Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 604 – 611

Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

An Improved DNA based Dual Cover Steganography

Prasenjit Das[a], Subhrajyoti Deb[a,*], Nirmalya Kar[a], Baby Bhattacharya[a]

[a]*National Institute of Technology Agartala, Tripura, 799046 India*

Abstract

Dual cover steganography is an evolving technique in the field of covert data transmission. This paper focuses on the concept of using a theoretical single stranded DNA (*ssDNA*) as a primary cover, which is extracted from an inconspicuous cover image. We have analyzed the security loopholes and performance issues of the existing algorithm and proposed an improved algorithm on the same basis. Performance of both the algorithms are tested against several visual and statistical attacks, and parameterized in terms of both security and capacity. The comparison shows that the proposed improvements provide better overall security.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Deoxyribo Nucleic Acid; DNA algebra; logistic map; LSB replacement; metrics of distortion; neighbourhood; primer; steganography; difference image histogram

1. Introduction

In today's world the best way to protect data is to wipe out the very trace of existence of that data. Hiding information within cover media using steganographic techniques is the best way to do that. To ameliorate the security of steganography systems we need stronger algorithms as well as new cover media. Recently DNAs are being used as one such cover because of their high information density. Adleman[1], Clelland[2], Leier[3] have done some pioneering work in the field of organic DNA steganography. Some more relevant works have been proposed in[4–7]. Despite being a highly secured technique, DNA steganography algorithms have some common drawbacks like the biological errors (e.g. mutation) and difficulty of implementation. The most feasible solution to this problem is- use of theoretical model of DNAs. By utilizing its natural properties we can strengthen the existing hiding techniques. Some of the worth mentioning works are represented here.

A combination of cryptography and DNA steganography is utilized in[8], where steganography is used for hidden symmetric encryption key distribution on every new communication. Hayam Mousa *et al.*[9] adopted the reversible contrast mapping technique to develop a reversible information hiding scheme for DNA sequence. Meenakshi S Arya *et al.*[10] proposed a DNA encoding based algorithm to embed watermarks in images with the help of DNA cryptography and spread spectrum watermarking. In[11] a secret image is hidden with a cover image by creating 256

* Corresponding author. Tel.: +91-381-2348047.
E-mail address: subhrajyotideb1@gmail.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.103

combinations of DNA bases using 4 nucleotides and replacing them with the original pixel values. Suman *et al.* in [12] proposed a double cover DNA based steganography using magic numbers as the forward tracking algorithm. Suman Chakraborty *et al.* in [13] has proposed a loss-less DNA based secret image hiding technique using sudoku solution matrix. Amal *et al.* in [14] illustrate a DNA-based steganography method combined with a DNA cryptography technique for secure exchange of data in DNA carriers. In [15], a secret message is hidden inside a reference DNA strand collected from a publicly available DNA database. Later the indices (locations) of message bases is sent to the receiver.

In [16,17] a dual cover steganography is proposed, in which a theoretical single stranded DNA (ssDNA) is extracted from the pixel information of a cover image. The secret message is hidden inside the ssDNA, which in turn is hidden inside the image. During the embedding process the mutated DNA is processed through primer addition, which increases the steganographic security but causes an active mutation to the DNA. This reduces its security against different steganalytic attacks for the cover image. In this paper we propose some of the improvements over the existing model to enhance both its performance and security.

2. Proposed algorithm

Our proposed algorithm improves the existing dual cover steganography by reducing the noise bits in the secondary cover - the oligonucleotide. Moreover the algorithm is modified to accommodate secret image message (both 24-bit and 8-bit) in the cover. The process overview is given in the Fig. 1. We have used the same set of 3 keys. They are: Key 1 (K_{E1}) with six (6) parameters of 2D logistic map ($x_0, y_0, \mu_1, \mu_2, \gamma_1$ and γ_2), Key 2 (K_{E2}) - a primer (short DNA sequence) and Key 3 (K_{E3}) - a variable length key for RC4 encryption. The entire process consists of some modules discussed next.

