Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol.5, No.7, 2014
www.iiste.org

Secure Data Transmission Using DNA ENCRYPTION

ANUPRIYA AGGARWAL[1] & PRAVEEN KANTH[2]
1Research Scholar, BRCM, Bahal, Haryana, India
2Assistant Professor, BRCM, Bahal, Haryana, India

Abstract
DNA Encryption is preferable biological technique for securing text/image because of its parallelism, vast storage and fast computing quality. The process involve biological molecule present in human body called DNA abbreviated as Deoxyribose Nucleic Acid .The DNA molecule is synthesized and protein component part is extracted and then converted to nitrogen base . This nitrogen base is used in Encryption/Decryption and formulated as A (Adenine), C (Cytosine), T (Thymine) and G (Guanine) characters.

DNA Cryptography components are ACTG characters only and how the message gets merged and located is known as DNA Cryptography. This ACTG characters create DNA Sequence S and merged with message M to produce new sequence S' and send to receiver where Sequence S' back converted to S.

The paper will introduce traditional methods of DNA cryptography in which there is need of key and proposed methods ,in which introduction to key is not required ,hence removing the tension of securing the key. The proposed method involves Complementary pair method.

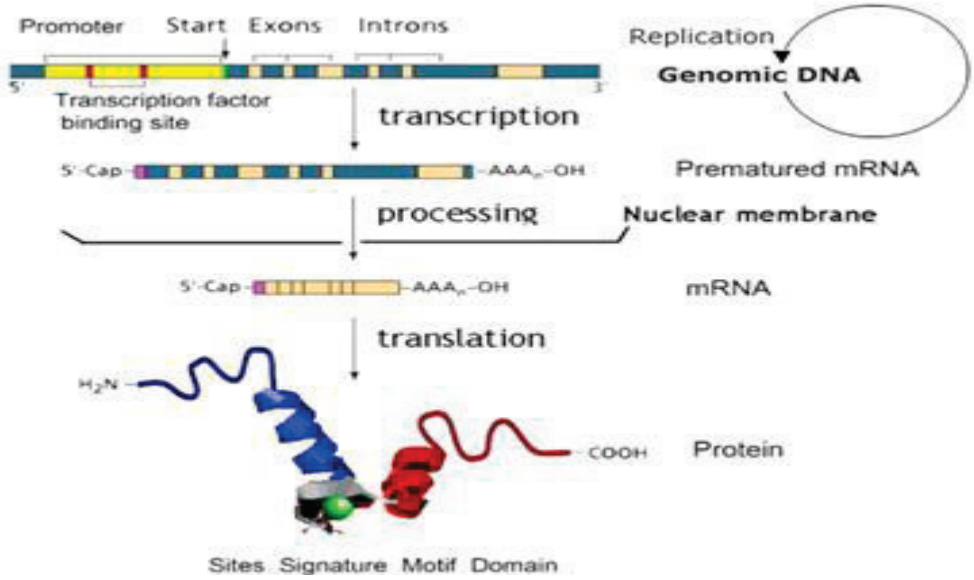
1. Introduction

The earliest form of cryptography was the simple writing of a message, as most people could not read (New World, 2007). In fact, the very word cryptography comes from the Greek words kryptos and graphein, which mean hidden and writing, respectively (Pawlan, 1998).

Early cryptography was solely concerned with converting messages into unreadable groups of figures to protect the message's content during the time the message was being carried from one place to another. In the modern era, cryptography has grown from basic message confidentiality to include some phases of message integrity checking, sender/receiver identity authentication, and digital signatures, among other things (New World, 2007).

DNA Cryptography

In human body to transform the genetic inform from one part to another part nucleic acids are present. There are two type of nucleic acid DNA and RNA which code for all type of instructions needed for the cell to perform different function. The DNA Stands for Deoxyribo Nucleic Acid. DNA is the molecule that contains the genetic code of organism. DNA is material that governs Genetic similarity of looks, nature in human and animals.



