International Journal of Engineering Research and Technology. ISSN 0974-3154 Volume 10, Number 1 (2017)

DNA Cryptography using Index-Based Symmetric DNA Encryption Algorithm

Nikita S. Kolte
ME (E & TC) student,
D. Y. Patil college of Engg. & Tech.
Kolhapur, Maharashtra, India.

Prof. Dr. K.V. Kulhalli
H.O.D. Dept. of Information Tech.,
D. Y. Patil college of Engg. & Tech.
Kolhapur, Maharashtra, India.

Samrat C.Shinde
Faculty, Electronics Engg. Dept.,
D. Y. Patil college of Engg. & Tech
Kolhapur, Maharashtra, India.

E-mails: nskolte@gmail.com, kvkulhalli@gmail.com, findsamrat@gmail.com

Abstract
DNA has a great cryptographic strength, its biding properties between nucleotides bases (A-T, C-G) offer the possibility to create self assembly structures which are an efficient means of parallel molecular computations. In this paper, a new index-based symmetric DNA encryption algorithm has been proposed. Adopting the methods of Block cipher and Index of strings, the algorithm encrypts the DNA-sequence based plaintext. First, the algorithm encodes each character into ASCII and their equivalent binary format. And then this binary format transformed into the DNA sequence and compared to special DNA sequence which is nothing but the symmetric key, finds the sequence which has no difference with it. Then, the algorithm will store the position as the cipher text. The algorithm is based on the idea to use DNA chromosomes as one-time-pad structures and index them in order to encrypt the plaintext.
Keywords: DNA Cryptography, Nucleotides, Index, Block cipher, Symmetric Key, One-Time-Pad.

I. Introduction
With the rapid development of modern communications technology and the Internet, the importance of information security becomes highlighted. DNA cryptology has progressed been put forward as a newly technique, which, together with traditional quantum cryptology, formed the three main branches of cryptology. DNA Encryption means combing DNA technique with cryptology, producing new cryptography to provide safe and efficient cipher services. The difference between DNA cryptography and traditional one is that the former is based on the limitation of biotechnology, which is unrelated to numeracy. Thus, it is immune to the attack from super computer. Certainly the security also needs serious mathematical justification. But it is undeniable that with the further development of biotechnology and cryptology, DNA's vast parallelism, extraordinary information density and exceptional energy efficiency could make large-scale data storage and encryption or decryption faster. When the information is encrypted which is less demanding for parallel real-time data or extremely demanding for mass data storage applications, it has a unique advantage. Index-based DNA symmetric encryption methods, definitely belongs to this field, is presented in this paper. The algorithm has a huge key space, high sensitivity to plain text, and extremely great effect on encryption.

II. Symmetric Cryptography And DNA Cryptography
2.1 Symmetric Cryptogaphy
Cryptography is constructed by five elements {M,C,K,E,D },among which Message space M is also called Plaintext space, Cipher text space C, Key space K, Encryption algorithm E and Decryption algorithm D. For each plaintext m in Message space M, Encryption algorithm E can encrypt the m to the Cipher text c with the secret key of Ke; and the Decryption algorithm D could decrypt the Cipher text c to the plaintext m with the key of Kd. The process of encryption could beexpressed as:
$EKe\ (P) = C$
The process of decryption could be expressed as:
$DKd\ (C) = P$
According to Key's feature, Cryptography can be divided into Symmetric Cryptosystem and Asymmetric Cryptosystem. Based on the different methods of encryption, Symmetric Cryptosystem falls into Stream Cipher and Block Cipher. In the case of Stream Cipher, in order to get cipher text, the user needs to compute the plaintext by-bit with the word created by

Pseudorandom bit generator. However, in the field of Block Cipher, user should divide plaintext into different groups, and then encrypt each group.

2.2 DNA Cryptography

DNA, as an information carrier, has an extraordinary storage density, which can better solve the problem of how to create or store the large quantity of Pad. DNA Cryptography is implemented by using modern biological techniques as tools and DNA as information carrier to fully exert the inherent advantages of high storage density and high parallelism to achieve encryption.

Based on the DNA double helix structure and the principle of Watson-Crick complementary, the encryption and decryption can be simulated as the DNA's biological or chemical process. It encodes the information and then stores plaintext, cipher-text or other information on DNA encoding nucleotide sequence after significant operation. By now, the password system has been built.

