International Journal of Computer Science and Engineering (IJCSE)
ISSN(P): 2278-9960; ISSN(E): 2278-9979
Vol. 3, Issue 3, May 2014, 51 - 66

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

DNA ENCRYPTION

ANUPRIYA AGGARWAL[1] & PRAVEEN KANTH[2]

[1]Research Scholar, BRCM, Bahal, Haryana, India

[2]Assistant Professor, BRCM, Bahal, Haryana, India

ABSTRACT

DNA Encryption is the technique for encrypting the secret message using Bio molecular computation which makes this unique from mathematical computation. DNA Cryptography provides parallelism and fast computation that it can break DES Encrypted message and less power comsumption .These four nucleotides of DNA and their positioning and their corresponding conversion into binary strands palys the major role in encryption .The DNA materials are stable and long lasting.

We presented traditional method and improved complimentary pair method. For each method, we secretly select a reference DNA sequence **S** and incorporate the secret message **M** into it such that we obtain **S'**. We send this **S'**, together with many other DNA, or DNA-like sequences to the receiver. The receiver is able to identify the particular sequence with **M** hidden in it and ignore all of the other sequences. He will also be able to extract **M**. The DNA Cipher text to be converted into plain text involves biological process of PCR (Polymerase Chain Reaction) as well. The enzymes and protein material plays the role in DNA Computation .Each DNA Cryptography methods first converts the plain text into ASCII Code and then furthur process starts .There are various applications od DNA other than security as used in inks and for analyzing the human behaviour and various others.

KEYWORDS: The DNA Materials, ASCII Code, RSA and DES

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.

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**, 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.

Recent study shows that DNA Computing can be used as a new efficient method to solve difficult mathematical problems (Adleman). Adleman, in 1994 solved Hamiltonian path problem (HPP) by using DNA computing and its advantages such as vast parallelism and extraordinary information density. Deoxyribonucleic acid (DNA) is a kind of molecule that encodes genetic information by cellular function. A single strand DNA (Leier et al.; Dove) consists of four different base nucleotides, adenine (A), thymine (T), cytosine (C) and guanine (G). Those nucleotides are able to be bound together in the long sequence. One of the important DNA roles was presented by Watson–Crick which is described in (Hegedu¨s et al.). Actually, DNA computing mostly includes three main steps (Cui et al.):

- Encoding of all candidate's solutions against computational problems.
- Reaction Control by enzymes and generating all types of data pools that include possible solution to the computational problem.
- Problem solution's mining by a Polymerase Chain Reaction (PCR).

Until now several cryptographic methods have been proposed and most of them are based on complex mathematical equations. In order to make them more secure, scientists are working to increase their complexity by changing their mathematical equations. Thus, an intruder could not find a quick solution to predict secret keys and break the cryptosystems. While complex equations have been used in the heart of traditional cryptosystems for several years, today, DNA computing breaks these cryptosystems by using its exclusive characteristics (i.e. parallel processing in molecular level). For example, RSA data encryption method has some computational disadvantages and DNA computing is able to attack into different parts of RSA algorithm simultaneously and breaks it in a short period of time. These computational disadvantages are described in the following (Xiao et al.):

- Reliability of RSA algorithm is based on produced factoring large numbers.
- Breaking RSA cryptosystem is infeasible on the assumption.

PROBLEM FORMULATION

Earlier many researchers have proposed various encryption algorithms such as AES, DES, Triple DES, RSA, Blowfish etc. Some of them are most popular in achieving data security at a great extent like AES and Blowfish. But, as security level is increased, the time and complexity of algorithm is also increased.

Broadly we categories these algorithms in two types:

- Symmetric encryption
- Asymmetric encryption

In symmetric encryption major disadvantages are:

- Need of secure channel for secret key exchange.
- Too many keys.
- Origin and authenticity of message cannot be guaranteed.

In asymmetric encryption major disadvantages are:

- Public key should be authenticated.
- Slow.
- More computer resources are required.
- Loss of private key may be irreparable.

The fundamental idea behind this encryption technique is the exploitation of DNA cryptographic strength, such as its storing capabilities and parallelism in order to enforce other conventional cryptographic algorithms.

RELATED LITERATURE

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.

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.

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.”

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.DNA is usually in the form of a right-handed double helix. The helix consists of two polydeoxynucleotide chains. Each chain is an alternating polymer of deoxyribose sugars and phosphates that are joined together via phosphodiesterlinkages. One of four bases protrudes from each sugar: adenine and guanine, which are purines, and thymine and cytosine, which are pyrimidines.