Comparisons between Different Cryptography Techniques

- DES is the old "data encryption standard" from the seventies. Its key size is too short for proper security (56 effective bits; this can be brute-forced, as has been demonstrated more than ten years ago). Also, DES uses 64-bit blocks, which raises some potential issues when encrypting several gigabytes of data with the same key (a gigabyte is not that big nowadays).
- 3DES is a trick to reuse DES implementations, by cascading three instances of DES (with distinct keys). 3DES is believed to be secure up to at least "2^{112}" security (which is quite a lot, and quite far in the realm of "not breakable with today's technology"). But it is slow, especially in software (DES was designed for efficient hardware implementation, but it sucks in software; and 3DES sucks three times as much).
- Blowfish is a block cipher proposed by Bruce Schneier, and deployed in some softwares. Blowfish can use huge keys and is believed secure, except with regards to its block size, which is 64 bits, just like DES and 3DES. Blowfish is efficient in software, at least on some software platforms (it uses key-dependent lookup tables, hence performance depends on how the platform handles memory and caches).
- AES is the successor of DES as standard symmetric encryption algorithm for US federal organizations (and as standard for pretty much everybody else, too). AES accepts keys of 128, 192 or 256 bits (128 bits is already very unbreakable), uses 128-bit blocks (so no issue there), and is efficient in both software and hardware. It was selected through an open competition involving hundreds of cryptographers during several years. Basically, you cannot have better than that.

Advantages of DNA Cryptography

- 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.
- A DNA molecular computer has low power consumption, only equal to one-billionth of a traditional computer.

Components of DNA

DNA is a polymer. The monomer units of DNA are nucleotides, and the polymer is known as a "polynucleotide." Each nucleotide consists of a 5-carbon sugar (deoxyribose), a nitrogen containing base attached to the sugar, and a phosphate group. There are four different types of nucleotides found in DNA, differing only in the nitrogenous base. The four nucleotides are given one letter abbreviations as shorthand for the four bases.