In recent years, researchers have proposed a large number of DNA-based encryption algorithms, as a result of that DNA cryptography is still in the initial stage, it don't have a complete model and an efficient verification mechanism.

III. The encryption protocol and Proposed Index based symmetric DNA Encryption Algorithm

3.1 The Encryption Protocol

In our work we used a cryptosystem with symmetric key named One-Time-Pad (OTP). It is an algorithm where each key is used just once where from the name of One- Time-Pad. OTP encryption uses a large non-repeating set of truly random key letters. Each pad is used exactly once, for only one message. The sender encrypts the message and then destroys the used pad. As it is a symmetric key cryptosystem, the receiver has an identical pad and uses it for decryption. The receiver destroys the same pad after decrypting the message. New message means new key letters. A cipher text message is equally likely to correspond to any possible plaintext message. Cryptosystems which use a secret random OTP are known to be perfectly secure. Introducing DNA into the common symmetric key Cryptography, it is possible to follow the pattern of Symmetric key cryptosystem, while also exploiting the Inherent massively-parallel computing properties and storage capacity of DNA in order to perform the encryption and decryption using OTP keys. The resulting encryption algorithm which uses DNA medium is much more complex than the one used by conventional encryption methods.

We developed an encryption algorithm which uses OTP as symmetric key and real chromosomal sequence as OTP. We extracted chromosomal sequence from publically available data bases(NCBI) and used it for implementation of this algorithm.

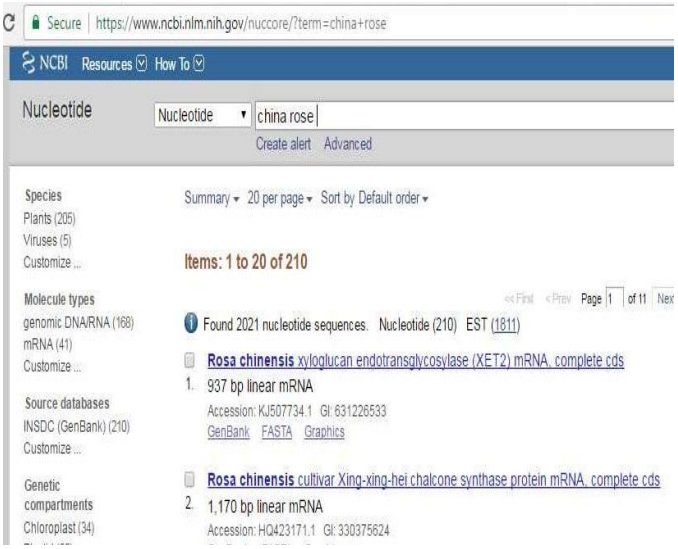

Figure 1. DNA sequence as OTP extracted from NCBI publically available data base.

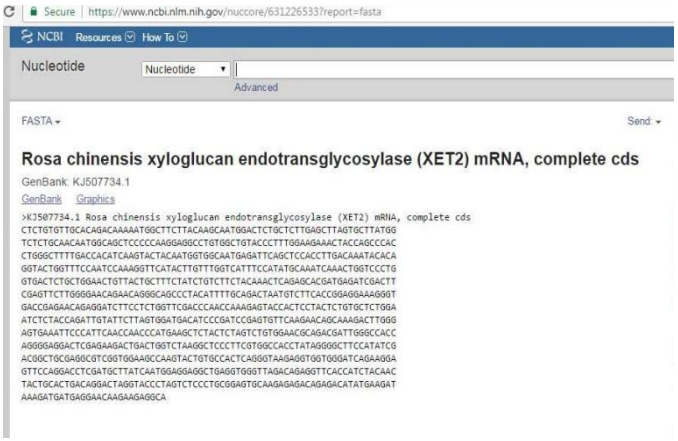

Figure 2. Fragment form DNA sequence file in FASTA format

3.2 Proposed Index based symmetric DNA Encryption Algorithm

3.2.1 *Encryption*

Plaintext message was transformed in bits and after that in DNA format. We used a text message for encryption so an encryption unit was a character and in ASCII cod it was represented on 7 bits. Transformation from a 2-letter (0, 1) alphabet to a 4-letter (A, C, G and T) alphabet was done using 2 bits to represent a letter:

A – 00
C – 01
G – 10
T – 11

Using this substitution a character was represented on 4 letters which is equivalent to a byte. Using MATLAB functions we obtained decimal ASCII codes of the plaintext message, and transformed them in binary form, each character on 8 bits. After that, using functions we transformed our message from binary to DNA alphabet. Each character was transformed in a 4- letter DNA sequence and searched in the chromosomal sequence, used as OTP. OTP sequence was scanned from

bases 1 till the end of the sequence FASTA format. At each step was analyzed a segment of 4 bases From the OTP sequence and compared to the characters DNA sequence. If 4-letter sequence representing a character from the message was retrieved in the Chromosomal sequence then the starting point (index in chromosome) of identical 4-letters was memorized in an array. At the next step was analyzed another 4 bases from OTP where first 3 of them are the last bases from previous step (Figure 3).

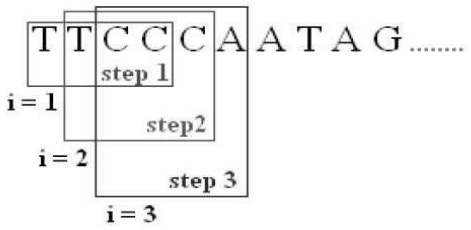

Figure 3. Exemplification of the OTP scanning process for message encryption.

For each character was obtained an array of indexes in chromosomal sequence. Number of indexes for a character depends on how often the character's DNA sequence retrieved in the chromosomal sequence. For each character was chosen a random index from its array of indexes. We obtained the final encrypted message: an array of random indexes, one for each character.

Example of implementation results:

Plain text" secret"

ASCII code 115 101 99 114 101 116

s- 115- 01110011 – CTAT – indexes:
166 258 789 927 1295 2954
3045 3098 3181 3207 3361 3763
……..till the end FASTA format of selected DNA sequence. For each character was chosen a random index from its array of indexes using MATLAB function. Below are established positions of random indexes inside character's arrays:
115- 70th index 23811(for alphabet s)
101- 26th index 13981(e)
99- 7th index 8011(c)
114- 57th index 21195(r)
101- 57th index 32741(e)
116 – 158th index 25264(t)

Final encrypted message is:
23811 13981 8011 21195 32741
25264

3.2.1 *Decryption*

At message decryption is used the same OTP as at encryption, because it is a symmetric key algorithm. The received encrypted data is in the form of index positions. These index positions are compared with the database which is considered as shared key to obtain DNA form of data. The DNA form of data is converted back to binary form, which is further converted to ASCII format and then to Plain text.

IV. Analysis

4.1 Theoretical analysis

Because DNA Encryption System does not have a uniform model, scientists can't find an existing comprehensive standard to analysis the security of the algorithm. In this paper, we analyses the security of the algorithm from the aspects of bio-security and math security.

4.1.1 Bio- Security

As a result of that the encryption algorithm is designed to search the sequence on the special DNA which is same as the cipher-text, the attacker can't decrypt it without the information of the special DNA sequence. Without the key including the information ofthe special DNA sequence, it is impossible to find the special DNA sequence. If the attacker tries to decrypt the cipher-text by searching the DNA sequence exhaustively, there is no difference from the attacker to decrypt the cipher text without key. Likewise, DNA has an extremely large data storage capacity, because one single chain of a chromosome has tens of millions of nucleotides. In order to find the correct starting position and end position, the attacker will spend numerous resources. In summary, the algorithm is safe enough from the aspect of bio-security.

4.1.2 Math- Security

In the simulation, we selects the gene sequence of any living being chromosome , whose length is more than million bp(base pair). It is no doubts that the key space could satisfy the all 8-bit binary coding. From the statistics, the number of each kinds nucleotides range from tens thousand to millions. And of course it decreases the possibility that encryption system will be attacked. If the researcher wants to achieve higher security, s/he could encrypt a longer DNA sequence. Besides, because of the relative action being operated on the DNA, the plaintext-oriented attack loss its meaning. In other words, the algorithm could assure the validity in preventing plaintext-oriented attack.

