DNA-based image encryption algorithm

V A Gasimov and J I Mammadov
Computer systems and networks department, Azerbaijan Technical University, H. Cavid av, 25, Baku, AZ 1073, Azerbaijan

E-mail: gasumov@yahoo.com

Abstract. The development of a new image encryption algorithm using real structures of deoxyribonucleic acid (DNA) molecules is considered. In the proposed algorithm, the encryption process is performed by confusing and rearranging the pixels of the image based on the coordinates of the chaotic points obtained by the chaos game of DNA symbols, the sequence of DNA symbols, and the encoding rule. The selection of a specific DNA sequence from the Databank and the determination of the encoding rule for DNA symbols is based on the secret key.

1. Introduction
Nowadays, to the use of features of nitrogen-based substances like adenine (A), guanine (G), cytosine (C) and thymine (T) in the deoxyribonucleic acid (DNA) molecule for cryptographic protection of information more attention is being paid. This is because of the high development of parallel calculation technologies based on DNA and the elaboration of technologies for storing large volume of information in DNA molecules, and also the creation of a large open gene bank, where a large volume of DNA sequence individually belonging to each of the living organisms is gathered together.

Analysis of the corresponding literature shows that a certain success has been achieved in the image encryption using DNA sequences. In one of the works [1], this problem was solved by transforming the plain image into DNA sequence, using the rules for encoding the DNA symbols, and translating it into an encrypted form using different DNA operations. In another algorithm given in work [2], first a secret key is created on the basis of the DNA sequence and certain calculating operations, and then on the basis of the key and DNA operations a corresponding encryption process on each pixel of the image is carried out. Due to the given algorithm, the primary key can be expanded to the required level, and this symmetric encryption method removes the need of transmitting a long key. In the work [3] for image encryption, a hybrid model consisting of a chaotic logical map, masking of DNA and an algorithm for replacing DNA was suggested. The main advantage of the proposed method is to obtain the best mask appropriate to the primary image. For that reason, the authors used a set of rules for encoding and decoding of DNA symbols. Saranya M.R. and others proposed in the work [4] an effective algorithm to encryption images using a chaotic logical map, a DNA sequence, and genetic algorithm. Using a chaotic logical map, this algorithm generates chaotic sequence in the required length, but for this purpose, the initial value based on the secret key is calculated. At the first phase of the algorithm, many masks of DNA are created with the use of a chaotic logical map. At the next phase, the best mask for encryption by implementing a genetic algorithm is obtained. In the algorithm proposed in the work [5], first of all, the image is encoded into the DNA matrix, and then the permutation operation is performed. A chaotic
sequence created by a two-dimensional chaotic logical map is used to permute rows and columns periodically. The initial meaning and parameters for the chaotic system are calculated from the plain image with the help of SHA-256 hash function. At the next phase, the process of line-by-line diffusion of the image at the DNA level is implemented. At the end of the process, with the decoding of DNA matrix, an encrypted image is obtained. As the rules of encoding and decoding of DNA and the key matrix are formed of the initial image, the algorithm depends much more on this image, and this image, in its term, increases the level of security of the algorithm greatly. In the work [6] an algorithm for encrypting images using DNA sequence of dynamic chaos, and hash-functions is described. The estimation of the resistance of this algorithm to statistical and linear cryptanalysis under various chaotic mappings is carried out. According to the author's view, encryption with a baker's mapping showed better efficiency compared with the use of other mappings, and therefore it is the most suitable for this algorithm in comparison with other algorithms. Work [7] approved that conventional encryption algorithms are not ideal for image encryption, and that DNA-based cryptographic methods play the role of a bridge between existing methods and new encryption technology. In work [8] the survey of the results of DNA cryptography, and also the positive aspects and problems of different tendencies of the DNA approach in cryptography are given.

The analysis of these given research methods for image encryption and also similar algorithms for encrypting text data approves the fact that both real DNA and DNA principles are a promising aspect of use in cryptography. In this research work, to encrypt images using DNA we suggest a new method, and the essence of this method is that, along with the binary codes of DNA symbols for masking pixels, sequences of pseudo-random numbers generated by chaos game representation (CGR) of the symbols are also used.

