A New DNA Cryptography Algorithm Based on the Biological Puzzle and DNA Chip Techniques

Yun-peng ZHANG[1], Zhen-zhen WANG[2a], Zhi-wen WANG[2b], Yasin Hasan KARANFIL[3] and Wei-di DAI[4]*

[1]Department of Information and Logistics Technology,University of Houston, Houston, U.S. yzhang119@uh.edu

[2]School of Software Engineering, Northwestern Polytechnical University, Xi’an, China. [a]poweryp@183.com, [b]rjxyjs@nwpu.edu.cn

[3]Imperial College London, London, SW3 6NP, UK, yasinkaranfil@gmail.com

[4]School of Computer Science and Technology, Tianjin University, Tianjin, China, davidy@tju.edu.cn

Keywords: DNA Cryptography, Algorithm, Biological Puzzle, DNA chip.

Abstract. Currently, a lot of theoretical and technical issues still need to be addressed regarding applications of DNA computing in the field of DNA cryptography. The current common DNA cryptography algorithms are basically based on traditional cryptography, integrate dintoexisting DNA technology, of which the feasibility has not yet been fully demonstrated. Through research and analysis, we select the biological puzzle that "DNA sequencing is difficult under the conditions of not knowing the correct sequencing primers and probes". On the basis of the DNA chip and biological technologies, we design a new DNA cryptography algorithm. The simulation verification about the feasibility and safety of the method is provided. The results show that, while ensuring feasibility, this method has stronger security with respect to traditional encryption methods.

Introduction

DNA computing is a pioneering academic research field that has arisen in recent years. It feature sa new calculation method of which the calculation medium is a biological DNA molecule and the calculation means are biochemical reactions. Compared with existing computer technology, DNA computing has the following advantages: a high degree of parallelism, low energy consumption, and a huge amount of information storage. Based on these characteristics, DNA cryptography has a unique advantage in the applications of massively parallel data encryption with less real-time demanding, secure data storage, authentication, digital signature, information hiding, etc.

DNA cryptography [1] is a new area of cryptography that has emerged in recent years along with the research area of DNA computing. The information carrier of DNA code is a DNA molecule, and the implementation tools are modern biotechnologies. It makes full use of the inherent high storage density and high parallelism advantages of DNA computing and DNA cryptography, completing cryptographic functions like encryption, authentication and signatures.

Related Technologies and Biological Difficult Problems

Andre' L,et.al,[1] introduced two different encryption methods based on a DNA binary string. The methods are carried out under certain assumptions. The feasibility of the method is limited, but it can be used as a reference for future research.

Sivan S, et. al,[2] achieve image encryption by combining the molecular automaton with the DNA chip. The advantages of the method in this paper are the use of molecular automaton techniques and DNA chip technology, while the downside of the method is that it can only be used for image encryption and the operation feasibility of the method is not sufficiently validated. But we can do more in–depth exploration and research on DNA encryption methods according to the authors' findings.

In paper[5], combining difficult biological problems with the conventional cryptographic theory, the authors proposed an encryption method. Difficult biological problems and encryption computational difficulties provide a double secure version. The validation experiments show that the method has a high level of security strength.

Luming X,et. al,[6] made use of DNA synthesis, DNA cloning, PCR amplification and DNA chip technologies, combined with computational complexity theory of cryptography, to propose an encryption method based on biotechnology. Due to the limitations of existing biological and computing technologies, it is difficult to get the plaintext encrypted by this method without knowing the correct decryption keys, which ensure the security of the method.

DNA Encoding and PCR Primer Designing

DNA Encoding

There are mainly three kinds of DNA encoding methods:

1) A base represents 2 binary numbers, A, T, G, C represent the 10, 01,11 and 00 respectively;[5]

2) Using two DNA short chain molecules to represent 0 and 1 respectively;[2]

3) DNA encoding as quaternary, using three bases to represent a letter or number.[7]

We choose the first encoding method, which uses a base to represent 2 binary numbers, A/T/G/C represent the 10/01/11/00 respectively, namely 0123 / CTAG.

PCR Primers

Currently, the common primer design software includes Oligo 6.0, Primer premier4.11 and others. In this paper, we designed the appropriate primers based on the template DNA strand using the Oligo Analyzer tool. Each DNA fragment is inserted into a different location of the DNA vector. In the decryption process, we first perform PCR amplification on the long strands of DNA according to the designed primer, so as to find the corresponding DNA fragments (P_1, P_2, P_3) in a long DNA chain.

Since the original DNA template has a great effect on the design of primers and the characteristics of the DNA sequences converted from plaintext through a series of steps have difficulty meeting the PCR primer design principles, the designed encryption method adds a pair of encryption key sat the beginning and the end of each DNA sequence. The encryption key pairs are selected from the DNA sequences corresponding to the PCR primers that can be used for PCR amplification. Thus, we can avoid a situation in which the PCR primers corresponding to the DNA sequences derived from the encrypted plaintext can't be designed. This has the added result thatthe encryption key and the decryption key correspond to one another.

The encryption process requires three pairs of encryption keys: (s_i, e_i), i = 1,2,3. The length of the encryption key is 20.The corresponding encryption keys are added to the beginning and to the end of each DNA segment.The corresponding primers of the encryption keys can be designed using the Oligo Analyzer tool. The encryption keys can be obtained from the DNA sequences that can be performed to design appropriate PCR primers and carry outPCR amplification in a biological laboratory. The DNA fragments after adding corresponding encryption key pairs at both ends are shown in Fig 1.

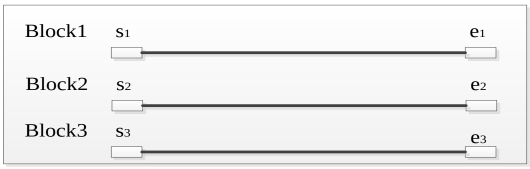
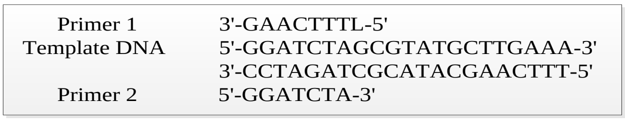


Fig 1.The schematic of DNA fragments after adding corresponding encryption primers at both ends

According to the designed encryption primer pairs (s_i, e_i), using the Oligo Analyzer to design corresponding decrypt primers (decode_s_i, decode_e_i), i=1,2,3.The schematic diagram of the combination of PCR primers and the template is shown in Fig 2.



Primer 1	3'-GAACTTTL-5'
Template DNA	5'-GGATCTAGCGTATGCTTGAAA-3'
	3'-CCTAGATCGCATACGAACTTT-5'
Primer 2	5'-GGATCTA-3'

Fig 2. The schematic diagram of the combination of PCR primers and the template

The schematic diagram of long DNA chain that contains the desired DNA fragmentsis shown in Fig 3.

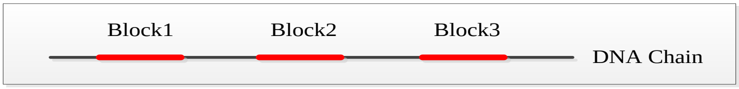


Fig 3. The schematic diagram of long DNA chain that contains the desired DNA fragments

The DNA Encryption Method based on the Biological Puzzle

Key Generation

Here the keys that can hybridize with the encryption keys used for designing DNA chips andproducing the signals that meet certain conditions can all be used as the decryption keys. The decryption keys can hybridize with different encryption keys, producingthe same hybridization signals. Similarly, all encryption keyscan also hybridize with different decryption keys and produce the same hybridization signals. The encryption keys and the decryption keys are different, sothe method is a kind of asymmetric encryption method. Here the decryption key contains the complementary sequence of all encryption keys.

Encryption keys and decryption keysboth have a lot of different probes, which are generally divided into two kinds:probe 0 and probe 1. For the plain binary matrix, we put one kind of probe 1 on the position where the value is 1, and put one kind of probe 0 on the position where the value is 0. In this waywe can make the desired DNA chips. The receiving party obtains the plaintext by using the decryption key to hybridize with the DNA microarray.The decryption keys contain the complementary sequence of all 0 and 1 probes in encryption keys.

In this paper, the encryption keys and the decryption keys are both divided into three types: two different kinds of probes corresponding to the two kinds of bases, and a kind of probe corresponding to useless information. Since the plaintext is separated into two matrices, it is necessary to make two DNA chips, these two DNA chips corresponding to two different encryption keys and decryption keys. For amatrix containing the nucleotide bases A and T, we need three kinds of keys:probe_A, probe_Tand probe_ATK (ATK corresponds to the location where the nucleotide base is neither A nor T). The location with the nucleotide base A is put in one kind of probe A among all encryption keys and the location with the nucleotide base T is put in one kind of probe T among all encryption keys. In the locations which contain no information one kind of probe ATK is placed among all encryption keys. In this way, we can make a DNA chip. For a matrix containing the nucleotide bases C and G, we also need three kinds of keys, probe_C, probe_G and probe_CGK (CGK corresponds to the location where the nucleotide base is neither A nor T). The location with the nucleotide base C is put in one kind of probe_C among all encryption keys and the location withthe nucleotide base G is put in one kind of probe_G among all encryption keys. In the locations which contain no information one kind of probe_CGK is placed among all encryption keys. In this way, we can make another DNA chip. The receiver can get the plaintext byusing the decryption keysto hybridize with the DNA microarray.

If the encrypted ciphertext is not composed of DNA chips but a mixture, the thing being placed in the test tube must be the DNA probe that corresponds to every bit of the plaintext binary matrix, i.e. every DNA probe corresponds to a bit of the matrix which is different because you do not know the hybridization signal intensity of each bit. If the cipher text is a DNA chip, the probe on each bit of chip may be the same. In the paper, the cipher text we make is a DNA chip. In order to take full advantage of the rich resources of DNA, as well as increase the key space, each probe being put on the chip is not the same.

Among the existing biological molecules material, the stable molecule like DNA, PNA [8] and protein can all be used as encryption keys. The only difference is the different experimental processes.

Encryption Scheme

The flowchart of the encryption process of the method is shown in Fig 4.

1)The plain text is converted to the binary sequence (N bits) that will be pretreated with the following steps. (1)Find the DNA sequence Q(such as HIV) from the online gene pool, and record the path of Q in the online gene pool as Path_Q; (2) Convert the Q into a binary stream Q_2; (3) Divide the binary stream of plaintext and the binary stream transformed from the DNA sequence Q with 8-bit binary as a unit. If the number of 1 in the 8-bit plaintext binary sequence is odd, we perform the XOR operation between the 8-bit plaintext binary sequence and the binary sequence transformed from the corresponding DNA sequence Q. Otherwise, we perform the XOR operation between the 8-bit plaintext binary sequence and the complement binary stream of binary sequence transformed from the corresponding DNA sequence Q.Or we can find a DNA sequence X which has N/16 or (N + 1)/16 nucleotides from the online gene pool and record the path of X in the online gene pool as Path_X. Then X is converted to a binary sequence X_2, of which each bit corresponds to an 8-bit binary block. If the value of one bit is 1, we perform the XOR operation between the corresponding 8-bit plaintext binary sequence and the binary sequence transformed from the corresponding DNA sequence Q. Otherwise; we perform the XOR operation between the corresponding 8-bit plaintext binary sequence and the complement binary stream of binary sequence transformed from the corresponding DNA sequence Q. It is equivalent to performing a stream encryption on the plaintext.

2)The DNA sequence obtained from step 1 is named as W and then W is segmented into n segments, each has the length of N/n. These n DNA sequence segments are named as P_1, P_2,... ,P_n. For the sequence P_i, we add s_i and e_i at both ends of P_i, i=1,2,…,n, so as to prepare for designing primers for PCR amplification. Then, the new sequences are plugged in a complex region of bio-genetic DNA respectively. The vector DNA can be selected from the biological experiments area or the gene pool on the network. Each DNA fragment is thus block-encrypted into new DNA blocks S_1, S_2, ..., S_n. Then these DNA blocks are rearranged and connected into a long DNA strand.

