Journal of Theoretical and Applied Information Technology
15th November 2012. Vol. 45 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NEW DNA CRYPTOGRAM SCHEME BASED ON PCR TECHNOLOGY

[1, 2]YUNPENG ZHANG, [1]DONGWEI ZHOU, [1]LIU HE,
[2, 3]YASIN HASAN KARANFIL, [1]BOCHEN FU

[1]School of Software and Microelectronics, Northwestern Polytechnical University, Xi'an, 710072, China

[2]Imperial College London, London, SW3 6NP, UK

[3]University of Wales, Cardiff, CF10 3NS, UK

ABSTRACT

In the condition of the traditional cryptogram which has been breached successively, this paper mainly studies the current cryptography-DNA cryptogram. Using biology question which has none solution in "if we don't know the correct primer situations from an unknown mixture of trying to isolate specific DNA of DNA is quite difficult" as the safety of constructing DNA encryption algorithm. Using the chaotic system which is constructed by using Logistic chaos mapping and Henon chaos mapping to produce the pseudo-random to process the plaintext and eliminate the statistical rule. Building the encryption algorithm, and analyzing the DNA encryption algorithm based on PCR technology. And improve the algorithm in the key space, security, and experiment. Doing exclusive or operation between the binary code of plaintext, we can get the new code, and then append the primers. By this way, we can increase the number of primers, and the attackers can never get the plaintext unless he get all DNA chain with information. In addition, it will lead the failure if the target DNA chain is too long when doing the PCR amplification. In this encryption algorithm, the PCR amplification can be successfully by dividing the binary code of plaintext into many short sequences. This paper uses encryption examples to describe the whole encryption algorithm, and then analyses the security and maneuverability of the whole system. And then we get the whole system.

Keywords: *Information Security, DNA Cryptogram, Chaos, PCR*

1. INTRODUCTION

Cryptography is a science and a technology which study encryption and code-breaking[1]. In recent years, with the study of DNA computing research, DNA cryptography, a new area of cryptography appear. Originally there are no connections between Cryptography and Molecular Biology (also known as genetics or genomics). But with the study of DNA, especially after Adleman put forward the DNA computing in 1994, we have had a in depth study of DNA computing[2,3]. And after this research has been using into the field of information security the DNA cryptography finally appeared[4]. DNA cryptography is built on the information carrier of DNA.

And we achieve the encryption process by using the modern biotechnology tools and the characteristics of massively parallelism and high storage density of DNA. The most important is that DNA cryptography use the limited nature of learning biology and abandon the traditional cryptography witch use the intractable of mathematical problem as the security guarantee. In theory, DNA cryptography is mainly based on the security of biology technique's limitations, and has no connection with the computing ability. This kind of cryptography can avoid the attack from both modern computer and even quantum computers in the future. Therefore, several groups have already been started to study the better encrypted effect of the DNA cryptography[5].

2. ANALYSIS DNA ENCRYPTION WHICH BASED ON THE PCR AMPLIFICATION

2.1 Pcr Technology

PCR Technology is one of the rapid Amplification Technology of DNA, and it is also called Polymerase Chain Reaction. The PCR Technology usually used to amplify the determined DNA because the difficulty of manipulation the small amounts of DNA. The PCR has a high efficiency of amplification, and can amplify a large number of targets DNA in a short time. In order to achieve PCR amplification, the experimenters

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

needs to know the sequences of the chain of target DNA, and then use it to design the primers for amplification. Actually, primer is also a DNA sequence which contains a number of nucleotides. It is clearly that the primers can amplify target DNA. In short, the PCR process can be divided into two stages: 1) design two primers, and loaded for target DNA in the beginning and the ending; 2) find target DNA with the use of polymerase and then amplification it.