Fig. 1. Overview of the proposed dual cover steganography

2.1. Prepare secret data and verify length constraints

Initially the secret message is transformed into an equivalent stream of DNA bases. It is performed in the following manner.

- Extract the byte information from the secret text/ image. Characters in text message are converted to their ASCII equivalent. Every pixel in a gray image contains single byte information, whereas 24-bit true color image pixels contain 3 bytes of information. The byte stream is encrypted using RC4 cipher and K_{E3}. Every encrypted byte is divided into 4 couples of 2 consecutive bits. Each couple is converted to a DNA base using DNA encoding technique [16]. For example, 00 is coded to T. Likewise, 01 is coded to A, 10 is coded to C and 11 is coded to G.
- Let A be a $(M \times N)$ cover image and f is a $(m \times n)$ secret image. Effective message length (L_E) is calculated as,

$$L_E = 4 \times (L + 3) \tag{1}$$

For text, L = number of characters in secret text. For 8-bit image, L = number of pixel in the image ($m \times n$) and for 24-bit image, $L = (m \times n \times 3)$. Extra 3 bytes considered for storing metadata like secret image dimension (width = height = 12 bits), or Logistic map iterator value N.

- We can hide 2 bits of data in a pixel. So to hide all the data, the following relation must exist.

$$L_E < M \times N \tag{2}$$

2.2. ssDNA extraction & primary cover capacity check

Prior to data hiding in the extracted DNA we need to check whether the DNA can contain the entire message or not. For this purpose we perform the following steps.

- Cover image pixels are used to construct the ssDNA codons. The pixels are chosen randomly based on the sequence generated by the 2D logistic map in (3) with the K_{E1} parameter values $2.75 < \mu_1 \leq 3.4$, $2.75 < \mu_2 \leq 3.45$, $0.15 < \gamma_1 \leq 0.21$, $0.13 < \gamma_2 \leq 0.15$ [16,20].

$$\begin{aligned} x_{i+1} &= \mu_1 x_i(1 - x_i) + \gamma_1 {y_i}^2 \\ y_{i+1} &= \mu_2 y_i(1 - y_i) + \gamma_2({x_i}^2 + x_i y_i) \end{aligned} \tag{3}$$

To improve the statistical properties (such as auto-correlation and cross-correlation), (3) is preprocessed by (4) and post processed by (5).

$$\begin{aligned} x_i &= 10^k x_i - floor(10^k x_i) \\ y_i &= 10^k y_i - floor(10^k y_i) \end{aligned} \tag{4}$$

$$\begin{aligned} x_i &= x_i \; mod \; 1, \text{if } x_i > 1 \\ y_i &= y_i \; mod \; 1, \text{if } y_i > 1 \end{aligned} \tag{5}$$

- Two Least Significant Bits from each color channel of every pixel indicated by (x_{i+1}, y_{i+1}) are converted to DNA bases by DNA encoding [9,17]. These bases are arranged sequentially to construct the DNA strand. Thus from a pixel we get a triplet of nucleotide bases, known as a codon and the 3rd base is obtained from blue channel.
- Now with K_{E2} we perform a DNA primer addition [18] between the ssDNA using base substitution (BS) box in Fig. 2. The primer bases are added to the first two bases of each codon [19]. If the two new bases combined with the 3^{rd} original base create a degenerative codon [17,19], then the original codon is marked for embedding and capacity counter is incremented.

Fig. 2. DNA primer codon selection (a) Base Substitution box; (b) Repeated Primer addition - degenerative codons are marked in blue and replaceable 3^{rd} bases are marked in red.

2.3. Embedding process of secret data

In this step 3^{rd} nucleotide bases of the marked codons are replaced by the encrypted message bases. Then the 2D logistic sequence is generated again and all the modified codons are embedded to their corresponding pixel locations. This stego image is sent over the insecure channel. This algorithm gives performance improvement because when the DNA is embedded back to the original cover image, only the blue channels of some pixels are affected due to bit flipping. As the blue channel has lowest impact on luminance, it gives better result against most of the attacks.