2. Basic theory of proposed algorithm

2.1. Structure of deoxyribonucleic acid

Deoxyribonucleic acid is a macromolecule which provides depot, transmission from generation to generation, and the realization of a genetic program for the development and functioning of living organisms. A DNA molecule contains biological information in the form of a genetic code that consists of a sequence of nucleotides.

From the chemical point of view, DNA is a long polymer molecule which consists of repeating blocks-nucleotides [8-10]. Each nucleotide consists of nitrogenous bases, sugar (deoxyribose), and a phosphate group. There are four types of nitrogenous bases (adenine (A), guanine (G), thymine (T) and cytosine (C)) in DNA. According to the principle of complementarity nitrogen bases of one of the chains are connected to the nitrogen bases of other chain by hydrogen bonds: adenine (A) is only connected to thymine (T), and guanine (G) only to cytosine (C). For example, if the sequence in one polynucleotide chain is $A$-$G$-$A$-$T$-$C$-$G$-$T$, the nucleotides in another chain will be lined up with the sequence $T$-$C$-$T$-$A$-$A$-$G$-$C$-$A$.

In image encryption algorithms, especially, the above mentioned model is used, named after its discoverers by the Watson-Crick model.

As there are 4 various symbols (A, C, G, T) in the DNA chain, we can replace them by the quaternary number system (0, 1, 2, 3). As in the binary system, the 0 and 1 computations complete each other, so in the quaternary system there is a pairwise complement. Here 0 with 3 and 1 with 2 are mutually complementary. Since adenine (A) with thymine (T) and guanine (G) with cytosine (C) mutually complete each other, in this case, according to Watson-Crick model, there are eight possible encoding of DNA symbols. These variants are also called rules for encoding and decoding of DNA symbols. As a rule, in practice, symbols of the binary system are used instead of the symbols of the quaternary system (table 1).
Table 1. The rules of encoding and decoding of DNA symbols (in the quaternary number system).

| Rule   | A | T | C | G |
|--------|---|---|---|---|
| 0 (R0) | 0 | 3 | 2 | 1 |
| 1 (R1) | 0 | 3 | 1 | 2 |
| 2 (R2) | 3 | 0 | 2 | 1 |
| 3 (R3) | 3 | 0 | 1 | 2 |
| 4 (R4) | 2 | 1 | 0 | 3 |
| 5 (R5) | 1 | 2 | 0 | 3 |
| 6 (R6) | 2 | 1 | 3 | 0 |
| 7 (R7) | 1 | 2 | 3 | 0 |

Table 2. The rules of encoding and decoding of DNA symbols (in binary system).

| Rule   | A | T | C | G |
|--------|---|---|---|---|
| 0 (R0) | 00| 11| 10| 01|
| 1 (R1) | 00| 11| 01| 10|
| 2 (R2) | 11| 00| 10| 01|
| 3 (R3) | 11| 00| 01| 10|
| 4 (R4) | 10| 01| 00| 11|
| 5 (R5) | 01| 10| 00| 11|
| 6 (R6) | 10| 01| 11| 00|
| 7 (R7) | 01| 10| 11| 00|

2.2. Chaos game representation of symbols
Chaos game representation is mathematically expressed by Iterated Function System. This system converts a sequence of symbols into a unique set of points (mostly, in the form of fractal figures) in a two-dimensional space [11-13]. The significance of the method is that, at first, in the dimensional coordinate plane the coordinates of the vertices of a given polygon are marked. Any point inside this polygon is adopted as the starting point (z_0) of the game. At the first iteration of the game, one of the vertices is selected randomly, and in the center of the line connecting the starting point and the selected vertex, a new point (z_1) is marked. This point is considered as a starting point at the next iteration, and the process is repeated until the required number of points is obtained.

The principle of chaos game representation is that; each next point is defined as a function of the current point:

\[ z_{k+1} = f_i(z_k), \]

where \( z_k \) itself represents the first point of the game, and \( f_i \) determines the position of the point in the i-th position. Each point \( z_k = f_{k,1}(f_{k,2} \ldots f_{2}(f_1(z_0)) \ldots) \) has the address \( s_k s_{k-1} s_{k-2} \ldots s_2 s_1 \) (where \( s_n \) indicates the selected vertex at step \( n \)).

