DNA Cryptography Based on DNA Hybridization and One Time pad scheme

Shreyas Chavan
MMM's College of Engineering, Pune

Abstract

Cryptography is one the major elements in data security and communications security .DNA Cryptography is a new field in cryptography which harnesses rise of DNA computing to provide high level of data encryption. However, DNA cryptography is not used in mainstream computing due to the computational complexity and requirement of a bio molecular laboratory. This paper proposes a new scheme for DNA cryptography using a combination of Hybridization of DNA oligonucleotides and the generic binary one time pad technique.

1. Introduction

DNA computing is a new emerging field that uses simulation of DNA biomolecular structure and using the biological technology for computation. Due to the rapid growth of internet application and overall requirement of data security, the need for new cryptographic technology and better algorithms has risen exponentially. Cryptography is the most important element of information security and communications security. With the mainstream cryptographic algorithms such as DES and RSA 768 being cracked, there is a need for a new and better cryptographic technology.DNA cryptography first appeared after research of Adleman[1] in DNA computing and Risca[2] in DNA Information project.Adleman used DNA computing to solve directed Hamiltonian path problem[1].DNA computing was further extended by Lipton to solve NP complete problem[3].In the research of Boneh,Dunworth and Lipton[4] it was seen that by using DNA computing, the Data Encryption Standard (DES) cryptographic protocol can also be broken.DNA molecular structure is used for computations because of its vast parallelism, low energy consumption and very high storage capacity[5].DNA chain has a very large scale of parallelism, its computing speed can reach 1 billion times per second.DNA molecular computer has a low consumes. It is only equal to one-billionth of a traditional computer. One gram of DNA can store up to 10^8 terabytes of information. But, DNA computing has some drawbacks. Some of them being requirement of huge computing time, very high computational complexity and need of a high tech biomolecular laboratory. There are other one time pad DNA cryptography schemes.

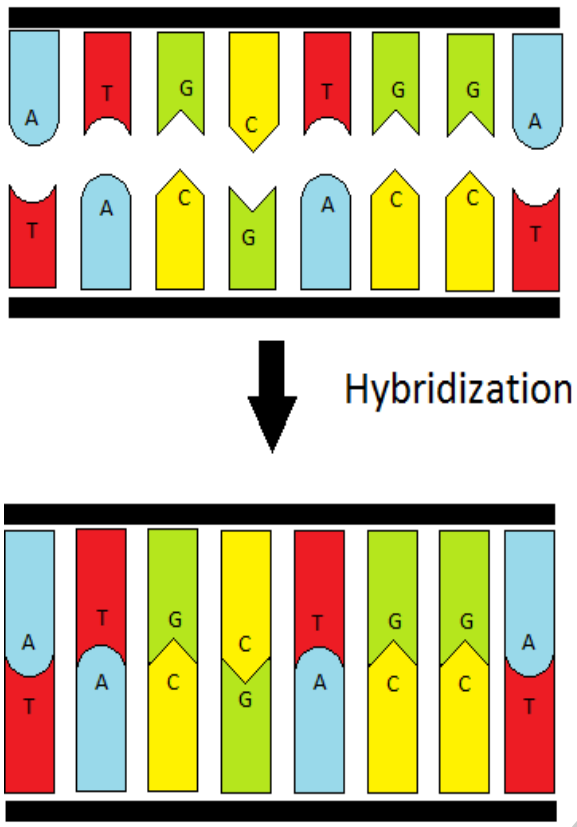
2. DNA Structure

Deoxyribonucleic acid(DNA) is a molecule that stores all genetic information of a living organism. This information is encoded as a sequence of four nucleotides Adenine (A), Cytosine (C), Guanine (G) and Thymine (T).Watson-Crick[6] proposed a pairing rule for these nucleotides that states that Adenine(A) can only pair and form a hydrogen bong with Thymine(T) and Cytosine(C) can only pair with and form a hydrogen bond with Guanine(G).DNA has a double helical stucture.DNA is a very long sequence and is responsible for transfer of all genetic information. If the DNA is separated in such a way that all hydrogen bonds between the nucleotides are dissolved, then we are left with two long single sided structures known as single sided DNA (ssDNA) or Oligonucleotides. There Oligonucleotides run in opposite directions of each other. Thus they are called anti-parallel. They are commonly referred to as 3'(three prime) and 5'(five prime).

