Journal of Network and Information Security
6 (1), June 2018, 30-33
http://www.publishingindia.com/jnis

DNA Cryptography: New Field of Cryptography

Ginu Saju[1*] and Deepika M. P.[2]
[1]PG Student, Adi Shankara Institute of Engineering and Technology, Kalady, Ernakulam District, Kerala, India.
Email: ginusaju199@gmail.com
[2]Associate Professor, Adi Shankara Institute of Engineering and Technology, Kalady, Ernakulam District, Kerala, India.
*Corresponding Author

Abstract: Cryptography technique is used to hide or represent information so nobody but authorized parties can decode it. DNA cryptography drawn attention after DNA computing was first proposed by Adleman in 1994, and now it has become the frontier of cryptography. DNA cryptography is a new born cryptographic field, in which DNA is used as information carrier. In this paper, we briefly introduce the biological background of DNA cryptography and techniques of DNA cryptography.

Keywords: Cryptography, DNA cryptography, Nucleotides, Oligonucleotides, RNA.

I. Introduction

Cryptography technique is used to hide or represent information so nobody but authorized parties can decode it. Cryptography is the studies of mathematical techniques of information security such as confidentiality, data integrity, and authentication. DNA cryptography is developed with the research of DNA computing (also called biological computing or molecular computing) [4]. The traditional cryptography made great progress in the 20th century with the development of electronic technology and is widely used currently. Quantum cryptography was invented in 1970s and has made some progress in recent decades, but there is still some distance from being used in practice. DNA computing proposed by Adleman in 1994 draws the attention to DNA cryptography and now it has become the frontier of cryptography. All the three kinds of cryptography might constitute the major fields of future cryptography.

Symmetric and Asymmetric key are two categories Cryptographic algorithms. In Symmetric algorithm, a common key is shared between the sender and the receiver. The other involves public and private keys that are mathematically related. Nowadays, many research papers are proposed based on DNA encryption schemes that use biological properties of DNA sequences [6, 7, 8]. In the last few years, DNA Cryptography seems to be a promising strategy for fulfilling the current information security needs.

There are different processes to encode data and different DNA techniques for secure data transmission like Polymerase chain reaction, bio-molecular, one-time-pad [9]. The most advantage of DNA cryptography is parallel processing capabilities.

II. Biological Background

DNA is the abbreviation for deoxyribonucleic acid which is unique for all living organisms. DNA is a kind of biological macromolecule and is made of nucleotides. Each nucleotide contains a single base and there are four kinds of bases, which are adenine (A) and thymine (T) or cytosine (C) and guanine (G), corresponding to four kinds of nucleotides. A single stranded DNA is constructed with one end is called 5′, and the other end is called 3′. Usually DNA exists as double-stranded molecules in nature [1]. The two complementary DNA strands are held together to form a double-helix structure by hydrogen bonds between the complementary bases of A and T (or C and G). This structure was discovered by Watson and Crick; thus the complementary structure is called Watson-Crick complementary. A DNA molecule has double-stranded structure obtained by two single-stranded DNA chains, bonded together by hydrogen bonds: A = T double bond and C ≡ G triple bond. The double helix structure is configured by two single anti parallel strands [3] (Fig. 1).

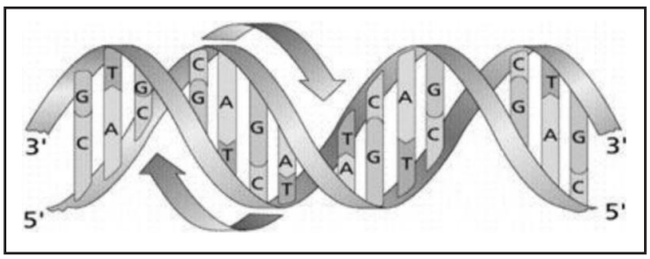

Fig. 1: DNA Structure

The DNA strands that bond each other through A-T and C-G bonds are known as complementary strands. The DNA strands can be chemically synthesized using a machine, known as DNA synthesizer. The single-stranded chains obtained artificially

with the DNA synthesizer are named oligonucleotides having usually 50-100 nucleotides in length.

Ribonucleic acid (RNA) is a single strand of ribonucleotides, identical to ssDNA strands except thymine (T) which is replaced with uracil (U). The genetic information flows from DNA into mRNA (messenger RNA) process known as transcription and from mRNA to protein, known as translation, defining what is known as the central dogma of molecular biology [5].

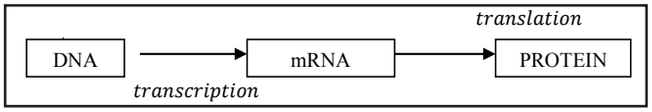

Fig. 2: Central Dogma of Molecular Biology

III. DNA Hybridization Technique

The strands or sequences of DNA obtained have 50 to 100 nucleotides in extent. These strands are termed as oligonucleotides. In this literature the single stranded DNA sequences are represented as ssDNA and the double stranded or helical form of the DNA sequences are represented as dsDNA. A single unique ssDNA under specific situations can combine with other matching or complementary ssDNA to form the double stranded DNA helix form dsDNA. The process of forming dsDNA is illustrated in the fig. 3[2]. Since the ssDNA from distinct sources which are considered to be hybrids, join together to form molecules of double strands. This process is termed as hybridization.