2.2 DNA Encryption Based On The PCR Technology

2.2.1 Encryption process

If someone (The authors call she the encrypter in the paper) wants to encrypt the plaintext, first she/he needs to transform the plaintext into DNA sequence by using the code rules, and now the Base sequence of DNA represented a special meaning. Then with DNA sequences we use the measure of biotechnology to synthesis DNA chain as the target DNA artificially. After that, we can design the appropriate primers as a key. When the sender obtains the key, they will load them for a target DNA on its beginning and ending. Base on that, we use the cutting and splicing technology of DNA to implantation this DNA into a long DNA chain. Finally, adding some confusion DNA chain which is common DNA chain and contains any meaning.

2.3 Analysis Of DNA Encryption Based On The PCR Technology

2.3.1 safety analysis

When the code-breakers intercepted the ciphertext, what she/he had obtained is a DNA mixture and has a lot of confusion DNA chain in it. This is because that the ciphertext use DNA chain for carrier and its message will represent by the Base sequence of DNA chain. As the character of PCR Technology, it needs a high requirement of the correctness of primer sequence. If starting amplification experiment, it is impossible to try to find out the target gene without knowing primer sequence [6]. In this case, it is because if we designed the primer by ourselves. Firstly, we don’t know the molecule length of the correct primer. The different length we have, the wrong message we get. Secondly even if the length is right, we suppose that there are 25 Base sequences and it is easy to know that there will be 4^{25} kinds of primers in theory. If experiment them one by one, we assume that take one PCR amplification needs 2 or 3 hours. Then we need 10^{27} years to finish it. This is impossible.

But only use the DNA Encryption which based on the PCR Technology is not always safety. That is because the plaintext and the converted DNA is one to one relationship, and this make the ciphertext contains the unique statistical properties which only plaintext has. In this case the cryptanalyst can decipher it though statistical attacks, and bring a security risk to cryptography.

2.3.2 Feasibility Analysis of Experimental Operation

Primers designed must comply with the following principles:

1) Specific

Primers should be arranged in specific, especially with the amplify target sequences between two primers, and make sure of at least 30% different and can not be arranged in a continuous 8 Bases are the same;

2) Length

Statistical calculations indicate that the 17 base sequence in human DNA is likely to occur in 1 time, so the primer length general controlled more than 17, but cannot unlimited long, at maximum for less than 30 Bases sequence. Usually the best length is 20 to 24 Bases. This length of DNA primer has a strong stability when reaction, and does not produce hybrids;

3) The content of C and G Bases

The content of C + G needs to control in 40% to 60%, and to avoid contains too many Bases Polymers, and the percentage of the C + G in the two primers should be similar;

4) Random Distribution of Bases

Bases distribution in the primer should be random, but to avoid more than three consecutive identical bases;

5) Primer Itself

The complementary sequence should not appear in the primer sequence itself, if it cannot avoid, ensure that there at least less than 3 Bases complementary situation;

6) Between the Primers

Each primer should be avoided to appear the complementary sequence, as the principles, if it cannot avoid, ensure that there at least less than 3 Bases complementary situation;

7) The End of Primer 3’

Trying not used Base A in 3' end, because A has a high rate of mismatch. And cannot add any modification in 3' end;

8) The End of Primer 5'

5' end of the primer is limit the length of PCR amplification product, but has less demanding, some fluorescent markings can be modified.

As the PCR primer design is a crucial part of technology. And the use of PCR technology is the core of this encryption algorithm, and also for its safety and security conditions. If use the inappropriate PCR primers will lead the experimental to failure easily. Therefore, in designed of the primers must comply with the above principles. Here we can use the biological expertise software to help designed primers. The software called -- Primer Premier 5.0.

