16

Research on DNA Cryptography

Yunpeng Zhang* and Liu He Bochen Fu
College of Software and Microelectronics, Northwestern Polytechnical University, Xi'an, China

1. Introduction

The 21st century is a period of information explosion in which information has become a very important strategic resource, and so the task of information security has become increasing important. Cryptography is the most important component part of the infrastructure of communication security and computer security. However, there are many latent defects in some of the classical cryptography technology of modern cryptography - such as RSA and DES algorithms - which have been broken by some attack programs. Some encryption technology may set a trap door, giving those attackers who understand this trap door the ability to decipher this kind of encryption technology. This information demonstrates that modern cryptography encryption technology based on mathematical problems is not so reliable as before.

The relation between cryptography and molecular biology was originally irrelevant, but with the in-depth study of modern biotechnology and DNA computing, these two disciplines begin to work together more closely. DNA cryptography and information science was born after research in the field of DNA computing field by Adleman; it is a new field and has become the forefront of international research on cryptography. Many scholars from all over the world have done a large number of studies on DNA cryptography. In terms of hiding information, there are such results as "Hiding messages in DNA microdots," "Cryptography with DNA binary strands" and so on. In terms of DNA algorithms, there are such results as "A DNA-based, bimolecular cryptography design," "Public-key system using DNA as a one-way function for key distribution," "DNASC cryptography system" and so on. However, DNA cryptography is an emerging area of cryptography and many studies are still at an early stage.

DNA Cryptography is based on biological problems: in theory, a DNA computer will not only has the same computing power as a modern computer but will also have a potency and function which traditional computers cannot match. First, DNA chains have a very large scale of parallelism, and its computing speed could reach 1 billion times per second; second, the DNA molecule - as a carrier of data - has a large capacity. It seems that one trillion bits of binary data can be stored in one cubic decimetre of a DNA solution; third, a DNA molecular computer has low power consumption, only equal to one-billionth of a traditional computer.

*Corresponding author

2. Technology and software

DNA cryptography is a subject of study about how to use DNA as an information carrier and it uses modern biotechnology as a measure to transfer ciphertext into plaintext. Thus, biotechnology plays an important role in the field of DNA cryptography. In this part we will introduce some of the DNA biotechnology and software of the field of DNA.

2.1 Gel electrophoresis

Electrophoresis is a phenomenon where one charge moves in the opposite direction of its electrode in an electric field. This is an important method for the separation, identification and purification of DNA fragments. At present, there are two kinds of medium: agarose and polyacrylamide. Both of these can be made for a gel with different sizes, shapes and diameter. In causing electrophoresis on different devices, we call it either agarose gel electrophoresis or polyacrylamide gel electrophoresis. When DNA molecules go through the sieves which are formed by the gel, the short DNA molecule moves faster than the longer one and so we can discriminate between them easily.

2.2 The technology of DNA fragment assembly

DNA fragment assembly is a technology which attempts to reconstruct a large number of DNA fragments into the original long chain of DNA. In order to solve the limit of the length of the sequence, the researchers developed this technology. The measures are as follows: First, the researchers amplified the DNA chain and got lots of backup; second, they obtained a large number of short DNA fragments by cutting the DNA long chain at random locations; finally, the researchers recombined the DNA fragments - which have an overlapping part - back into the original DNA chain. This strategy is called “shotgun sequencing.”

2.3 DNA chip technology

DNA chip technology is to the manuscript should be presented without any additional comments in the margins.synthesis oligo probe on solid substrates or else directly solidifies a large amount of a DNA probe in an orderly fashion on the surface of substrates using the method of micro-printing. It then hybridises with the labelled sample, through the testing and analysis of the hybridised signal, so as to get the genetic information (the gene order and the information it gives) about the sample. Since silicon computer chips are usually used as solid substrates, it is called a DNA chip.

DNA chip encryption technology has two layers of security: one layer is provided by the limitations of biotechnology and it is also the security that the system primarily based on. The other layer is that of computing security - even if an attacker breaks through the first layer of security - in the case where they do not have the decipher key - they must have strong computing power and data storage capacity in order to decipher the DNA chip. Now, the encryption progress of DNA chip technology will be presented.

2.4 PCR technology

PCR Technology is also called “polymerase chain reaction” and it is a rapid amplification technology of DNA. Because it is very difficult to manipulate small amounts of DNA, PCR

Technology usually used to amplify the DNA which has been determined. In practice, DNA amplification techniques include cloning. The amplification efficiency of PCR is very high, and can amplify a large number of chosen DNA in a short period of time. Moreover, PCR will achieve the amplification by using natural nucleotide molecules. In order to achieve PCR amplification, the experimenter needs to know the sequence of the chosen DNA chain, and use it to design primers for amplification. Actually, the primer is also a DNA sequence which contains a number of nucleotides. It is certain that the primer can be amplified for the chosen DNA. In short, the PCR process can be divided into two stages:

1. The design of two primers, separately loaded onto the target DNA in the beginning and at the end;
2. The finding of the target DNA under the action of the polymerase and its amplification.

2.5 The DNA code

DNA is the genetic material of eukaryotes, with a double-helix molecular structure and two single-strands parallel to each other. DNA is something which is called a polymer, which composed of many small nucleotides. Each nucleotide consists of three parts:

1. The Nitrogenous bases;
2. Deoxyribose;
3. Phosphate.

DNA coding is a new area of cryptography which has appeared in recent years along with DNA computing research. Originally there was no connection between these two disciplines -- cryptography and molecular biology (also known as genetics or genomics). However, with the study of DNA - especially after Adleman put forward DNA computing in 1994 - and with more in-depth study, this research can be used in the field of information security. Ultimately, DNA cryptography appeared only gradually. DNA cryptography is built on DNA - which is an information carrier - and modern biotechnology for its tools, and it achieves the encryption process by the use of the characteristics of DNA of massive parallelism and high storage density. In addition, the reason why we can combine cryptography and molecular biology is the encoded plaintext, which can combine the computer and the use of molecular biological techniques, such as polymerase chain reactions, polymerisation overlapping amplification, affinity chromatography, cloning, mutagenesis, molecular purification, electrophoresis, magnetic bead separation and other techniques of molecular biology, and then obtain the final ciphertext. Most importantly, DNA code abandons that traditional cryptography which uses the intractable mathematical problem of the security guarantee, instead using the limited nature of the learning of biology. In theory, DNA code is mainly based on the biology's limitations for security, and has nothing to do with computing ability; as such, it is immune to the attacks of both modern computers and even the quantum computers of the future. Therefore, many scholars have already started to study the better encryption effect of DNA code.

2.6 The chaos code

Chaos will be included in the example of the chapter, and so we discuss the chaotic system only simply, leading to two tracks from two initial points concerning such systems.

Sometimes these tracks will infinitely close, and sometimes they are away from each other. Both cases will appear numerous times - this indicates that the system's long-term behaviour has no rules. It is a pseudo-random phenomenon which can be used in cryptography.