In the literature there are the results of some similar researches on the properties of DNA molecules [14-15]. The main feature of these works is that a rectangular coordinate system uses a square, the vertices of which are denoted by DNT symbols, and the choice of a vertex at each iteration determined by the sequence of real DNA symbols. Features of DNA are investigated as the obtained results of the chaos game representation in the form of different figures consisting of sets of chaotic points. In this article we suggest to use the randomness of points obtained by game of chaos sequence of a DNA symbols color image encryption algorithm.

3. Proposed image encryption algorithm
In this algorithm, in order to encrypt 24-bit BMP-images, it is expected to have a 128-bit secret key, which is known to transmitting part and to whom the secret information is intended. On the basis of this
A real block of DNA symbols from DNA-bank is determined and selected, for example from a source [16]. At the same time, one of the eight rules of encoding DNA symbols is determined using a secret key. Based on the sequence obtained by the secret key, the encryption process is performed by the encoding rule, and the chaos game representation of the DNA symbols.

3.1. Obtaining the chaos game representation of the DNA sequence
An image of chaos game representation of the DNA sequence is obtained in the following way:

1. In a rectangular coordinate system, a square is taken whose side is equal to one, and the vertices of the square are denoted by one of the rules indicated in table 2 (for example, let’s choose the rule R0 and denote the points A(0,0), G(0,1), C(1.0) and T(1.1));

2. The initial point in square is determined as \( z_0(x_0, y_0) \);

3. The first symbol of DNA chain, taken from the gene bank, is checked, depending on the symbols A, G, C, T, the coordinates \( x_1 \) and \( y_1 \) of the point \( z_1 \) are calculated on the basis of the following formulas:

\[
\begin{align*}
A: & \quad x_1 = \frac{x_0}{2}; \quad y_1 = \frac{y_0}{2}; \quad G: \quad x_1 = \frac{x_0}{2} + \frac{1}{2}; \quad y_1 = \frac{y_0}{2}; \\
C: & \quad x_1 = \frac{x_0}{2}; \quad y_1 = \frac{y_0}{2} + \frac{1}{2}; \quad T: \quad x_1 = \frac{x_0}{2} + \frac{1}{2}; \quad y_1 = \frac{y_0}{2} + \frac{1}{2}.
\end{align*}
\]  

4. By checking the following symbols of the DNA chain, the coordinates of all points are calculated using the following iterative function:

\[
\begin{cases}
x_{i+1} = \frac{x_i}{2} + \frac{1}{2}a_j, & a_j \in (0,1) \\
y_{i+1} = \frac{y_i}{2} + \frac{1}{2}b_j, & b_j \in (0,1)
\end{cases}
\]  

Here the index \( j \) shows the character to be selected and, depending on it, the following values are obtained:

\( a_j = 0, b_j = 0 \), at \( j = A \); \( a_j = 1, b_j = 0 \), at \( j = G \); \( a_j = 0, b_j = 1 \), at \( j = C \); \( a_j = 1, b_j = 1 \), at \( j = T \).

Let’s note that, when using the other rules given in table 2, the calculations are carried out in a similar way. You can set or calculate the rule you use a certain way. In our work, based on the secret key, we calculate the rule used by the following expression:

\[
R = (\sum_{i=1}^{16} K_i) \mod 8,
\]  

where \( K_i \) is binary code of the first (i-th) symbol of the encryption key.

On the other hand, the address of the used real DNA sequence, which is selected from the gene bank, is determined on the basis of the secret key by the following expression:

\[
A_S = K_1K_2 \oplus K_3K_4 \oplus K_5K_6 \oplus K_7K_8 \\
A_{SB} = K_9K_{10} \oplus K_{11}K_{12} \oplus K_{13}K_{14} \oplus K_{15}K_{16}
\]  

where, \( A_s \) and \( A_{SB} \) are, respectively, the address of the DNA chain and the initial symbol of the used block. By expression (4), we can address up to 65,536 DNA strands and in the same number of blocks of the DNA sequence used. Of course, if necessary and while using a larger database, you can easily expand the addressing field.
The chaos game representation, given in a rectangular coordinate system for symbols of the real DNA chain of the Dickeya chrysanthemi plant, taken from the Bank of genes, is shown in figure 1. As you see, the dots are rather chaotic. The graphical representation of the time sequence of the coordinates of the points also confirms the high randomness (figure 2).