3. DNA Hybridization

DNA Hybridization is a process by which two ssDNA segments are combined to form a DNA sequence. In hybridization process it is necessary that a hydrogen bond is formed only between two Watson-Crick pairs. If a Watson-Crick pair is not found at the corresponding position then the hybridization of that pair fails. In DNA hybridization, it necessary that the length of both ssDNA segments should be same. If not, fragmentation occurs and fragment assembly has to be done in order to repair the DNA molecule.

Figure 1. Hybridization



4. Cryptographic Algorithm

In this paper, encryption of plaintext is done by combination of DNA cryptography and Binary One Time Pad(OTP) scheme. This algorithm uses two keys for two stages of encryption. These keys are used for encryption on sender side and for decryption on the receiver side. Hence this algorithm is a Symmetric key cryptographic algorithm. First of these keys is a random string of nucleotides forming a ssDNA sequence. Length of this key depends on length of the plaintext. The second key is a Binary sequence that is used for OTP.The length of this binary key is twice the length of the ssDNA sequence key

4.1 Sender's algorithm

Steps of algorithm are as follows.

1.Convert the plaintext message to its corresponding ASCII values. Convert these ASII values into binary bits of plaintext forming binary plaintext.
Length of binary plaintext = Length of plaintext * 7

2.Decide the scale of oligonucleotides with the receiver (n).The scale of oligonucleotides refers to number of nucleotides per bit in binary sequence.

3.Create a random ssDNA sequence and a random binary sequence. These are the keys for the algorithm.
Length of ssDNA = Length of binary plaintext * n
Length of binary sequence = Length of ssDNA * 2

4.Establish a secure channel between the sender and receiver for transmission of these keys.

5.After transmission of keys, take Watson-Crick complementaries of all nucleotides corresponding to bit 1 in the plaintext binary sequence. This is the cipher ssDNA.

6.Convert this cipher DNA into intermediate binary sequence by mapping nucleotides according to values in Table 1.

Nucleotide	A	C	G	T
Code	00	01	10	11

Table 1.Binary codes

7. XOR this intermediate binary sequence with the binary OTP key generated in step 3 of the algorithm to generate the cipher binary sequence.

8. Send this cipher binary sequence to the receiver.

4.2 Receiver's algorithm

Steps of algorithm are as follows.

1. Decide the scale of oligonucleotides with the sender (n)

2. Receive the ssDNA key and binary key through the secure channel.

3. Receive the cipher binary sequence whose length is same as that of the binary key.

4. XOR the cipher key with binary key.

5. Convert the resultant binary sequence into ssDNA by grouping in pairs of two and mapping according to Table 1.

6. Hybridize this resultant ssDNA sequence with the ssDNA key. When Watson-Crick pairs are found in groups of the scale(n),note down the bit as 1.If hybridization fails due to occurrence of same nucleotide, note down the bit as 0.

7.Group this sequence into 7 bits in each group and convert each group to its corresponding decimal number.

8.These decimal number are ASCII values. Hence convert the ASCII into characters to reveal the original plaintext.

5. Experiment

An experiment is conducted to demonstrate the working of the algorithm and Security analysis. In this experiment, the plaintext to be encrypted is taken as "DNA".It is converted into its ASCII code which is 69 79 66.Then it is converted into binary code which is 100010110011111000010.For this experiment let us assume that the scale of oligonucleotides is decided as 4.Hence there will be 4 nucleotides per bit in this plaintext. Length of plaintext is 21.Hence the length of random ssDNA will be 21*4=84.This random ssDNA is.

ACGT CTAG CATT AGCC CTAT CATT ACAG TCGA ATCG AGGC TACG TTAG CAGC CAGT TCGA CTAC TAGT AGCT TGCA ATCT TGGC

Also a random OTP key of length 84*2=168 is generated.

