# A Hybrid Encryption Scheme using DNA Technology

**Grasha Jacob[1], A. Murugan[2]**

[1] Research and Development Centre, Bharathiar University, Coimbatore – 641046, India
Dept. of Computer Science, Rani Anna Govt College for Women, Tirunelveli]
(e-mail: grasharanjit@gmail.com)
[2] Dept. of Computer Science, Dr. Ambedkar Govt Arts College, Vyasarpadi, Chennai

**Abstract:** Sensitive information such as financial transactions, medical and personal records are transmitted through public communication facilities. The security of the sensitive information poses a great threat by an unintended recipient. With the growth of technological advancements, the threats grow exponentially and security has become a critical issue in data storage and transmission. The concept of using DNA Cryptography has been identified as a possible technology that brings forward a new hope for unbreakable algorithms as traditional cryptographic systems are now vulnerable to certain attacks. This paper outlines a hybrid encryption scheme using DNA technology

**Keywords:** DNA computing, DNA Cryptography, OTP, Watson-Crick complementarity

## 1. Introduction

With advancements in digital communication technology and the growth of computer power and storage, internet provides essential communication between tens of millions of people and is being increasingly used as a tool in the fields of medicine and commerce. The sensitive data transmitted is vulnerable to attack during transmission across the network and the difficulties in ensuring security become increasingly challenging. Though cryptography enables in ensuring security of sensitive information, code breakers have come up with various methods to crack the cryptographic system. The concept of using DNA computing in the field of cryptography has been identified as a possible technology that brings forward a new hope for unbreakable algorithms as traditional cryptographic methods built upon mathematical and theoretical models are vulnerable to attacks.

DNA stands for DeoxyriboNucleic Acid. DNA represents the genetic blueprint of living creatures. DNA contains instructions for assembling cells. Every cell in the human body has a complete set of DNA. DNA is unique for each individual. DNA is a polymer made of monomers called deoxyribo nucleotides. Each nucleotide consists of three basic items: deoxyribose sugar, a phosphate group and a nitrogenous base. The nitrogenous bases are of two types: purines (Adenine and Guanine) and pyrimidines (Cytosine and Thymine). The key thing to note about the structure of DNA is its inherent complementarity proposed by Watson and Crick. A binds with T and G binds to C. All DNA computing applications are based on Watson-Crick complementarity. DNA computing is an interdisciplinary area concerned with the use of DNA molecules for the implementation of computational processes.

The main features of DNA are massive parallelism, intense storage capacity and energy efficiency. Adleman's pioneering work gave an idea of solving the directed Hamiltonian Path

Problem (Travelling Salesman Problem) of size n in O(n) using DNA molecules. The principle used by Adleman lies in coding of information (nodes, edges) in DNA clusters and in the use of enzymes for the simulation of simple calculations. The various operations performed on DNA are synthesised, cutting, ligation, translation, substitution, polymerase chain reaction, detection using gel electrophoresis and affinity purification. Lipton extended the work of Adleman and investigated the solution of Satisfiability of Propositional Formula pointing to new opportunities of DNA computing.

Gehani et. al., introduced the first trial of DNA based Cryptography in which a substitution method using libraries of distinctly one time pads, each of which defines a specific, randomly generated, pairwise mapping and an XOR scheme utilizing molecular computation and indexed, random key strings were used for encryption[9]. Research work is being done on DNA Computing either using test tubes biologically or simulating the operations of DNA using computers (Pseudo or Virtual DNA computing). The constraints of its high technology lab requirements and computational limitations, combined with the labor intensive extrapolation means prevent DNA computing from being efficiently used in today's security world [12].

