Chaotic Image Encryption Technique using S-box based on DNA Approach

Anchal Jain
Inderprastha Engineering College, Ghaziabad, India.

Pooja Agarwal
Inderprastha Engineering College, Ghaziabad, India.

Rashi Jain
Inderprastha Engineering College, Ghaziabad, India.

Vyomesh Singh
Inderprastha Engineering College, Ghaziabad, India.

ABSTRACT
In recent years many DNA approach based encryption algorithms have been suggested but many of them are vulnerable to attacks or have slow execution time. Recently DNA Index based techniques have also been suggested. In this paper a new image encryption is proposed which is effective and has improved execution time compared to index based symmetric cryptographic algorithms having DNA approach. The algorithm presented in the paper is an integrated approach coupling the power of S-Box's and the two dimensional logistic map. Theoretical analysis and experimental results show that the proposed algorithm is reasonably secure and faster in run time compared to indexed based approach.

Keywords
Cryptography, DNA, Logistic map, S-box.

1. INTRODUCTION
The rapid advancement in the field of information technology has increased the capability to exchange a large volume data across the widespread network of computer over the globe. Security is a prime concern when data is transmitted across the network of computers. Various types of attacks are possible on the data being transmitted so protecting the data against such unauthenticated access and keeping its integrity intact is an important task.

Cryptography is mainly divided in two fields the symmetric key cryptography and asymmetric cryptography depending upon the no of keys used in the process of encryption and decryption. Since traditional cryptographic algorithms are susceptible to various kinds of attack therefore the thrust to explore new ideas to enhance the security of data has led to the emergence of new era of data security based upon the analogy of the biological sciences with the computer sciences called DNA cryptography. Deoxyribose Nucleic Acid (DNA) is the genetic material inside the living creatures responsible for carrying the hereditary features from parents to the offspring. DNA is polymer which is made up of several monomer units called nucleotides. Each nucleotide is further composed of following three units: Deoxyribose sugar, phosphate group and the nitrogenous base. The nitrogenous bases are Adenine, Guanine, Thymine and Cytosine. Watson-Crick proposed the complementary DNA structure and this inherent complementarily is used in DNA computing. The ability to store large amount of data, inherent complementarily and massive parallelism makes it suitable choice towards security. DNA computing does not actually involve the real biological DNA strands rather it uses the central principle behind DNA. Initially work in DNA computing was done by Adleman [1] where Hamiltonian path problem was solved using DNA approach. The principle which Adleman used is the ability to represent information in DNA clusters. Gehani et al. [2] also proposed DNA based cryptography. Since then several algorithms have been proposed by various others which include the DNA based implementation of YAEA algorithm [3]. Various authors have taken the advantage of DNA based operations such as addition, complementation, deletion, insertion, truncation, transformation for proposing encryption scheme. Qiang Zang et al. [4] proposed an image encryption using DNA addition combining with chaotic maps. But Houcemeddine Hermassi et al. [5] proved that this algorithm was weak against known plain text attacks. The complementary property of DNA nucleotides is also utilized in DNA based cryptography. Hongjun Liu et al. [6] proposed the image encryption scheme based on DNA complementary rule combined with chaotic map. Qiang Zhang [7] also proposed an encryption scheme using DNA operations like as deletion, insertion, truncation, and transformation combined with chaotic maps. Chaotic maps have been widely used by authors to propose effective encryption schemes. In paper [8] also, one dimensional chaotic logistic map is used to propose encryption scheme using neural network.

Recently Grasha Jacob et al. [9, 10] proposed two image encryption schemes using DNA gene bank based indexing. Zhang et al. [11] also proposed the index based symmetric DNA encryption algorithm using chaotic key generator. In DNA based indexing approach the cipher generation is based upon searching the DNA encoded original plain text pattern in gene bank, the search is sequential and execution is slow in generating cipher. All these papers [9, 10, 11], have time complexity issue. In this paper a new effective and comparatively faster image encryption technique based on DNA substitution box is proposed. The proposed algorithm is compared with the techniques [9, 10] and found to be faster in execution time. The paper is organized in total 5 sections. In section 2, the step by step process for encryption is described. In section 3, decryption algorithm is described. In section 4, the algorithm is analyzed against various attacks such as, brute force attack, key space analysis, statistical analysis including correlation and histogram analysis and in section 5 paper is concluded.

