A DNA Based Image Steganography using 2D Chaotic Map

Prasenjit Das
Dept. of Computer Sc. & Engineering.,
National Institute of Technology Agartala, India
e-mail: pj.cstech@ gmail.com

Nirmalya Kar
Dept. of Computer Sc. & Engineering,
National Institute of Technology Agartala, India
e-mail: nirmalya@nita.ac.in

Abstract—**This paper presents a novel DNA based image steganography algorithm. DNA based security systems have drawn the attention of many researchers for more than a decade and lots of work have been done. Still it is in its evolutionary phase. The proposed work contributes to such a motive to improve the overall security of age old image steganography technique. It provides a two layer security to the secret data by wrapping it inside two layers of cover media, one being a theoretical DNA strand, and the other is a cover image. The DNA strand is constructed from the cover image using a 2D logistic map, thus having the attributes of the cover media. The process of hiding or embedding data inside the theoretical DNA strand requires degenerative genetic code replacement. The obtained mutated DNA is hidden back into the cover. Thus it supports dual layer hiding, making the system more reliable.**

Keywords-Deoxyribo Nucleic Acid; DNA algebra; logistic map; LSB replacement; primer; steganography

I. INTRODUCTION

Steganography is the art of hiding information imperceptibly in a cover medium. The sole purpose of steganography is to hide the very existence of the message in the cover medium [1]. Different types of covers (like image, audio, video) are uses based on several parameters like, capacity, security, ease of use, encoding and decoding time etc. Recently scientists work on different kind of new steganography algorithms and cover media to ameliorate the security of system. DNA Steganography is one of the cutting edge techniques in this area. The vast parallelism, exceptional energy efficiency and extraordinary information density inherent in DNA molecules are explored for cryptographic purposes.

Adleman [2] started DNA computing in 1994 by solving a small instance of the Hamiltonian path problem in wet lab. Clelland *et al.* [3] proposed a technique to hide a message-carrying DNA sample in the punctuation mark period (called microdot) in the contents of an ordinary innocuous letter. Representing DNA nucleotides by binary bits emerged in 1999 by Rauhe and *et al.* [4]. Some more robust data hiding techniques using DNAs were proposed in [5]-[12].

Although, these works using natural DNAs are very worthwhile and introduced new area of data hiding approaches, they have some drawbacks such as the biological errors like mutation and difficulty of implementation of DNA system [14]. One solution to this problem is to use theoretical model of DNAs and utilize its natural properties to strengthen the existing hiding techniques. In this paper we are proposing one such algorithm in which we are extracting a DNA strand from an image cover. Later we hide the secret message in the DNA strand. This is why we call it a dual cover steganography.

II. BIOMOLECULAR TECHNOLOGY BACKGROUNDS

This section gives a brief review of the DNA technology related terms for better visualization of proposed algorithm.

A. DNA

In cells genetic information is stored in DNAs made by four nucleotides: adenine (A), guanine (G), cytosine (C) and thymine (T). The ordered sequence of these four bases reflects the genetic information. The standard situation of nucleotides allows to make a hydrogen bind between C and G; or A and T. This complementarily standard rule is known as Watson-Crick base pairing [14]. Artificially synthesized single-stranded DNA chains are named oligonucleotides.

B. DNA Encoding Technique

It is a binary coding scheme to represent the 4 nucleotides introduced for the ease of theoretical DNA computation. For example A is represented by 0(00), similarly T = 1(01), C = 2(10), G = 3(11) [15]. So there are 4! = 24 possible coding patterns by this encoding format.

C. Codon

A codon is a triplet of three bases (T, C, A and G). With these four letters, 4^3 = 64 combinations are possible. With three exceptions, each codon encode for one of the 20 amino acids, used in the proteins synthesis. The three exceptions are TAA, TAG and TGA, indicate codons STOP [12].

D. Degenerative Codons

When two or more codons (synonymous codons) code for the same amino acid they are called degenerative codons. They typically differ in their 3rd base (e.g: GAA, GAC for glutamine). So by replacing third base, we can hide data without affecting the resulting amino acid [12]. This event is known as silent mutation. Fig. 1 shows the mapping of codon to amino acid with yellow ones being the degenerative codons.