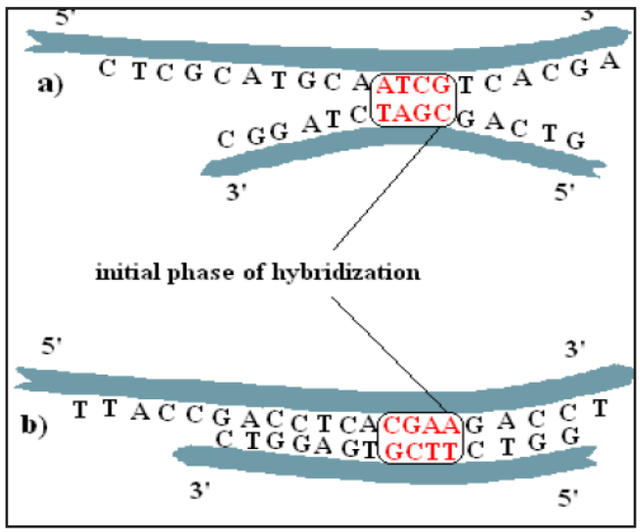

Fig. 3: DNA Hybridization

As in all the cryptographic methods, the DNA hybridization technique also involves the encryption and decryption processes in converting the plaintext into the cipher text and then retrieving back the original message.

Plain Text:

The original message is taken as plain text. Let us consider the plain text to be 'THE'. The hybridization technique will be described with the example given below:

Conversion of Plain Text to the Binary Form:

The plain text is initially converted to the ASCII code and it is again converted to the binary message as follows:

The plain text THE is converted to the corresponding ASCII code.: THE → 84 72 69

The ASCII code is converted to its binary form of the data:

84 72 69 → 010101000100100001000101

OTP Key:

The OTP key should be lengthier than the size of the message. The random generation of OTP key makes the DNA hybridization technique a secure method. The OTP key is to be DNA bases. For each bit in the binary message, a key length of 10 bits is generated. For example we have '15' binary bits. Thus, a key length of 15*10 = 150 bases is to be produced.

Encryption:

During the encryption process, the Hybridization technique performed only for the binary '1' in the data. If the binary bit is found to be '0' no operation is functioned. For example the binary data: 010101000100100001000101The randomly generated OTP key:

TATTTGTCCA, CCCACAGTCG GAACCTTCTC,

TATGAGTTTG, CCGAGACCTC, GTCGATCTCT, AAGATCACAA, ATGGCCTTCT, AGGCCGTACA, CTGTACCCTA, CTACAAAAGT, CTTAGAATAA, TGATCAGTCG, GATTAACTGG, CTTGACGAGG

If the first bit of OTP key is 1 then the 10 bases of the key from last is choosen as follows:

- The first bit of the binary data: 0(so no operation)
- The second bit of binary data:1
- The last 10 bases of the OTP key: CTTGACGAGG
- The encrypted message: GAACTGCTCC

Decryption:

Now, the data decryption is performed by comparing the OTP key. Let us consider below given is the small part of the encrypted message:

Encrypted message:

TATTTGTCCA, GAACTGCTCC GAACCTTCTC, etc...

OTP Key:

TATTTGTCCA, CCCACAGTCG GAACCTTCTC,

TATGAGTTTG, CCGAGACCTC, GTCGATCTCT,
AAGATCACAA, ATGGCCTTCT, AGGCCGTACA,
CTGTACCCTA, CTACAAAAGT, CTTAGAATAA,
TGATCAGTCG, GATTAACTGG, CTTGACGAGG

In decryption process, the first 10 bits of the encrypted message is compared with the last 10 bits of the OTP key, if they are found to be complementary then a binary '1' is formed. If the complementary matches are not found, it is simply replaced with a zero, '0'.

- First 10 bases of the encrypted message: TATTTGTCCA,
- The last 10 bases of the OTP key: CTTGACGAGG (no complementary matches)
- Decrypted message in binary form: 0
- The next 10 bases of the encrypted message: GAACTGCTCC
- The next 10 bases of the OTP key from reverse: CTTGACGAGG
- The decrypted message: 1
- So, the total decrypted message is: 01…

The next comparison will be done for the next 10 bases of the encrypted message and the next 10 bases of the OTP key taken in reverse. Thus, the process continues in this manner and the decrypted message is obtained as 01010100010010000 1000101.

From this binary form of the data, the message can be converted to its corresponding ASCII code and the original data will be obtained as 'THE'. This is the process carried out in DNA hybridization method.