A chaotic system has three key advantages:

- The sensitive dependence on initial conditions;
- The critical level. This is the point of non-linear events;
- The fractal dimension, which shows the unity of order and disorder.

Usually, it is a self-feedback system and so this leads to the system itself being unable to forecast for the long-term.

At present, many chaotic cryptosystems have been used in the iterative process in order to complete data encryption or decryption. The security of ciphertext mainly benefits from the effect of chaotic dynamics. The more dimensions the equation has, the greater the security that will be obtained. However, the time of encryption or decryption will increase, and the ciphertext will soon become longer. Chaotic encryption mainly uses the random sequence - generated by the chaotic system's iteration - as an impact sequence of the encryption transform. This sequence inherits the pseudo-randomness of the chaotic system. Moreover, it can make and spread confusion and it does not identify characteristics of the obtained ciphertext after the use of this sequence to treat the plaintext. This is a great challenge for cryptanalysts. Therefore, the chaos code has been used in some encryption recently.

2.7 Software

DNA fragment stitching software - the DNA Baser Sequence Assembler. The DNA Baser Sequence Assembler is used for splicing DNA fragments fatly. It should be noted that we must prepare some DNA fragments for splicing before using this software.

3. Biological problems

An unintelligible problem in biology is due to the limits of human cognitive and experimental means as well as the problems which have resulted from other scientific laws and which will not be solved in the visible future.

The known biological problems are, mainly:

1. That we do not know the proper primers at present: it is difficult in that we have to separate the unknown and specific sequences of DNA from the unknown mixed liquids of DNA and then sequence them. In the literature, by using DNA synthesis, PCR amplification and DNA digital coding adequately, and with the combination of traditional cryptography, Guangzhao Cui proposed a DNA-based encryption scheme. Unfortunately, the author did not make an adequate difficulty of this biological problem. Therefore, the lack of difficult problems in the literature does not provide sufficient reliability and theoretical support.
2. We have to perform completely accurate sequencing in order to decipher the unknown hybrid DNA (PNA) probe information where the DNA chip (microarray) is only a different nucleotide arrangement. This is the second biological problem.

Now there are two main types of sequencing method:

1. The Maxam-Gilber method, which has also been known as the "chemical degradation method;"
2. The Sanger method, which is also known as the "enzyme method."

Neither of the two methods are suitable for sequencing a little of the unknown mixed sequence of a DNA chip.

In the literature, the author had a discussion as to this problem. He proposed a non-deterministic symmetric encryption system – DANSC-based on this problem. Generally speaking, the biological problem in the literature depends on the sequencing technology, which is still in the primary stages and has its own weaknesses. This will generate a hidden danger when we build the encryption scheme; what is more, the DANSC will also likely face a fate of being cracked in the future.

Of course, there are other difficult biological problems that can be used in DNA cryptography which will be discovered in the future.

4. Analysis DNA encryption which is based on PCR amplification technology

4.1 DNA encoding scheme

In the field of information science, the most basic encoding method is binary encoding. This is because everything can be encoded by the two states of 0 and 1. However, for DNA there are four basic units:

1. Adenine (A);
2. Thymine (T);
3. Cytosine (C);
4. Guanine (G).

The easiest way to encode is to represent these four units as four figures:

1. A(0) –00;
2. T(1) –01;
3. C(2)–10;
4. G(3)–11.

Obviously, by these encoding rules, there are 4! = 24 possible encoding methods. For DNA encoding, it is necessary to reflect the biological characteristics and pairing principles of the four nucleotides. Based on this principle, we know that:

A(0) – 00 and G(3) – 11 make pairs,
T(1) – 01 and C(2) – 10 make pairs.

In these 24 programs, there are only 8 programs

0123/CTAG,
0123/CATG,
0123/GTAC,
0123/GATC,
0123/TCGA,

0123/TGCA,
0123/ACGT,

0123/AGCT match the DNA pair of a complementary principle. The coding scheme should be consistent with the weight of a molecular chain, so we get that 0123/CTAG is the best encoding scheme.

4.2 Encryption process

If the encrypter wants to encrypt the plaintext, he first needs to transform the plaintext by using the code rules. Next, he obtains the DNA sequence with its base sequence represented a special meaning and he then uses the biotechnology and - according to DNA sequences - artificially synthesises the DNA chain as the target DNA. After this, he can design the appropriate primers as the key. When the sender has the key, he loads them onto the target DNA for its strand and end according to the sequence synthesis primers of the primer. On this basis, we use DNA technology to cut and splice, and implant this DNA to a long DNA chain. Finally, he adds an interfered DNA chain, namely the common DNA chain. The sequence of these chains does not contain any meaningful information.

4.3 Analysis of DNA encryption based on PCR technology

4.3.1 Safety analysis

For this encryption scheme - and because the ciphertext includes the DNA chain for the carrier, its message will be represented by the base sequence of the DNA chain. When the cryptographers intercept the ciphertext, what is obtained is a DNA mixture in which there is a lot of confusion in the DNA chain. As with the technology of PCR itself, this technology has high requirements for the correctness of the primers of the sequence. If starting amplification experiment, then it is impossible to try to find out the target gene without knowing of the primer sequence. Because, in this case, (if) cryptographers designed the primer by themselves, then first, they do not know the molecule length of the correct primer. For any different length that they have, they will get the wrong message. Even if the length is right, and supposing there are 25 base sequences, in theory there will be 4^{25} kinds of primers. If cryptographers experiment on them one by one - and they assume that taking one PCR amplification requires 2 or 3 hours - they would need 10^{27} years to finish it. This is impossible.

However, only using DNA Encryption based on PCR Technology is not always safe, because the plaintext and the converted DNA are in a one-to-one relationship, and the ciphertext contains the plaintext's unique statistical properties. In this case, the cryptanalyst can decipher it though statistical attacks, giving the password a security risk.

4.3.2 Feasibility analysis of the experimental operation

The primers that are designed must comply with the following principles:

1. Specificity.

Primers should be arranged in a specific way - especially with regard to the amplified target sequences between the two primers - and we should make sure of at least a 30% difference and the arrangement of 8 consecutive Bases cannot be the same;

2. Length.

Statistical calculations indicate that the 17 base sequences in the human DNA are likely to occur at one time, and so the primer length general controls more than 17; however, it cannot have unlimited length and at most it cannot longer than 30 Bases sequence. Usually, the best length is 20 to 24 Bases. This length of DNA primer has a strong stability when reacting, and does not produce hybrids;

3. The content of C and G bases.

The content of C + G needs to be controlled at 40% to 60% so as to avoid containing too many bases polymers, and the percentage of the C + G in the two primers should be similar;

4. Random Distribution of bases.

The distribution of bases in the primer should be random so as to avoid more than three consecutive identical bases;

5. The primer Itself.

The complementary sequence should not appear in the primer sequence itself, and if it cannot be avoided we must ensure that there are less than 3 bases in a complementary situation, at the very least;

6. Between the Primers.

Each primer should avoid appearing in the complementary sequence;

7. The End of Primer 3′.

Not using Base A at the 3′ end, because A has a high rate of mismatch, and it cannot make any modification at the 3′ end;

8. The End of Primer 5′.

The 5′ end of the primer limits the length of PCR amplification's product, but it is less demanding and some fluorescent markings can be modified.