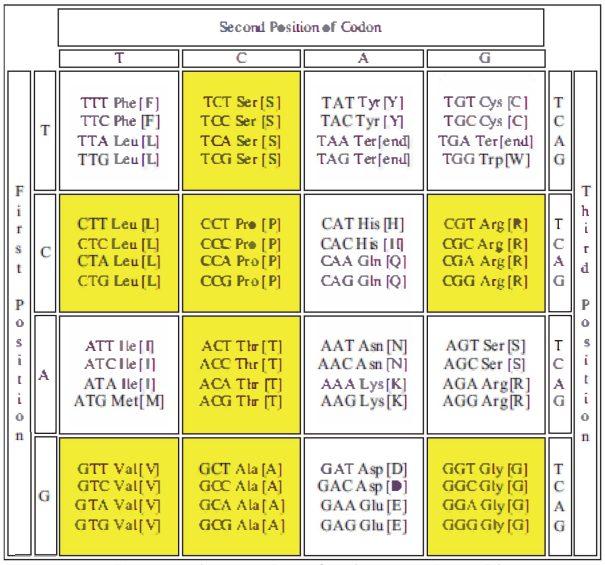

Figure 1. The mapping of codon to amino acid

E. Addition and Subtraction Algebra for DNA

Addition and subtraction operation can be performed on DNA sequences according to traditional addition and subtraction in the Z_2 (i.e. mod 2) [13]. For example, 11 + 10 = 01, so if (C, A, T, G) = (0, 1, 2, 3), we have G + T = A. Table I & II shows addition and subtraction matrices.

III. RELATED WORK

For more than a decade researchers are working on a new paradigm of DNA-based steganography, in which we can work on theoretical or mathematical models of a DNA and utilize the several properties of an actual DNA to render faster and more efficient data hiding. Some of them are discussed here.

Md. Reza Najaf *et al.* [14] proposed an algorithm combining the concepts of cryptography and DNA steganography. They used steganography for hidden key distribution for every establishment of new communication. On receiving the stego key the receiver sends the symmetric encryption key hiding into a background DNA. For all next communications the symmetric session key is used to encrypt the secret data. Hayam Mousa *et al.* introduced a reversible information hiding scheme for DNA sequence based on reversible contrast mapping in [15]. Meenakshi *et al.* in [16] proposed an algorithm to embed an offline handwritten signature in the form of a watermark as binary information using the principles of spread spectrum watermarking, which is further encrypted as DNA sequences attached to particular proteins or amino acids to enhance redundancy of secret data.

Samir *et al.* in [1] described a method of image steganography in which an secret image) is hidden within another image by creating 256 combinations of DNA bases using 4 nucleotides and replacing them with the original pixel values. Suman *et al.* in [17] proposed an approach in which secret data is covered by more than one object, one is outer cover and other is inner cover. They used the magic number as the forward tracking algorithm and embed secret information bits within DNA sequence by DNA sequence Complementary rules indexed by the magic number sequence. Suman *et al.* in [18] propose a lossless steganography method in which DNA sequence is used to represent secret image. DNA sequence is converted into decimal form and the secret image is hidden by the cover image using camouflaging procedure.

Amal *et al.* in [19] illustrate a DNA-based steganography method combined with a DNA cryptography technique for secure exchange of data in DNA carriers. It first encrypts the plaintext using Amino acid and DNA based Playfair cipher & then applies a novel complementary substitution method to hide the encrypted DNA cipher text into some reference DNA. Md. Reza *et al.* [20] proposed a simple yet secured DNA based steganography method in which secret message is hidden inside a reference DNA and the indices (locations) of message is sent to the receiver.

TABLE I. ADDITION OPERATION FOR DNA SEQUENCE [13]

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

TABLE II. SUBTRACTION OPERATION FOR DNA SEQUENCE [13]

-	T	A	C	G
T	C	G	T	A
A	A	C	G	T
C	T	A	C	G
G	G	T	A	C

IV. PROPOSED ALGORITHM

The proposed algorithm uses images as primary cover media for message transfer between two interested parties. A single stranded DNA (ssDNA) or oligonucleotide is extracted from the image which is used as the secondary cover providing a huge amount of background noise for the secret data. By selecting the appropriate image pixels we can hide 2 bits of secret data in a single pixel. The entire process consists of two algorithms - embedding and extraction.

A. Embedding Algorithm

Using this algorithm the secret message bits are hidden inside the ssDNA, which in turn is hidden inside the cover image. The secret data bits are encrypted prior to embedding providing another layer of security. The entire procedure is explained in the flowchart shown in Fig 2.