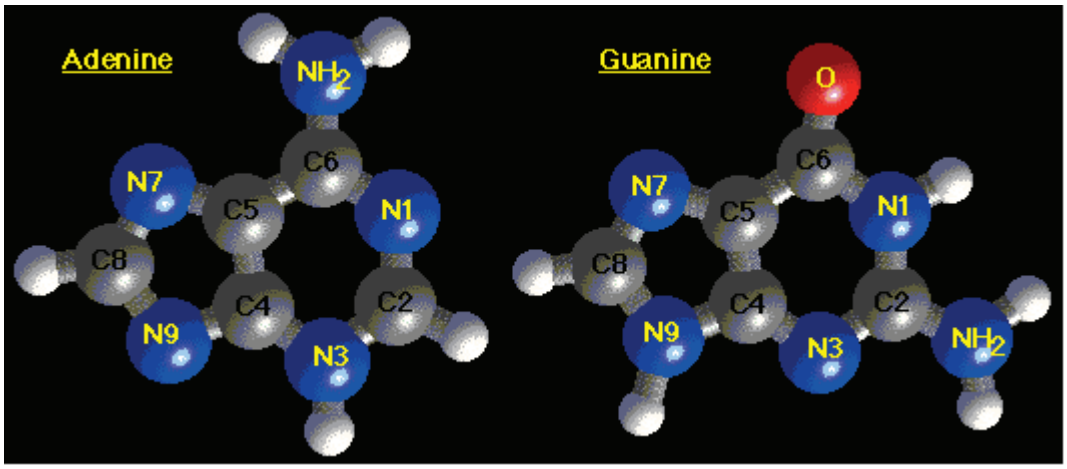
A is for adenine

G is for guanine

C is for cytosine

T is for thymine

Structure of A and G

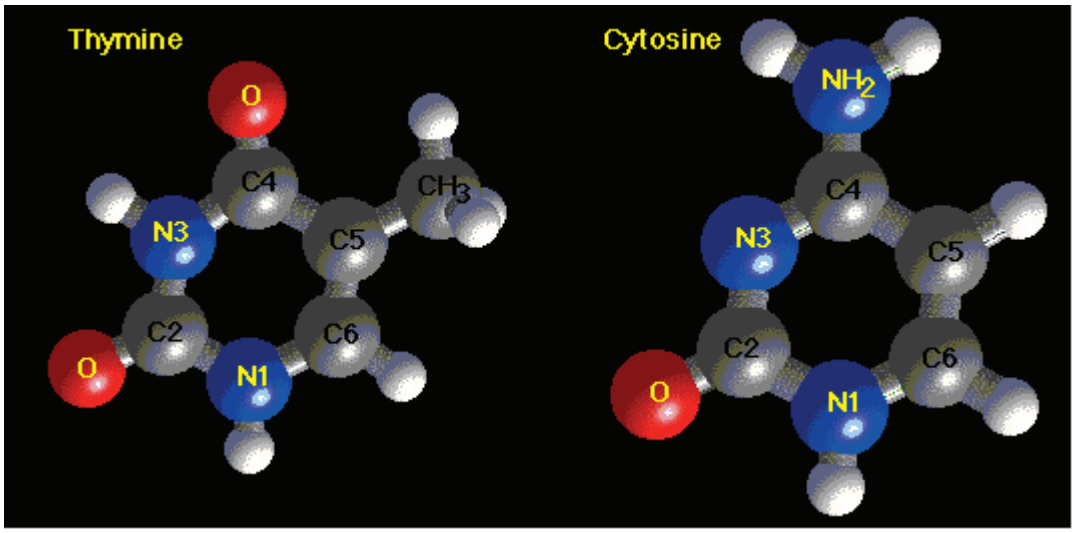


Purine Bases

Adenine and guanine are purines. Purines are the larger of the two types of bases found in DNA. Structures are shown below:

The 9 atoms that make up the fused rings (5 carbon, 4 nitrogen) are numbered 1-9. All ring atoms lie in the same plane.

Structure of C and T



Pyrimidine Bases

Cytosine and thymine are pyrimidines. The 6 Atoms (4 carbon, 2 nitrogen) are numbered 1-6. Like purines, all pyrimidine ring atoms lie in the same plane.

2. 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.

DNA Coding Scheme

The easiest way to encode is to represent 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.

Traditional DNA Encryption Algorithm

In this section we propose two methods to encrypt the plaintext using DNA, so that it could be send securely over a network.

A. Method I

Encryption

Step1: The binary data, text or image, is used under the form of ASCII code (in decimal format).

Step2: These numbers are then grouped in blocks and encrypted in using a traditional method (eg. DES, will form a 2 level encryption).

Step3: This encoded message is then changed to binary format.

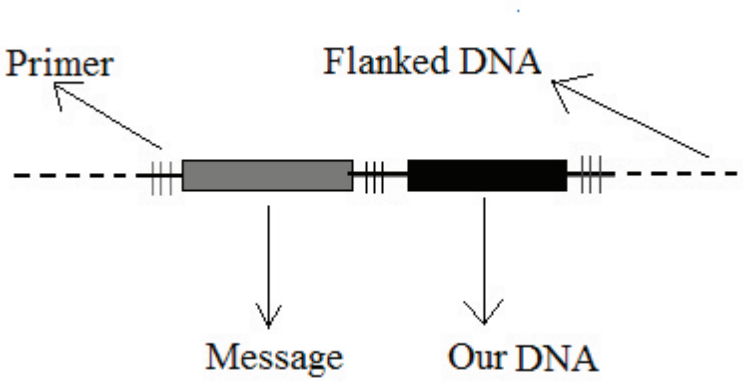
Step4: Then these digits are grouped into two and substituted as A for 00, T for 01, G for 10, and C for 11.

Step5: We then fit the primers on either side of this message. Primers will act as stoppers and detectors for the message. This has to be given to the receiver prior to the communication.

Step6: This message is followed by our own DNA sequence followed by another stopper/primer.

Step7: This message is then flanked by many sequences of DNA or by confining it to a microdot in the micro-array.

Step8: If considered as a pseudo method: this sequence is transferred to the receiver through the Internet. Else the micro-array is sent physically (though time consuming).



Decryption

This message can then be recovered only by an intended recipient who both can find it, and who knows the sequences of the PCR primers employed, and also the encryption key (2 level encryption used). For pseudo method:

Step1. The DNA sequence is searched for the primers (start primer and end primer). The message in-between them is retrieved and the next DNA sequence before the next primer (our DNA) is retrieved.

Step2. The ATGC characters are substituted accordingly (00,01,10,11 respectively).

Step3. They are then converted into ASCII code and then the message is retrieved.

For the use of actual DNA:

Step1. We have the DNA. What we want is the message flanked by the primers. To accomplish this we can use a technique called Polymerase Chain Reaction (PCR), which allows you to produce many copies of a specific sequence of DNA. PCR is an iterative process that cycles through a series of copying events using an enzyme called polymerase. Polymerase will copy a section of single stranded DNA starting at the position of a primer, a short piece of DNA complimentary to one end of a section of the DNA that you're interested in. By selecting primers that flank the section of DNA you want to amplify, the polymerase preferentially amplifies the DNA between these primers, doubling the amount of DNA containing this sequence. After many iterations of PCR, the DNA we are working on is amplified exponentially.