3. THE UNITED CHAOS ENCRYPTION ALGORITHM OF LOGISTIC MAP AND HENON MAP – BASED

3.1 Research For Logistic Map

Logistic map is most widely used in chaotic map. It is one-dimensional chaotic map with high efficiency and simplicity advantages. Logistic map is defined as:

$$x_{n+1} = \lambda \times x_n \times (1 - x_n), \quad \lambda \in (0,4), n = 1,2,\cdots \tag{3-1}$$

We use Parameter λ and the initial value x0 as a key. Parameter λ can be divided into three parts, and start parameter validation. Make x0 equal to a random value of 0.79284, and then take the above data into the formula 3-1, let it iteration 100 times. And make a picture to analysis for each x. We have the situations is: when λ∈ (0, 3.6), if we give a random value of λ. Then we iterative it for 100 times, the value is shown that it is not a chaos system. But when λ∈ (3.6, 4), and we have a random value for λ= 3.8374666542. The value of x after iterative 100 times is shown in Figure 1. We can see that the value of x has a more significant fluctuation. After analysis, it is shown that this result is not a circulation. So this system will be a chaotic system [7].

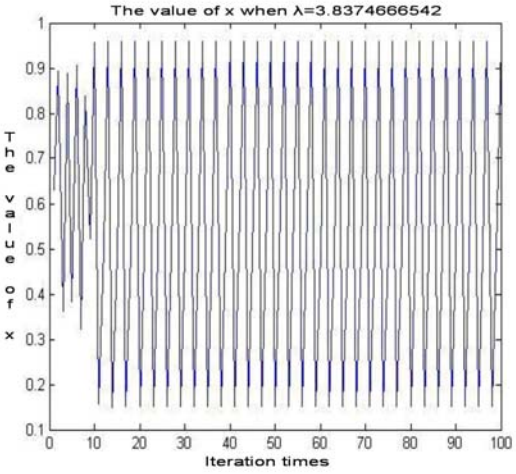


Figure 1: Logistic Experiment

3.2 The United Chaos Encryption Algorithm Of Logistic Map And Henon Map

We can add a two-dimensional chaotic map under the situation that to ensure the efficiency is good. This chaotic map is called Henon map. We can use it to start encryption united whit Logistic map. And this can be achieved without losing efficiency, while strengthening its security.

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

$$\begin{cases} X_{n+1} = 1 + Y_n - a \times X_n^2 \\ Y_{n+1} = a \times X_n \end{cases} \tag{3-2}$$

When use this map, we need to set initial values x0 and x1, and the parameters a and b. And the algorithm flow is shown in Figure 2.

This chaotic system is mainly to generate a chaotic sequence of random numbers. It could have characteristics of chaotic. The purpose we use this chaotic system is to pretreatment the encrypted plaintext. The whole algorithms flow of chaotic preprocessing is:

1) Make a encoding conversion for the encrypted plaintext, transfer the ASC II code which corresponding the plaintext character into n-bit binary code;

2) Use the n-bit pseudo-random number sequence which produced by the chaotic system to conduct XOR with plaintext's binary sequences. All these sequences is 0,1 sequences. And obtain the binary sequences after treatment;

3) Obtain the DNA chain by using digital coding rules of DNA to transfer these binary sequences into DNA base sequence.

