chosen randomly.
2. Next the gene sequence is divided into substrings of length four. Each substring is then converted into binary using DNA coding. Mathematically, if there are total n number of substrings then the binary number generated from the DNA sequence are denoted as $N:\{N_0, N_1, N_2, \ldots\ldots\ldots\ldots, N_{n-1}\}$.
3. The key is generated as in (1)

$$k_i = [((y_i \oplus N_i) \times A) + B] \bmod 2^8, \qquad 0 \le i \le n \qquad (1)$$

Where, y_0 is 8-bit seed value, $y_i = k_{i-1}$ for i = 1 to n-1, A&B are two large prime numbers and $K = \{k_0, k_1, k_2, \ldots\ldots\ldots, k_{n-1}\}$ is the key sequence generated.

3.1.2 Key Expansion

The key expansion is as follows:

1. At first, elements of the key sequence are encoded into the DNA sequence using DNA encoding method.
2. The resulting DNA sequence is copied four times and then substring generation takes place from each copied string.
3. The first string is subdivided into the substrings of length four and the division takes place from the first character.
4. Similarly, the division of the second string takes place leaving the first character.
5. Similarly it is done with the third and fourth string leaving first two and first three characters respectively.
6. Now from all the strings only the substrings of length four are chosen as the elements of expanded key.
7. Finally each substring is converted back into numerical value. For example, let the DNA code of the key sequence is "ACTCCTGCTACATATC". Now copy it four times to generate the substrings as below.

   ACTC/CTGC/TACA/TATC (here none of the character is discarded)

   A/CTCC/TGCT/ACAT/ATC (first character is discarded)

   AC/TCCT/GCTA/CATA/TC (1st two characters are discarded)

   ACT/CCTG/CTAC/ATAT/C (1st three characters are discarded)

3.2 Generating Pseudo Random Sequence

1. If there are M×N pixels in the original gray scale image then it is required to generate the pseudo random sequence of length MN.
2. For this purpose a state array S is initialized of length 256 and the key sequence K.
3. The values of the state array S are filled from 0 to 255, i.e. s[0] = 0, s[1] = 1, ……s[255] = 255.
4. Now a 256 byte temporary array T is created and the values of K are copied into T and the remaining positions of T are filled again with the values of K.
5. Finally, the pseudo random sequence Z = {z[0], z[1], ……..z[MN]} of length MN is generated as below.

   Initialize j = 0;

```
for i = 1 to MN
    for j = 0 to 255
      j = ( j + s[i] + t[i]) mod 255;          (2)
      z[i] = j mod 8;
      s[i] = s[j];
```

3.3 Encryption

The encryption process uses CBC mode of encryption. At first, the original gray scale image is converted into a matrix of pixel values. This is then converted into a one dimensional sequence used for encryption process. In CBC mode before doing encryption a randomly chosen 8-bit value, called initialization vector (IV) is XOR with the first pixel value of the first block of the plain text. Here, the purpose of IV is to make each message unique.

The encryption steps are as follows:

Input : grayscale image, key sequence, pseudo random sequence.

Output : encrypted image.

1. Convert the image into pixel value matrix and the transform it into one dimensional sequence P.
2. Now perform the operation as in (3).

$$c_i = \begin{cases} [IV \oplus p_0) + k_0] \bmod 2^8, & i = 0 \\ [(c_{i-1} \oplus p_i) + k_{i \bmod n}] \bmod 2^8, & 1 \le i < MN \end{cases} \quad (3)$$

where $p_i \in$, P = {p_0, p_1, p_2, ……, p_{MN-1}}, plain text sequence and $k_i \in$ K, K= {k_0, $k_{1,,.....,}$ k_{n-1}}, key sequence and resultant intermediate cipher values forms the sequence C = { $c_{0,}$ $c_{1,}$ ……., c_{MN-1}}.

3. Encode each value of the sequence C into DNA code and compute the encrypted sequence X = { x_1, x_2, ……., x_{MN}} using (4).

$$x_i = \begin{cases} DnaAdd\,(c_1, Comp(Rotate\,(c_{MN}))), & i = 1 \\ DnaAdd\,(c_i, Comp(Rotate\,(x_{i-1}))), & 2 \le i \le MN \end{cases} \quad (4)$$

4. Here, three functions DnaAdd, Comp, Rotate are described as
   a. DnaAdd : The function DnaAdd is used for performing DNA addition of two values according to addition table.
   b. Comp : The function comp is used to generate the complement of the DNA code using complementary rule.
   c. Rotate : The function Rotate is used to do bitwise left rotation of the DNA code, the number of bits rotated at iteration i is equal to the value of the pseudo random sequence.
5. Convert the sequence X into a two dimensional matrix and the DNA code is converted back into the decimal pixel values to generate encrypted image.

4. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

4.1 Histogram Analysis

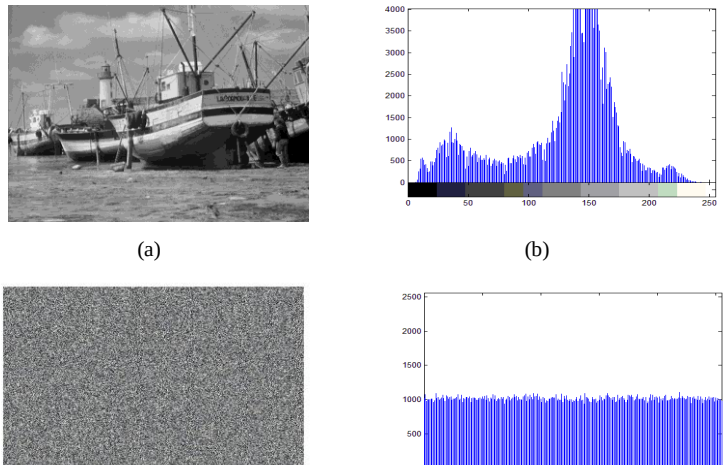
An image–histogram illustrates how pixels in an image are distributed at each color intensity level. The histogram of both the original image "boat.tiff" and encrypted image are plotted. Figure 1(a) shows the original image and 1(b) is its histogram respectively. Histogram of encrypted image is uniform in nature significantly different from original image histogram. Hence, any statistical attack is unlikely in the proposed encryption technique.

4.2 Correlation Coefficient Analysis

In addition to the histogram analysis, the correlation between two vertically as well as horizontally adjacent pixels in the original image and its encrypted image has also been analyzed. Correlation is a statistical measurement of the relationship between two variables which ranges from +1 to -1. As it is well known that in any image the correlation of adjacent pixels is very high, i.e. a good encryption algorithm is require to lower the correlation between adjacent pixels. Table 3 shows the correlation coefficient between the adjacent pixels of original and encrypted image both horizontally and vertically.

Table 3. Correlation coefficient between adjacent pixels of original image boat.512.tiff and corresponding encrypted image

	Original image	Encrypted image
Horizontal	.9713	-.0008
Vertical	.9381	.0003



(a) (b)

Fig 1: (a) Original image boat. 512.tiff (b) histogram of original image. Frame (c) encrypted image of original image of fig.1 (a) using proposed method (d) histogram of encrypted image in fig.1 (c).

4.3 Key Space Analysis

Key space is one of the important criteria of performance analysis of image encryption algorithm. A good encryption algorithm should have a large key space to prevent brute force attack. Above algorithm have the following secret keys: (1) the DNA sequence used, (2) initial value y_0 and IV (3) DNA coding and complementary rule. The information of selected gene sequence, as one part of key will be shared between sender and receiver through secure channel. In order to enhance the security of the algorithm, researcher sets start point and end point of the search position. The advantage of this is that there is no need to send the full long expression of the gene sequence over the secure channel keeping the original key length small that is shared on the secure channel. Here, a DNA sequence of length 256 has been taken i.e. there are total $4^{256} = 2^{512}$ possible combinations of gene sequence. The seed value $y_0 = 199$ and IV = 191, 8-bit value each i.e. there are total $2^8 \times 2^8$ possible combinations, also two prime numbers are used which remain constant for generating key sequence. The key expansion method used here increases the size of key up to four times approximately of the original key. There are only 8 kind of DNA coding to meet the complementary rule. So, the total key space will be $2^{512}\times2^8\times2^8\times4\times8\times6$ which ensures that the key space is sufficient enough to prevent any kind of brute force attack.

4.4 Sensitivity Analysis

A good image encryption algorithm should produce good avalanche effect. The avalanche effect is evident if the input is changed slightly, the output is changed significantly. It usually depends upon the sensitivity of the algorithm with respect to secret key and plaintext both, i.e. the change of single bit either in key or in plaintext should produce a significantly different encrypted image. To measure the sensitivity of the proposed method, both the key sensitivity analysis and plain text sensitivity analysis have been performed.

4.4.1 Key sensitivity analysis

To perform the key sensitivity analysis the original image is encrypted at first with the secret key sequence resulting in 0encrypted image A and after that with a slight change in the key sequence: (a) by changing most significant bit of the key sequence resulting in encrypted image B and (b) by changing least significant bit of the key sequence resulting in encrypted image C. Now for comparison, the correlation between the pixels of the three encrypted images is calculated as shown in Table 4.