Because PCR primer design is a crucial part of the technology, and because the use of PCR technology is at the core of this encryption algorithm - as well as for its safety and security conditions - if we use inappropriate PCR primers, it will lead to experiment failure. Therefore, the design of the primers must comply with the above principles. Here, we can use the biological expertise software to help design the primers. The software called - Primer Premier 5.0.

5. The united chaos encryption algorithm based on the logistic map and the henon map

5.1 Research for the logistic map

The logistic map is the most widely used chaotic map. It is a one-dimensional chaotic map with the advantages of a high efficiency and simplicity. A logistic map is defined as:

$$x_{n+1} = \lambda \times x_n \times (1 - x_n)\text{，}\ \lambda \in (0,4), n = 0,1,... \qquad (1)$$

We use Parameter λ and the initial value x_0 as a key. Parameter λ can be divided into three parts and start parameter validation. Make x_0 equals to a random value of 0.79284, and then take the above data into formula 1 which is as the defination of a logistic map, and making it iterate 100 times. Next, make a picture to analyse each x. There are three kinds of situations, as follows:

When λ ∈ (0,1) and where we have a random value for λ=0.5789757497. Then we iterate it 100 times and the value is shown in Figure 1. We can see that after 10 times, the values of x have tended to 0. Here, it is already doesn't have any random features which the chaos should have.

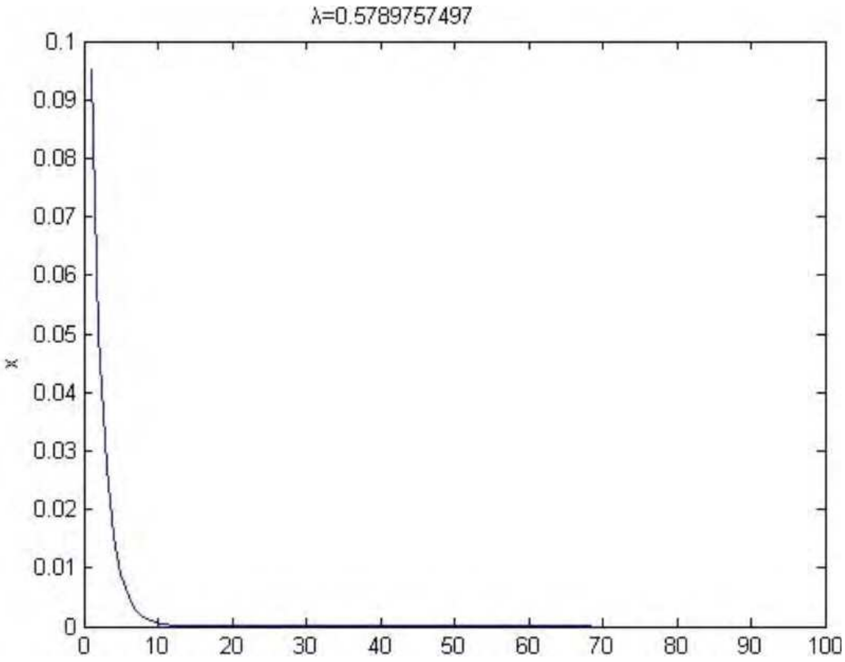


Fig. 1. Logistic experiment 1

When λ ∈(1,2) and where we have a random value for λ= 1.8438643285. As shown in Figure 2, in the case of 100 iterations, the value of x after 10 times is little changed. However, the data shows that if we take 17 decimal places after the decimal point for x, the top 15 are identical, but only the last two have subtle differences. And the following value of x became periodicity,(And always became periodicity,) these values are 0.45766074838406956, 0.45766074838406962, 0.45766074838406973. It is always these three numbers, and so the overall system does not appear to have the features of chaos.

When λ ∈(2,3) and where we have a random value for λ= 2.4829473982. As shown in Figure 3, it is a similar situation for λ ∈(1,2) when, after 10 iterations, the figure tends to be stable. The data shows that there are two numbers in circulation: 0.59725284525763811 and 0.59725284525763822, and the overall system does not appear to have the features of chaos.

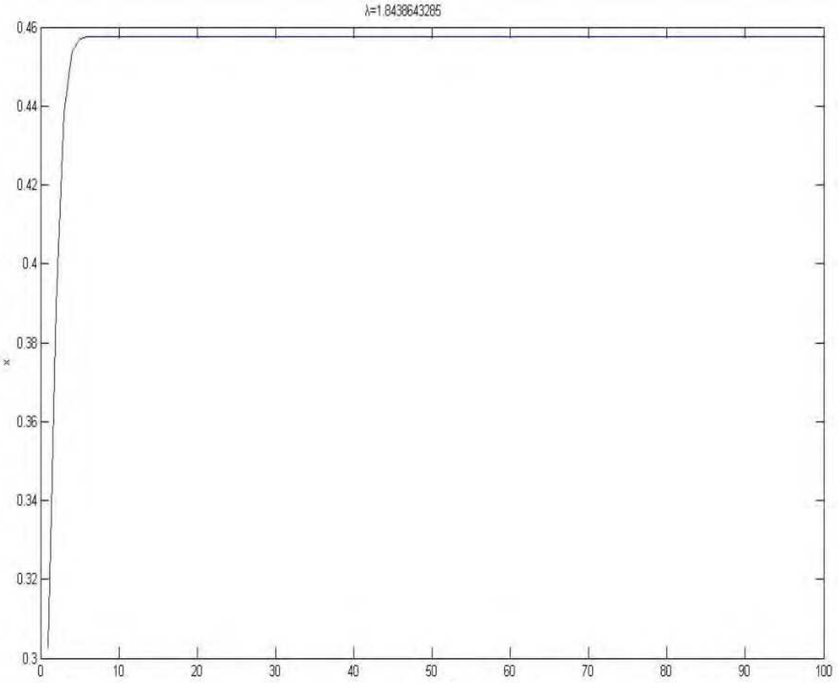


Fig. 2. Logistic experiments 2

When λ ∈(3,3.6) and where we have a random value for λ= 3.3483997432. It is iterated 100 times, as is shown in Figure 4: the value of x has relatively large fluctuations and becomes a discrete state. However, the data shows that although the value of x is volatile, it is still a circulation. Moreover, although this periodicity is not as obviou as the former two have, it still has some implications for encryption security.

When λ ∈(3.6,4) and where we have a random value for λ= 3.8374666542. The value of x after it is iterated 100 times is shown in Figure 5. We can see that the value of x has a more significant fluctuation. After analysis, it was shown that this result is not a circulation. As such, this system will be a chaotic system.

