Secure Data Communication and Cryptography Based on DNA Based Message Encoding

Atanu Majumder[1], Abhishek Majumdar[2], Tanusree Podder[3], Nirmalya Kar[4], Meenakshi Sharma[5]
[1]Dept. of Electronics and Information Technology, NIC Tripura State Centre, Agartala, INDIA
[3,4]Dept. of Computer Science and Engineering, NIT Agartala, Tripura, INDIA
[2,5]Dept. of Computer Science and Engineering, SSCET, Badhani, Punjab, INDIA
[1]atanu.cse21@gmail.com, [2]abhishekmajumdar91@gmail.com, [3]nirmalya@nita.ac.in, [4]tanusreepodder29@gmail.com, [5]hod.csebadhani@srisaigroup.in

***Abstract*—Secure data communication is the most important and essential issue in the area of message transmission over the networks. Cryptography provides the way of making secure message for confidential message transfer. Cryptography is the process of transforming the sender's message to a secret format called cipher text that only intended receiver will get understand the meaning of the secret message. There are various cryptographic or DNA based encoding algorithms have been proposed in order to make secret message for communication. But all these proposed DNA based encryption algorithms are not secure enough to provide better security as compared with the today's security requirement. In this paper, we have proposed a technique of encryption that will enhance the message security. In this proposed algorithm, a new method of DNA based encryption with a strong key of 256 bit is used. Along with this big size key various other encoding tools are used as key in the encoding process of the message like random series of DNA bases, modified DNA bases coding. Moreover a new method of round key selection is also given in this paper to provide better security in the message. The cipher text contains the extra bit of information as similar with the DNA strands that will provide better and enhanced security against intruder's attack.**

Keywords—Encoding tools; DNA sequence; Final Cipher; Ciphertext; Coded message

I. INTRODUCTION

Security in data communication is required when message transfer between sender and receiver is needed to be kept confidential. Cryptography is the process of achieving confidentiality in message transfer. Cryptography can be thought of as a process of secret writing in order to protect data or message from various attacks of the intruder. Secret writing is achieved through the process of transforming a message called plaintext into cipher text by means of a cryptographic algorithm. Security is concerned with the protection of message or data while transmitting over the networks. But now-a-days to achieve complete data security is a challenging issue of data or message transfer. In order to get better security in message transfer several DNA based encryption scheme had been proposed [5][7]. In order to enhance data security and make the data more confidential effective encryption algorithms are required.

DNA based encryption method is one of the recent technique embedded into cryptographic field, lot of researchers are working on this. Some of them used DNA computing, while some other applied biological property of DNA strands and DNA sequence after making few modification [2][6]. DNA complementary rule substitution, message embedding within a DNA sequence, makes cipher text much larger in size compare to plaintext size [3]. In order to reduce the cipher size a modified DNA substitution has been adopted in this paper using the properties of DNA strands and DNA sequences.

The encryption algorithm proposed here is based on the combination of concept of DNA based cryptography and conventional cryptography. The algorithm proposed here works on block cipher with a key of 256 bit. The encryption algorithm has the four round of coding and each round has used the concept of cipher block chaining coding. Moreover the proposed system has a new scheme of key selection for round operation for better security aspects.

II. DNA BIO-LOGICAL THEOREM AND RELATED WORKS

Bio-logical DNA is made up of two strands of nucleotides, each of the strands are coded with four DNA bases namely A, T, C, G (A – adenine, G – guanine, C – cytosine, T – thymine). In DNA double-stranded structure, DNA bases are bonded each other by hydrogen bonds: A with T and C with G, which is called the complementary pairs of DNA strands.

Most of the security related researchers are working on data security and they are trying to implement DNA concept into their proposed algorithms directly or indirectly. Some of them are using DNA computing in their algorithms while others are working on incorporating DNA properties into their algorithms. They had pointed out that the properties of DNA sequences which they can utilize to encrypt data by incorporating the message into the DNA sequence [3] and working also on incorporating extra coding within the cipher text.