## 2. DNA based image representation

Silicon computers are based on binary logic and any decimal can be represented in its binary equivalent. As there are four bases A, C, T, G in DNA sequence, according to the DNA digital coding technology [7], C denotes the binary value 00, A denotes 01, T denotes 10 and G denotes

Received Nov. 25, 2012, Reviewed January 15, 2013

11. Therefore any binary number or image can be represented in its equivalent DNA form. In general, the term image refers to a two-dimensional light intensity function, denoted by f (x, y), where the value of f at spatial coordinates (x, y) gives the intensity (brightness) of the image at that point. As light is a form of energy f (x, y) must be nonzero and finite, that is

$$0 < f(x, y) < \infty \quad \ldots\ldots\ldots (1)$$

A digital image is an image f (x, y) that has been discretized both in spatial coordinates and brightness. A digital image can be considered a matrix (two dimensional array) whose row and column indices identify a point in the image and the corresponding matrix element value identifies the gray level at that point. The elements of such a picture array are called pixels. All images considered in this paper are rectangular arrays of size m x n, where m, n ≥ 1. In image processing applications, operations are generally carried out pixel by pixel. Arithmetic and logic operations can be done "in place" in the sense that the result of performing an operation can be stored in that location in one of the existing images. The commutative property of the Ex-OR operation is used for one-time padding in cryptographic applications. The Ex-Or operation when applied once returns a meaningless information and when applied twice returns the original information. Table 1 gives the characteristic table of Ex-OR operation in DNA based computing.

**Table 1 Ex-OR Characteristic Table**

| ⊕ | **A** | **C** | **G** | **T** |
|---|---|---|---|---|
| **A** | C | A | T | G |
| **C** | A | C | G | T |
| **G** | T | G | C | A |
| **T** | G | T | A | C |

## 3. Hybrid Encryption Scheme

Symmetric encryption algorithms use an identical secret key for encryption and decryption process and the key is sent to the receiver through a secure communication channel.

$$Ke = K_d = K \quad \ldots\ldots\ldots (2)$$

The requirement of a symmetric algorithm is that both the sender and the receiver know the secret key, so that they can encrypt and decrypt the information easily. The proposed hybrid encryption scheme is a combination of a cryptosystem using XOR-OTP and substitution proposed by Gehani et. al[9] and the DNA-based Implementation of YAEA Encryption Algorithm proposed by Amin et. al[14].

In the hybrid encryption scheme, the key image is sent to the receiver through a secure communication channel. The image to be encrypted and the key image are synthesized - transformed into DNA images and Ex-OR One-Time Padding is performed using the substitution operation (a complex operation which is a combination of cutting and ligating operations). Both cutting and ligation use the same enzymes that organisms use for the maintenance of their own DNA. One of the many positions of the quadruple DNA nucleotide sequences of the translated image which is randomly obtained from the gene sequence binary file is detected and substituted and the resultant encrypted image is sent to the receiver.

Both the sender and receiver should download the same gene sequence GI| 417839630| ref| NZ_AFQN01000062.1| Haemophilus haemolyticus M19107 M19107_062, whole genome shotgun sequence, 81211 bp linear DNA from GenBank and store the sequence as a binary file. The encryption algorithm can be summarized as follows:

```
ALGORITHM HYBRID_CRYPT

Input: X [image file] to be encrypted, Y [image file] – key image, R[Binary file that contains DNA nucleotides sequence], RND [Z]

Output: Encrypted image E

1. SYNTHESIS
     a. Convert image file X into its DNA sequence
          $X \leftarrow DNA [X]$
     b. Convert key image file Y into its DNA sequence
          $Y \leftarrow DNA [Y]$
2. SUBSTITUTION                    // OTP
          $X \leftarrow X \oplus Y$
3. DETECTION and SUBSTITUTION
     For each quadruple DNA nucleotide sequence in X search starting from a random location RND (Z) in a binary file are represented in the form of a single strand DNA sequence. If the correct pattern is found, its location I is then recorded

     If the search is successful
          Then store I in E;
     Else
          Repeat step 4.

End Algorithm
```

To ensure the security of a cryptosystem, the outputs of a cryptosystem must be unpredictable in the absence of knowledge of the inputs. The encrypted image, the key image, the DNA sequence file and the decryption algorithm are necessary for decrypting the encrypted image.