![Figure 1](image1.png)

**Figure 1.** The chaos game representation based on 500 (a), 2500 (b) and 10,000 (c) symbols of the DNA chain of the plant Dickeya chrysanthemi.

![Figure 2](image2.png)

**Figure 2.** The graphical representation of the time sequence of points obtained by 500 symbols of DNA chain of the Dickeya chrysanthemi plant.

In this way, we get a set of X (abscissa) and Y (ordinates), pseudo-random numbers, which consist of the coordinates of points obtained from the chaos game representation of the symbols of the DNA chain. In the given algorithm, the set X is used for confusion, and Y is used for permutation of the pixels of the encrypted image.

### 3.2. Changing the meanings (confusion) of image pixels

To change the pixel meanings of the image, a new array X1 is created by re-ordering the elements of the X-array obtained in the above mentioned method. Though the arrays X and X1 are composed of the same elements, the numbers of these elements in the arrays are different. We use this sequence for chaotic selection of DNA-symbols, with the help of which the image pixels are later confused.

Changing the pixel meanings of the image is carried out according to the following expression:

\[
P_i = K_i \oplus P_i \oplus S_i
\]

(5)

here, \(K_i\) - is binary code of 3 least bytes \((K_{14}K_{15}K_{16})\) of the encryption key (at the end of each iteration, the binary code of the secret key is cyclically shifted to the right one bit), \(P_i\) - is binary code for the numerical expression of the i-th pixel of the plain image, \(S_i\) - is binary code chosen from the DNA chain.
of 12 consecutive symbols. The address of the first of the 12 selected DNA chain symbols is specified according to the sequence number of the elements in arrays X and XI. In this way, if the element of the i-th position in the array X is in the j-th position in the array XI, then to mix the i-th pixel, 12 symbols from the DNA-chain are selected starting from j-th position.

3.3. Permutation of pixels
Permutation of image pixels (diffusion process) is based on changes in the position of the elements of the array Y as a result of sorting. As mentioned above, the elements of the array Y consist of the ordinates of the chaotic points taken as a result of the chaos game of the DNA chain symbols. By sorting the elements of the array Y, we get the array Y1. In the process of image permutation, the position of the elements of the array Y are taken as the initial positions of the pixels, and the position of these elements in the array Y1 show the new positions to which the pixels move. So, in this order, all the pixels of the plain image are moved, and this minimizes the correlation dependence of the adjacent in rows, along columns and diagonally, pixels of the encrypted image.

3.4. Description of the algorithm
According to the scheme described above, the encryption process is carried out in the following sequence:

- based on the initial 128-bit secret key K consisting of ASCII code symbols, expression (4) calculates the addresses of the used DNA chain and its block of symbols;
- using key K, expression (3) defines the encoding rule R;
- on the calculated address the block of symbols of a DNA chain is calculated, in accordance with this block and rule R, by the CGR algorithm arrays X and Y;
- from the arrays X and Y arrays X1 and Y1 are obtained by sorting;
- based on the array X1, the DNA block and in accordance with the rule R, the pixel meanings of the image are changing (a process of confusion happens);
- based on the array Y1, the pixels of the image are rearranged (the diffusion process is carried out), and an encrypted image is obtained.

We must note that, the algorithm refers to symmetric encryption methods, and the decryption process is carried out similarly to the encryption algorithm using the same key K, the only difference is that during the decryption process, the permutation process is performed before the process of pixel confusion.

4. Results analysis and discussion
To test the cryptographic resistance of the proposed encryption system, standard security analysis - such as key space analysis, statistical analysis and differential analysis were carried out. Specific statistical tests were carried out on the distribution of pixels in the encrypted image, the correlation between adjacent pixels, by the definition of the encrypted images. These tests were carried out on various 24-bit BMP format images. For a deeper analysis, the tests were carried out individually on each of the color components (red, green, blue) that constitute the pixels of the image.