While the sugar phosphate backbone is regular, the order of bases is irregular and this is responsible for the information content of DNA. Each chain has a 5_ to 3_ polarity, and the two chains of the double helix are oriented in an anti parallel manner—that is, they run in opposite directions. Pairing between the bases holds the chains together. Pairing is mediated by hydrogen bonds and is specific: Adenine on one chain is always paired with thymine on the other chain, whereas guanine is always paired with cytosine. This strict base-pairing reflects the fixed locations of hydrogen atoms in the purine and pyrimidine bases in the forms of those bases found in DNA. Adenineandcytosine almost always exist in the amino as opposed to the iminotautomeric forms, whereas guanine and thymine almost always exist in the keto as opposed toenol forms. The complementarity between the bases on the two strands gives DNA its self-coding character.

The two strands of the double helix fall apart (denature) upon exposure to high temperature, extremes of pH, or any agent that causes the breakage of hydrogen bonds. Upon slow return to normal cellular conditions, the denatured single strands can specifically reassociate to biologically active double helices (renature or anneal).DNA in solution has a helical periodicity of about 10.5base pairs per turn of the helix. The stacking of base pairs upon each other creates a helix with two grooves. Because the sugars protrude from the bases at an angle of about 120°, the grooves are unequal in size.

The edges of each base pairare exposed in the grooves, creating a pattern of hydrogenbond donors and acceptors and of van der Waals surfaces that identifies the base pair. The wider—or *major*—groove is richer in chemical information than the narrow (*minor*) groove and is more important for recognition by nucleotide sequence-specific binding proteins. Almost all cellular DNAs are extremely long molecules, with only one DNA molecule within a given chromosome. Eukaryotic cells accommodate this extreme length in part by wrapping the DNA around protein particles known as nucleosomes.

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:

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

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:

- The Nitrogenous bases.
- Deoxyribose.
- 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 cipher text. 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.

The 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.

DNA Coding 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:

- Adenine (A).
- Thymine (T).
- Cytosine (C).
- Guanine (G).

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

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

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

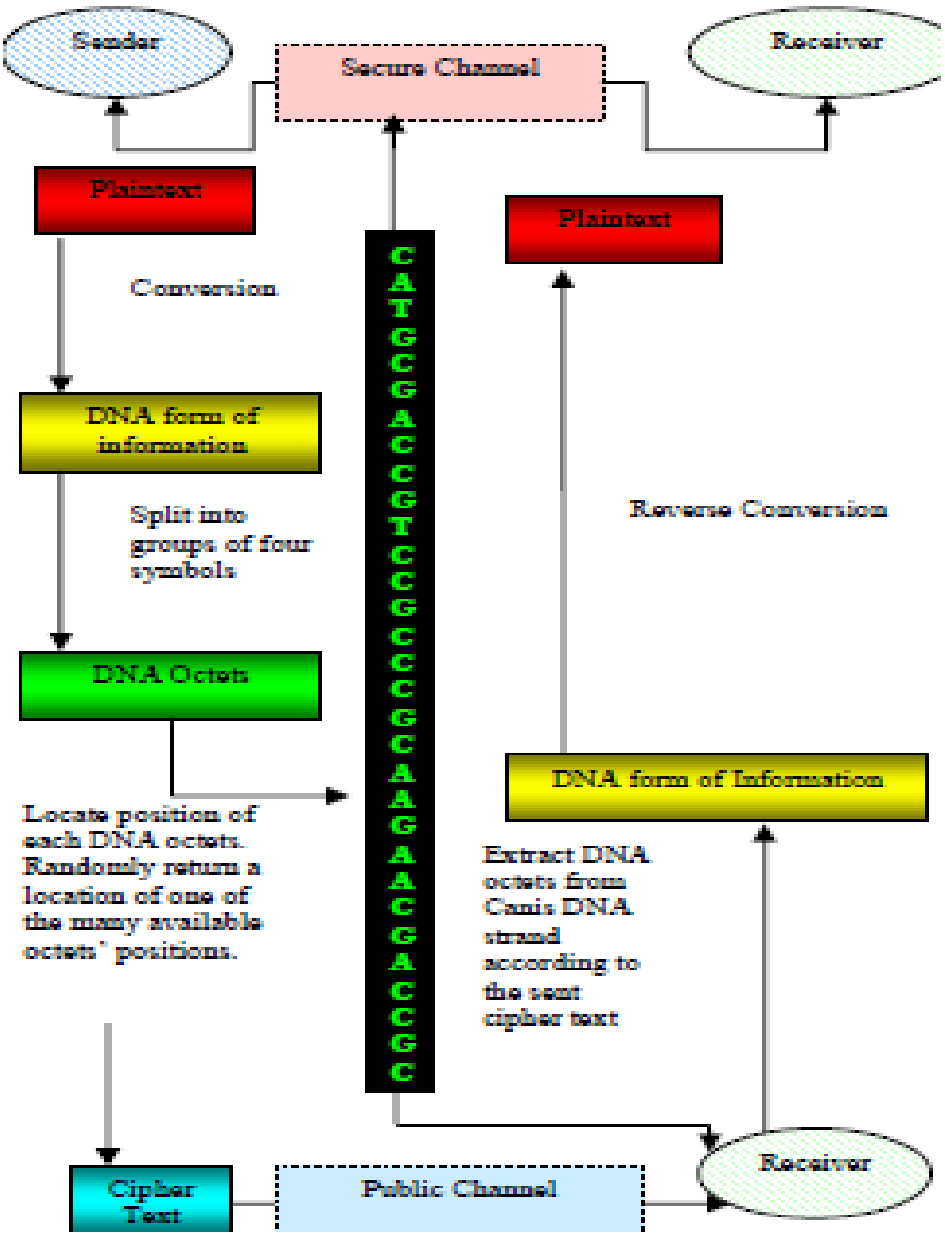

Figure 1: The Computational Grap

Traditional DNA Encryption Algorithm

Encryption Process

Input: A reference DNA sequence S, a secret binary message M and a binary coding scheme to code A, C, G and T into binary digits.

Output: An encrypted DNA sequence S'.

Step 1: Code S into a binary sequence S_1 by using the binary coding scheme.

Step 2: Generate k is by using a random number generator to divide S_1 into segments and generate r is to divide the secret message Minto segments. Denote S_1 by $s_1s_2s_3\ldots\ldots s_n$ and M by $m_1m_2m_3\ldots\ldots m_p$.

Step 3: Insert each m_i of M before S_i of S_1 in order to produce a new binary sequence.

Step 4: Denote this new binary sequence as S_2 and convert it into fake DNA sequence and denote it as S_3.

Step 5: Return S_3.

Get s2, s3, …,sn n-1 sequences and then l1, s2, s3, …, sn, 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

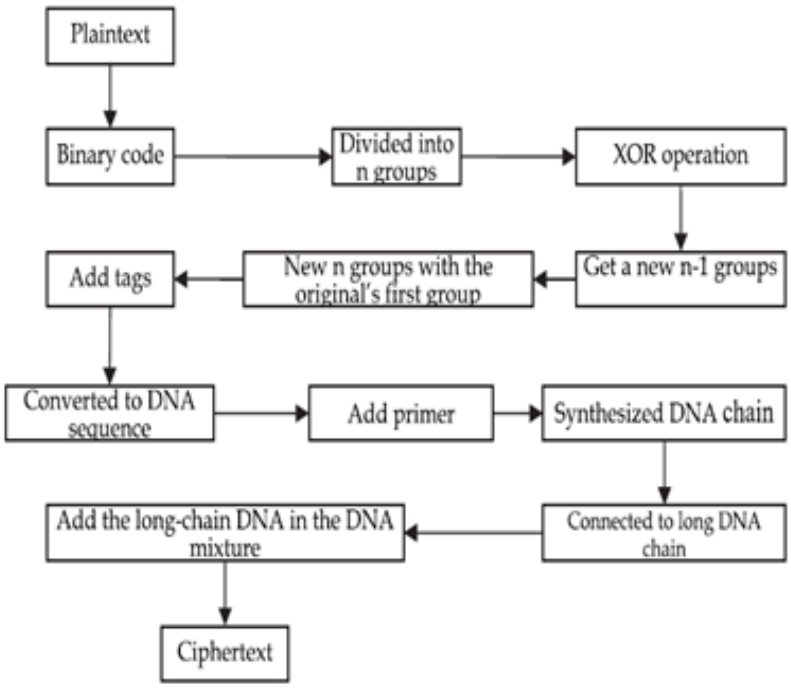

Figure 2: Encryption Process

DNA encryption algorithm containing technologies of DNA synthesis, PCR amplification, DNA digital coding, XOR operation as well as traditional cryptography The intended PCR two primer pairs was used as the key of this scheme that not independently designed by the sender or receiver. This operation could increase the security of encryption method.