2. PROPOSED ENCRYPTION PROCESS
The image is first encoded into DNA sequence and the resulting sequence is substituted using DNA based S-Box which generates the intermediate cipher and finally a chaotic sequence is generated using 2 D logistic mapping which

permute the rows and columns of intermediate cipher to produce final cipher. Encryption process is divided into following steps:

2.1 DNA Encoding of Image

A gray scale image has the pixel range from 0 to 255. The input image of size M×N is taken and each pixel value is converted into its 8 bit binary equivalent. Now, the 8 bit binary equivalent is converted into a quadruple nucleotide sequence using the following scheme:

C→00
T→01
A→10
G→11

The process is repeated for all the pixels in the image and a matrix of size M×4N is generated.

2.2 Generation of DNA based S- Box

The inception behind S-box traces back to cryptographic substitution introduced by Shannon [12] S-box can be classified in two groups namely static S-box where the input vector does not changes and dynamic S-box where the change input vales result in the changes of corresponding output as well [13].

A good design of nonlinear S-box ensures better cipher complexity. The proposed algorithm in this paper uses a DNA based static S-box for the encryption scheme. To create DNA S-box a gene sequence is downloaded from the gene bank. This sequence is a binary file that contains a single DNA strand.

Canis_familiaris chromosomal sequence (genome) from the gene bank is sufficiently large enough to encode image. For sake of simplicity the algorithm has been implemented on gray scale images. Intensity values of gray scale image lies in the range $0 \leq f(x, y) \leq 255$. Therefore, the static S-box will be of size 16×16 having total 256 entries in it. Each such entry will correspond to a gray scale value in image. Start the search in the downloaded genome from location x which is part of secret key for the quadruple nucleotide sequences and pick 256 quadruple with equivalent distinct decimal values . For example suppose one of the quadruple DNA nucleotide sequence is "ATGC" the equivalent decimal value will be 156. Therefore 256 distinct values are chosen and filled in S-box sequentially in row major fashion. The generated s box is shown in figure 1.

2.3 Generation of Chaotic Sequence and Final Cipher

For generating the initial condition method described in [4] is used. Calculate two parameters k_1 and k_2 as in (1)

$$\left.\begin{aligned} k_1 &= \frac{1}{256} mod\left(\sum_{i=1}^{\frac{m}{2}} \sum_{j=1}^{n} P_{ij}\ , 256\right) \\ k_2 &= \frac{1}{256} mod\left(\sum_{i=\frac{m}{2}}^{m} \sum_{j=1}^{n} P_{ij}\ , 256\right) \end{aligned}\right\} \quad (1)$$

where P_{ij} is the value of the image pixel at location (i, j) in the image. Further let $x_0' = 0.59$ and $y_0' = 0.15$. Calculate initial conditions as in (2)

$$\left.\begin{aligned} x_0 &= mod\left(\left(x_0' + k_1\right), 1\right) \\ y_0 &= mod\left(\left(y_0' + k_1\right), 1\right) \end{aligned}\right\} \quad (2)$$

The proposed algorithm uses 2 D logistic map based on chaotic sequence is defined as in (3)

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

where $2.75 < \mu_1 \leq 3.4$, $2.75 < \mu_2 \leq 3.45$, $0.15 < \gamma_1 \leq 0.21$, $0.13 < \gamma_2 \leq 0.15$, for example values are set a $\gamma_1 = 0.16$ and $\gamma_2 = 0.14$,

INDEX	CC	CT	CA	CG	TC	TT	TA	TG	AC	AT	AA	AG	GC	GT	GA	GG
CC	131	208	50	13	64	216	15	129	32	211	0	67	96	248	38	14
CT	108	54	128	48	52	2	99	40	97	11	56	220	62	12	224	66
CA	8	205	136	24	20	100	6	1	202	3	212	180	244	112	140	178
CG	139	195	160	61	192	176	58	34	179	227	53	254	232	246	35	63
TC	190	23	76	200	55	252	80	240	222	127	207	141	184	29	206	134
TT	217	47	143	51	59	162	4	104	16	218	72	22	135	45	98	172
TA	132	44	120	204	28	247	253	88	163	255	115	189	223	17	199	230
TG	237	126	241	251	60	119	245	147	65	243	137	138	130	74	117	57
AC	94	168	221	101	122	148	10	103	153	69	175	79	242	239	191	111
AT	219	125	95	249	114	49	7	43	209	146	33	18	42	157	78	151
AA	234	90	183	82	165	89	81	87	235	152	250	75	118	124	113	226
AG	92	68	164	36	167	83	145	106	37	86	173	70	169	85	142	166
GC	236	109	203	174	19	21	229	213	197	84	210	93	228	171	39	5
GT	156	121	185	158	30	46	187	215	186	116	283	149	27	123	225	144
GA	150	177	233	170	231	196	159	188	25	161	41	26	31	77	194	9
GG	155	201	107	133	181	71	91	73	214	110	102	105	154	198	182	193