2.4. Extraction process of secret data

At the receiver end the ssDNA is extracted from the cover using 2D logistic map and K_{E1}. After applying DNA primer addition using the self-invertible BS-box and K_{E2}, the degenerative codons are identified and their 3^{rd} bases are extracted. The base sequence is processed by DNA decoding and later secret message bytes are extracted after RC4 decryption using K_{E3}.

3. Simulation and performance analysis

We conducted the following tests on both our proposed algorithm and existing algorithm to check and compare the effectiveness of hiding against some of the very common as well as most sophisticated attacks available today. Our secret message length varies from 2500 bases to 160000 bases. The algorithm hides 2 bits (1 base) per pixel. The key primer length is 20-40 bases. As a result distinct number of possible primers ranges from 4^{20}-4^{40}. The primer, along with the 6 parameters for 2D map provides a huge key-space for steganalysis purpose. Scheme-I or S-I represents the existing algorithm, Scheme-II (S-II) represents proposed algorithm.

3.1. Metrics of distortion calculation

To assess the quality of our stego images we calculated the following Full Reference metrics - Mean Squared Error (MSE), Peak signal to Noise Ratio (PSNR), Average Difference (AD), Laplacian Mean Squared Error (LMSE), Normalized Absolute Error (NAE), Structural Content (SC), Normalized Cross-Correlation (NCC) and Maximum Difference (MD). The metrics are calculated in terms of embedding capacity and distortion made to the image. We used the YCbCr color space and calculated the metrics over luminance channel because the human eye is most sensitive to luma information. Results in Table 1 & 2 show that, change in image distortion is quite less for our proposed algorithm with respect to payload amount.

Table 1. Metrics of distortion values at different payloads (scheme-I)

Payload	MSE	PSNR	AD	LMSE	NAE	MD	NCC	SC
10%	0.0410	61.993	-0.0029	0.00126	0.000213	3.0	1.0	0.999994
20%	0.0816	59.008	-0.0058	0.00251	0.000423	3.0	1.0	0.999988
30%	0.1228	57.236	-0.0086	0.00376	0.000635	3.0	1.0	0.999985
40%	0.1631	56.005	-0.0112	0.00498	0.000844	3.0	1.0	0.999982
50%	0.2038	55.036	-0.0136	0.00620	0.001054	3.0	1.0	0.999978
60%	0.2443	54.250	-0.0160	0.00740	0.001264	3.0	1.0	0.999979
70%	0.2800	53.659	-0.0179	0.00847	0.001446	3.0	1.0	0.999980
80%	0.3251	53.010	-0.0205	0.00981	0.001679	3.0	1.0	0.999980
90%	0.3656	52.500	-0.0232	0.01100	0.001888	3.0	1.0	0.999974

3.2. Visual attack

We applied the LSB enhancement method[21] to check any visually perceivable change between the carrier and the stegogramme. Fig. 3(a) - 3(c) shows the result. As we can see that even with 74.6% payload, difference between

Table 2. Metrics of distortion values at different payloads (scheme-II)

Payload	MSE	PSNR	AD	LMSE	NAE	MD	NCC	SC
10%	0.0024	74.290	0.00007	0.00007	0.000017	1.0	0.9999	1.000001
20%	0.0049	71.276	0.00009	0.00015	0.000035	1.0	0.9999	1.000003
30%	0.0073	69.515	0.00023	0.00022	0.000052	1.0	0.9999	1.000004
40%	0.0097	68.256	0.00035	0.00030	0.000069	1.0	0.9999	1.000006
50%	0.0121	67.267	0.00009	0.00037	0.000087	1.0	0.9999	1.000005
60%	0.0147	66.430	0.00046	0.00045	0.000106	1.0	0.9999	1.000010
70%	0.0172	65.760	0.00048	0.00053	0.000123	1.0	0.9999	1.000009
80%	0.0195	65.215	0.00064	0.00060	0.000140	1.0	0.9999	1.000012
90%	0.0222	64.662	0.00087	0.00068	0.000159	1.0	0.9999	1.000014