Step2.Then the ATGC sequence in this DNA strand is read.

Step3. The ATGC characters are substituted accordingly (00,01,10,11 respectively).

Step4. They are then converted into ASCII code and then the message is retrieved.

B. Method II

Encryption

Step 1: Sender encodes his message in the original DNA sequence and allows the message to be transcripted to mRNA. During transcription, a DNA segment that constitutes a gene is read, starting from the promoter (starting position) of the DNA segment. The non-coding areas (intron) are removed according to certain tags, and the remaining coding areas (extron) are rejoined and caped. Then the sequence is transcribed into a single stranded sequence of mRNA (messenger RNA). The mRNA moves from the nucleus into the cytoplasm

Step 2: At this stage, the mRNA is translated to protein. During translation, the mRNA sequence is translated into a sequence of amino acids as the protein is formed. During translation, the ribosome reads the fragment starting from certain three-bases, and then the ribosome reads three bases (a codon) at a time from the mRNA and translates them into one amino acid; there are also certain ending three-bases to sign the end of the translation.

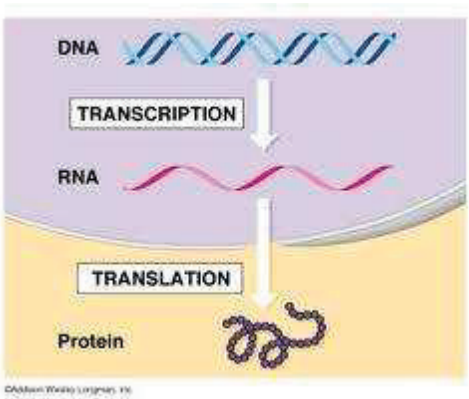


Fig: Transcription and Translation of the encoded message

Step 3: The resulting protein then behaves like the public key and can be sent to the receiver using public channel.

Step 4: At the same time, the sender sends the shared secret key to the receiver which consists of the information he needs to reassemble the DNA such as the location of the non coding regions that need to be reinserted.

Decryption

When the receiver received the protein form of data and the keys, he uses the keys to recover mRNA form of data from protein form of data, and then recover DNA form of information, in the reverse order as the sender encrypts the information. He can then recover then binary form of information, and finally gets what the sender sent him.

3. 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.

3. Hardware & Software Requirements

Languages Used: JAVA

Platform: Windows 7

4. 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.

5. 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.

6. 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.

7. REFERENCES

[1] Introduction to DNA Structure,
http://www.blc.arizona.edu/molecular_graphics/dna_structure/dna_tutorial.html

[2] Yashaswita R. Bhoir, R.Mathangi ," DNA CRYPTOGRAPHY with BINARY STRANDS" , Fr. C. Rodrigues Institute of Technology

[3] 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.*

[4] S.Jeevidha, Dr.M.S.Saleem Basha, Dr.P.Dhavachelvan , "Analysis on DNA based Cryptography to Secure Data Transmission", *International Journal of Computer Applications (0975 – 8887) Volume 29– No.8, September 2011*

[5]. L. M. Adleman, "Molecular computation of solutions to combinational problems," *Science, vol. 266, pp. 1021–1024, 1994.*

[6] S.V. Kartalopoulos, "DNA-inspired cryptographic method in optical communications," in *authentication and data mimicking Military Communications Conference, 2005, pp. 774–779.*

[7]. T.Kawai and Y.Hayashizaki, *"DNA BOOK"*.

[8] M. Saeb, A. Baith, "An Encryption Algorithm for Data Security," Recent Advances in Information Science & Technology, N.E. Mastorakis, (editor), World Scientific Publishing Company, pp. 350-354, 1998.

[9] "DNA-Based Cryptography. DIMACS DNA Based Computers," V, American Mathematical Society, 2000.

[10] 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.

[11] Advantages and Disadvantages of Symmetric and Asymmetric Key Encryption Methods
http://voices.yahoo.com/comparing-symmetric-asymmetric-key-encryption-6329400.html?cat=15

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar