IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 6, June 2015.
www.ijiset.com
ISSN 2348 – 7968

A New Cipher Text Generation Technique by Digitizing the Genetic DNA Code Using Random Number

Isha Yadav[1], Nipun Gupta[2] and M.K. Beniwal[3]

[1] Department of Computer Science Engineering,Swami Keshvanand Institute of Technology Management and Gramothan,
Jaipur, Rajasthan 302017/Asia, India

[2] Department of Computer Science Engineering, Modern Institute of Technology & Research Center & Xtreme Infosoft Pvt. Ltd.,
Alwar, Rajasthan 301001/Asia, India

[3] Reader-Department of Computer Science Engineering,Swami Keshvanand Institute of Technology Management and Gramothan,
Jaipur, Rajasthan 302017/Asia, India

Abstract

DNA cryptography is a new branch of cryptography in which DNA is used as an information carrier and advanced biological technology is acclimated as ability tool. DNA is an appealing media for data storage due to very large amount of data can be stored in small volume of DNA. DNA can be used for storing and transmitting the information in cryptography field. Although it is in primitive stage it has shown its effectiveness in the field of data transmission. Much research is going on to make the computational process more complex to the unauthorized user. It uses DNA as the computational apparatus forth with several molecular techniques to dispense it. By using the D.N.A we can make our method principally unbreakable due to random nature of DNA. DNA has four types of nucleotides bases Adenine (A), Thymine (T), Cytosine (C) and Guanine (G). By the computational model of DNA we come to know that these bases can be represent by binary numbers A↔00, C↔01, G↔10 and T↔11. DNA provides massive parallelism.

Keywords: *DNA, Nucleotides Bases, Cryptography, PCR Amplification, Random Number Generator.*

1. Introduction

“DNA cryptography could be a cryptographical technique within which every letter of the alphabet is regenerated into a unique binary combination of the four ester bases, that forms the human desoxyribonucleic acid (DNA). A bit of deoxyribonucleic acid writing system out the message to be encrypted synthesized, and therefore the hold on is slipped into a traditional fragment of human deoxyribonucleic acid of comparable length. The tip results dried out on paper and turns over little dots. As just one deoxyribonucleic acid strand is concerning thirty billion can contain the message, the detection of even the existence of the encrypted message is unlikely”.

The word cryptography comes out from ancient Greek. Cryptography is a mishmash of two words: (a) krypto means “hidden” and (b) grafo means to “write”. So the literal meaning of cryptography is “hidden writing”. It is the very old science of encoding messages so that only the sender and receiver can understand it. Cryptography is a science which uses mathematics to encrypt and decrypt the data. Cryptography enables us to store sensitive information and transmit it across insecure networks (like the Internet) so that it can be kept unreadable by anyone except the intended recipient.

2. Biological Background

DNA is found within the nucleus of each human cell. The knowledge in DNA: guides the cell (along with RNA) in creating new proteins that verify all of our biological individuality and gets passed (copied) from one generation to ensuing. Thus, it carries style data between generations, and therefore accounts for hereditary biological behavior. These DNA molecules contain the styles for all the fabric or elements that a living organism desires for growth, development, and daily living. It is the major source of the genetic information of any living organism in the biosphere. DNA is composed of two long strands of nucleotides bases arranged in a helix sort of structure as shown in figure below.

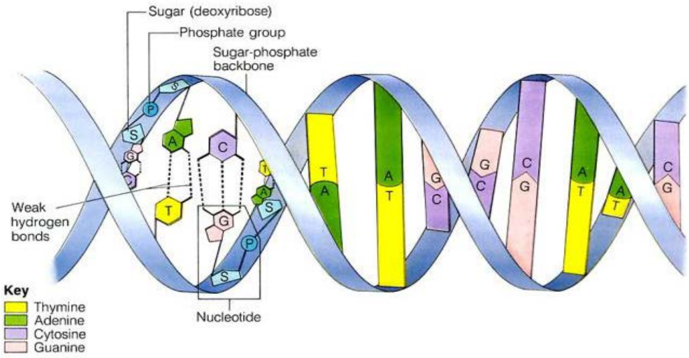


Fig. 1 Basic DNA Structure.

2.1 Watson Crick Model

This is a basic and well known model of DNA. James D. Watson and Francis H. C. Crick deduced double-helix structure of DNA in 1953 and got Nobel Prize in 1962. Adenine and Thymine always bond together as a pair, and Cytosine and Guanine bond together as a pair. The pairs are linked together like rungs in a ladder. Watson and Crick discovered that deoxyribonucleic acid had 2 sides, or strands, these strands were twisted along sort of a twisted ladder the spiral

Desoxyribonucleic acid is formed of chains of chemical subunits referred to as nucleotides, every of that contains one gas base: A (A), T (T), C (C), or G (G).

