International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

An Improved Concept of Cryptography Based on DNA Sequencing

Bibhash Roy, Atanu Majumder

***Abstract* — Cryptography is the process of hiding the actual information from a message or proving security in message while transmitting through network by encoding the original message into a new form of message. Data security is concerned with the areas of data transmission where data or information is encoded into a human non-understandable format, so that the intruder would not be able to get the actual meaning of the message. One of the promising direction of achieving data security is the DNA based Cryptography. This paper is concerned with the way how DNA sequencing can be used in cryptographic works and how to make the messages more secure and effective for transmitting over networks. In this paper we have proposed a new method of encryption which has two levels of encryption. Moreover it has a new scheme of key generation and key sharing. The encryption scheme is designed by using the technologies of DNA sequencing.**

***Keywords* — Encryption Key, Decryption Key.**

I. INTRODUCTION

Cryptography is the technique of secret writing which is used for data and information security and to protect that information from various attacks. The process of converting a message or plaintext into a human unreadable format called cipher text by encoding the original message using an encryption algorithm. In recent years, most of the research work has been going on DNA based encryption schemes [3]. Most of them use biological properties of DNA sequences. The encryption algorithm proposed in this paper does not make use of the biological properties. Instead, the properties of DNA sequences have been used.

A DNA sequence consists of four alphabets namely: A, C, G and T. Each alphabet is related to a nucleotide. A large number of DNA sequences publicly available in various web-sites. There are almost 55 millions of DNA sequences available publicly (approx).

By using the DNA sequencing property, DNA based encryption methods are designed. All of these methods would secretly select a reference sequence **S** from publicly available DNA sequences. Only the sender and the receiver have the information about this selected reference sequence. The sender would transform this selected DNA sequence **S** into a new sequence, say $\mathbf{S_q}$ by incorporating the selected DNA sequence **S** with the secret message **M**. This transformed sequence $\mathbf{S_q}$ is sent by a sender to the receiver along with many other extra sequences.

II. DNA BASED CRYPTOGRAPHY

DNA computing is more generally known as molecular computing. Computing with DNA offers a completely new paradigm for computation. The main idea of computing with DNA is to encode data in a DNA strand form in order to simulate arithmetical and logical operations. In DNA computing, designing and synthesizing information in the DNA sequence form is an important issue where wrong design might leads to wrong result.

Now a large number of researcher groups are trying to implement DNA concept in the solutions of various applications like cryptography, scheduling, clustering, forecasting. Moreover they are trying to apply this in signal and image processing application [2, 8]. Some other researchers in this field are also working on implementing of DNA algorithm for information security technology. The DNA based cryptography methods may have the following advantages over the conventional cryptographic methods:

1. Sender does not require having much key information to encrypt the information. Initially a part of the key is sufficient to encode the information.
2. It requires a little information (only the private keys or part of the key) to be communicated through the secure channel.

III. PROPOSED SYSTEM

A new method of encryption process is proposed here for providing better security and effectiveness in dada transmission. A private key is shared through a secure channel between sender and receiver before the communication establishment. This key can also be shared in between two parties using various public key encryption processes like RSA, Diffie-Helman key exchange etc. The level 1 key for encryption is calculated based on the response of a random number generator and the information about the key is send to the receiver end through a secure channel [11]. Sender will generate a random number using a random number generator function, and then this number along with the private key will go through a procedure that will produce level 1 key (K_1). The same procedure will generate session key for sending at the receiver end as for the key that is being used.

This proposed encryption algorithm is works on the byte values of the message or plaintext file [1], [8]. The byte values are extracted from the plaintext. These byte values are signed numbers (-128 to +127). In order to apply effective mathematical operations while encrypting the message, we add 128 to each byte value (0 to 255) [2], [8]. Then each modified byte value will go through the level 1 encryption by level 1 key (key1). After the completion of encoding of all the modified byte values, all the encrypted byte values are concatenated and that will produce the intermediate cipher text.