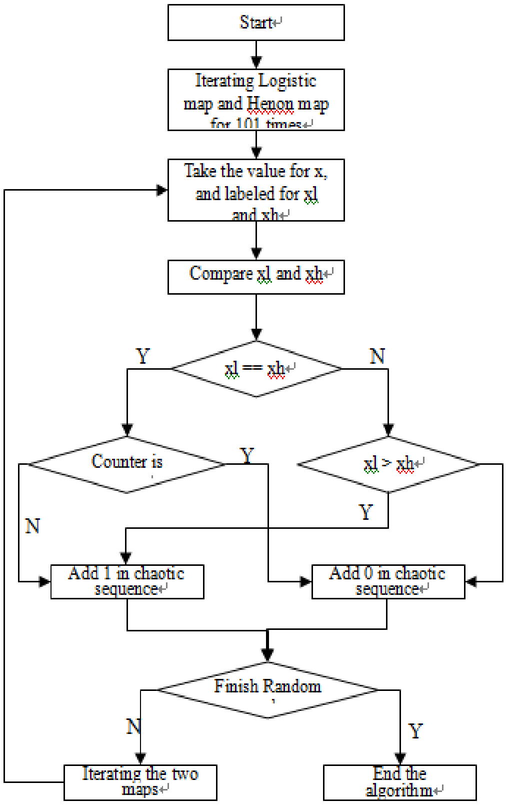

Start
Iterating Logistic map and Henon map for 101 times
Take the value for x, and labeled for xl and xh
Compare xl and xh
xl == xh
Y
N
Counter is
Y
N
xl > xh
Y
N
Add 1 in chaotic sequence
Add 0 in chaotic sequence
Finish Random
N
Y
Iterating the two maps
End the algorithm

Figure 2: The Algorithm Flow

3.3 Security Verification

•Key Analysis

In this encryption system, we give xl0 = 0.3, xh0 = 0.5, xh1 = 0.4 and the two parameters of chaotic maps λ = 3.8264775543, a = 1.3649226742, b = 0.3. The initial value range of this two parameters is (0, 1), and the value is real numbers. Logistic map parameter has a value in range of (3.6, 4). In the two parameters of Henon map, one is the fixed value for b=0.3, the other parameter we assume it to a. And its value range will better be in (1.07, 1.4), so by this we can reflect a better characteristic of chaos. Sensitivity can be reflected in the key, now in encrypt system keep all of the parameters have a correct value, only change the λ = 3.8264775543 to λ = 3.8264775544 of the Logistic map. Add 10-10, it is means that only change the tenth number after the decimal. And take this kind of key into the chaotic system to have decrypted. The result is shown in Figure 3.

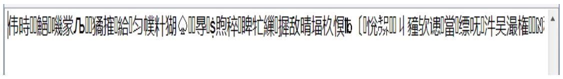
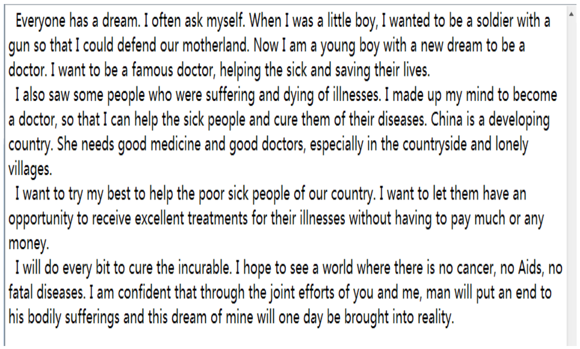

伟時鯧嘰蒙ЛЬ獝摧給匀噗籵猢♤孶s煦稡睥忙繅握敌晴塥杦惧℔（悅ｷﾞｯﾄ 丬 蘊欤谑當缥呒汼吴濊榷

Figure 3: Decrypt Results Of Wrong Key

• Statistical analysis

Generally the message of plaintext is text or other information, and they have a certain statistical laws, such as in the English words, letter r, a, e, etc there are have a high frequency to use, but letter q, z, u, etc are low because of the law of English words. So it bring some security risk to password. If the encrypted ciphertext still include the characteristics of these statistics, it is easy to have the statistical attacks. Now we use encrypted to analysis an English article -- Martin Luther King's speech "I have a dream". The original is shown in Figure 4.


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.

Figure 4: Plaintext Examples

We analysis this article, and statistics the letters on the number of occurrences. As is shown in Figure 5, we found that the frequency of letters that appear in each word is not the same. First is letter e, and second is letter o. And in this article we cannot find letter q and z. So the statistical regular is very clear to see, and the cryptanalyst can make attack by the times of the number of the character appear in ciphertext.