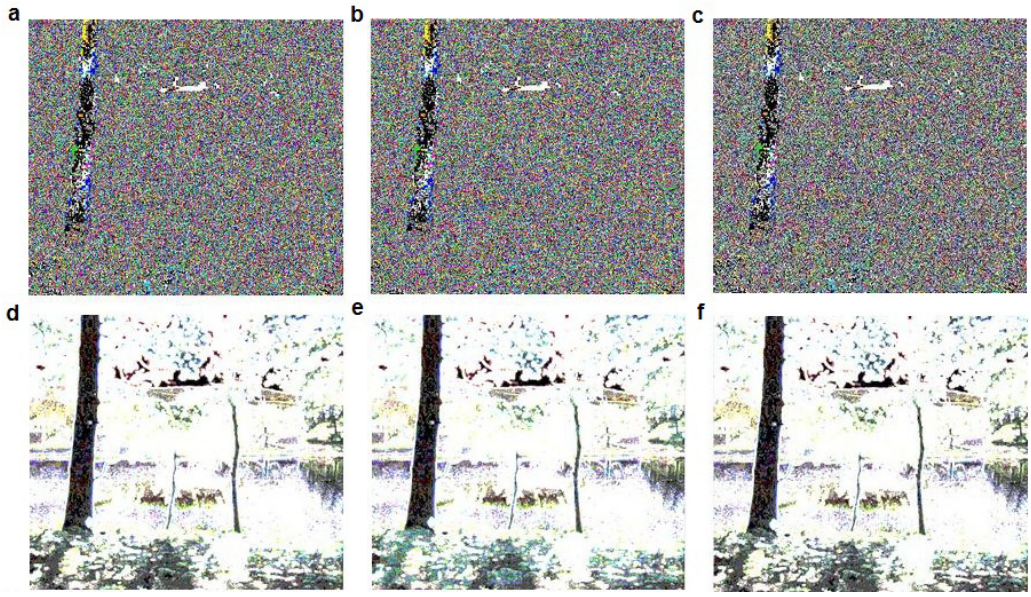

Fig. 3. (a), (d) LSB planes of cover; (b),(e) LSB planes of stegogramme with S-I; (c),(f) LSB planes of stegogramme with S-II.

the two LSB planes is barely distinguishable. Fig. 3(d) - 3(f) shows results of another representation of LSB planes by plotting the pixel values. By applying pixel-by-pixel comparison we can say that S-II stegogrammes have less changes in their LSB planes.

3.3. Statistical attacks

We applied 3 types of histogram analysis on stegogrammes created by both scheme-I and scheme-II. The histograms are computed over data from luminance channel, neighbourhood pixel property and pixel difference. First two attacks are used to detect distortion in spatial domain, whereas the last one is used in frequency domain.

3.3.1. Histograms Analysis

Histogram analysis is performed on luminance channel in YCbCr color space to detect significant changes in frequency of brightness, between the cover image with the stego-image. From the results shown in Fig. 4, we can say that with 65.4% of payload this algorithm preserves the general shapes of the histogram. It also exhibits minimal changes in several other parameters (such as mean, standard deviation, energy and relative entropy) as compared to scheme-I. For example entropy variation of luminance channel is 0.0007 (for S-I), whereas for S-II the value is 0.0003.