4.1. Key space analysis
The algorithm of image encryption should have a sufficiently large key space to endure the brute-force approach, that is by method of checking all possible variants of keys [17]. Since a 128-bit key is used in the proposed system, the total key system, thus the number of possible keys is $2^{128}$. Even using the most modern computer system, a person's life is not enough to check so many options. Furthermore, if necessary, it is possible to expand the key space by making small changes in the algorithm.
4.2. Key sensitivity analysis

Encrypted information should be sensitive to minor changes in the secret key. In systems with sufficient cryptographic resistance, the encrypted versions of the same image with keys with a slight difference should be significantly different. To test the key sensitivity, in our work, the same image was encrypted with the $K_s$ key, and 16 other keys, each of which differed only by one binary bit in different bytes of the $K_s$ key. The numerical meanings of the pixel bytes for the color categories encrypted with the image $K_s$ key were compared with the numerical meanings of the bytes of the corresponding pixels of each image encrypted by the close keys. The results of these tests were positive, since the difference were quite large (figure 3).

We must note that the problem of analyzing the sensitivity of the cryptosystem key also includes performing the corresponding operations in the decryption process. In this way, it must be confirmed that the image decrypted with a key with a small difference from the correct key is completely different from the plain image. This process is carried out by decrypting the encrypted image with one correct key and with 16 very similar (similar to encryption process) to it incorrect keys. As a result of testing, the plain image was restored, only using the correct key, restoring the plain image with other keys failed (figure 4).

Thus, it can be concluded that the proposed encryption algorithm is very sensitive to the keys, and a minor change in the secret key leads to a completely different image from the original decrypted image.

![Figure 3](image3.png)  

**Figure 3.** a – original image (Lena), b – image encrypted with the key $aT/c7dlfP4\%vMxsE$, c – image encrypted with the key $aT/c7dlfP4\%vMxsF$, d – graphical performance of the difference of encrypted by close-up image keys.

![Figure 4](image4.png)  

**Figure 4.** Image encrypted with the key $aT/c7dlfP4\%vMxsE$ (a), decrypted with a valid key image (b), decrypted by wrong $aT/c7dlfP5\%vMxsE$ key image (c).
4.3. Statistical analysis
To evaluate the resistance of the proposed algorithm against statistical cryptanalysis, histograms and entropy of the encrypted images were investigated, and also the correlation dependence between adjacent pixels along the horizontal, vertical and diagonal lines were checked.

4.3.1. Histogram analysis. The most commonly used tool to visualize pixel distribution by color is image histograms for each color channel, which reflect the frequency of the pixel meanings by brightness. Figure 5 and 6 show histograms for each color channel, plain and encrypted images respectively. As you see from the figures, in the plain image, the colors are irregularly distributed, and there is a uniform distribution in the encrypted image, which shows that, based on the encrypted image, it is almost impossible to obtain statistical information about the plain image and the encryption key used.

4.3.2. Information entropy analysis. Taking into consideration that, the pixel meanings of the image is a quantitative indicator of the information contained in the image, the uncertainty of the content of this information can be estimated by the concept of entropy proposed by Shannon [17]. Entropy is expressed in bits. The entropy of the source S is determined by the expression:

\[ H(s) = -\sum_{i=0}^{255} P(i) \log_2 P(i), \]

where \( P(i) \) - is probability of occurrence of a price \( i \) \((0 \leq i \leq 255)\) by color channels of the encrypted image. 5 images were tested, the entropy calculated by expression (5) were between 5.32 and 7.75. After the encryption, the entropy of these images increased to the limit of 7.9831 - 7.9904, which is close to the maximum theoretical value – 8, and this is a high security feature for the encryption process.

4.3.3. Correlation coefficient analysis. In normal images, pixels have a strong correlation with adjacent pixels horizontally, vertically and diagonally [18], which is reflected in the meaning of the Pearson correlation coefficient.

In order to analyze the quality of encryption of the proposed algorithm, the correlation of adjacent pixels in the plain and encrypted images was estimated using the following formula:
\[ \rho_c(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{D(X)D(Y)}}, \]  

where \( C \) – is channel color, \( D(X), D(Y) \) – is variation and \( \text{cov}(X, Y) \) – covariance of two random variables. The average value of the correlation coefficient calculated for 10 encrypted images was in the limit \([-0.01823, 0.031427]\), which shows the elimination of a strong correlation between adjacent pixels.