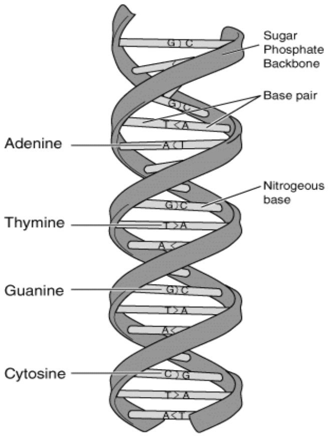


Fig. 2 Watson & Crick Model of DNA.

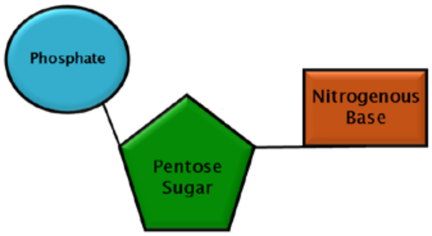


Fig. 3 Nucleotide structure

A G G - C T C - A A G - T C C - T A G
T C C - G A G - T T C - A G G - A T C

Fig. 4 Complimentary DNA strands

2.2 Biological Operations

The biological model of DNA cryptography is that the main activity distributed by the cell in our body that is integrated within the field of DNA computing. Here we will shortly discuss basic operations that area unit adopted from bio-chemistry operations as a computing tool within the field of DNA cryptography.

***A.* Synthesis**

Synthesis is a process of designing and reorganizing the information in DNA sequence form. In DNA computing, designing and synthesizing information in the DNA sequence form is an important process where a slight off beam design might leads to wrong result.

***B.* Denaturation**

The double stranded DNA molecule can be heated up to 90 degree Celsius, so that it is resolved into two single stranded DNA.

***C.* Ligation**

DNA ligation is a process of combining two single linear DNA fragments together. DNA ligation involves creating a phosphodiester bond between 3’ -hydroxyl of one nucleotide and the 5’ -phosphate of another.

***D.* Hybridization**

Hybridization is a process where the double stranded DNA is formed by combining single stranded DNA sequences; Nucleotides will bind to their complement. In this process A always pairs with T and G always pairs with C according to Watson crick complement under appropriate condition.

***E.* Polymerase Chain Reaction(PCR)**

PCR is a process that rapidly amplifies the amount of specific molecules of DNA in a given solution using primer extension by a polymerase.

***F.* Gel Electrophoresis**

Gel electrophoresis is a technique to sort DNA strands based on their length or weight through a gel such as agarose gel, in electrical field based on the fact that DNA is negative charge.

3. DNA Cryptography

Some of key technologies in deoxyribonucleic acid analysis have developed and well accepted in recent years. These technologies are polymerase Chain Reaction (PCR), deoxyribonucleic acid synthesis and deoxyribonucleic acid digital cryptography.

3.1 Random number Generation

Random Numbers are "A sequence of integers or group of numbers which show absolutely no relationship to each other anywhere in the sequence. At any point, all integers have an equal chance of occurring, and they occur in an unpredictable fashion".

3.2 One Time Pad (OTP)

The essential Security of the OTP(OneTimePad) is entirely because of the randomness of the key. The one-time pad is that the solely cryptosystem that exhibits what's mentioned as good secure.

Table 1: Conversion of Binary Data to DNA Format and Vice Versa

DNA	Binary	ASCII- 7 bits Decimal	ASCII- 8bits Decimal
A	00	0	0+1=1
C	01	1	1+1=1
G	10	2	2+1=3
T	11	3	3+1=4

3.2 Diffi- Helman key Exchange

Whitefield Diffie and Martien Hellman deviced an incredible answer to a tremendous of key agreement or key exchange in 1976. This solution is known as the Diffie-Hellman Key Exchange/Agreement Algorithms.