Table 4. Correlation coefficient between the corresponding pixels of the three different encrypted images obtained by slightly different secret key on image boat.512.tiff

Image 1	Image 2	Correlation coefficient
Encrypted image A	Encrypted image B	.0034
Encrypted image B	Encrypted image C	.0007
Encrypted image C	Encrypted image A	-.0015

4.4.2 Plaintext sensitivity analysis

To perform the Plain text sensitivity analysis at first the original image is encrypted with the secret key resulting in "encrypted image P" and after that with a slight change in the pixel value: (a) by changing most significant bit of the plain text (pixel value) resulting in "encrypted image Q" and (b) by changing the least significant bit of the plain text resulting in "encrypted image R". Now for comparison, the correlation between the pixels of the three encrypted images is calculated as shown in Table 5.

Table 5. Correlation coefficient between the corresponding pixels of the three different encrypted images obtained by slightly different pixel value of image Boat.512.tiff

Image 1	Image 2	Correlation coefficient
Encrypted image P	Encrypted image Q	.0017
Encrypted image Q	Encrypted image R	-.0009
Encrypted image R	Encrypted image P	-.0032

4.5 Theoretical Analysis

In this paper, the security of the algorithm is analysed from the aspects of bio-security and math-security.

4.5.1 Bio- security

In this paper a special DNA sequence is selected to generate key is of the same length as of the original key sequence, here key expansion method is used to generate the shotgun gene sequences, it provides a very large key space as the DNA shotgun sequences are generally very large thus, it resists all possibilities of the brute force attack. In this paper, we need only to refer to the GENE BANK once for key generation in comparison with indexed based method et al. [4], where every time the database is required to be referenced for DNA sequence. There are total 8 kinds of the DNA coding to meet the complementary rule, and there are altogether 6 kinds of legal complementary rules to be applied, which make the algorithm more secure.

4.5.2 Math -security

Availability of a larger key space makes the algorithm preventive from any kind of brute force attack. Also the DNA addition and complementary rule makes it more secure. A small change in the key value or plaintext will change the resultant cipher text completely, thus producing good avalanche effect. The algorithm is also secure against the known plaintext attack; it will be difficult for the cryptanalyst to detect the key from the given plaintext and cipher text values.

5. CONCLUSION

In this paper, symmetric-key encryption algorithm based on the DNA approach is proposed. The original key sequence can be expanded to desired length using proposed key expansion method guided by the pseudo random sequence. The advantage is that there is no need to send a long key over the channel. The variable key expansion in encryption process combined with DNA addition and complement makes the technique sufficiently secure. The proposed technique has been experimentally evaluated in terms of brute-force attack, sensitivity analysis, avalanche effect and statistical analysis and acceptable results have been found.

6. REFERENCES

[1] Guangzhao Cui #1, Limin Qin #2, Yanfeng Wang #3, Xuncai Zhang, "An Encryption Scheme Using DNA Technology", 2008 IEEE.

[2] Qiang Zhang, Shihua Zhou and Xiaopeng Wei, "An Efficient Approach for DNA Fractal-based Image Encryption", Applied Mathematics & Information Sciences 5(3) (2011), 445-459 – An International Journal.

[3] Hongjun Liu, Xingyuan Wang, Abdurahman kadir,"Image encryption using DNA complementary rule and chaotic maps", Applied Soft Computing 12 (2012) 1457–1466.

[4] Qiang Zhang, Xianglian Xue, Xiaopeng Wei, "A Novel Image Encryption Algorithm Based on DNA Subsequence Operation ", The Scientific World Journal Volume 2012, Article ID 286741.

[5] Nirmalya Kar, Atanu Majumder, Ashim Saha, Anupam Jamatia, Kunnal Chakma, Dr. Mukul Chandra Pal, "An Improved Data Security using DNA Sequencing", 2013 ACM 978-1-4503-2207-2.

[6] Aradhana Soni, Anuja Kumar Acharya, " A Novel Image Encryption Approach using an Index based Chaos and DNA Encoding and its Performance Analysis", International Journal of Computer Applications (0975 – 8887), Volume 47– No.23, June 2012.

[7] Qiang Zhang, Ling Guo, Xiaopeng Wei, "Image encryption using DNA addition combining with chaotic maps", Mathematical and Computer Modelling 52 (2010) 20282035.

[8] Houcemeddine Hermassi, Akram Belazi, Rhouma Rhouma, Safya Mdimegh Belghith, "Security Analysis of An Image Encryption Algorithm Based on A DNA Addition Combining with Chaotic Maps", Multimedia Tools Application, june 2013 Springer Science.

[9] Zhang Yunpeng, Zhu Yu, Wang Zhong, Richard O.Sinnott, "Index-Based Symmetric DNA Encryption Algorithm", 2011 4th International Congress on Image and Signal Processing, IEEE.

[10] Grasha Jacob, A. Murugan, "An Encryption Scheme with DNA Technology and JPEG Zigzag Coding for Secure Transmission of Images", 2013 IEEE.