On the other hand, the traditional encryption method and DNA Digital Coding are used to preprocess operation we can get completely different cipher text from the same plaintext, which can effectively prevent attack from possible word as PCR primers.

Code

```
import java.io.*;
importjava.util.*;


class Encrypt
{
	public static void main(String ...a)throws Exception
	{
		char[] S = {'A','C','G','G','A','A','T','T','G','C','T','T','C','A','G'};
		char[] M = {'0','1','1','1','1','0','1','1','0'};
		char[] SS = new char[15];
		int i, j;
		int[] A = {3, 4, 6, 7, 8, 11, 13};
		j = 0;
		for(i = 0; i < 15; i++)
		{
			if(i == A[j] - 1 && M[j] == '1')
			{
				if(S[i] == 'A') SS[i] = 'C';
				else if(S[i] == 'C') SS[i] = 'G';
				else if(S[i] == 'G') SS[i] = 'T';
				else if(S[i] == 'T') SS[i] = 'A';
				if(j < 6)j++;
			}
			else if(i == A[j] - 1 && M[j] == '0')
			{
				if(S[i] == 'A') SS[i] = 'A';
```

```
                    else if(S[i] == 'C') SS[i] = 'C';
                    else if(S[i] == 'G') SS[i] = 'G';
                    else if(S[i] == 'T') SS[i] = 'T';
                    if(j < 6)j++;
                }
                else if(i != A[j] - 1)
                {
                    if(S[i] == 'A') SS[i] = 'G';
                    else if(S[i] == 'C') SS[i] = 'T';
                    else if(S[i] == 'G') SS[i] = 'A';
                    else if(S[i] == 'T') SS[i] = 'C';
                }
            }
            for(i = 0; i < 15; i++)
            {
                System.out.println(SS[i]);
            }
        }
}
```

Decryption Process

Input: Fake DNA sequence k and r.

Output: Plaintext M.

Step 1: Generate binary sequence from fake DNA sequence.

Step 2: Divide that binary sequence into r + k size of segments.

Step 3: With each segment of size r + k extract first r bits and store them into M.

Step 4: Return M.

The basic problem with this approach is security of key if anyone has find the key then he/she can decrypt the plain text.