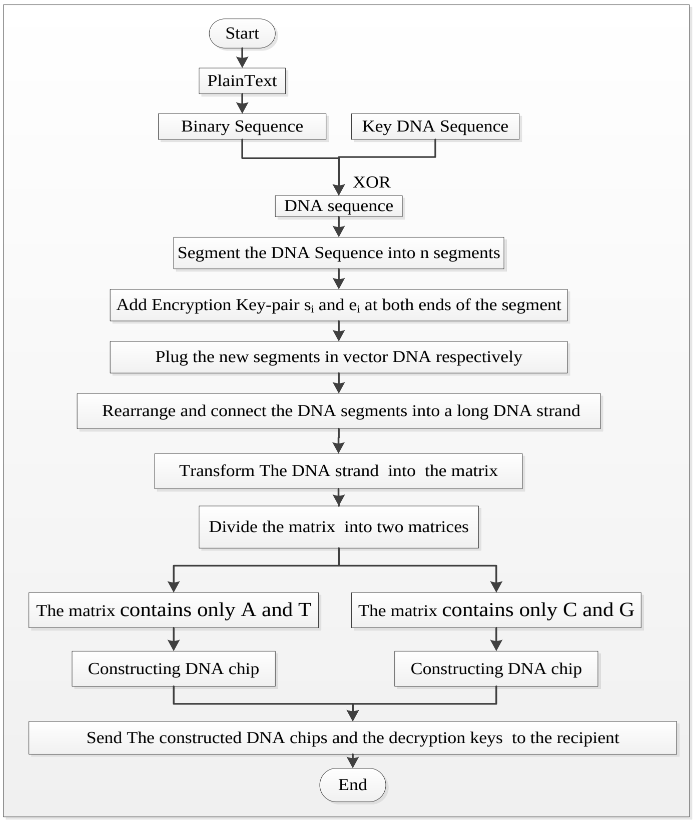

Start
PlainText
Binary Sequence
Key DNA Sequence
XOR
DNA sequence
Segment the DNA Sequence into n segments
Add Encryption Key-pair s_i and e_i at both ends of the segment
Plug the new segments in vector DNA respectively
Rearrange and connect the DNA segments into a long DNA strand
Transform The DNA strand into the matrix
Divide the matrix into two matrices
The matrix contains only A and T
The matrix contains only C and G
Constructing DNA chip
Constructing DNA chip
Send The constructed DNA chips and the decryption keys to the recipient
End

Fig 4. Encryption flowchart

3) The long DNA strand obtained in step 2 is transformed into the matrix form (p*q) according to a certain arrangement of the composition, which will be encrypted according to the designed encryption method to get the cipher text. Firstly, the matrix is divided into two matrices, one containing only A and T, and another containing only G and C. And then we design two kinds of DNA sequences corresponding to each position on the encryption DNA chip(For the matrix containing only A and T, we design two kinds of sequences corresponding to A and T and for the matrix containing only C and G, we design two kinds of sequences corresponding to C and G).

4)Constructing a DNA chip according to the two kinds of DNA sequence Keys designed in step 3. Herein the size of the DNA chip is p*2q. The left half of the matrix p*q, corresponds to the matrix that contains only A and T. For the position where the base is A, we select the DNA sequence corresponding to A as the DNA probe and put it in the corresponding position on the DNA chip. If the base on a position is T, we select the DNA sequence corresponding to T as the DNA probe and put it in the corresponding position on the DNA chip. The right half of the matrix p*q corresponds to the matrix that contains only C and G. And the manufacturing operation is the same as the left half. Thereby, we complete the construction of the two DNA chips.

5)The constructed DNA chips and the decryption keys are sent to the receiver. The encryption process is finished.

Decryption is the inverse process of encryption.

Simulation

In this paper, the cryptography algorithm is simulated using VC ++ 6.0. We simulate the encryption and decryption process. The obtained plaintext is the same as the original information and the decryption is successful.

Conclusions

In this paper, we propose a new encryption method which takes full advantage of characteristics of the biological puzzle. The main contributions of this paper are:(1) We analyze the feasibility of designing a new encryption method based on the biological puzzle, which provides a theoretical basis for the security and operability of the proposed encryption method;(2) We propose a new DNA cryptography algorithm based on DNA chip technology and the biological puzzle.

Through this study, we have found an effective cryptography algorithm, and the simulation results show that the algorithm is feasible and secure, which can be a reference for future research into DNA encryption methods.

Acknowledgements

The research was funded by the University of Houston - National Research University Fund (NRUF, 110664 LINKED TO 105433) and the Seed Foundation of Tianjin University (2016XZC-0016).

References

[1] Jiang J, Yin Z. The Advantages and Disadvantages of DNA Password in the Contrast to the Traditional Cryptography and Quantum Cryptography[C],the Eighth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), 2013. Springer Berlin Heidelberg, 2013:307-316.

[2] Andre´ L, Christoph R, Wolfgang B, Hilmar R.Cryptography with DNA binary strands [J].BIOSYSTEMS,2000,57(1): 13-22.

[3] Sivan S, Ron P, Yoav A, Ehud K. A Molecular Cryptosystem for Images by DNA Computing [J]. AngewandteChemie, 2012, 124(12): 2937-2941.

[4] Zhang Z, Shi X, Liu J. A method to encrypt information with DNA computing[C]// Bio-Inspired Computing: Theories and Applications, 2008. BICTA 2008. 3rd International Conference on. IEEE, 2008:155-160.

[5] GuangzhaoC, LiminQ, YanfengW.An Encryption Scheme Using DNA Technology[J]. 3rd International Conference on Bio-Inspired Computing - Theories and Applications,2008:37-41

[6] Xin L M, Yuan C, Lei Q, et al. An encryption scheme based on DNA [J]. Journal of Xidian University, 2006, 33(6):939-942.

[7] Morford L. A theoretical application of selectable markers in bacterial episomes for a DNA cryptosystem[J]. Journal of Theoretical Biology, 2011, 273(1): 100-102.

[8] Nielsen P E, Egholm M, Berg R H, et al. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide[J]. Science, 1991, 254(5037): 1497-1500