The decryption algorithm can be summarized as follows:

```
ALGORITHM HYBRID_DECRYPT
Input: Encrypted image E, Y [image file] – key image,
 R [Binary file that contains DNA nucleotides sequence]
Output: Decrypted image X
1. SYNTHESIS
     Convert E into DNA sequence
2. DETECTION and SUBSTITUTION
     $F \leftarrow$ DNA sequence represented by E [I] in the binary file R
3. . SUBSTITUTION
```

```
    F ← F ⊕ Y                              // OTP
4. SUBSTITUTION
    Convert F into its binary equivalent and display the image X.
End Algorithm.
```

## 4. Experimental Results

Matlab R2008a is used to simulate the DNA operations on a MiTAC Notebook PC with Intel® Core™ 2Duo CPU T6400 @ 2.00 GHz, 2 GB RAM, 32 bit operating system. Experiments are performed using different images of different sizes to prove the validity of the proposed algorithm.

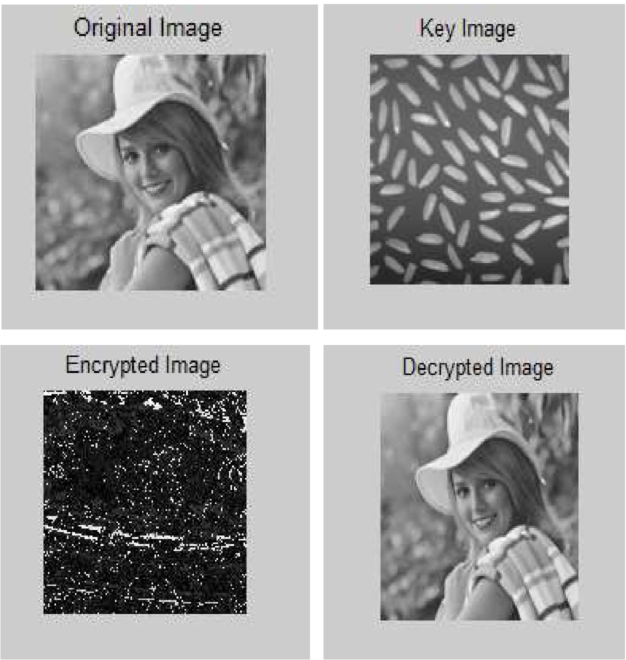


Fig 1 a) Original image of size 128 x 128
b) Key image of size 128 x 128
c) Encrypted image d) decrypted image

## 5. Cryptanalysis

A good information security system should be able to protect confidential images. The level of security that the hybrid encryption algorithm offers is its strength. Cryptanalysis is the art of deciphering encrypted data. Cryptographic attacks are a part of cryptanalysis and are designed to subvert the security of cryptographic algorithms, and they are used to attempt to decrypt data without prior access to a key. A good encryption technique should be robust against statistical, cryptanalytic and brute-force attacks. The proposed method is examined through statistical analysis, sensitivity to key changes and key space analysis.

### 5.1 Statistical Analysis

The encrypted image should not have any statistical similarity with the original image to prevent the leakage of information. The stability of the proposed method is examined via statistical attacks - the histogram and correlation between adjacent pixels.

### 5.1.1 Histogram Analysis

An image histogram is a graphical representation of the number of pixels in an image as a function of their intensity. An image histogram describes how the image-pixels are distributed by plotting the number of pixels at each intensity level. The histograms present the statistical characteristics of an image. If the histograms of the original image and encrypted image are different, then the encryption algorithm has good performance. An attacker will find it difficult to extract the pixels statistical nature of the original image from the encrypted image and the algorithm can resist a chosen plain image or known plain image attack. Fig 2 a) and b) are the histograms of the original Elaine.bmp of size 128 x 128 and the encrypted image. The histogram of the encrypted image is fairly uniform and significantly different from the original image.