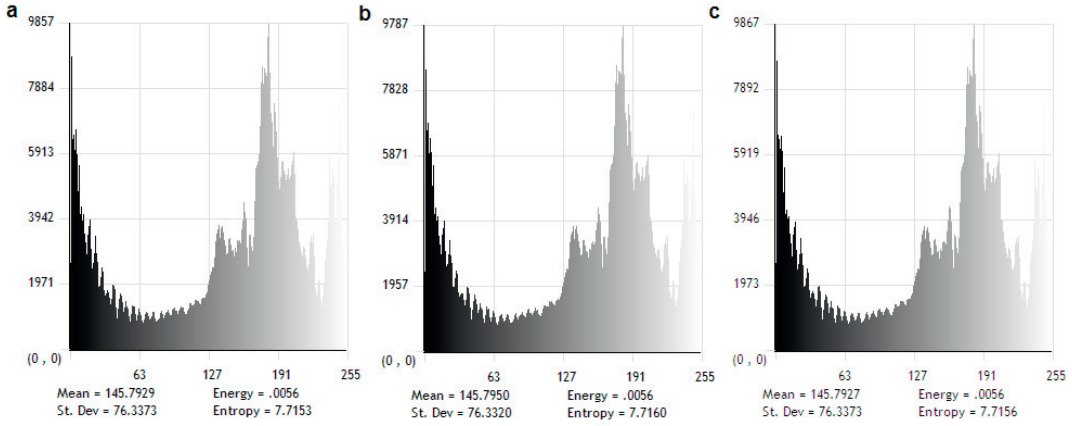

Fig. 4. Luminance histogram for (a) cover image; (b) stegogramme of scheme-I; (c) stegogramme of scheme-II

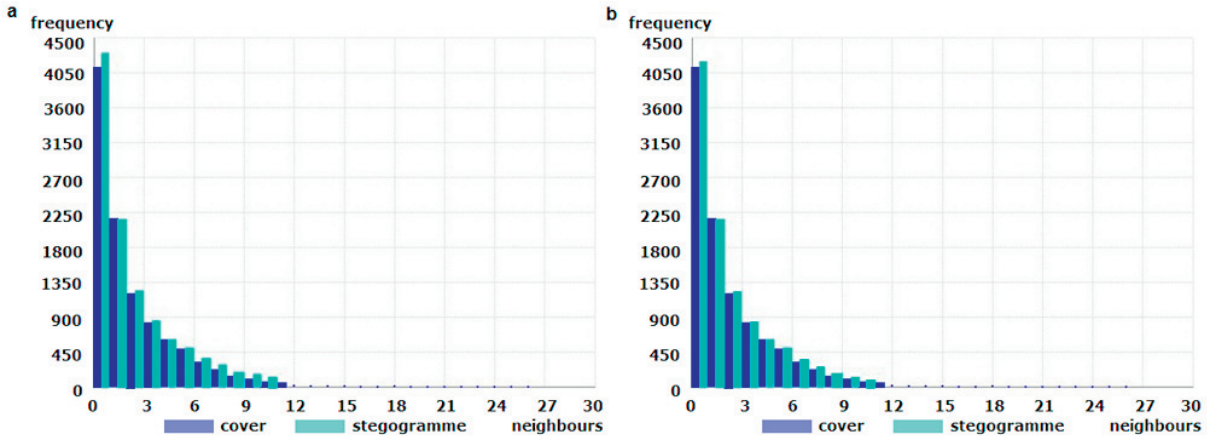

Fig. 5. Neighbourhood histogram for stegogramme with 71% payload for stegogramme with (a) scheme-I; (b) scheme-II.

3.3.2. Neighbourhood Histogram

Neighbourhood histogram[22] analysis shows the effect of LSB embedding on neighbourhood of each pixel. An inefficient steganographic technique increments or decrements LSB in such a fashion that it produces up to 26 neighbours for each modified pixel. From Fig. 5 we can see, with 70.62% payload, change in neighbours count is negligible for both algorithms, yet for scheme-II the frequencies of neighbours are very close to the original cover. Hence with moderate payload this test will fail to detect any changes to the cover for scheme-II.

3.3.3. Difference Image Histogram

We analyzed the difference image histograms[23] to identify any detectable changes in frequency domain. For a clean image the histogram illustrates a linear distribution to the frequencies of the DCT coefficients across zero (see Fig. 6) and no visible Pairs of Values (PoVs) are identified. For our stego images in Fig. 7 (a) & (b) the same properties are preserved. By comparing the stegogrammes from S-I and S-II we can see that S-II is slightly better than S-I, as it retains the natural Gaussian shape by maintaining the slope of bars in each range - very close to the original cover.