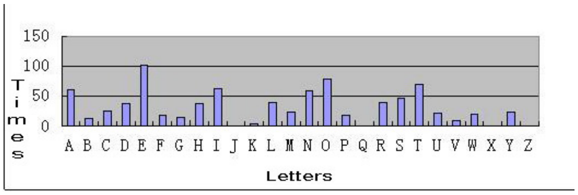
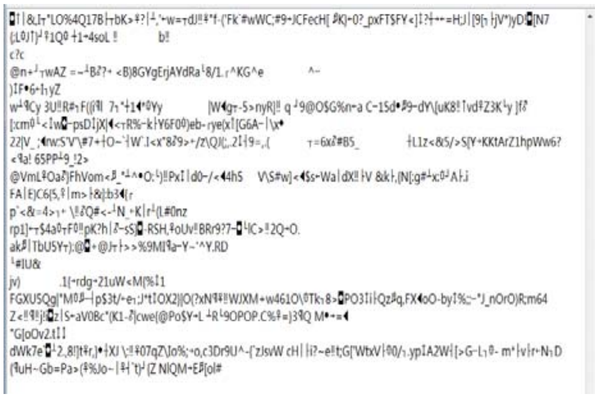


Figure 5: Statistical Law Of The Letters In Plaintext

In this case, we use the chaotic system and its key to encryption the article. And the encrypted file is shown in Figure 6. The figure shown that the article which also called ciphertext include a lot of confused characters after encrypted,. And they does not have any statistics features, all the characters were randomly and there is no regular. So this kind of encrypt have the ability to avoid statistical attacking.



```
▮↑|&Jr"LO%4Q17B├bK>‡?|┴,'÷w=┬dJ‼‡"f-('Fk'#wWC;#9÷JCFecH[ ßK)÷0?_pxFT$FY<]↕?╪÷÷=H;J|[9[┐├jV")yD▮[N7
(;L0J↑)┘‡1Q0 ┼1÷4soL !          b!
c?c
@n+┘┬wAZ =÷┴B♂?÷ <B)8GYgErjAYdRa└8/1.r^KG^e         ^÷
)↕F•6÷┐yZ
w┴¶Cy 3U‼R#┐F((¶l  7┐"┼1◄"0Yy          |W◄g┬-5>nyR]‼ q ┘9@O$G%n÷a C÷15d•ß9÷dY\[uK8!↑vd‡Z3K└y ]f♂
[:cm0└<↕w▮÷psD↕jX|◄<┬R%÷k├Y6F00)eb-rye(x↕[G6A÷|\x•
22|V_;◄rw:S'V'\#7+┤O÷'┤W'.1<x"8♂9>+/z\QJ(;.2↕┤9=,.(        ┬=6x♂#B5_          ┼L1z<&5/>S[Y+KKtArZ1hpWw6?
<¶a! 65PP┴9_!2>
@VmL‡Oa♂)FhVom<ß_*┴^•O:└)‼Px↕|d0÷/<◄4h5    V\S#w]<◄$s÷Wa|dX‼├V &k├,(N[:g#┴x:0┘A├.i
FA|E)C6(5,‡|m>├&┤b3◄(r
p'<&=4>┐+ \‼♂Q#<÷┘N_+K|r┘(L#0nz
rp1]÷┬$4a0┬F0‼pK?h|♂÷sS▮-RSH,‡oUv!BRr9?7÷▮└IC>‼2Q÷O.
akß|TbU5Y┬):@▮+@J┬├>>%9MI¶a÷Y÷'^Y.RD
└#IU&
jv)          .1(÷rdg÷21uW<M(%↕1
FGXU5Qg|"M0ß÷┤p$3t/÷e┐;J*t↕OX2||O(?xN¶‡‼WJXM+w461O\0Tk┐8>▮PO3↕i├Qzßq,FX◄oO-by↕%;÷"J_nOrO)R;m64
Z<‼¶‼j!▮z|S÷aV0Bc"(K1-♂]cwe[@Po$Y÷L ┴R└9OPOP.C%‡=)3¶Q M•÷=◄
"G[oOv2.t↕↕
dWk7e▮┴2,8!]t‡r,]•┤XJ \:‼‡07qZ\Jo%;÷o,c3Dr9U^-('zJsvW cH||i?~e‼t;G['WtxV├0/┐.ypIA2W┤[>G÷L┐0- m*├v├r÷N┐D
(¶uH~Gb=Pa>(‡%Jo÷|‡┤'t)┘(Z NIQM÷Eß[ol#
```