IV. DNA Indexing Technique

In the DNA indexing method of the DNA cryptography, the original message i.e. the plain text is converted into the DNA form of the data. This DNA form of the data is then compared with the OTP key and the data is compared for its match and an index of array is generated for that particular. A random number is chosen from the index array as the encrypted. Therefore, for each single character an index is generated. The decryption process is done in reverse to obtain the original data.

Plain Text:

The original data is considered as the plain text. The plain text is first converted to the ASCII code and then to the corresponding binary code of the data. From the binary form of the data it is again converted to the DNA form of the data [9]. The below example explains further:

- Plain text: 'set'
- Conversion of plain text to ASCII code:
  - 115 101 116
- So, for the alphabet, s in the plain text the corresponding ASCII code is 115.
- The conversion of ASCII to binary data for the alphabet s: 01110011
- Thus, now we have s → 115 → 01110011. It is again converted to the DNA form of the data. The DNA form is converted the substitutions are A for 00, C for 01, G for 10 and T for 11.Thus, letter s in the plain text is further converted to the DNA form as,:

$$s \rightarrow 115 \rightarrow 01110011 \rightarrow CTAT$$

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 code 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 above substitution each character is represented by a 4 letter word. The transformed 4- letter DNA nucleotide bases used as OTP and it is scanned till the end of the sequence. 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 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.

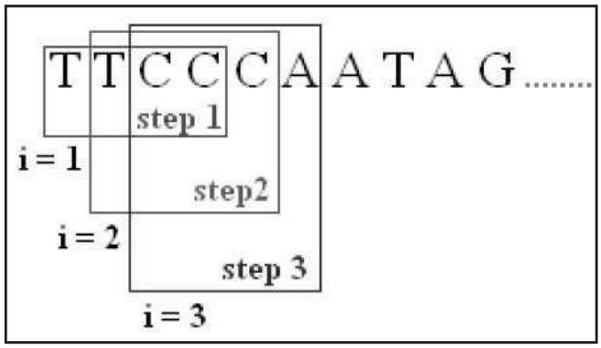

Fig. 4: Scanning Procedure of OTP Key [6]

Each character is obtained as an array of characters and the number of characters depends on the DNA sequences. For each character was chosen a random index from its array of indexes. Thus we obtain the encrypted message as an array of characters, one for each character.

For example Plain text" set" is converted:

ASCII code 115 101 116

s- 115- 01110011 – CTAT –

indexes: 166 258 789 927 1295 2954

3045 3098 3181 3207 3361 3763

……..till selected DNA sequence. For each character was chosen a random index from its array of indexes using. Below are established positions of random indexes inside character's arrays:

115- 70th index 23811(for alphabet s)

101- 26th index 13981(e)

116 – 158th index 25264(t)

Final encrypted message is:

23811 13981 25264

Decryption

In both encryption and decryption the same OTP is, 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 to get the original plain text.

V. Conclusions

In this paper, we got into DNA Cryptography, biological background; basic notions of DNA Cryptography were talked. Afterwards, two proposed techniques in DNA Cryptography were investigated.

References

[1] S. Pramanik, and S. K. Setua, "DNA cryptography," *7th International Conference on Electrical & Computer Engineering*, 2012.

[2] S. Jain, and V. Bhatnagar, "Bit based symmetric encryption method using DNA Sequence," *5th IEEE International Conference on Confluence: The Next Generation Information Technology Summit*, 2014.

[3] E. M. S. Hossain, K. M. R. Alam, M. R. Biswas, and Y. Morimoto, "A DNA cryptographic technique based on dynamic DNA sequence table," *19th International Conference on Computer and Information Technology*, 2016.

[4] D. Kumar, and S. Singh, "Secret data writing using DNA sequences," *International Conference on Emerging Trends in Networks and Computer Communications*, 2011.

[5] O. Tornea, and M. E. Borda, "DNA cryptographic algorithms," *International Conference on Advancements of Medicine and Health Care through Technology*, pp 223-226, 2009.

[6] M. Borda, and T. Olga, "DNA secret writing Techniques," *8th International Conference on Communications,* 2010.

[7] Z. Yunpeng, Z. Yu, W. Zhong, and R. O. Sinnott, "Index-based symmetric DNA encryption algorithm," *4th International Congress on Image and Signal Processing*, pp. 2290-2294, 2011.

[8] G. Cui, L. Qin, Y. Wang, and X. Zhang, "An encryption scheme using DNA technology," *In IEEE International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 37-42, 2008.

[9] C. T. Celland, V. Risca, and C. Bancroft, "Hiding messages in DNA microdots," *Nature: Interntional Journal of Science*, vol. 399, no. 6736, pp. 533-534, 1999.

[10] G. Jacob, and A. Murugan, "A hybrid encryption scheme using DNA technology," *The International Journal of Computer Science and Communications Security*, vol. 3, pp. 61-65, 2013.