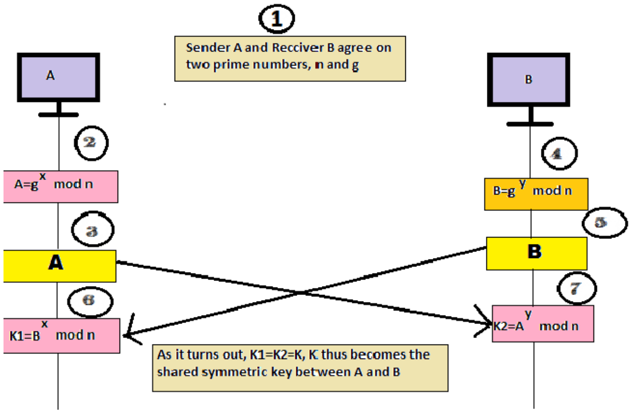

1
Sender A and Recciver B agree on two prime numbers, n and g
A
B
2
A=g^x mod n
4
B=g^y mod n
3
5
A
B
6
7
K1=B^x mod n
K2=A^y mod n
As it turns out, K1=K2=K, K thus becomes the shared symmetric key between A and B

Fig.5 Diffie-Hellman key exchange illustration

4. System Design of Proposed Encryption and Decryption Scheme

Suppose there's a sender A, who owns associate coding key K_A, and a supposed receiver B who owns a decipherment key K_B ($K_A = K_B$ or $K_A \neq K_B$). A uses K_A to translate a plaintext M into ciphertext C by a translation E. B uses K_B to translate the ciphertext C into the plaintext M by a translation D.

The encryption process is:

$$\mathbf{C = E_{KA}(M)}$$

The decryption process is:

$$\mathbf{D_{KB}\ (C) = D_{KB}\ (E_{KA}\ (M)) = M}$$

4.1 General Process of Scheme

Here we describe the general process of the encryption and decryption scheme as follows:

The scheme will be completed in three phases first is Key Generation, second is encryption and third is decryption phase.

4.2 Key Generation

After a pair of PCR primers is respectively designed and exchanged over a secure communication channel by using Diffe-Hellman, we can get an encryption key K_A that is a pair of PCR primers and B's public key e, as well as and encryption key K_B that is a pair of PCR primers and B's secret key d.

So we have these keys

Encryption key $\mathbf{K_A}$**= first pair of PCR Primers.**
Decryption key $\mathbf{K_B}$**= second pair of PCR Primers.**
e is a B's public key
d is a B's Secret key

4.3 Encryption

The original text message will be processed by pretreatment of data. After the information pretreatment we will get fully totally different ciphertext from an equivalent plaintext, which might effectively stop attack from a potential word as PCR primers.

4.4 Decryption

After that receiver B amplifies the secrete-message polymer sequence, he may retrieve the plaintext M sanded from sender A from the reverse preprocess (post process) operation victimization his secret key d. This coding method isn't solely a mathematical computation, however additionally a biological process.

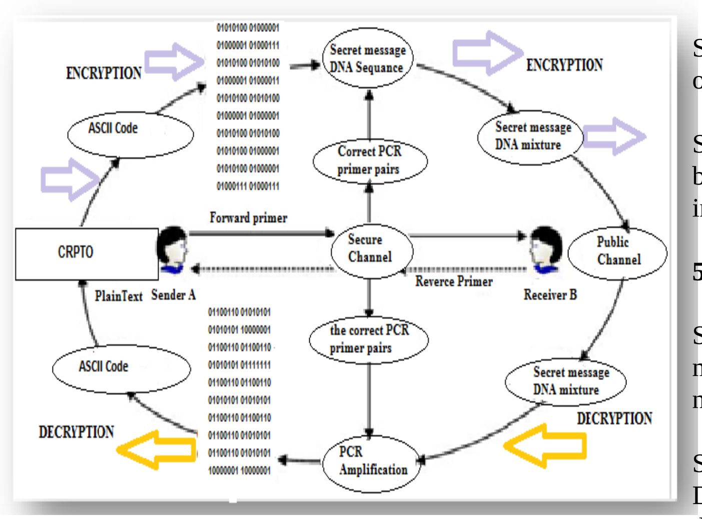


Fig. 6 Flow chart of Scheme

5. Proposed Algorithm

The algorithm to implement data security in binary representation of DNA sequence is done by using the random number generator as well as using encryption and decryption algorithm, based on the method of binary addition and binary subtraction rule.

5.1 Algorithm for Encryption

Step-1: The original message is converted into ASCII code (converted into numerical format)

Step-2: ASCII code is converted into binary code (number conversion is applied)

Step-3: Binary code is converted into DNA base equivalent by using table 1(DNA sequence)

Step-4: Generate a random numbers for each nucleotide of DNA sequence within the range of
1-99.
If random number is greater than 99 then number should be subtracted by 99.

Step-5: Convert DNA sequence of step 3 into standard ASCII value. Then convert into binary value.

Step-6:Convert random numbers of step 4 into binary numbers and generate a sequence of binary numbers for DNA sequence.

Step-7:Perform binary addition for the values of step 5 & step 6.

Step-8: Repeat step 7 for entire length.

Step-9: give out the complete binary encrypted message of original message.