Figure 6: Example Of Ciphertext

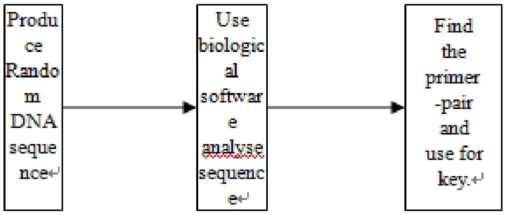
4. A NEW DNA CRYPTOGRAPHY RESEARCH BASED PCR AND CHAOS OPTIMIZATION

4.1 Encryption System Design

4.1.1 Key generation

n this encrypt system we use united keys instead of single key. And this key can divided into two parts: First is the primers which will use in PCR technique as a key of KeyA; The second part is the initial conditions and parameters which can be used in the chaotic system as key KeyB.

This cryptography system is based on bio-security mainly. So there require a high quality of keys of DNA code. But keys KeyA is related to DNA code in united keys. KeyA is a string of Bases sequence of DNA, which is used as primers for PCR amplification, and this is very important. If the key is designed strictly according to the design principles of primer, it will cause the limited of key space. Therefore, we can use the software -- Primer Premier 5.0 to design the primer in this encryption system. The design process shown in Figure 7:



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

Figure 7: Key Preparation Process

We have already described the measure of production KeyB at the earlier time, here we do not say it more.

4.1.2 Encryption process

The message sender (also called encrypter) can began to encrypt the plaintext and made ciphertext after he/she completed the design of the key. Firstly, we will converting the plaintext into binary code. And then using the DNA encoding rules to deal with the binary code for chaos. We obtain the chaotic pseudo-random number sequence by bring the key KeyB into the chaotic system. Then make the binary of the sequence and the plaintext sequence to have a XOR compute and obtain the processed binary sequence. After that divided this binary sequence into an n sub-sequences. And pair it whit the numbered l_1, $l_2 \ldots l_n$ and compute them by the following operations:

$$l_1 \oplus l_2 = s_2$$

$$s_2 \oplus l_3 = s_3$$

$$\cdots$$

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

Obtain an n-1 of s_2, s_3, …, s_n and then add this n sequences, l_1, s_2, s_3, …, s_n, into the beginning of each sequences. After that we convert this sequence into the DNA Base sequence by the rules of DNA coding. Then select n primers from which we obtained in the previous step, and added them into the front of this n sequences. After all of these steps the ciphertext sequence will be made successfully.

Using the biological experimental technique, mainly we need to use the technology to synthesis tht DNA artificially, then making these DNA sequences synthesis into a short DNA chain artificially. And then use the technology of cutting and splicing to add these n short-chains of DNA into a long DNA template chain. We completed this long-chain DNA system and add it into DNA mixture. To here we finished product of the cipher text.

4.1.3 Decryption process

First of all, cracker will find out KeyA to have the PCR amplification from the information which he/she has got. Then is the second step, pick out the amplified DNA by using the technology of electrophores and these DNA were the DNA which we found, and they will have the information. After that will be the third step, through sequencing the

DNA chain we can draw the corresponding DNA sequence. The fourth step, restore the DNA sequence into a binary sequence by the use of DNA encoding rules. At this time, the binary sequence we obtained is l_1, s_2, s_3, …, s_n which is also in the encrypted process. After sorting it is 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, $l_2 \ldots l_n$. The fifth step, splicing the binary sequences together, then we can get the sequence that is clear binary sequence after sequence of the pre-treated. The sixth step, building chaotic system and bring the parameters of the key KeyB into the chaos system. After operations, we can obtain a binary sequence corresponding plaintext. Seventh step, through the transcending and to restore a character data, we can get the plaintext.