01011010 11010101 01010101 00111010 11010111 01100101 01100101 01101010 10101010 10110101 10110101 01010101 01010101 11010000 00100110 10101011 10110111 11011011 01010110 10010010 11000111

These sequences are sent to the receiver through a secure channel. Now, Watson-Crick pairs of nucleotides resembling bit 1 are taken. Result is

TGCA CTAG CATT AGCC GATA CATT TGTC AGCT ATCG AGGC ATGC AATC GTCG GTCA AGCT CTAC TAGT AGCT TGCA TAGA TGGC

By mapping according to Table 1, we get the following.

11100100 01110010 01001111 00100101 10001100 01001111 11101101 00100111 00110110 00101001 00111001 00001101 10110110 10110100 00100111 01110001 11001011 00100111 11100100 11001000 11101001

XOR this sequence with binary key to generate cipher sequence. Cipher sequence is as follows.

10111110 10100111 00011010 00011111 01011011 00101010 10001000 01001101 10011100 10011100 10001100 01011000 11100011 01100100 00000001 11011010 01111100 11111100 10110010 01011010 00101110

This cipher sequence is now sent to the receiver. When the receiver receives the cipher binary sequence, its XORs it with the OTP key. The result is

11100100 01110010 01001111 00100101 10001100 01001111 11101101 00100111 00110110 00101001 00111001 00001101 10110110 10110100 00100111 01110001 11001011 00100111 11100100 11001000 11101001

Now by mapping according to Table 1, we get

TGCA CTAG CATT AGCC GATA CATT TGTC AGCT ATCG AGGC ATGC AATC GTCG GTCA AGCT CTAC TAGT AGCT TGCA TAGA TGGC

Now Hybridize this ssDNA sequence with ssDNA key. When a Watson-Crick pair occurs in group of 4, bit is noted as 1 and if Hybridization fails due to occurrence of same nucleotide, then bit 0 is noted. The result is 1000101 1001111 1000010.Converting to decimal in groups of 7, we get 69 79 66.Converting back from ASCII to characters, we get "DNA" which is the original plaintext.

6. Security analysis

The algorithm presented in this paper is scalable. The amount of encryption to be applied to the message is variable. Hence messages containing important and classified information can be transmitted with better security than regular messages. On analysing the security features of the algorithm it is seen that in case of a Brute Force attack to crack the algorithm, the chances of getting the correct combination are 1 in $4^{(m*7*n)}$.Where m is the length of message and n is the scale of oligonucleotides.Hence for a message of length 5 and 4-mer oligonucleotides, the chances of getting the combination right are 1 in $1.94x10^{84}$ combinations. This is a very small chance, but the odds of cracking the algorithm can be further decreased by switching values in Table 1.Also, constraints can be applied on the number of tries to further the security.

7. Conclusions

This paper illustrates an original DNA cryptography technique by using DNA Hybridization and further supplemented by OTP. The main advantages of this technique are its scalability and reusability. The scale

of encryption can be changed easily according to the level of encryption required for the data. Thus the complexity of the algorithm varies with the data and hence large amount of processing is not required for data which need not be protected immensely. If the random ssDNA which is generated is large enough, then the same ssDNA can be used again for transmission of different data. Thus redundant generation of random ssDNA can be avoided. The length of the ssDNA can be calculated before transmission according to length of multiple messages and a single ssDNA can be used to encrypt all messages. It can be concluded from the security analysis of the algorithm that this method provides a very high level of security. However, high computational complexity is still a problem which can be diminished as technology progresses.

8. References

[1] L. M. Adleman, "Molecular computation of solution to combinatorial problems", Science, 1994.11, (266): 1021-1024.

[2] C. Taylor, V. Risca, and C. Bancroft, "Hiding messages in DNA microdots", Nature, 1999, (399): 533-534.

[3] Lipton R J "Using DNA to solve NP complete problems".Science,1995,268:542-545

[4] Boneh D,Dunworth C,Lipton R. "Breaking DES using molecular computer". Technical report CS-TR-489-95,Princeton University,1995

[5] G. Z. Cui, "New direction of Data Storage: DNA molecular Storage Technology", Computer Engineering and Applications,vol. 42, pp. 29-32, 2006.

[6]J.D.Watson, F.H.C Crick "A structure for deoxyribose nucleic acid",Nature,vol 25,pp 737-738,1953

.