III. PROPOSED METHOD

In order to provide better security and reliable data transmission an efficient method of DNA based cryptography is proposed here. The algorithm has two phases; these are round key selection and message encryption.

In the round key selection phase, a key of 256-bit is chosen randomly for encryption. This 256-bit of key is then transformed into an 8X8 matrix with each cell having 4-bit key value. Transform the 256-bit key into the matrix in row wise.

ISBN No. 978-1-4799-3914-5/14/$31.00 ©2014 IEEE

Then read the key bits in column wise, two columns at a time. Thus every two columns will produce a sub-key of 64-bit that is a total of four blocks of sub-keys will generate. Each block of sub-keys is labeled with one of four DNA bases namely A, T, C, G. Select randomly a DNA sequence of length four with no repetition of DNA bases. There are 24 number of possible combination like TGCA, AGTC etc., randomly one sequence will be selected for round key selection. At the time of encryption these sub-keys blocks are used in four round of encryption operation. The secret key, selected DNA sequences, position of extra coding are to shared between the sender and receiver through a secret channel.

In message encryption phase, the byte values are extracted from the input file or message [1]. The encryption process is works on the unsigned byte values of the input file or text called plaintext [4]. Then these byte values will be transformed into 8-bit binary. Plaintext is then divided into 256-bit blocks; each block of plaintext will go through the encryption process. Then in round operation 256-bit block of plaintext is divided into four 64-bit plaintext blocks. Now perform 64-bit Ex-OR operation between fourth block of plaintext and the round 1 key, K1. The result is then getting Ex-ORed with the third block of the plaintext. Then this result will go for second block and so on. Then these four 64-bit results will go through a straight D-Box. The output of the D-Box will be used as input for the second round. In second round, round key 2, K2 will get Ex-ORed with the first block of plaintext, the result of this will be used as key for second block as like as cipher block chaining method. The four round of encryption operations are shown in following figure:

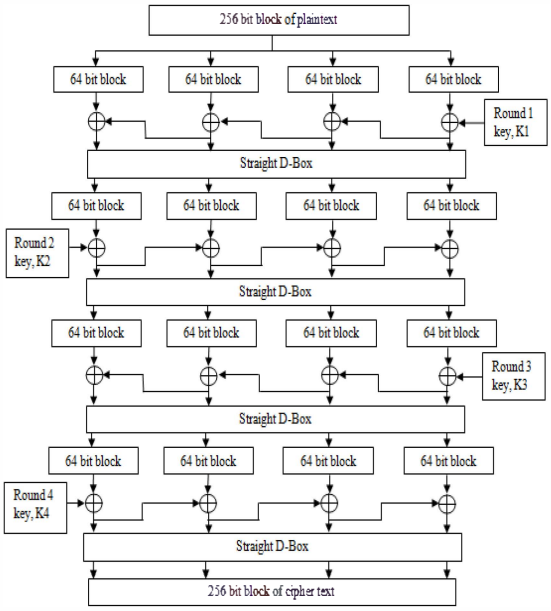


Figure 1. Schematic diagram of proposed encryption technique

In order to embed more security extra bits are padded at both the end of the cipher text and also within the cipher text [4]. The final cipher contains the extra coding namely starting primer (SP), file type code (FTC), CRC code and ending primer (EP) as like as biological DNA strand contains introns, primers as extra information.

SP	CTB_1	FTC	CTB_2	CRC	CTB_3	EP

Figure 2. Cipher text in the form of DNA strands

The cipher text created after round key operation is hidden in the final form of the cipher text.

IV. ALGORITHMIC PRESENTATION

The encryption scheme proposed in this paper has the following phases:

A. Round Key Selection

In this phase, a key of size 256-bit is selected randomly. This selected key is transform into an 8X8 matrix.