Now, the information transmission process is over. The sender send the message successfully and the receiver got it safely.

5. ANALYSIS OF CRYPTOGRAPHIC ALGORITHMS

5.1 Key Space

The size of key space is very important for security of the encrypt system. A good cryptographic algorithm should have enough key space to ensure its safety. The traditional encryption schemes of PCR amplification technology of DNA encrypted existent a problem that is do not have enough key space. So here are three ways to improve the security problem of the system.

1. This encryption using a method for combining PCR technology and chaos technology.

2. If we don't know the correct primers, we cannot start PCR amplification and we cannot obtain the DNA which has plaintext information at the same. This is the feature of encrypt system we described above. And this method is more enhanced on security.

3. This encrypt system is a common encrypt system of the combination of DNA code and chaotic encryption. Here we use chaotic system to pretreatment the plaintext.

This encrypt system have the above features, and it can adjust the size of the entire key space dynamically. Especially to the key adjust of DNA code.

5.2 Features And Benefits Of The System

In this paper, we use chaotic encryption for encryption systems to deal with plaintext. This encrypted systems eliminating the statistic rules in plaintext, and loading the chaotic encryption to DNA code. This will let it have the same advantage that the traditional encryption has. So the security has been improved. Even if the attacker deciphered the DNA code, then he will still face a lot of chaos code to decrypt. And it is increase the difficultly of decrypt. To be a new tape of encryption system, DNA code base on a different security with the traditional code. So we can obtain a complementary effect when we combined these two systems.

6. CONCLUSION

In this paper we use the encryption instance to described all the encrypt algorithm. And we have had the analysis of each encryption effect. Finally, we analyze the security and operability of the entire system. And use biology software to demonstrate the bio-security of analog of the amplification primers. Using computer to analyze statistical and demonstrate the effect of chaotic system. Nowadays theory and application of DNA cryptography is not perfect, so we still have a lot of work to do.

ACKNOWLEDGEMENTS

This work is supported by Science and Technology Development Project of Shaanxi Province Project (2010K06-22g), Industrial application technology research and development project of Xi'an (CXY1118(1)), Basic research fund of Northwestern Polytechnical University(GAKY100101), R Fund of College of Software, Microelectronics of Northwestern Polytechnical University (2010R001), and 2012 Graduation design (Thesis) key support fund of Northwestern Polytechnical University.

REFRENCES:

[1] Stallings, William, Cryptography and Network Security, 4thEd, Lodon, Prentice Hall, 2005.

[2] Adleman L, "Molecular computation of solutions to combinatorial problems", Science, Vol. 266, 1994, pp. 1021-1024.

[3] S. Ravinderjit, R. Braich, N. Chelyapov, et al., "Solution of a 20-variable 3-SAT problem on a DNA computer", Science, Vol. 266, 2002, pp.499-502.

[4] C. T. Celland, V. Risca, C. Bancroft, "Hiding messages in DNA microdots", Nature, Vol. 399, 1999, pp. 533-534.

[5] C. Popovici, "Aspects of DNA Cryptography", Annals of the University of Craiova Mathematics and Computer Science Series, Vol.37, No. 3, 2010, pp. 147-151.

[6] M. Borda, O. Tornea, "DNA secret writing techniques", Proceedings of IEEE COMM 2010, 2010, pp. 451-456.

[7] Xingyuan Wang, Lei Yang, Rong Liu, Abdurahman Kadir, "A chaotic image encryption algorithm based on perceptron model", Nonlinear Dynamics, Vol. 62, No. 3, 2010, pp.615-621.