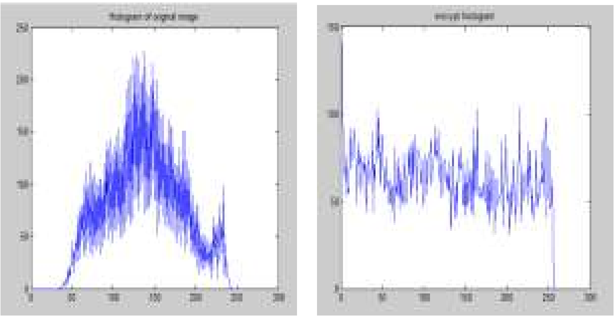


Fig 2 a) histogram of original image
b) histogram of encrypted image

As the encrypted image doesn't provide any information regarding the distribution of gray values to the attacker, the proposed algorithm can resist any type of histogram based attacks and strengthens the security of encrypted images significantly.

### 5.1.2 Correlation Coefficient Analysis

In most of the plaintext-images, there exists a high correlation among adjacent pixels, while there is a little correlation between neighboring pixels in the encrypted image. It is the main task of an efficient image encryption algorithm to eliminate the correlation of pixels. Two highly uncorrelated sequences have approximately zero correlation coefficient. The Pearson's Correlation Coefficient is determined using the formula:

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}} \quad \ldots\ldots\ldots \ (3)$$

Where x and y are the grayscale values of two adjacent pixels in the image and N is the total number of pixels selected from the image for the calculation. Table 2 tabulates the correlation coefficient calculated for the original and encrypted images. If there is no linear correlation or a weak linear correlation, r is close to 0. A value near zero means that there is a random, nonlinear relationship between the two adjacent pixels. It is clear from Table 2 that there is a negligible correlation between the two adjacent pixels in the encrypted image. However, the two adjacent pixels in the original image are highly correlated.

**Table 2 Correlation Coefficient Analysis**

| *Correlation Coefficient* | |
|---|---|
| ***Original Image*** | ***Encrypted Image*** |
| ***0.9666*** | ***0.0031*** |

## 5.2 Cryptanalysis

### 5.2.1 Differential attacks

Attackers often make a slight change for the original image, use the proposed algorithm to encrypt the original image before and after changing, and compare two encrypted images to find out the relationship between the original image and encrypted image. This is known as differential attack. Number of pixels change rate (NPCR) measure is used to test the influence of one pixel change on the whole encrypted image. For two encrypted images $C_1$ and $C_2$ of the plain images which vary by only one pixel difference, the NPCR is defined as follows:

Where

$$D(i,j)=\begin{cases}0, & if \quad C_1(i,j)=C_2(i,j), \quad , \\ 1, & if \quad C_1(i,j)\neq C_2(i,j).\end{cases} \qquad \ldots\ldots\ldots \quad \mathbf{(4)}$$

Results obtained from NPCR calculations show that the hybrid encryption scheme's sensitivity to one pixel change in the input images is under 0.01%.

### 5.2.2 Known-Plaintext and chosen plaintext attacks

For encryption with a higher level of security, the security against both known-plaintext and chosen-plaintext attacks are necessary. Chosen/Known-plain text attacks are such attacks in which one can access/choose a set of plain texts and observe the corresponding encrypted texts.

A mask image, **M** is obtained by XOR-ing the plain image **C** with its corresponding encrypted image $C_1$.

Let $Z_1$ be the encrypted image of the plain text image **Z**.

$$\mathbf{M} \leftarrow \mathbf{C} \oplus \mathbf{C_1} \qquad \ldots\ldots\ldots (5)$$

```
If $Z = M \oplus Z_1$ then
    Unknown encrypted image is decrypted
Else
    Hybrid encryption scheme resists Chosen/ Known
            Plaint Text attack
End if
                                        ……….(6)
```

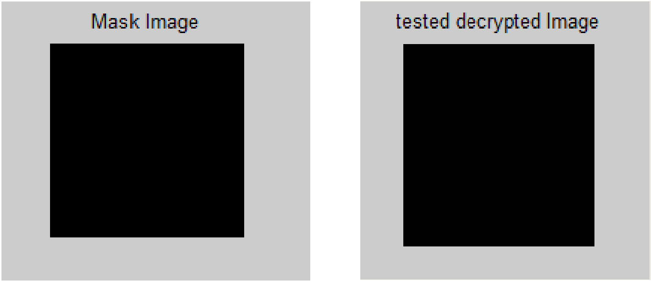
Fig 3 b) shows an unsuccessful chosen/known-plain text attack using the proposed algorithm.