The embedding procedure uses 3 keys (see Fig 3). Key 1 is a combination of 6 parameters for 2D logistic map. Key 2 is a primer (short DNA sequence). Key 3 is a variable length key for RC4 encryption.

This procedure has the following sub procedures.

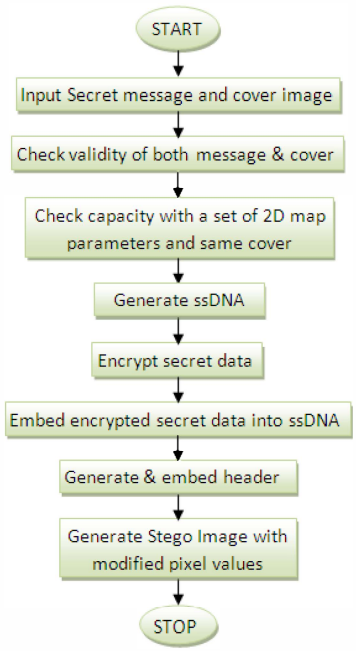

Figure 2: Flowchart for embedding process summery

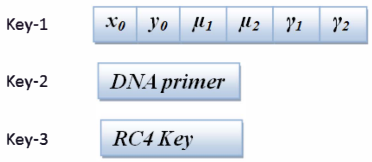

Figure 3: Keys used in the embedding process

1) Validating cover image & secret message: For getting robust output from the algorithm both the cover and secret media should possess some properties as discussed next.

Like all spatial domain image steganography the cover image-

- should be inconspicuous, thus not drawing attention to the fact that it could be concealing information.
- should not contain large blocks of one color or smooth homogeneous areas to avoid repetitive codon sequences in the extracted DNA strand.
- Should not be a 'lossy' file format such as JPEG. because on reconstruction of the embedded file, bits could be lost due to the compression.

The cover contains both data and meta-data. The meta-data contains value of N = number of iterations for 2D logistic map (3 bytes) and message authentication code. As the algorithm embeds 2 bits/pixel, so considering 1 character converted to an equivalent ASCII value of 8 bits, we have,

$$4 \times L + meta\text{-}data\_length < cover\_image\_size \quad (1)$$

Where, $cover\_image\_size = cover\_image\_width \times cover\_image\_height$. L = number of characters in text message. So, we can say that, Effective message length (L_E)

$$L_E = 4 \times L + meta\text{-}data\_length \quad (2)$$

2) Perform capacity check with respect to effective message length (L_E): During this step we construct the ssDNA from the cover image and check whether the message along with its meta-data can be hidden inside the DNA. The ssDNA can be considered as a series of codons. A single DNA codon is constructed from the 2 least significant bits of all the three color components of a pixel, using DNA encoding technique. Capacity counter is incremented by 1 if the generated codon is degenerative one. The flowchart for the process is shown in Fig. 4.

The pixel used to construct DNA codon is chosen based on the sequence generated by the 2D logistic map. The 2D Logistic map [21] is given by (3) & (4).

$$x_{i+1} = \mu_1 x_i (1 - x_i) + \gamma_1 y_i^2 \quad (3)$$

$$y_{i+1} = \mu_2 y_i (1 - y_i) + \gamma_2 (x_i^2 + x_i y_i) \quad (4)$$

When $2.75 < \mu_1 \leq 3.4$, $2.75 < \mu_2 \leq 3.45$, $0.15 < \gamma_1 \leq 0.21$, $0.13 < \gamma_2 \leq 0.15$; the system generates two chaotic sequences in the region (0, 1]. The scaled values of the values give the pixel locations in the cover image. To make generated chaotic sequence more random, the map is preprocessed using the following [21],

$$X_i = 10^k X_i - floor(10^k X_i) \quad (5)$$

Where $X_i = x_i$ or y_i. The chaotic sequence $S' = \{(x_i, y_i) \mid i = 1,2,...\}$ generated by (3)-(5) sometimes cross the positive bounds of 1. Hence it is further transformed using (6).