A graphic presentation of the correlation between adjacent pixels in the plain and encrypted images is shown in figure 7-9. The figure 7a shows the relationship between the pixels of the horizontally adjacent plain image, and figure 7b shows the similar dependence of the encrypted image. As you see, there is a strong correlation between the adjacent pixels on the plain image, but this dependence is almost completely eliminated in the encrypted image. It should be noted that similar graphics of the correlation dependence (figure 8a, b and figure 9a, b) were obtained between the pixels adjacent vertically and diagonally.

**Figure 7.** Graphics of correlation dependence of the horizontally adjacent pixels of the original (a) and encrypted (b) images.

**Figure 8.** Graphics of correlation dependence of the adjacent vertical pixels of the plain (a) and encrypted (b) images.
Thus, the calculated values of the correlation dependence between adjacent pixels in the plain and encrypted images and the visual display of these dependence once again prove that the proposed system is stable to statistical cryptanalytic attacks.

5. Conclusion
To encrypt images, it is proposed to use binary character codes of the real DNA chain, as well as the sets obtained by chaos game representation of DNA symbols. One of the two sets obtained together with the secret key and the binary code of the selected DNA symbols is used for confusion, and the second set is used to permute the pixels of the original image. The method performed using the natural DNA sequence obtained from the gene bank was tested using standard safety analyzes, such as key domain analysis, statistical analysis and differential analysis, and results comparable to the cryptographic resistance of similar algorithms were obtained.

References
[1] Grasha J and Annamalai M 2013 DNA based Cryptography: An Overview and Analysis International Journal of Emerging Sciences 3 36-42
[2] Gupta R and Jain A A 2014 New Image Encryption Algorithm based on DNA Approach International Journal of Computer Applications 85 27-31
[3] Soumya P, Pranjal D, Prabir K and Atal C 2017 Secured image encryption scheme based on DNA encoding and chaotic map Review of Computer Engineering Studies 4 70-5
[4] Saranya M R, Mohan A K and Anusudha K 2015 Algorithm for Enhanced Image Security Using DNA and Genetic Algorithm IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems pp 1-5
[5] Xiuli C, Yiran C and Lucie B 2017 A novel chaos-based image encryption algorithm using DNA sequence operations Optics and Lasers in Engineering 88 197-213
[6] Sidorenko A V and Shishko M S 2016 Image encryption using DNA-sequences, chaotic dynamics and hash functions Reports of the BSUIR 5 42-7
[7] Siyamol C and Sheena M 2013 Summarization on DNA cryptography IJERT 2 303-306
[8] Ashish K K and Verma A K 2014 DNA Based Cryptography: A Review International Journal of Information & Computation Technology 4 693-8
[9] Kornberg A 1959 The biologic synthesis of deoxyribonucleic acid Nobel Lecture Retrieved from: https://www.nobelprize.org/uploads/2018/06/kornberg-lecture.pdf
[10] Hasanov A S, Rzayev N A, Islamzadeh F and Efendiyev A M 1989 Biological chemistry (Baku) p 562
[11] Joseph J and Sasikumar R 2006 Chaos game representation for comparison of whole genomes BMC Bioinformatics 7 243
[12] Tino P 1999 Spatial representation of symbolic sequences through iterative function systems IEEE Transaction on Signal Processing A 29 386-93

Figure 9. Graphics of the correlation dependence of the diagonally adjacent pixels of the original (a) and encrypted (b) images.
[13] Aswathy K V 2015 A survey on chaos game representation for online signature verification
*International journal of modern trends in engineering and research* **2** 52-6

[14] Jeffrey H J 1990 Chaos game representation of gene structure *Nucleic Acids Research* **18** 2163-70

[15] Changchuan Y 2017 Encoding DNA sequences by integer chaos game representation *Journal of computational biology: a journal of computational molecular cell biology* **26(2)**

[16] The DNA sequence of Homo sapiens *NCBI Resources NCBI Reference Sequence* (NT_024524.12) Retrieved from: www.ncbi.nlm.nih.gov/nuccore/NT_024524.12?txt=on &report=fasta&from=15982005&to=16044755

[17] Shannon C E 1949 Communication Theory of Secrecy Systems Retrieved from: http://kek.ksu.ru/student/crypto/files/shannon.pdf

[18] Tropchenko A Y and Tropchenko A A 2009 *Compression methods of images, audio and video* (St. Petersburg) p 106