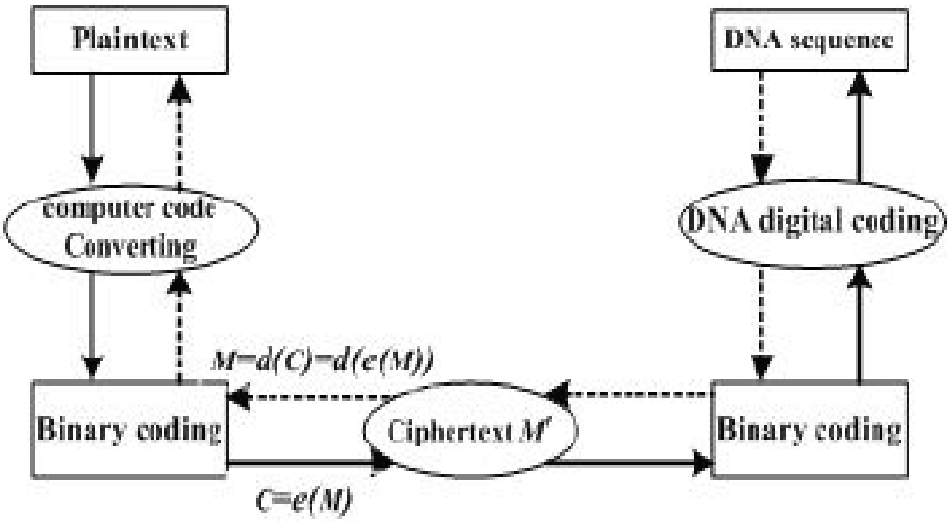

Figure 3

Data Pre (post)treatment flow chart Data

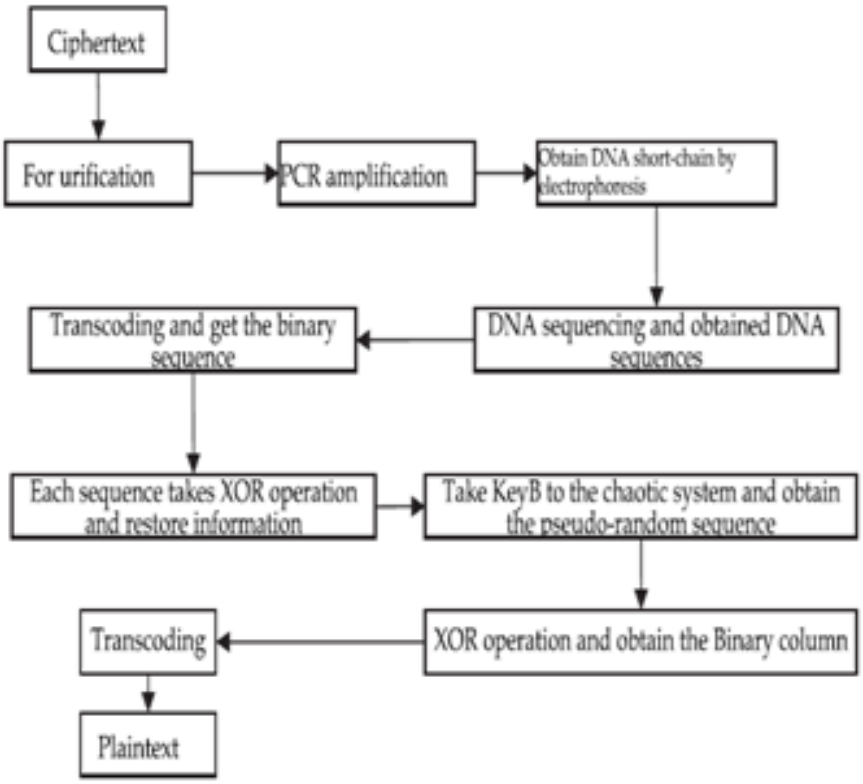

Figure 4: Decryption Process

The vast parallelism, exceptional energy efficiency and extraordinary information inherent in DNA molecules are being explored for computing, data storage and cryptography. DNA cryptography is a emerging field of cryptography. In this paper a novel encryption algorithm is devised based on number conversion, DNA digital coding, PCR amplification, which can effective elyprevent attack. Data treatment is used to transform the plain text into cipher text which provides excellent security.

code

```
import java.io.*;

importjava.util.*;


class Decrypt

{

        public static void main(String ...a)

        {

                char[] SS = {'G','T','G','T','G','C','A','T','A','T','A','C','C','G','A'};

                char[] S = {'A','C','G','G','A','A','T','T','G','C','T','T','C','A','G'};

                char[] M = new char[7];

                int i = 0, j = 0;

                for(i = 0; i < 15; i++)

                {

                        if(SS[i] == S[i])

                        {

                                M[j] = '0';

                                if(j < 6) j++;

                        }

                        else

                        {

                                if(S[i] == 'A' && SS[i] == 'C')

                                {

                                        M[j] = '1';

                                        if(j < 6) j++;

                                }

                                else if(S[i] == 'C' && SS[i] == 'G')

                                {

                                        M[j] = '1';
```

```
                    if(j < 6) j++;
                }
                else if(S[i] == 'G' && SS[i] == 'T')
                {
                    M[j] = '1';
                    if(j < 6) j++;
                }
                else if(S[i] == 'T' && SS[i] == 'A')
                {
                    M[j] = '1';
                    if(j < 6) j++;
                }

            }
        }
        for(i = 0; i < 7; i++)
        {
            System.out.println(M[i]);
        }
    }
}
```

PROPOSED METHODOLOGY

As we discussed above the problem with traditional DNA encryption method is with security of key. An another approach to solve that problem is complimentary pair approach like DNA structure we are not going to detail for this and using our own complimentary pairs.

A → T

C → A

G → C

T → G

Let us consider a reference sequence:

S = ACGGAATTGCTTCAG

Using the complimentary pair approach the new sequence S' will be:

S' = TACCTTGGCAGGATC

But in our methodology we will combine the complimentary approach with substitution approach and we will generate S' from S with help of plain text (M) steps may be as follows:

- Take any reference sequences S.
- Using complimentary pair approach and plain text generate fake DNA sequence S'.
- Send both S and S' by using any stagenography technique in order to generate more security.
- Receiver will generate plain text from S and S'.
- There is no need to send any keys like k and r in traditional cryptography, therefore key security problem is not there and we are choosing different reference sequence.