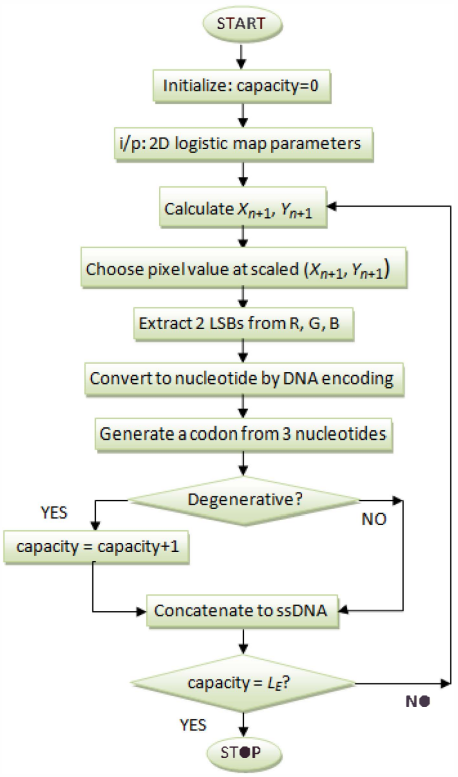

Figure 4: Flowchart for capacity check

$$X_i = X_i \; mod \; 1, \text{ if } X_i > 1 \qquad (6)$$

3) Encrypt the secret message: The secret message along with the meta-data is encrypted using the RC4 encryption algorithm.

4) Embed stego data: In this step 2 bits of secret data are hidden by replacing the bits of third base of a codon from the ssDNA if it shows degeneracy property. After all the message bits are embedded in the DNA, we perform an algebraic addition between the mutated ssDNA and the primer input by the user, using Table I. The same 2D map is generated once again and pixel LSBs are replaced with the new codons. The process is explained in flowchart Fig. 5 & 6.

B. *Extraction Algorithm*

The extraction procedure requires all the 3 keys used in the embedding procedure. Key 1 is used to regenerate the entire 2D chaotic sequence from which we can reconstruct the modified ssDNA. After primer subtraction using key 2, the degenerative codons are identified from the mutated ssDNA and the 3rd base is extracted to reconstruct the cipher text. The cipher text is decrypted using Key 3 to get the original the secret message. Entire procedure is explained in Fig. 7.

C. *Example for the proposed hiding algorithm*

Let us consider the DNA digital coding scheme: T=0(00), A = 1(01), C = 2(10), G = 3(11).

For 2D map parameters, $x_0 = 0.45$, $y_0 = 0.37$, $\mu_1 = 2.93$, $\mu_2 = 3.17$, $\gamma_1 = 0.197$ and $\gamma_2 = 0.139$, the cover image with selected pixels are shown in Fig. 8(a). Let value at i^{th} interval be, $\boldsymbol{x_i}$ = 0.6342, $\boldsymbol{y_i}$ = 0.2143. If the image size is 1024 × 768 then the scaled pixel location is, (floor (1024 × x_i), floor (768 × y_i)) = (floor (1024 × 0.6342), floor (768 × 0.2143)) = (649, 164).

If Fig. 8(b) shows the pixel value at (649, 164), then 1st base of codon comes from Red color 2 LSB, 10 = C; 2nd base from Green color 2 LSB, 10 = C; 3rd base from Blue color 2 LSB, 11 = G; Hence, the constructed codon is CCG.