Let, K be the key, K='1011 1010 0011 0011 1100 1100 1010 0011 0000 0000 0000 0000 1111 1111 1111 1111 1110 1110 1001 0011 0000 1010 1111 0100 1011 1100 0101 1001 0011 1011 0001 1010 0011 1001 0011 0100 1010 1100 1001 1010 0000 0001 1000 1010 1111 0001 1010 0101 0000 0001 1010 1100 1000 1111 1000 1111 0001 1111 0011 0010 1100 0001 1111 1000'

Transformation of key values into matrix row wise (in tabular form):

1011	1010	0011	0011	1100	1100	1010	0011
0000	0000	0000	0000	1111	1111	1111	1111
1110	1110	1001	0011	0000	1010	1111	0100
1011	1100	0101	1001	0011	1011	0001	1010
0011	1001	0011	0100	1010	1100	1001	1010
0000	0001	1000	1010	1111	0001	1010	0101
0000	0001	1010	1100	1000	1111	1000	1111
0001	1111	0011	0010	1100	0001	1111	1000

Read the key values column wise (two columns at a time), that generates four sub-keys. Label the sub-keys with DNA bases (A, T, C, G) as follows:

A='1011 0000 1110 1011 0011 0000 0000 0001 1010 0000 1110 1100 1001 0001 0001 1111'

T='0011 0000 1001 0101 0011 1000 1010 0011 0011 0000 0011 1001 0100 1010 1100 0010'

C='1100 1111 0000 0011 1010 1111 1000 1100 1100 1111 1010 1011 1100 0001 1111 0001'

G='1010 1111 1111 0001 1001 1010 1000 1111 0011 1111 0100 1010 1010 0101 1111 1000'

Let, randomly selected DNA sequence with DNA bases be 'TGCA' then,

Round 1 key, K1=T

Round 2 key, K2=G

Round 3 key, K3=C

Round 4 key, K4=A

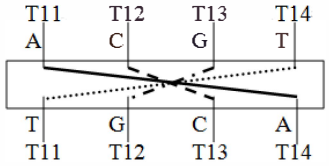
B. Message Encryption

Every 256-bit plaintext block will go through the four round of encryption process. After each round coded blocks will go through a straight D-Box. The D-Box has four input and output terminals. The inputs terminals are labeled with DNA bases (A, T, C, and G). The D-Box will works on the randomly selected DNA sequence of length four. The encryption algorithm is given below:

Step 1: Read the byte values from the input file called plaintext and transform each byte value into 8-bit binary representation.
Step 2: Make 256-bit plaintext blocks from the binary representation
Step 3: Repeat step 4 and 5 for each block of plaintext
Step 4: Split the 256-bit block into four 64-bit blocks, namely P1, P2, P3, P4

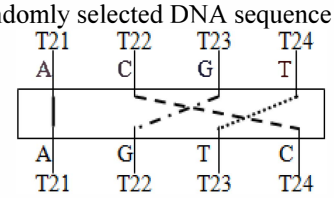
Round 1:

Temporary variables T11, T12, T13, T14
Compute T14= P4 ⊕ K1, T13= P3 ⊕ T14, T12= P2 ⊕ T13, T11= P1 ⊕ T12
Let, randomly selected DNA sequence be 'TGCA'



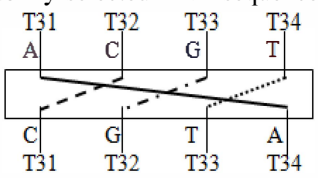
Round 2:

Temporary variables T21, T22, T23, T24
Compute T21= T11 ⊕ K2, T22= T12 ⊕ T21, T23= T13 ⊕ T22, T24= T14 ⊕ T23
Let, randomly selected DNA sequence be 'AGTC'



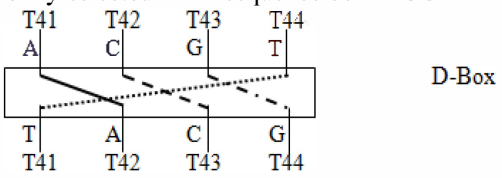
Round 3:

Temporary variables T31, T32, T33, T34
Compute T34= T24 ⊕ K3, T33= 23 ⊕ T34, T32= T22 ⊕ T33, T31= T21 ⊕ T32
Let, randomly selected DNA sequence be 'CGTA'



Round 4:

Temporary variables T41, T42, T43, T44
Compute T41= T31 ⊕ K4, T42= T32 ⊕ T41, T43= T33 ⊕ T42, T44= T34 ⊕ T43
Let, randomly selected DNA sequence be 'TACG'



Step 5: Combine all 64-bit cipher blocks to form 256-bit cipher text block
Step 6: Club together all the 256-bit cipher text blocks

After that a fixed number of bits are to be added both the end of the coded message and two specific positions within the coded message. After embedding extra coding the final form of the cipher text is mapped to a modified DNA sequence. In order to form modified DNA coding, 16 characters are randomly selected for making the modified DNA coding and transform all the 16 characters into 4X4 matrix form as follows:

	00	01	10	11
00	E	L	G	F
01	R	N	P	A
10	T	Q	C	M
11	D	S	B	H

Taking final form of ciphertext as, FnlCT='1000 1101 0011 0110'

To make the final coded form, Take 4-bit at a time, first 2-bit for selecting column while last two for row. Thus '1000' is mapped to 'G'

Therefore, after mapping all the bits, final form becomes, FnlCT='GADQ'

V. Result Analysis

This algorithm is applicable for almost all documents, image, audio, video file type such as .doc, .mp3, .avi, .mp4, .txt, .flv and so on. The results of generation of cipher text after encrypting the plaintext on few data sets are given below:

File	File size (KB)	Cipher size(KB)	Encrypt time (ms)	Decrypt time (ms)
.doc	1126.4	2253.4	9531	10672
.jpeg	2304	4609.5	18966	21345
.mp3	5990.4	11981.6	46936	53406
.avi	14950	299002	119281	133500

VI. CONCLUSION

The proposed technique of encoding is far better than the conventional cryptography like DES and DNA based encryption algorithms. The large key size, randomly selected sequence, D-Box, extra coding and modified DNA sequence coding makes the cipher text more secure against intruder's attack. Intruders will face more difficulty for decrypt the cipher text and get the information about the plaintext. The proposed method of DNA based encryption is applicable to text, audio, video rather almost all type of file. The encryption algorithm proposed here is based on a modified concept of DNA based cryptography methods. The algorithm proposed in this paper is more secure and reliable. It will be able to provides stronger protection against the various intruder's attacks like cipher text only, chosen cipher text etc. In this proposed algorithm both confusion and diffusion are embedded while making the cipher text.

REFERENCES

[1] Bibhash Roy, Gautam Rakshit, Pratim Singha, Atanu Majumder, Debabrata Datta, "A DNA based Symmetric key Cryptography", ICSSA-2011, G H Patel College of Engineering and Technology, Gujarat, India, pp. 68-72 Jan.2011.

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

[3] H.Z. Hsu and R.C.T.Lee, "DNA Based Encryption Methods", The 23rd Workshop on Combinatorial Mathematics and Computation Theory, National Chi Nan University Puli, Nantou Hsies, Taiwan 545.

[4] Bibhash Roy, Gautam Rakshit, Pratim Singha, Atanu Majumder, Debabrata Datta, "An improved Symmetric Key Cryptography with DNA Based Strong Cipher", ICDeCom-2011, BIT Mesra, Ranchi, Jarkhan, India, Feb 2011.

[5] G. Xiao, M. Lu, L. Qin and X. Lai, "New field of cryptography: DNA cryptography", Chinese Science Bulletin, vol.51, no.12, pp.1413-140, 2006.

[6] G. Cui, L. Qin, Y. Wang and X. Zhang, "Information security technology based on DNA computing", Proc. of the 2007 IEEE International Workshop on Anti-counterfeiting, Security, Identification, Xiamen, China, pp.288-291, 2007.

[7] "A Pseudo DNA Cryptography Method" Section 3- Motivation and method Kang Ning, Email: albertnk@gmail.com.

[8] D. Liu and P. Ning. "Establishing pairwise keys in distributed sensor networks", Proc. of the 9th ACM conference on computer and communications security (CCS'03), Oct. 2003.

[9] K. Tanaka, A. Okamoto and I. Saito, "Public-key system using DNA as a one-way function for key distribution", Bios stems, vol.81, no.1, pp.25-29, 2005.