Hardware & Software Requirements

Languages Used: JAVA

Platform: Windows 7

RESULTS & ANALYSIS

The aim of project was to develop a system that could compute the fundamental idea behind this encryption technique is the exploitation of DNA cryptographic strength, such as its storing capabilities and parallelism in order to enforce other conventional cryptographic algorithms. In this study, a binary form of data, such as plaintext messages, and images are transformed into sequences of DNA nucleotides. Subsequently, efficient searching algorithms are used to locate the multiple positions of a sequence of four DNA nucleotides. These four DNA nucleotides represent the binary octet of a single plaintext character or the single pixel of an image within, say, a Can is Familiaris genomic chromosome.

We call the file containing the randomly selected position in the searchable DNA strand for each plain text character, the ciphered text. Since there is negligible correlation between the pointers file obtained from the selected genome, with its inherently massive storing capabilities, and the plain-text characters, the method, we believe, is robust against any type of cipher attacks.

CONCLUSIONS

We have pointed out that the DNA sequences have the special properties which we can utilize for encryption purposes. We have proposed the algorithm and this is based upon a reference sequence known only to the sender and the receiver. This reference sequence can be selected from any web-site associated with DNA sequences. Since there are many websites and roughly 55 million publicly available DNA sequences, it is virtually impossible to guess this sequence.

FUTURE SCOPE

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.

REFERENCES

1. 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.
2. Adleman L, Molecular computation of solutions to combinatorial problems [J]. Science, 1994, 266: 1021-1024.
3. Kazuo T, Akimitsu O, Isao S. Public-key system using DNA as a one-way function for key distribution [J]. Bio systems, 2005, 81: 25-29.
4. Akutsu, T., Dynamic Programming Algorithms for RNA Secondary Structure Prediction with Pseudoknots, Dirsrete Applied Mathematics, Vol. 104, 2000, pp. 45-62.
5. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D., Molecular Biology of the Cell, New York & London: Garland Publishing, 1994.
6. Lehninger, A. L., Nelson, D. L. and Cox, M. M., Principles of Biochemistry, New York, Worth, 2000.
7. G. Z. Cui, "New Direction of Data Storage: DNA Molecular Storage Technology," *Computer Engineering and Applications*, vol. 42, pp.29–32, 2006
8. M. Amosa, G. Paun and G. Rozenbergd. "Topics in the theory of DNA computing," *Theoretical Computer Science*, vol. 287, pp. 3–38, 2002.
9. T. Kamei, "DNA-containing inks and personal identification system using them without forgery," *Jpn. KokaiTokkyoKoho*, 2002, p.8.
10. V. B. Semwal, V. B. Semwal, M. Sati and S. Verma, "Accurate location estimation of moving object in Wireless Sensor network," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 1, no. 4, pp. 71-75, 2011.
11. Semwal, Vijay Bhaskar, K. Susheel Kumar, Vinay S. Bhaskar, and Meenakshi Sati. "Accurate location estimation of moving object with energy constraint & adaptive update algorithms to save data." arXiv preprint arXiv: 1108.1321 (2011).
12. Pinki Kumari and Abhishek Vaish, "A Comparative study of Machine Learning algorithms for Emotion State Recognition through Physiological signal", Advances in Intelligent Systems and Computing, Vol.236-Springer; ISBN 978-81-322-1601-8
13. Pinki Kumari and Abhishek Vaish "Brainwave's Energy feature Extraction using wavelet Transform" proceeding of IEEE SCEECS, 2014, MANIT, Bhopal ISBN: 978-1-4799-2526-1.

14. Semwal, Vijay Bhaskar, K. Susheel Kumar, Vinay S. Bhaskar, and Meenakshi Sati. "Accurate location estimation of moving object with energy constraint & adaptive update algorithms to save data." *arXiv preprint arXiv: 1108.1321* (2011).

15. V. B. Semwal, V. B. Semwal, M. Sati and S. Verma, "Accurate location estimation of moving object in Wireless Sensor network," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 1, no. 4, pp. 71-75, 2011.

16. J. P. Gupta, N. Singh, P. Dixit, V. B. Semwal, and S. R. Dubey, "Human Activity Recognition using Gait Pattern," *International Journal of Computer Vision and Image Processing*, vol. 3, no. 3, pp. 31 – 53, 2013.

17. K. K. Susheel, V. B. Semwal and R. C. Tripathi, "Real time face recognition using adaboost improved fast PCA algorithm," *arXiv preprint arXiv*:1108.1353 (2011).

18. K. S. Kumar, V. B. Semwal, S. Prasad and R. C. Tripathi, "Generating 3D Model Using 2D Images of an Object," *International Journal of Engineering Science*, 2011.