In level 2 of encoding, a DNA sequence will be selected randomly by sender from publicly available DNA sequences. This DNA sequence will be used for level 2 key (key2). Receiver will have the information about the selected DNA sequence. Then the selected DNA sequence will be transformed into binary string using binary substitution method. This binary string is then divided into n-bit segments. The value of n is determined from the level 2 key. Integrate each of the bits of the intermediate cipher with every segment of DNA sequences. The DNA sequence will be repeated circularly, until all the bits of the intermediate cipher text are integrated with the DNA sequence. After doing this operation for all the bits of intermediate cipher with the DNA sequence, a new form of DNA sequence will be generated $\mathbf{S_q}$. Then this S_q is again goes through an encoding procedure using level 1 key. In the final form of cipher we add the extra information namely starting and ending primers that are not linked up with the original message [6], [7], like as biological DNA strand containing introns as extra information which are omitted by splicing process of biological DNA synthesizing. This extra information and DNA sequence can be used for providing integrity and authentication of the message [9], [10], [12].

IV. Sharing of Key

A. Level 1 Key Sharing

Begin

Sender and receiver both will agree on a common master key (MK) prior to communication [4], [8]. This key is of 32-bit size.

Sending end Computation:

Step 1: Sender will generate one random number (R). (32-bit)

Step-2: The random number R will get Ex-OR with Master key (MK). The result will be used as the Encryption key (K_1).

e.g.: let the Master Key, MK=137

and the random number R=157

Encryption key ($\mathbf{K_1}$) = (137 Ex-OR 157) = 20

Step-3: sender will compute the level1 session key as follows:

Random number 'R' will be divided by 16 and remainder will be kept in 'KS_n'

The quotient will once again divided by 16 keeping remainder in 'KS_{n-1}'

… Continue until quotient becomes less than 16 (KS_{n-2}….., KS_2, KS_1). Then club together all KSs in hexadecimal form.

e.g. 157 % 16 =13

(Remainder) hexadecimal form= D

157 / 16 = 9

hexadecimal form= 9

Concatenating these two values,

Session key $\mathbf{K_S}$ = 9D

Step-4: Sender will send $\mathbf{K_S}$ as level1 session key through private Channel with level2's session key. These two keys (level1 & level2) are sending in a digest form through private channel.

End

Receiving End Computation:

Begin

Step-1: Receiver will receive session key $\mathbf{K_S}$ and separate all the KSs and convert these to equivalent decimal form.

Step-2: Receiver will compute the decryption key as follows:

$T_i = (KSi*16) + KS_{i+1}$

$T_{i+1} = (Ti*16) + KS_{i+2}$ continue up to KSn

(i=1, 2, 3…..n)

T_{n-1} will be the session key

T_1= (9*16) + 13 =157

Decryption key: $K_d=K_1$= (MK X-OR T1) =20

End

B. Level 2 Key Sharing

Randomly generate a number that will be used as private key2 for level 2 encryption. Select randomly a DNA sequence from publicly available DNA sequences. The private key2 and the DNA sequence is to be known to the receiver end prior to message transfer.

V. Algorithm

Data encoding (cryptography) is used for making data secure for transmitting through public networks. The conventional encoding techniques are not secure enough today. In order to make data more secure against certain attacks DNA based encryption strategy is adopted. The message will go through the two rounds of encryption. Two levels of encoding are used for making the message more secure. Level 1 encryption is on the byte value of the plain text and the level 2 is concerned with the integrating intermediate cipher text into a DNA sequence. The idea that has implemented is given below in brief.

A. Encoding Technique

Four kinds of bases are there in DNA, which are adenine (A) and thymine (T) or cytosine (C) and guanine (G) in DNA sequence. The simplest coding patterns to encode the 4 nucleotide bases (A, T, C, G) is by means of 2-bit binary: 0(00), 1(01), 2(10), 3(11). As we all know that in a double helix DNA string, two DNA strands are held together complementary in terms of sequence, A is complement of T and C is complement of G according to Watson-Crick complementarities rule. Take DNA digital coding into account, it should reflect the biological characteristics of 4 nucleotide bases, the complementary rule that (~0=1), and (~1=0) is proposed in this DNA digital coding. According to this complementary rule, 3(11) is the complement of 0(00) and 2(10) of 1(01). Since A and T are complement to each other they are coded as '00' and '11' respectively and for C and G, '01' and '10' respectively.