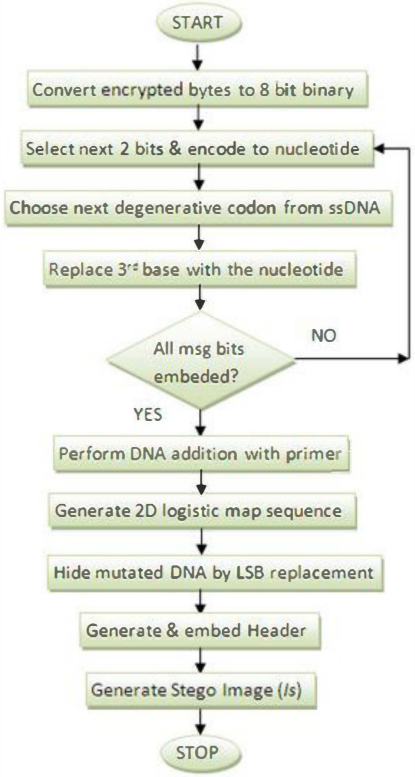

Figure 5: Flowchart for text embedding

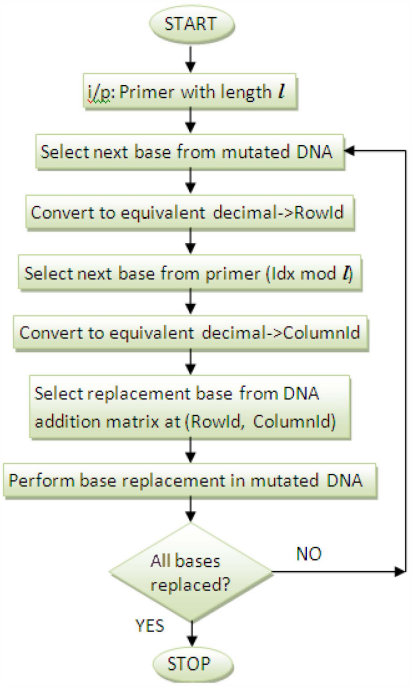

Figure 6: Flowchart for primer addition

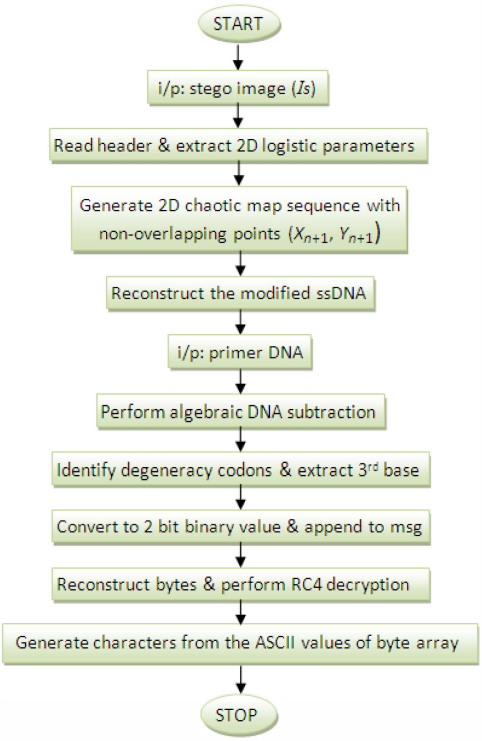

Figure 7: Flowchart for text extraction

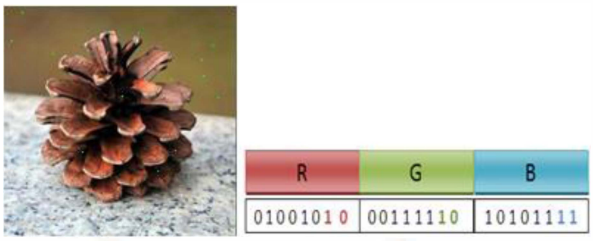

Figure 8: (a) sample cover Image, (b) pixel value at (649,164)

Let the secret message be "F22" and corresponding bit stream is 01000110 00110010 00110010. Using DNA coding the corresponding DNA strand is converted to - ATAC TGTC TGTC.

The generated DNA strand is shown in Fig. 9. The blue codons are the degeneracy codons (see map in Fig. 1).

GAAGACAACTTATGAGTCCGGCGCGAT
GGCGACCCACGATGCCTCACACTCTTG
TAGCCTAGCAATAACATGGTTGTT

Figure 9: Constructed DNA strand with degenerative codons.

After replacing the 3rd bases of degeneracy codons with the message nucleotides, the mutated DNA is shown in Fig. 10. The red nucleotides are the replaced bases.

GAAGACAACTTATGAGTACGTCGAGAT
GGCGACCTCGCTGCCTTACCCTTTTG
TAGCCGAGCAATAACATGGTTGTC

Figure 10: Mutated DNA after replacement with data bits

The primer DNA for addition is: AGGTCGCAAT. The primer addition operation is shown in Fig. 11. The final DNA strand is shown in Fig. 12 which is embedded back to the image pixels. The bases are marked to their corresponding color component.

GAAGACAACT TATGAGTACG TCGAGAT...
AGGTCGCAAT|AGGTCGCAAT|AGGTCGC...

TTTAAGACAC GTAAACTCAA GGCGGTT...

Figure 11: Primer addition with mutated DNA

TTTAAGACACGTAAACTCAAGGCGGTT
TTTTTGTCACTTCTGGCTTCAACGAAA
TTGAAACCGGAAACAGGCCCTCTA

Figure 12: Final DNA to embed back to the cover image

V. Conclusion & Future Work

The proposed work concentrates on hiding secret data in multiple layers of cover media. The DNA is attributed by the pixel properties of the image. Thus it makes it more secured than the methods using reference DNAs from public databases. The several parameters of 2D logistic map used, makes the algorithm further impenetrable. Yet the fully implemented system is required to be tested with several parameters of imperceptibility, which will be the focus of our future work.