Fig. 1: Static S box generated from the gene bank having 256 distinct values

μ_1=3.2 and μ_2=3.3 these values are part of the secret key. Use x_0 and y_0 as initial conditions for (3) and obtain two matrices of size 1×256 as in (4).

$$\left.\begin{aligned} X_m &= (x_1, x_2, x_3 \ldots . x_m) \\ Y_m &= (y_1, y_2, y_3 \ldots . y_m) \end{aligned}\right\} \quad (4)$$

Using (4) permute the rows and columns as in (5):

$$\left.\begin{aligned} R(i) &= R((x_i \times m) mod\ i) \\ C(j) &= C((y_j \times n) mod\ j) \end{aligned}\right\} (5)$$

The pseudo code for encryption

Input: X (image) to be encrypted, G (Binary file which contains the nucleotide sequences) and Secret Key (start location x for search in gene bank, end location y in gene bank, x_0', y_0', μ_1, μ_2, γ_1, γ_2)

Output: Encrypted image Y.

Step 1: DNA encoding of input image resulting in a matrix P of size (m×4n) as described in subsection 2.1.

Step 2: Create a DNA based S-box as described in subsection 2.2.

Step 3: Now substitute the pixel values of image with the values from S-box as:

For example: If X (i, j) = 109

Then equivalent DNA encoded quadruple is M= 'TAGT' Now this M can be split into 2 parts, m= 'TA' and n= 'GT'.

The new substituted values X (i, j) are calculated using m and n as index in substitution box. Where m is the row index and n is the column index.

X (i, j) =S_box (decimal (m), decimal (n)).

This generates an intermediate cipher image of the same size as that of the original image.

Step 4: Now permute the rows and columns as described in (5). This step generates the final cipher image Y.

3. DECRYPTION

Input: Encrypted image Y, G (Binary file which contains the nucleotide sequences) and Secret Key (start location x for search in gene bank, end location y in gene bank, x_0', y_0', μ_1, μ_2, γ_1, γ_2).

Output: Original image X.

Step 1: Perform the reverse permutation operation on the rows and columns of the cipher image Y using the reversible mapping of (5).

Step 2: Create an inverse S-box and perform reverse substitution to obtain the original image.

```
%%Creation of inverse S-Box
For i=1:256
        in_sbox (s_box (i) +1) =i-1
end
inverse sbox1=reshape (in_sbox, 16, 16)
```

4. SECURITY ANALYSIS OF THE ALGORITHM

The algorithm was simulated on Matlab R2012a to determine the efficiency and security against various kinds of attacks. Standard 256×256 gray image Lena.tiff and Peppers.png are used and the initial parameters as x_0=0.59, y_0=0.15, μ_1=3.2, μ_1=3.3, γ_1=0.16 and γ2=0.14. The encrypted image is shown in figure (2).

4.1 Efficiency of the Algorithm

The algorithm was tested on different images and the run time for the encryption scheme was noted down and tabulated along with the run time of the previous algorithm that used indexed based sequential search for cipher generation. The results in table 1 shows that proposed algorithm has better run time than the run time of encryption scheme in [9, 10].The run time result has been calculated by finding the average of the values obtained by applying the algorithm on different images.

Table 1: Comparison of Time Complexities of the Proposed Algorithm and Indexed Based Approach (in milliseconds)

Algorithm	***Time Complexity***	
Proposed Algorithm	Generation of S Box	Cipher Generation
	4950.408	23488.311
Indexed Based Algorithm	36078.380	

4.2 Security Analysis

The resistance of the algorithm against various categories of attacks defines its ability to protect the data in transit. The process of cryptanalysis includes various cryptographic attacks and the immunity of the cryptographic algorithm against such attacks should be at par without the access of keys. An encryption technique must be able to resist against cryptanalytic and brute force attack.