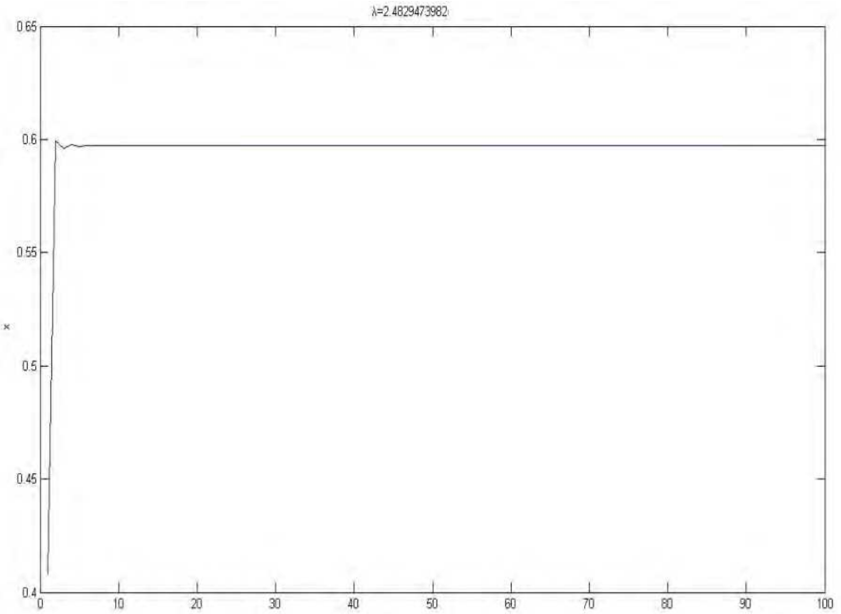


Fig. 3. Logistic experiments 3

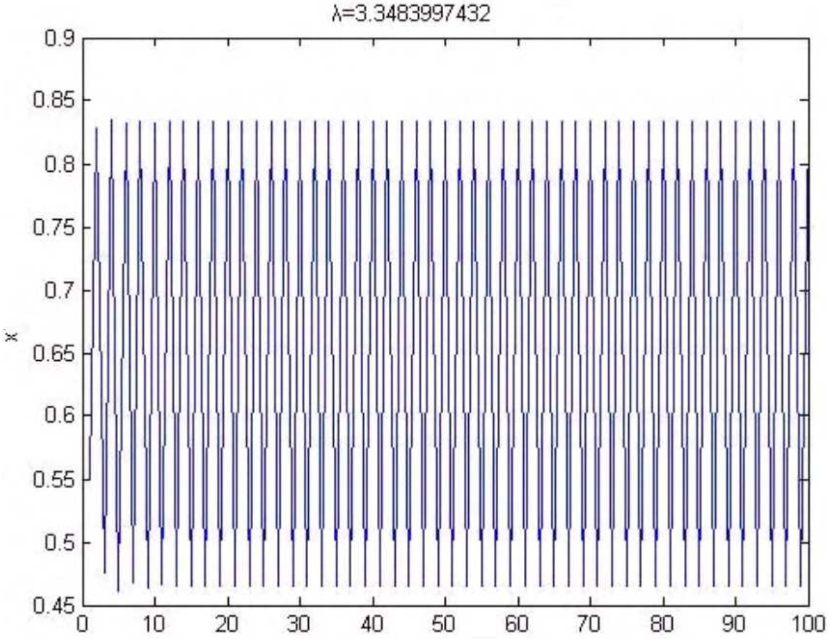


Fig. 4. Logistic experiments 4

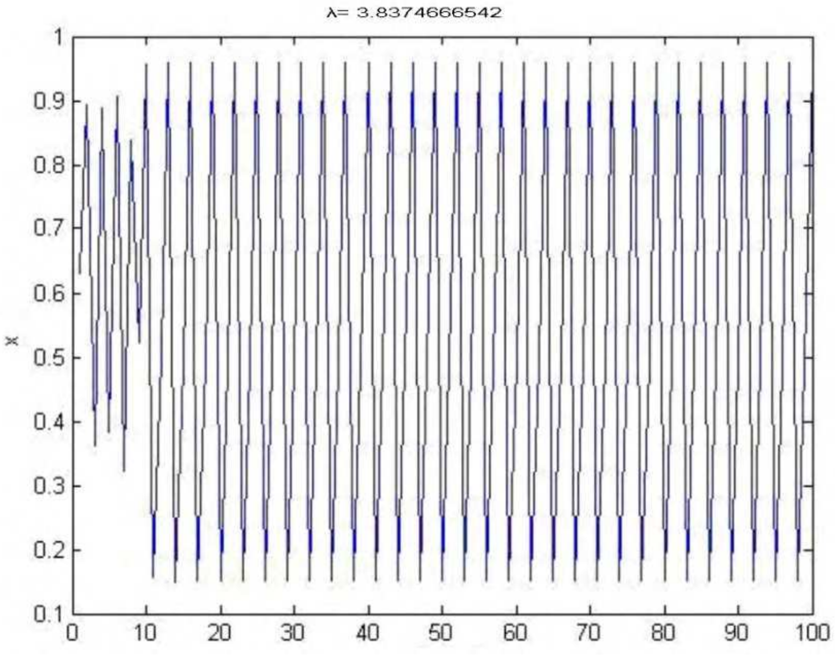


Fig. 5. Logistic experiments 5

5.2 The united chaos encryption algorithm based on logistic map and henon map

We can add a two-dimensional chaotic map in the circumstances that ensures that the efficiency is not too bad. This chaotic map is called a Henon map. We can use it to start encryption united with a Logistic map. Moreover, this can be achieved without losing efficiency while strengthening its security.

A Henon map as a two-dimensional chaotic map, and its equation is:

$$\begin{cases} X_{n+1} = 1 + Y_n - a \times X_n^2 \\ Y_{n+1} = b \times X_n \end{cases} \tag{2}$$

When using this map, we need to set initial values for x_0 and x_1 and the parameters *a* and *b*. The algorithm flow is shown in Figure 6.

This chaotic system is used mainly to generate a chaotic sequence of random numbers. It could have chaotic characteristics. The purpose of using this chaotic system is in the pre-treatment of the encrypted plaintext. The whole of the algorithm's flow of chaotic pre-processing is:

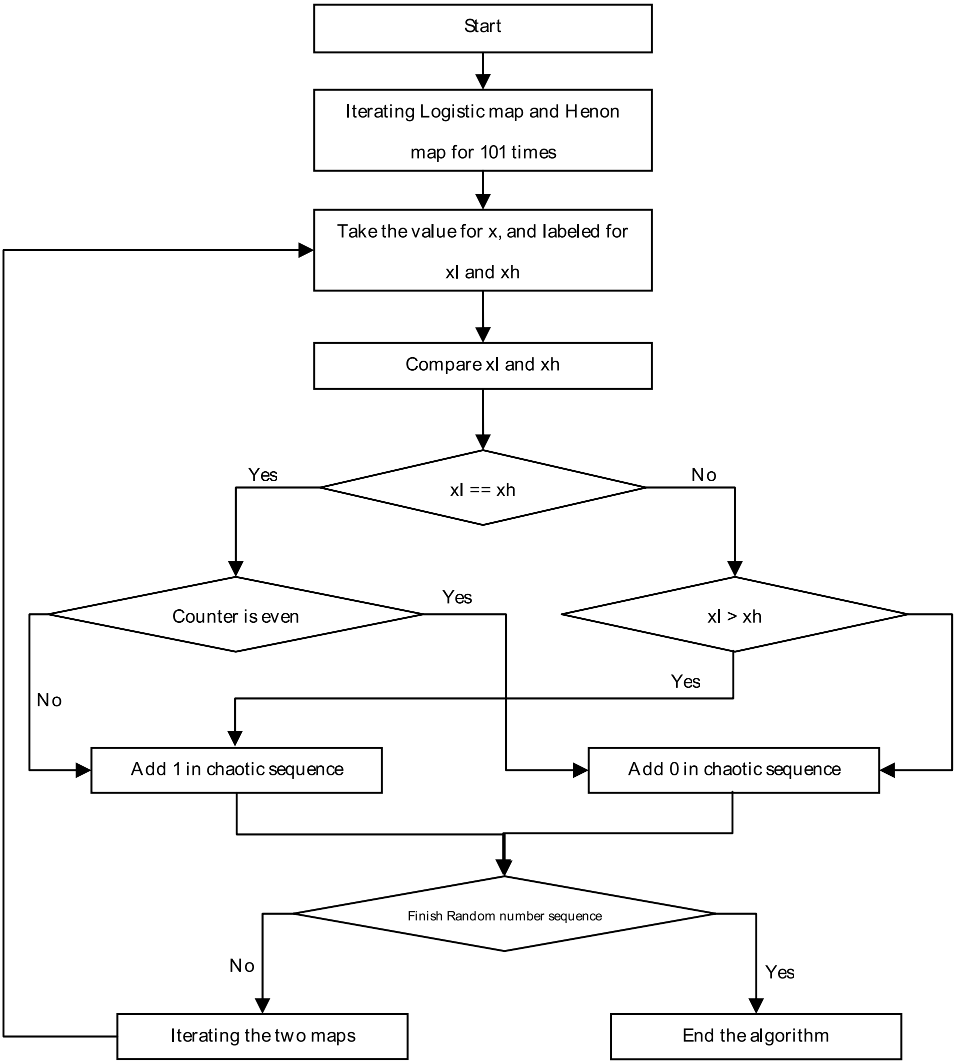

Start
Iterating Logistic map and Henon map for 101 times
Take the value for x, and labeled for xl and xh
Compare xl and xh
Yes
xl == xh
No
Counter is even
Yes
xl > xh
No
Yes
No
Add 1 in chaotic sequence
Add 0 in chaotic sequence
Finish Random number sequence
No
Yes
Iterating the two maps
End the algorithm

Fig. 6. The algorithm flow

1. Make an encoding conversion for the encrypted plaintext; transfer the ASCII code which corresponds to the plaintext character into n-bit binary code;
2. Use the n-bit pseudo-random number sequence which is produced by the chaotic system to conduct XOR with the plaintext's binary sequences. All of these sequences are 0, 1 sequences. Obtain the binary sequences after treatment;
3. Obtain the DNA chain by using the digital coding rules of DNA to transfer these binary sequences into a DNA base sequence.

The entire process shown in Figure 7:

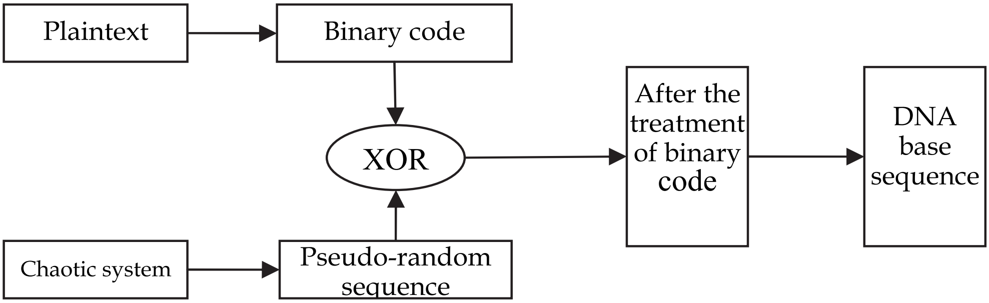
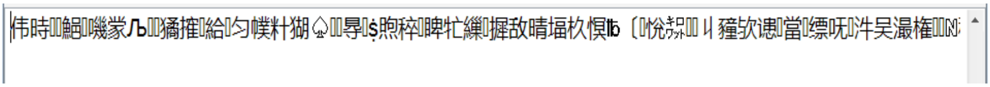


Fig. 7. XOR processing

5.3 Security verification

1. Key Analysis

In this encryption system, as a key, the initial values are xl_0 = 0.3, xh_0 = 0.5, xh_1 = 0.4 and the three parameters of the chaotic maps are λ = 3.8264775543, a = 1.3649226742, b = 0.3. The initial value range of these two parameters is (0, 1) and the value is a real number. The logistic map's parameter has a value in the range of (3.6, 4). In the two parameters of the Henon map, one is the fixed value for b=0.3, the other parameter we assume it to a. Moreover, its value range had better be in (1.07, 1.4), as this range can better reflect the characteristics of chaos. Sensitivity can be reflected in the key, and now we keep all of the parameters of the encryption system at a correct value, only changing λ = 3.8264775543 to λ = 3.8264775544 for the logistic map. We add 10^{-10}, which means that we only change a tenth of a decimal number. Next, we take this kind of key into the chaotic system in order to have it decrypted. The result is shown in Figure 8.



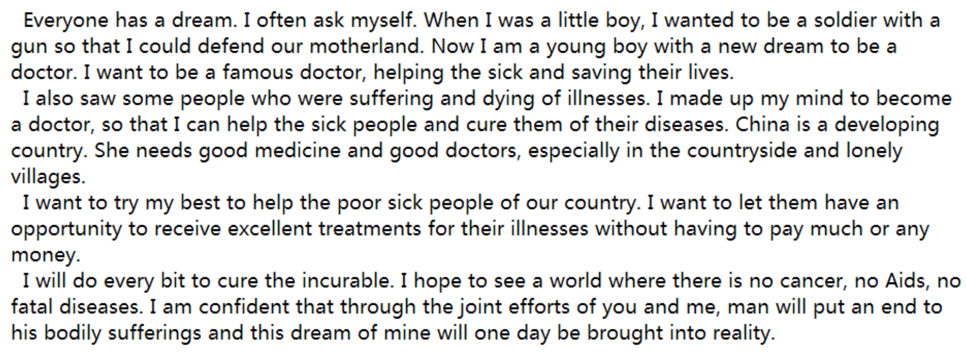
伟時□鯧□嘰蒙љ□□豨摧□給□匀業村瑚♤□□尋□ş煦稡□睥忙纝□握敌晴塩杦惧㏗〔□悅□□ 丬 癰欤谌□當□缥呒□汼吴灄権□□ℕ

Fig. 8. Decrypt results of the wrong key.

2. Statistical analysis

Generally the message of plaintext is text or other information and they all follow certain statistical laws, such as in English words the letters r, a, e, etc. have a high frequency of use, but letters q, z, u, etc. do not. It is a law of English words, and so it brings some security risk to the password. If the encrypted ciphertext still has the characteristics of these statistics, it is easy for statistical attacks. Next, we use encryption to analyse an English article -- Martin Luther King's speech "I have a dream."

The original is shown in Figure 9.



Everyone has a dream. I often ask myself. When I was a little boy, I wanted to be a soldier with a gun so that I could defend our motherland. Now I am a young boy with a new dream to be a doctor. I want to be a famous doctor, helping the sick and saving their lives.

I also saw some people who were suffering and dying of illnesses. I made up my mind to become a doctor, so that I can help the sick people and cure them of their diseases. China is a developing country. She needs good medicine and good doctors, especially in the countryside and lonely villages.

I want to try my best to help the poor sick people of our country. I want to let them have an opportunity to receive excellent treatments for their illnesses without having to pay much or any money.

I will do every bit to cure the incurable. I hope to see a world where there is no cancer, no Aids, no fatal diseases. I am confident that through the joint efforts of you and me, man will put an end to his bodily sufferings and this dream of mine will one day be brought into reality.

Fig. 9. Plaintext examples

We analyse this article, and add up the letters in terms of the number of their occurrence. As is shown in Figure 10, we found that the frequency of letters that appear in each word is not the same. The letter e occurs the most, the letter o is second, and so on. In this article, we cannot find the letters q and z. So, the statistical law is very clear and the cryptanalyst can make attack according to the number of times the characters appear in the ciphertext.

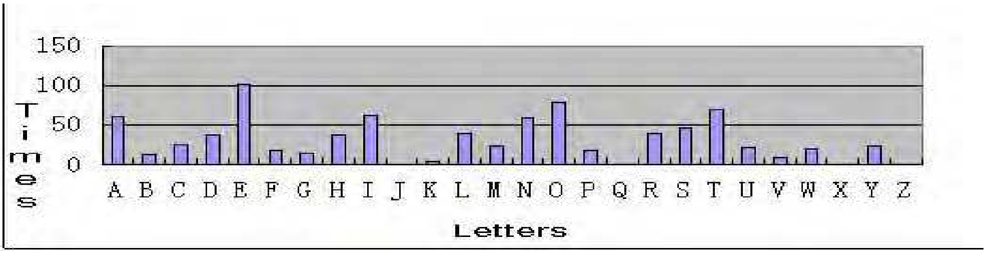
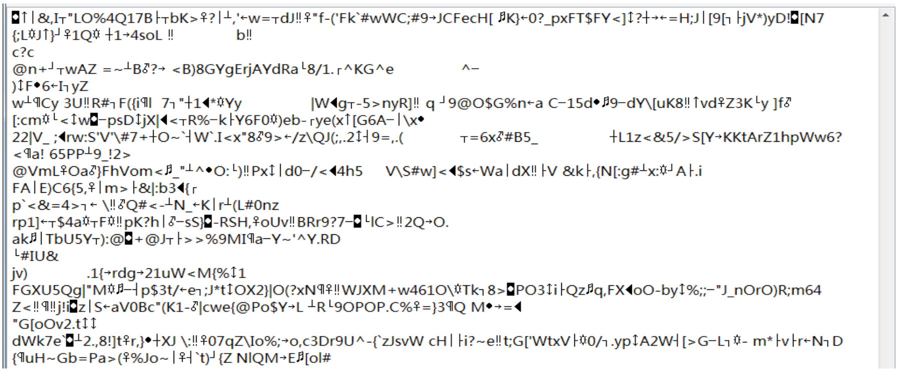


Fig. 10. Statistical laws of the letters in plaintext.

In this case, we use the chaotic system and its key to encrypt the article. The encrypted file is shown in Figure 11. The figure told us that after encryption the article - which is also called the ciphertext - has a lot of confusing characters. Equally, they do not have any statistical features: all of the characters are randomly distributed and they do not follow any law. So, this kind of encryption has the ability to avoid statistical attacks.



```
◘↑│&,I┬"LO%4Q17B├┬bK>♀?│┴,'←w=┬dJ‼♀"f-('Fk`#wWC;#9→JCFecH[ ♫K}←0?_pxFT$FY<]↕?┼→←=H;J│[9[┐├jV*)yD!◘[N7
{;L☼J↑}┘♀1Q☼ ┼1→4soL ‼              b‼
c?c
@n+┘┬wAZ =~┴B♂?→ <B)8GYgErjAYdRa└8/1.┌^KG^e              ^–
)↕F•6←I┐yZ
w┴¶Cy 3U‼R#┐F({i¶I  7┐"┼1◄*☼Yy              |W◄g┬-5>nyR]‼ q ┘9@O$G%n←a C–15d•♫9–dY\[uK8‼↑vd♀Z3K└y ]f♂
[:cm☼└<↕w◘–psD↕jX|◄<┬R%–k├Y6F0☼)eb-┌ye(x↑[G6A–│\x•
22|V_ ;◄rw:S'V'\#7+┼O~`┤W`.I<x"8♂9>←/z\QJ(;,.2↕┤9=,.(            ┬=6x♂#B5_             ┼L1z<&5/>S[Y→KKtArZ1hpWw6?
<¶a! 65PP┴9_!2>
@VmL♀Oa♂}FhVom<♫_"┴^•O:└)‼Px↕│d0–/<◄4h5     V\S#w]<◄$s←Wa│dX‼├V &k├,{N[:g#┴x:☼┘A├.i
FA│E)C6{5,♀│m>├&|:b3◄{┌
p`<&=4>┐← \‼♂Q#<-┴N_←K│r┴(L#0nz
rp1]←┬$4a☼┬F☼‼pK?h│♂–sS}◘-RSH,♀oUv‼BRr9?7–◘└IC>‼2Q→O.
ak♫│TbU5Y┬):@◘+@J┬├>>%9MI¶a–Y~'^Y.RD
└#IU&
jv)           .1{→rdg→21uW<M{%↕1
FGXU5Qg|"M☼♫–┤p$3t/←e┐;J*t↕OX2}|O(?xN¶♀‼WJXM+w461O\☼Tk┐8>◘PO3↕i├Qz♫q,FX◄oO-by↕%;;–"J_nOrO)R;m64
Z<‼¶‼j!i◘z│S←aV0Bc"(K1-♂|cwe{@Po$Y→L ┴R└9OPOP.C%♀=}3¶Q M•→=◄
"G[oOv2.t↕↕
dWk7e`◘┴2.,8!]t♀r,}•┼XJ \:‼♀07qZ\Io%;→o,c3Dr9U^-{`zJsvW cH│├i?~e‼t;G['WtxV├☼0/┐.yp↕A2W┤[>G–L┐☼- m*├v├r←N┐D
{¶uH~Gb=Pa>(♀%Jo~│♀┤`t)┘{Z NIQM→E♫[ol#
```

Fig. 11. Example of ciphertext.

6. A new cryptographic algorithms based on PCR and chaos optimisation

6.1 Encryption system design

6.1.1 Key generation

In this encryption system, we use the united keys instead of a single key. The key is divided into two parts: the first part is a PCR technique used in the primers, with the primer sequences as a key - KeyA; The second part concerns the initial conditions and parameters which are used in the chaotic system, and the system is called KeyB.

The password system is the most important which relies on bio-security. As such, the DNA code of the key has the requirement of high quality. However, in the united key, key KeyB is related with the DNA code. For the generation of KeyA, KeyA is a string of bases of the DNA sequence, which is used for the PCR amplification primers. Password security and systems can be realised, which is determined by the success of the primer design system. Accordingly, the design of this key is very important. If the key is designed strictly according to the design principles of the design primer, it will cause limited limitation of primer shortage space. Therefore, the primer design of the encryption system is designed by software Primer Premier 5.0, which is used in biological simulation.

The design shown in Figure 12:

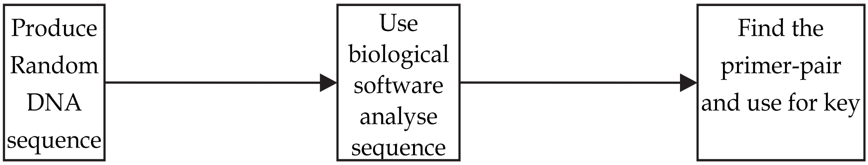

Produce Random DNA sequence
Use biological software analyse sequence
Find the primer-pair and use for key

Fig. 12. Key preparation processes

For the production of KeyB, we select the appropriate parameters in the chaotic system as keys. The parameter selection rules have been talked about in the preamble, so it need not be repeated. For the median of the parameters selected, this can be based on the security of encryption strength in order to develop the key's length.

6.1.2 Encryption process

The message sender is also called the encrypter: after completing the key design it begins to encrypt the plaintext and makes a ciphertext.

1. Explicating that which is converted into binary code;
2. Using the DNA encoding rule pre-treatment the binary code for chaos;
3. Bringing KeyB into the chaotic system to produce the chaotic pseudo-random number sequence;
4. Operating the sequence and the plaintext sequence corresponding to the binary by XOR so as obtain the processed binary sequence.

This binary sequence is divided into n sub-sequences and the specific number is decided by the length of the ciphertext. The pair sequence is numbered l_1, $l_2 \ldots l_n$ and is followed by the following operations:

$$l_1 \oplus l_2 = s_{2,}$$

$$s_2 \oplus l_3 = s_3$$

$$\ldots$$

$$s_{n-1} \oplus l_n = s_n$$

Get s_2, s_3, …, s_n n-1 sequences and then l_1, s_2, s_3, …, s_n, and its subscript number of these sequences. The sequences were added to each sequence at the beginning. Next, the sequence was transformed into a DNA base sequence according to DNA coding. The coding rules are 0123/CTAG (it has been illustrated in the fourth part of this chapter). Afterwards, select the stand-n-primer from that obtained in the previous primer sequence step added to the front of the sequence. The ciphertext sequence propagated successfully. It is shown in Figure 13.

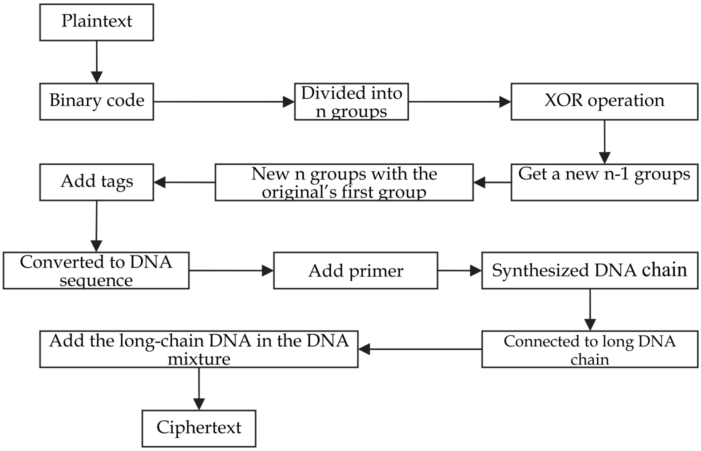

Plaintext
Binary code
Divided into n groups
XOR operation
Get a new n-1 groups
New n groups with the original's first group
Add tags
Converted to DNA sequence
Add primer
Synthesized DNA chain
Connected to long DNA chain
Add the long-chain DNA in the DNA mixture
Ciphertext

Fig. 13. Encryption Process

The use of biological experimental techniques - using mainly artificial DNA synthesis technology - see the formation of DNA sequences into short-chain DNA synthesis. Next, using cutting and splicing, the DNA technology is used to make short-chain n-DNA, splicing into a long DNA template chain. We complete this long-chain DNA system and add it to the DNA mixture. In the DNA mixture there are many different lengths of DNA, such as interference DNA. The ciphertext is thereby produced.

6.1.3 Decryption process

First, the cracker has to get KeyA using key information that is obtained from safe prior sources and then carry out PCR amplification. For the second step, the DNA to be amplified will be selected by using electrophorus and these DNA have the information we need. For the third step, through the sequencing of the DNA chain, we can draw the corresponding DNA sequence. For the fourth step, the DNA sequence was restored to a binary sequence by the DNA encoding. At this time, the obtained binary sequence is l_1 , s_2 , s_3 , … , s_n in the encrypted process. After sorting it is then calculated:

$$s_{n-1} \oplus s_n = l_n$$

$$\cdots$$

$$s_2 \oplus s_3 = l_3$$

$$l_1 \oplus s_2 = l_2$$

We can get l_1 , $l_2 \ldots l_n$. For the fifth step, the binary sequences are spliced together, and we can get a sequence that is a clear binary sequence after the sequence of the pre-treated. For the sixth step - the building of the chaotic system - we bring the parameters of KeyB into the chaotic system. After these operations, we can obtain a binary sequence corresponding to the plaintext. For the seventh step, through transcending and the restoration of the character data, we can get clear.

The entire process is shown in Figure 14.