References

[1] S. K. Bandyopadhyay and S. Chakraborty, "Image Steganography using DNA Sequence," Asian Journal Of Computer Science And Information Technology, vol. 1, No.2, pp. 50 – 52, 2011.

[2] L. Adleman, "Molecular computation of solutions to combinatorial problem Science," 266: pp. 1021-1024, 1994.

[3] C. Clelland, V. Risca, and C. Bancroft, "Hiding messages in DNA microdots," Nature, 399(6736), pp. 533-534, 1999.

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

[5] H. Shiu, K. Ng, J.F. Fnag, R. Lee, C. Huang, "Data hiding methods based upon DNA sequences," Information Sciences, 180, 2196–2208, 2010.

[6] A. Figureau, M. Soto, and J. Toha, "Biocryptography," Medical Hypotheses, vol. 54, No.3, pp. 394-396, 2000.

[7] P. Wong, K. Wong, and h. Foote, "Organic data memory using the DNA approach," *Communications of the ACM*, vol.46, No. 1, pp. 95-98, 2003.

[8] M. Arita, and Y. Ohashi, "Secret signatures inside genomic DNA," Biotechnology Progress, vol. 20, No.5, pp. 1605-1607, 2004.

[9] B. Shimanovsky, J. Feng, M. Potkonjak, "Hiding Data in DNA," In Petitcolas, F.A.P. (ed.) IH 2002. LNCS, vol. 2578, pp. 373–386. Springer, Heidelberg, 2003.

[10] C. Chang, T. Lu., Y. Chang, and C. Lee, "Reversible Data Hiding Schemes for DEOXYRIBONUCLEIC ACID Medium," International Journal of Innovative Computing, Information and Control, vol. 3, no. 5, pp. 1-16, 2007.

[11] M. Saeb, Eman El-Abd, and M. E. Ek-Zanaty, "On Covert Data Communication Channels Employing DNA Recombinant and Mutagenesis-based Steganographic Techniques," 2007.

[12] S. Jiao, R. Goutte, "Code For Encryption Hiding Data into Genomic DNA Of Living Organisms," ICSP2008 Proceedings, IEEE, 2008.

[13] P. Wasiewicz, l.J. Mulawka, W. R. Rudnicki, B. Lesyng, "Adding Numbers with DNA," International Conference on Systems, Man and Cybernetics, 2000, 265-270.

[14] Md R. N. Torkaman, P. Nikfard, N. S. Kazazi, Md R. Abbasy, and S. F. Tabatabaiee, "Improving Hybrid Cryptosystems with DNA Steganography," DEIS 2011, pp. 42–52, 2011.

[15] H. Mousa, K. Moustafa, W. Abdel-Wahed, and M. Hadhoud, "Data Hiding Based on Contrast Mapping Using DNA Medium," The International Arab Journal of Information Technology, Vol. 8, No. 2, April 2011.

[16] M. S. Arya, N. Jain, J. Sisodia, N. Sehgal, "DNA Encoding Based Feature Extraction for Biometric Watermarking," International Conference on Image Information Processing (ICIIP 2011), 2011.

[17] S. Chakraborty and Prof. S. K. Bandyopadhyay, "Two Stages Data-Image Steganography Using DNA Sequence," International Journal of Engineering Research and Development, Volume 2, Issue 7, PP. 69-72, August 2012.

[18] S. Chakraborty, S. Roy and Prof. S. K. Bandyopadhyay, "Image Steganography Using DNA Sequence and Sudoku Solution Matrix," International Journal of Advanced Research in Computer Science and Software Engineering, Vol 2, Issue 2, February 2012.

[19] A. Khalifa, A. Atito, "High-Capacity DNA-based Steganography," The 8th Internationa.l Conference on INFOrmatics and Systems (INFOS2012), Bio-inspired Optimization Algonthms and Their Applications Track, May, 2012.

[20] Md R. Abbasy, P. Nikfard, A. Ordi, and Md R. N. Torkaman, "DNA Base Data Hiding Algorithm," International Journal on New Computer Architectures and Their Applications (IJNCAA), vol. 2, No. 1, pp. 183-192, The Society of Digital Information and Wireless Communications, 2012.

[21] H. Liu, Z. Zhu, H. Jiang, B. Wang, "A Novel Image Encryption Algorithm Based on Improved 3D Chaotic Cat Map," The 9th International Conference for Young Computer Scientists, 3016-302l, 2009.