Step-10: Convert the result of step 10 into form of ATCG, by using table2. This will give out the encrypted message into DNA sequence.

5.2 Algorithm For Decryption

Step 1: Read the sequence of encrypted DNA mixture message with random numbers. And separate the random number from DNA sequence.

Step 2: Generate the sequence of decimal numbers for DNA sequence and parallel generate the sequence of decimal numbers for DNA sequence with the help of random numbers.

Step 3: Perform binary conversion on both decimal number sequence.

Step 4: Perform subtraction of random number binary sequence from binary sequence of DNA code.

Step-5: Convert the result of step 4 into decimal number and convert to ATGC using table 2.

Step-6: Convert the result of step 5 into binary sequence using table 1.

Step-7: Perform 8 bit division of bytes for the result of step 6 and convert into ASCII value and get original text message.

6. Algorithmic Analysis

Regardless of several differences between deoxyribonucleic acid and ancient cryptography, they each satisfy equivalent characteristics of cryptography. The security requirements should base only on the secrecy of decryption key.

In this scheme three keys are used. These are K_A and K_B and pairs of PCR primer pairs. This system provides **two levels** of security. First level is **Biological difficulties** used and second level is **Mathematical difficulties** used.

- First level – Difficult Biological Securities.
- Second level – Difficult Mathematical Securities.

6.1 First Level Securities

First of all Adversary must have the good knowledge of biology and chemistry as well as cryptographic computation. Because of this scheme uses the complexity of difficult biological problems and operations. It is very difficult to amplify the encoded message without knowing the correct two primer pairs.

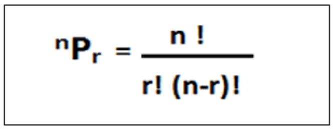
6.2 Second Level securities

The second level of securities are difficult mathematical problems used in this encryption scheme like traditional cryptography.

Our algorithm provides strong mathematical security level comparative to similar previously availablealgorithms, becauseour algorithm is based on **Random-One-Time-Pad**. This means that the secret key or Random key is used only once in one transaction. After one transaction this key will be destroyed.

6.3 Key Strength Analysis

In our algorithm we are using Random numbers for each nucleotide as key. Strength of the key can be calculated by permutation formula.
We are using four random numbers for each nucleotide within the range of 1 to 99.
Permutation formula is



$$^{n}P_{r} = \frac{n\,!}{r!\,(n-r)!}$$

Figure 7: Permutation Formula

7. Conclusions

In this Research Paper we have proposed a new algorithmic scheme named**, “A New Cipher Text Generation Technique by Digitizing the Genetic DNA Code Using Random Number”**. This scheme is based on Random one-time-pad key cryptographic system.
One of the major problems while using network is data security. This dissertation focuses on the data security issues to provide a secure and effective encryption and decryption method by using Random numbers key generation. We are considering DNA characteristics and Random key to achieve new idea in data security.This scheme uses the DNA digital coding technique, DNA synthesis and PCR amplification, Random number generation and Arithmetic operations as well as traditional cryptography.

The planned algorithmic program scheme continues to be far away being an ideal scheme but it has some specific benefits and meets cryptographic principles. But I hope this scheme will be good for some or particular techniques.
It is hoped that the “A New Cipher Text Generation Technique by Digitizing the Genetic DNA Code Using Random Number” encryption algorithm as well as dissertation work will make a positive contribution towards data security, communications, privacy and efficiency of data storage and transmission.

References

[1]. Leonard M. Adleman, “Molecular Computation of Solutions to Combinational Problems”

[2]. Online- http://www.scribd.com/doc/71935055/Monica-Borda-Fundamentals-in-Information-Theory-and-Coding

[3]. D.Prabhu and M.Adimoolam, “Bi-serial DNA Encryption Algorithm (BDEA)”

[4]. Monica Borda and Olga, “DNA Secret Writing Techniques”.

[5]. Naya and Selim, “Aspects of DNA Computing”.

[6]. Yunpeng Zhang and Liu He Bochen Fu, “Research on DNA Cryptography”

Mrs. Isha Yadav is working at Xtreme Infosoft Pvt. Ltd. as Director, & related to Cryptographic Implementations and Research field. Along with this she is currently active as Instructor to students of Engineering field.

Mr. Nipun Gupta is Technology Evangelist at Xtreme Infosoft Pvt. Ltd. and a Certified Ethical Hacker(CEH) & Profound knowledge of Networking ,Linux & Embedded System. He is currently pursuing his M.Tech at MITRC,Alwar.

Mr. M.K. Beniwal is working as Reader at Swami Keshvanand Institute of Technology Management and Gramothan in Computer Science Department.