For level1: A-00, T-11, C-01 and G-10

B. Encryption Steps:

Level 1:

Step-1: Input the file that is to be transfer at the receiving end.

e.g. nit1.doc. (Plain Text)

Step-2: Read the byte value from the input file (range of the byte values from -128 to +127).

Step-3: In order to transfer the byte values into positive domain for effective computation add 128 to each of the byte values. (range of the byte values 0 to 255)
Step-4: Key operation on the modified byte values to get the intermediate cipher.
IC=f(PT,K_1) (Block wise Ex-OR operation)
IC- Intermediate cipher
K_1- Private Key (level 1encryption key)
PT- Plain text
Step-5: Take one byte value and convert it into 8-bit binary equivalent then substitute it by A,T,C,G DNA bases.
E.g. if after key operation byte value is 76,
76=01001100 (00-A, 11-T, 01-C, 10-G)
Level 1 Cipher text(CT_{L1}) = CATA
Level 2:
Input: Level 1 Cipher text, CT_{L1} = 'CATA'
Let S be a randomly selected publicly available DNA sequence. S=AGCGTACCAGTGC (Let)
Step-1: Code 'S' into a binary sequence by using the binary coding scheme. Thus the sequence 'S' is 001001101100010100101111001
Step-2: Divide S into multiple segments. The size of the segments will be determined by the private key2.
e.g. private key2 = 4
Then the segments as follows: 0010, 0110, 1100, 0101, 0010, 1110, 01
Step-3: Insert bits from CT_{L1}, once at a time, into the beginning of segments of S.
CT_{L1}='01001100'
The result is as follows:
00010, 10110, 01100, 00101, 10010, 11110, 00010, 00110, 01
(Ignoring non-coding segments)
Final output:
00010, 10110, 01100, 00101, 10010, 11110, 00010 and 00110
Concatenating each of the above segments a new sequence of DNA is being produced: 000101011001100001011001011110000100011 0
Step-4: Now the above sequence of bits will go through an encryption algorithm with the level 1 encryption key (K_1). The same function can be used here that has been used in level 1 encryption.
Step-5: We use the binary code scheme to produce a newly coded DNA sequence S_q.
Step-6: Send the above sequence S_q to the receiver end as cipher text, along with many other irrelevant sequences.

C. Decryption Technique:

At the receiving side the encrypted message is decoded in order get back the original message. Decryption process has two levels decoding using the keys that has been sent by the sender. Decryption steps are as follows:
Level 1:
Step-1: Final cipher text will be converted to binary sequence using binary coding scheme.
Substitute, A-00, T-11, C-01, G-10
Step-2: Apply the level 2 key (key1) for getting the coded DNA sequence.
Step 3: Extract the intermediate cipher text and DNA sequence using the shared private key2.
Step-4: Match the extracted DNA sequence with the DNA sequence that has been shared between sender and receiver.
Step-5: If both the sequences are same then only message will go for level2 decryption (Authentication, Integrity).
Level 2:
Step-1: The output of the level 1 decryption is converted into binary string by substituting A-00, C-01, G-10, and T-11
Step-2: Take 8-bit code from the binary string and convert those into decimal which becomes byte value for all the binary string.
Step-3: Decryption key operation for all the byte values
PT=f'(IC, K_1)
Step-4: Subtract 128 from each of the byte values to get the actual byte values
Step-5: Write the byte values to a file and save it according to file format.

D. Format of Cipher Text

From plain text (PT) the intermediate cipher text (IC) is obtained by using the encryption algorithm and the 1st level key (K_1).

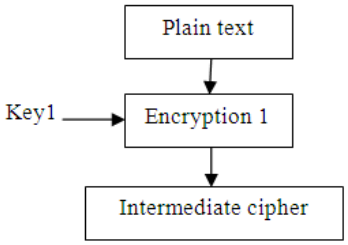

Fig: Level 1 cipher text format

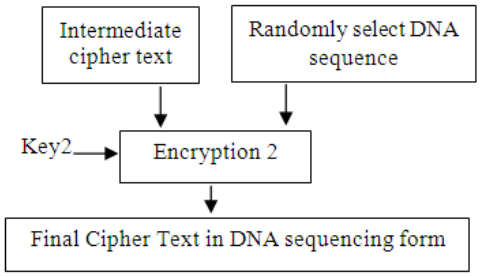

Fig: Level 2 cipher text format

The final cipher text has starting primer, ending primer, authentication code, integrity code and file type as like as introns in DNA strand. After decrypting the final cipher using private key2 we will get the actual final text. Then applying private key1 it will be decrypted to plain text.