4.2 Simulation Analysis

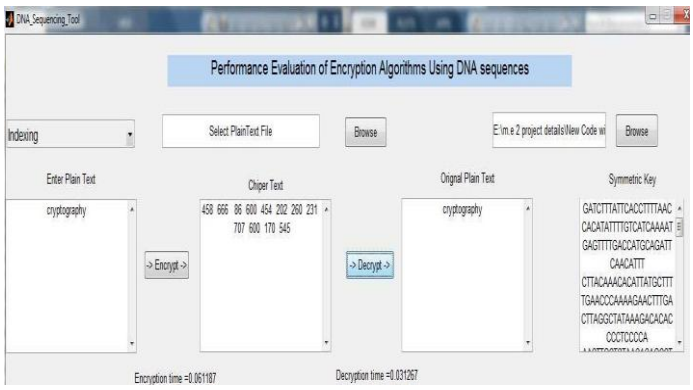

Figure 4. Encryption and decryption simulation output for a single word

Figure 5. Simulation output for proposed algorithm

V . Complexity of proposed Encryption Algorithm

With an OTP, an adversary has no information about how to cryptanalyze the cipher text, since every key is equally likely to correspond. Cryptosystems which use a secret random OTP are known to be perfectly secure and are applicable primarily for transmission of ultra-secure information. The problem of this cryptosystem is that the key letters have to be truly random and the key sequence can not be reused ever again. Any genetic sequences, of any living matter can be used as a secret key for encryption. We can exploit great storing capabilities and variety of DNA sequences for the usual OTP cryptosystem.

This encryption algorithm treats also the problem of the vulnerability to frequency attack. For the same character from the plaintext message we obtained a number of different indexes which are used as values in the cipher text by a random choice. Another advantage of this algorithm is that at encryption of another message, indexes for each character will be different from the previous encryptions values. The same character will appear in ciphertext under different values at encryption of two different messages. Chromosome or any DNA sequence which was chosen to be the encryption key dictates what kind and how many indexes for ciphertext will have a character.

VI . Conclusion

We use one-time-pad principle and DNA chromosomes storing capabilities for message encryption. Implementation was performed in MATLAB and genetic database maintained by NCBI. One single chromosome from any species is composed from thousand of bases and it is perfect to be used in this algorithm to address the characters from the plaintext.

Each character from the plaintext message was transformed into a unique sequence of 4 bases and searched in the chromosome, used as OTP. A random index of a character in chromosome becomes part of the cipher text. The strength of this algorithm is based on the secrecy of the OTP and protection from frequency attack.

The aim of this paper is to find useful and practical DNA cryptographic algorithm and to study its applicability in DNA technology. Laboratory implementations are possible but are still expensive and time consuming. Despite of this, simple and effective algorithms are necessary in order to bring DNA computing on digital level and use it on a large scale.

REFERENCES

[1]. L. M. Adleman, "**Molecular computation of solutions to combinational problems**," *Science*, vol. 266, pp. 1021–1024, 1994.

[2]. A. Gehani, T. H. LaBean and J. H. Reif, "**DNA-based cryptography**," *DNA Based Computers V. Providence: American Mathematical society***,** vol. 54, pp. 233–249, 2000.

[3]. Souhila Sadeg ,"**An Encryption algorithm inspired from DNA**", *IEEE*, pp 344 - 349 November 2010.

[4]. Xing wang, Qiang Zhang ,"**DNA computing based cryptography**", *IEEE*, pp 37-42, 2009.

[5]. Qinghai Gao, "**A Few DNA-based Security Techniques**", *IEEE*, 2011

[6]. Deepak Singh Chouhan, R.P. Mahajan, "**An Architectural Framework for Encryption & Generation of Digital Signature Using DNA Cryptography**", *IEEE*, pp 743-748, 2014

[7]. Ashish Kumar Kaundal, A.K Verma**, "DNA Based Cryptography: A Review",** *IJICT,* Volume 4, pp. 693-698, 2014

[8]. B A Mitrans, A Kh Aboo, **"Proposed Steganography Approch Using DNA Properties",** vol.14, 2013

[9]. William Stallings, *"***Cryptography and Network Security***"*, Third Edition, Prentice Hall International -2003.