4. 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 this procedure makes it more secured than the methods using reference DNAs

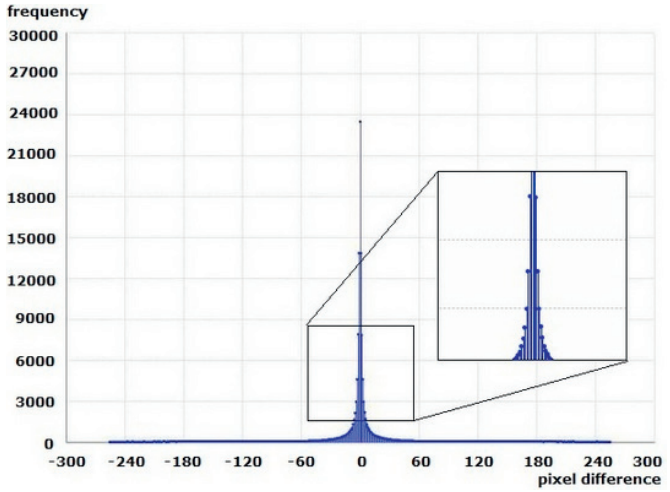

Fig. 6. Difference image histogram with DCT coefficients for original cover.

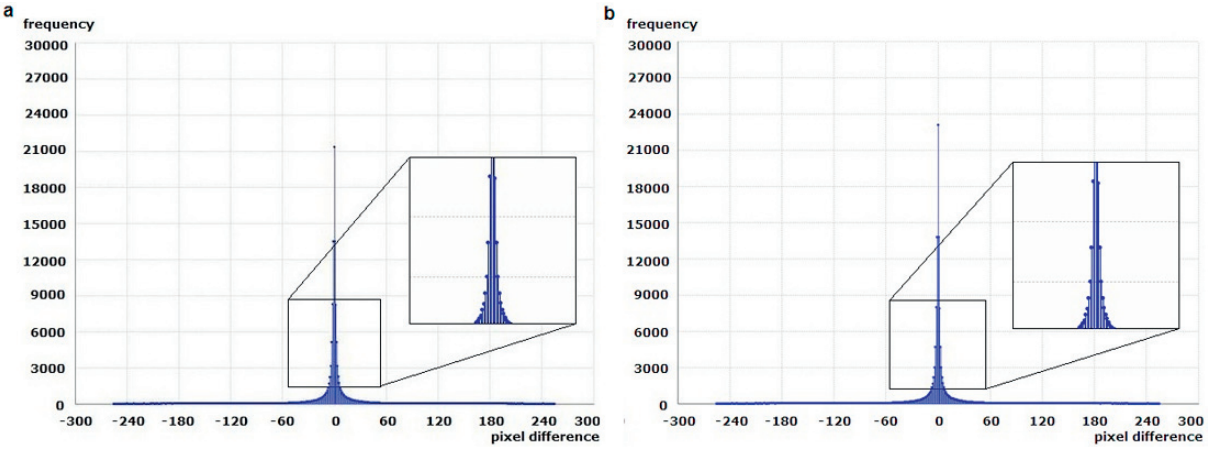

Fig. 7. Difference image histogram for stegogramme with (a) scheme-I; (b) scheme-II.

from public databases. The several parameters of 2D logistic map make the algorithm further impenetrable. Yet the system can be made more secured by avoiding the fixed 2 bpp capacity approach, which will require us to use the same pixel for upto 3rd level of overlapping, enabling a non predefined capacity ranging from 0 to 6 bpp. Multiple keys are required for the entire process starting from DNA construction to data embedding, and, their transfer between sender and receiver requires a secure key exchange protocol. The aforementioned two refinements will be the focus of our future work.

References

1. Adleman LM. Molecular computation of solutions to combinatorial problem. *Science* 1994; **266**:5187, p. 1021-1024.
2. Clelland C, Risca V, Bancroft C. Hiding messages in DNA microdots. *Nature* 1999, **399**:6736, p. 533-534.
3. Leier A, Richter C, Banzhaf C, Rauhe H. Cryptography with DNA binary strands. *BioSystems* 2000, **57**, p. 13-22.
4. Shiu H, Ng K, Fnag JF, Lee R, Huang C. Data hiding methods based upon DNA sequences. In: *Information of Sciences*, **180**:11, p. 2196-2208, 2010.
5. Shimanovsky B, Feng J, Potkonjak M. Hiding Data in DNA. In: *Procs. of the 5th International Workshop in Information Hiding*, LNCS, **2578**, p. 373-386, 2002.
6. Arita M, Ohashi Y. Secret signatures inside genomic DNA. *Biotechnology Progress*, **20**:5, p. 1605-1607, 2004.