VI. CONCLUSION

The technique that is proposed in this paper is based on a new scheme of symmetric key cryptography where actual keys are not directly shared between the sending and receiving end. Sender and receiver only share such things that bear the information about the keys. In order to facilitate the understanding of principles and some

techniques of the newly born field of DNA cryptography this concept is proposed. We believe that the application of DNA computation in cryptology will have great potential on information field in the future. The DNA computing has become a large area of interests in the research domain of cryptography; the proposal can surely be enhanced with much more advanced concepts such as realization in several security technologies of encryption, steganography, signature and authentication. This method is efficient, and it is powerful against certain attacks; the useless extra information that contained in the cipher text makes the algorithm much more difficult for cryptanalysis. Moreover this improved concept can be used in the security concerned of wireless networks.

REFERENCES

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

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

[3] G. Xiao, M. Lu, L. Qin and X. Lai, New field of cryptography: DNA cryptography, *Chinese Science Bulletin, vol.51, no.12, pp.1413-1420, 2006.*

[4] L.Eschenauer and V. D. Gligor. A key-management scheme for distributed sensor networks. *Proceedings of the 9th ACM conference on Computer and communications security, Washington, DC, USA*, pp. 41–47, November 18-22 2002.

[5] W. Du, J. Deng, Y. S. Han and P. K. Varshney. A pairwise key predistribution scheme for wireless sensor networks. *Proceedings of the Tenth ACM Conference on Computer and Communications Security (CCS 2003)*, pp. 42–51, October 2003.

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

[7] W. Du, J. Deng, Y. S. Han and P. K. Varshney. A pairwise key predistribution scheme for wireless sensor networks. *Proceedings of the Tenth ACM Conference on Computer and Communications Security (CCS 2003)*, pp. 42–51, October 2003.

[8] 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.

[9] Guangzhao Cui Limin Qin Yanfeng Wang Xuncai Zhang; Information Security Technology Based on DNA Computing; Anticounterfeiting, Security, Identification, 2007 IEEE International Workshop on 16-18 April 2007,page(s): 288-291, ISBN: 1-4244-1035-5, Location: Xiamen, Fujian.

[10] S. V. Kartalopoulos, DNA-inspired cryptographic method in optical communications, authentication and data mimicking, Proc. of the IEEE on Military Communications Conference, vol.2, pp.774-779, 2005.

[11] 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.

[12] 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.

[13] Garfinkel Simson, Web Security, Privacy & Commerce, 2nd Edition, O'Reilly Publisher, November 2001.

[14] Dan Boneh, Cristopher Dunworth, and Richard Lipton. *Breaking DES Using a Molecular Computer*. Technical Report CS-TR-489-95, Department of Computer Science, Princeton University, USA, 1995.

AUTHOR PROFILE

Bibhash Roy
Email: bibhashroy10@yahoo.co.in
Bibhash Roy has received his M.Tech in Computer Science & Engineering from Tripura University in the year 2009 and was a gold medalist. He has supervised more than 40 projects of B.E./MCA/ BCA students. He has more than 9 years of teaching experience and currently he is an Assistant Professor in the department of CSE in Tripura Institute of Technology, Narsingarh, Tripura, India. He is pursuing PhD from Calcutta University in QoS security issues in ad hoc networking.

Atanu Majumder
Email: atanu.cse21@gmail.com
Atanu Majumder is pursuing M.Tech in Computer Science and Engineering at National Institute of Technology, Agartala. He has completed his B.E. in Computer Science and Engineering from Tripura Institute of Technology, Narsingarh, Tripura, India on 2011 and was university Gold Medalist. He is working on cryptography and various attacks and security aspects of data and networks.