4.2.1 Key Space Analysis

In proposed algorithm, location of gene bank, the initial value and the system parameter of the chaotic maps can be seemed as secret key. Thus, there are eight secret keys. (Start location x, end location y, x_0, y_0, μ_1, μ_2, γ_1, γ_2). If the precision is 10^{-12}, the secret key's space is ($10^{-12} \times 10^{-12} \times 10^{-12} \times 10^{-12} \times 10^{-12} \times 10^{-12}$) .The secret key's space is large enough to resist exhaustive attack.

4.2.2 Statistical Analysis

The cipher generated by the algorithm must not show any predictable statistical relation to the original image in order to prevent the attacker to use that information and to use it to decipher the information being transmitted. Statistical attacks include histogram analysis and correlation coefficient analysis.

4.2.2.1 Histogram Analysis

The histogram of an image is the profile of the image describing the distribution of the image pixels against the various intensity levels. A good encryption algorithm is the one which generates a cipher image whose histogram is

completely different from that of an original image. Thus it will be difficult for attackers to identify the pixels from the encrypted image having similar nature to that of the original image. The histogram for the cipher and the original image has been shown in figure 2.

4.2.2.2 Correlation Coefficient Analysis

There exists a high correlation between the pixel values of image. The adjacent pixels are highly correlated. A good encryption algorithm must produce an encrypted image where there exists a very little correlation between the existing pixels. Pearson's correlation coefficient is given by the formula:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \qquad (6)$$

Table 2: Comparison of Correlation Coefficients

File Name		Peppers.png	Lena.tif
γ_{col}	Original image	0.9478	0.9201
	Encrypted image	0.0271	0.0200
γ_{row}	Original image	0.9482	0.9560
	Encrypted image	0.0330	0.0267

Where x and y are the grayscale values of adjacent pixels in the image and n is the number of image pixels chosen for calculation of correlation coefficient. A value near to zero for the correlation coefficient shows a little or no correlation. The correlation coefficient for cipher image and original image has been shown in table 2. And the corresponding scatter diagrams are shown in fig.2

4.2.3 Differential Attacks

The sensitivity of the algorithm against slight changes is a key issue that should be kept in mind while designing the encryption algorithm. For a slight change in the original image the extent to which the encrypted image changes represents the strength of the encryption algorithm against differential attacks. Pixel change rate measures the change in the entire encrypted image corresponding to the change in the single pixel value in the original image. Let E_1 and E_2 be the two encrypted images of the original images which vary only by difference of one pixel value. Then define Number of Pixel Change Rate (NPCR) as in (7)

$$NPCR = \frac{\sum_{i,j} S(i,j) \times 100\%}{W \times H} \qquad (7)$$

Where W=Width of the image, H=Height of the image

$$S(i,j) = \begin{cases} 1 & E_1(i,j) \neq E_2(i,j) \\ 0 & \text{otherwise} \end{cases}$$

NPCR values of Lena.tiff and Pepper.png is found out to be 99.03% and 99.01% respectively, which is consistent with the results of the algorithm.