7. Chang C, Lu T, Chang Y, Lee C. Reversible Data Hiding Schemes for Deoxyribonucleic Acid Medium. In: *International Journal of Innovative Computing, Information and Control*, **3**:5, p. 1-16, 2007.
8. Torkaman MRN, Nikfard P, Kazazi NS, Abbasy MR, Tabatabaiee SF. Improving Hybrid Cryptosystems with DNA Steganography. *DEIS 2011*, p. 42-52, 2011.
9. Mousa H, Moustafa K, Abdel-Wahed W, Hadhoud M. Data Hiding Based on Contrast Mapping Using DNA Medium. In: *The International Arab Journal of Information Technology*, **8**:2, p. 147-154, 2011.
10. Arya MS, Jain N, Sisodia J, Sehgal N. DNA Encoding Based Feature Extraction for Biometric Watermarking. In: *International Conference on Image Information Processing (ICIIP 2011)*, 2011.
11. Bandyopadhyay SK, Chakraborty S. IMAGE STEGANOGRAPHY USING DNA SEQUENCE. In: *Asian Journal Of Computer Science And Information Technology*, **1**:2, p. 50-52, 2011.
12. Chakraborty S, Bandyopadhyay SK. Two Stages Data-Image Steganography Using DNA Sequence. In: *International Journal of Engineering Research and Development*, **2**:7, p. 69-72, August 2012.
13. Chakraborty S, Roy S, Bandyopadhyay SK. Image Steganography Using DNA Sequence and Sudoku Solution Matrix. In: *International Journal of Advanced Research in Computer Science and Software Engineering*, **2**:2, February 2012.
14. Khalifa A, Atito A. High-Capacity DNA-based Steganography. In: *The 8th International Conference on INFOrmatics and Systems (INFOS2012)*, Bio-inspired Optimization Algonthms and Their Applications Track, May 2012.
15. Abbasy MR, Nikfard P, Ordi A, Torkaman MRN. DNA Base Data Hiding Algorithm. In: *International Journal on New Computer Architectures and Their Applications (IJNCAA)*, **2**:1, p. 183-192, 2012.
16. Das P, Kar N. A DNA Based Image Steganography using 2D Chaotic Map. In: *proceedings of International Conference on Electronics and Communication Systems (ICECS-2014)*, p. 149-153, 13th-14th February, 2014.
17. Das P, Kar N. A Highly Secure DNA Based Image Steganography. In: *IEEE International Conference On Green Computing, Communication And Electrical Engineering (ICGCCEE'14)*, 6th-8th March, 2014.
18. Wasiewicz P, Mulawka IJ, Rudnicki WR, Lesyng B. Adding Numbers with DNA. In: *International Conference on Systems, Man and Cybernetics*, p. 265-270, 2000.
19. Jiao S, Goutte R. Code For Encryption Hiding Data into Genomic DNA Of Living Organisms. In: *9th International Conference on Signal Processing (ICSP2008)*, p. 2166-2169, 2008.
20. Liu H, Zhu Z, Jiang H, Wang B. A Novel Image Encryption Algorithm Based on Improved 3D Chaotic Cat Map. In: *The 9th International Conference for Young Computer Scientists*, p. 3016-302l, 2009.
21. Westfeld A, Pfitzmann A. Attacks on steganographic systems. In: *Third International Workshop on Information Hiding*, **1768**, p. 61-76, 1999.
22. Westfeld A. Detecting low embedding rates. F.A.P. Petitcolas (Ed.), Information Hiding. *5th International Workshop*, Springer-Verlag Berlin Heidelberg; 2003. p. 324-339.
23. Zhang T, Ping X. Reliable detection of lsb steganography based on the difference image histogram. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 03)*, **3**, 2003.