### 5.3 Brute Force Attack

A Brute Force Attack or exhaustive key search is a strategy that can be used against any encrypted data by an attacker who is unable to take advantage of any weakness in an encryption system that would otherwise make his task easier. It involves systematically checking all possible keys until the correct key is found.

Fig 3 a) XOR Mask

b) Failed attack to crack the Encrypted image



The key to this encryption algorithm is an image which is about the same size as that of the image to be encrypted. Moreover, the aspect of bio-molecular environment is more difficult to access as it is extremely difficult to recover the DNA digital code without knowing the correct coding technology used. When an intruder gets the encrypted image and tries to decrypt the encrypted image without knowing the correct DNA digital coding technology, it would not be decrypted at all. An incorrect coding will cause biological pollution, which would lead to a corrupted image.

## 6 Conclusion

DNA based encryption is the beneficial supplement to the existing mathematical encryption. DNA binary strands support the feasibility and applicability of DNA-based Cryptography. The security and the performance of the DNA based cryptographic algorithms are satisfactory for multi-level security applications of today's network. The proposed hybrid encryption scheme using DNA technology can resist brute-force, statistical and differential attack.

REFERENCES

[1] L. M. Adleman, "Molecular Computation of Solutions to Combinatorial Problems", Science, 266, November 1994, pp. 1021-1024.

[2] M. Amos, G. Paun, and G. Rozenberg, "Topics in the theory of DNA computing", Theoretical Computer Science 2002, 287, pp. 3-38.

[3] C. Chelland , V.Risca, C.Bancroft, "Hiding messages in DNA microdots", Nature 1999, 399:533-534

[4] J. Chen, “A DNA-based, biomolecular cryptography design,” in IEEE International Symposium on Circuits and Systems (ISCAS), 2003, pp.822–825

[5] G.Z. Cui, L.M. Qin, and Y.F. Wang, “An Encryption Scheme using DNA Technology”, *Computer Engineering and Applications*, (2008), pp. 37-42.

[6] G. Cui, L. Qin, "Information Security Technology Based on DNA Computing", IEEE International,2007

[7] Dominik Heider and Angelika Barnekow , “DNA-based watermarks using the DNA-Crypt algorithm” , BMC Bioinformatics, May 2007

[8] Donald Nixon, “ DNA and DNA Computing in Security Practices – Is the Future in Our Genes?” GSEC Assignment Version 1.3.

[9] Gehani, Ashish La Bean, Thomas H. Reif, H.John, “DNA - BasedCryptography”, Dimacs Series In Discrete Mathematics and Theoret*ic*al Computer Science 2000, 54:233-249.

[10] A.Leier , C.Richter, W.Banzhaf, H. Rauhe, “Cryptography with DNA binary strands”, BioSystems 2000, 57:13-22

[11] Monica BORDA “DNA secret writing Techniques” , IEEE conferences 2010

[12] Ning Kang, “A pseudo DNA cryptography Method”,http://arxiv.org/abs/0903.2693 ,2009

[13] Shihua Zhou, Qiang Zhang, Xiaopeng Wei, “Image Encryption Algorithm Based on DNA Sequences for the Big Image”, 2010 International Conference on Multimedia Information Networking and Security

[14] Sherif T. Amin, Magdy Saeb, Salah El-Gindi, "A DNA based Implementation of YAEA Encryption Algorithm," IASTED International Conference on Computational Intelligence,2006

[15] Souhila Sadeg “An Encryption algorithm inspired from DNA”, IEEE pp 344 – 349, November 2010G.

[16] Xiao, M. Lu, L. Qin, X. Lai, " New field of cryptography: DNA cryptography," Chinese Science Bulletin, vol. 51(\0), pp. 1139-1144, Jun. 2006.