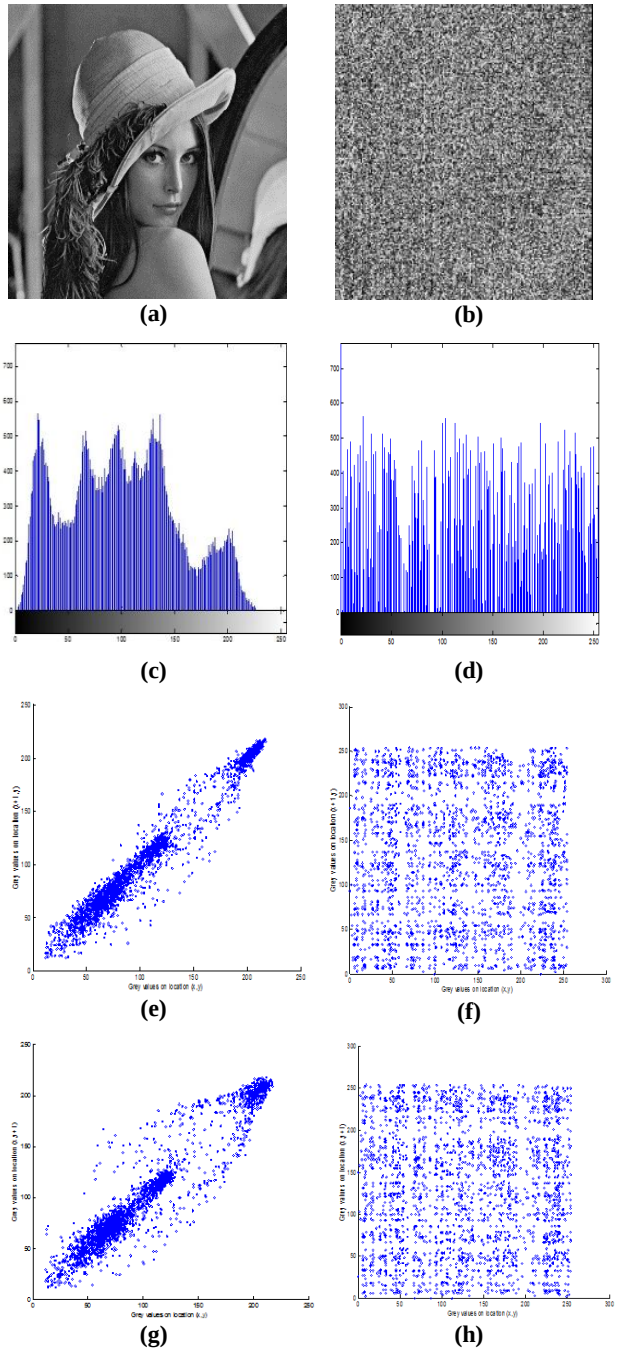

Fig. 2 (a) Original Image Lena.tiff, (b) Cipher Image of Lena.tiff, (c) Histogram of Original Lena image, (d) Histogram of Ciphered Lena Image, (e) Horizontal Scatter Plot of Plain Lena Image, (f) Horizontal Scatter Plot of Ciphered Lena Image, (g) Vertical Scatter Plot of Plain Lena Image, (h) Vertical Scatter Plot of Ciphered Lena Image.

5. CONCLUSION

A Symmetric image encryption algorithm based upon DNA S-Box and chaotic sequence is proposed in this paper. First, the unique DNA based s-box performs the substitution on the DNA encoded image and then confusion is achieved by shuffling rows and columns of cipher. The simulation of the algorithm shows that it is has faster execution time than the algorithms proposed in [9, 10]. The proposed technique is analyzed in terms of brute-force attack, key space analysis, statistical analysis and satisfactory results have been found.

Further, in future the technique can be tested on various other attacks. It can also explore the use of dynamic S- box for better computing security.

6. REFERENCES

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

[2] Ashish Gehani, Thomas LaBean and John Reif. "DNA-Based Cryptography. DIMACS DNA Based Computers," V, American Mathematical Society, 2000.

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

[4] Qiang Zhang, Ling Guo, Xiaopeng Wei, "Image encryption using DNA addition combining with chaotic maps", Mathematical and Computer Modeling 52 (2010) 2028_2035.

[5] 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", Journal of Multimedia Tools and Applications, DOI 10.1007/s11042-013-1533-6.

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

[7] Qiang Zhang, Xianglian Xue, and Xiaopeng Wei, "A Novel Image encryption Algorithm Based On DNA Subsequence Operation" The Scientific World Journal, Volume 2012(2012) , Article ID 286741.

[8] Anchal Jain, Navin Rajpal, "A Two Layer Chaotic Neural Network based Image Encryption Technique", IEEE National conference on computing and Communication systems, ISBN 978-1-4673-1952-2, 2012.

[9] Grasha Jacob, A. Murugan, "An Encryption Scheme with DNA Technology and JPEG Zigzag Coding for Secure Transmission of Images", arXiv: 1305.1270v1, 2013.

[10] Grasha Jacob, Murugan A, "A Hybrid Encryption Scheme using DNA Technology", IJCSCS Vol 3, Feb 2013.

[11] Zhang Yunpeng, Zhu Yu, Wang Zhong, Richard O.Sinnott, "Index-Based Symmetric DNA Encryption Algorithm", IEEE (4th CISP), DOI 10.1109/CISP.2011.6100690.

[12] C.E. Shannon, "Communication theory of secrecy systems, Bell System" Technical Journal 28–4 (1949) 656–715.

[13] Xingyuan Wang · Qian Wang "A novel image encryption algorithm based on dynamic S-boxes constructed by chaos", Nonlinear Dyn (2014) 75:567-576.