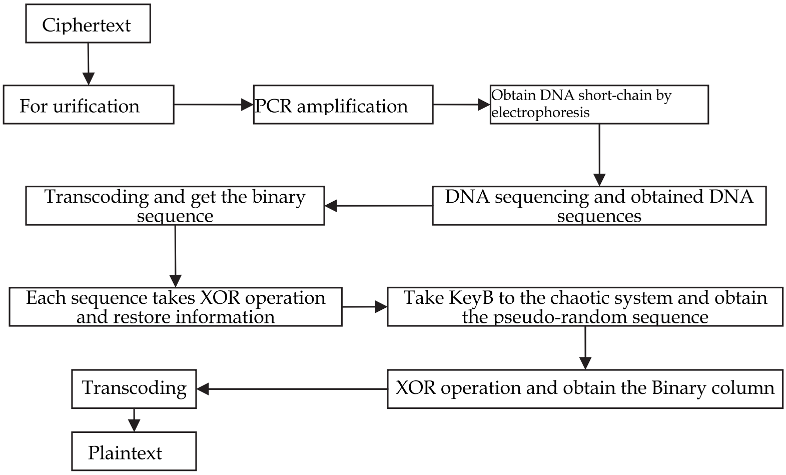

Ciphertext
For urification
PCR amplification
Obtain DNA short-chain by electrophoresis
Transcoding and get the binary sequence
DNA sequencing and obtained DNA sequences
Each sequence takes XOR operation and restore information
Take KeyB to the chaotic system and obtain the pseudo-random sequence
Transcoding
XOR operation and obtain the Binary column
Plaintext

Fig. 14. Decryption Process

Now, the information transmission process is over. When the sender sends a successful message, the receiver will get safe information and they will get plaintext.

7. Analysis of cryptographic algorithms

7.1 Key space

The size of the key space is very important for the security of the encryption system. A good cryptographic algorithm should have enough key space to ensure its safety. The traditional encryption schemes of the PCR amplification technology of encrypted DNA witnesses a problem where it does not have enough key space. As such, we present three ways for improving the security problem of the system.

1. Using a method for combining PCR technology and chaos technology.
2. If we do not know the correct primers, we cannot start PCR amplification and, at the same, we cannot obtain the DNA which has the plaintext information. This is the feature of encryption system we described above and on security issues this method will be more stronger than others.
3. This encryption system is a common encryption system for the combination of DNA code and chaotic encryption. Here, we use a chaotic system to pre-treat the plaintext.

This encryption system has the three above features and it can adjust to the size of the entire key space dynamically, and especially to the key of the adjusted DNA code.

7.2 Features and benefits of the system

In this system, we use chaotic encryption for encryption systems dealing with plaintext. This encrypted system eliminates the statistic rules in plaintext and loads chaotic encryption into DNA code. This means that the DNA code has the same advantages that traditional encryption has. As such, security has been improved. Even if the attacker deciphered the DNA code, he will still face a lot of chaos code that it would be necessary to decrypt. This increases the difficulty of decryption. In order to be a new type of encryption system, DNA code is based on a different security to the traditional code. Accordingly, we can obtain a complementary effect when we combined these two systems.

8. Conclusion

This paper mainly discusses DNA Cryptography and one example algorithm, analysing the encryption algorithm of the PCR-based amplification technology of DNA, improving security and the key space, and it provides an operational test of it. In order to solve the key space-constrained problem that the PCR amplification technology of DNA has, the authors used a method for building a chaotic system. This system includes a logistic chaotic map and a Henon chaotic map. We can generate a chaotic pseudo-random sequence which could handle the plaintext for eliminating the statistical rules in it with the two maps. On the one hand, it makes the encryption algorithm immune of statistical attack. On the other hand, it increases the key space. After using the binary code of the message of plaintext to make an XOR operation, we can obtain a new binary code. We can then ensure an increase in the number of primers, and we add some primers to it; this is one of the primers' features. After all of this, we have increased the security of the entire system.

In addition, during the PCR amplification experiment, if the amplified target DNA is too long, it may lead to a failure of the amplification. In this encryption algorithm, we separate the binary code of the plaintext into many small sequences. In this manner, we guarantee that the amplification could be carried out smoothly during its operation.

This chapter used the encryption instance to describe all of the encryption algorithm. Moreover, we have analysed each encryption effect. Finally, we analysed the security and operability of the entire system, and used biology software to demonstrate the bio-security of the analogue of the amplification primers, using computer to analyse the statistics and demonstrate the effect of the chaotic system.

9. Acknowledgements

This work is supported by the Aero-Science Fund of China (2009ZD53045), Science and Technology Development Project of Shaanxi Province Project (2010K06-22g), Basic research fund of Northwestern Polytechnical University (GAKY100101), and the R Fund of College of Software and Microelectronics of Northwestern Polytechnical University (2010R001), Xi'an science and technology plan (CXY1118).

10. References

Leier A et al. Cryptography with DNA binary strands [J]. Biosystems, 2000, 57(1): 13-22.

Beenish Anam et al. "Review on the Advancements of DNA Cryptography", eprint arXiv:1010.0186, 10/2010

Cui G et al. DNA computing and its application to information security field [C]. IEEE Fifth International Conference on Natural Computation, Tianjian, China, Aug. 2009.

Xiong Fuqin, Cryptography Technology and Application [J]. Science, 2010.

Luque G et al. Metaheuristics for the DNA Fragment Assembly Problem. International Journal of Computational Intelligence Research, 2005, 1(2), 98–108.

Hayashi et al. Anonymity on paillier's trap-door permutation[C]. Springer Verlag, 2007 , 200-214.

Huo J-J et al. Encoding Technique of DNA Cryptography [J]. Information Security and Communications Privacy, 2009, 7: 90-92.

Chen J. A DNA-based, biomolecular cryptography design [J]. ISCAS, 2003, 3:822-825.

Adleman L, Molecular computation of solutions to combinatorial problems [J]. Science, 1994, 266: 1021-1024.

Limin Qin. The Study of DNA - Based Encryption Method [D]. Zheng Zhou: Zheng Zhou University of Light Industry, 2008.)

Borda M. & Tornea O. DNA secret writing techniques [C]. In COMM(2010), Chengdu: IEEE, June 10-12, 2010: 451-456.

C Popovici. Aspects of DNA Cryptography [J]. Annals of the University of Craiova Mathematics and Computer Science Series, 2010, 37(3).

Limin Qin. The Study of DNA - Based Encryption Method [D]. Zheng Zhou: Zheng Zhou University of Light Industry, 2008.

Kazuo T, Akimitsu O, Isao S. Public-key system using DNA as a one-way function for key distribution[J]. Biosystems, 2005, 81: 25-29.

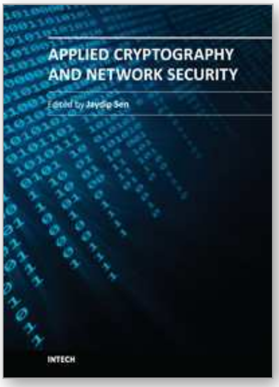
Celland C T et al. Hiding messages in DNA microdots [J]. Nature, 1999, 399: 533-534.

Xing-Yuan Wang et al. A chaotic image encryption algorithm based on perceptronmodel [J]. Nonlinear Dyn, 2010, 62: 615-621.

Luo Ming Xin et al. A Symmetric Encryption Method Based On DNA Technology [J]. Science in China (Series E:Information Sciences),2007,37(2): 175-182.

http://baike.baidu.com/view/107254.htm, 2011.7

http://baike.baidu.com/view/25110.htm, 2011.7



ISBN 978-953-51-0218-2
Hard cover, 376 pages
Publisher InTech
Published online 14, March, 2012
Published in print edition March, 2012

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821