International Journal of Information Technology and Business Management
29th June 2013. Vol. 14 No.1

ISSN 2304-0777 www.jitbm.com

PROPOSED STEGANOGRAPHY APPROACH USING DNA PROPERTIES

Ban Ahmed Mitras **Adeeba Kh. Aboo**

Prof. Dr. **M. Sc. Student**

Dept. of Computer Science, College of Computer & Mathematic Science-Mosul Univ./ IRAQ
dr.ban_mitras@yahoo.com dalyadiamond@yahoo.com

Abstract

In this research one method has been proposed to hide the secret message in a sequence of DNA. The secret message after it has been encrypted using the RSA algorithm hid in the DNA sequence known using complementary character. The DNA reference sequence can be selected from the different DNA database. One very important DNA database is EBI that provides fundamental genetic information.

Keywords:

DNA sequence, RSA , Secret Information, complementary pairing rules, Steganography.

1-Introduction

Information security is of utmost importance in today’s fast developing era. Information or messages are being exchanged over various types of networks. With the huge growth of computer networks and advancement in technology, a huge amount of information is being exchanged. A large part of this information is confidential or private which increases the demand for stronger encryption techniques. Security has become a critical feature for thriving networks. Communication is not secure due to the presence of hackers who wait for a chance to gain access to confidential data[2]. Cryptography is derived from the Greek words “kryptos” (meaning “hidden”) and “graphein” (meaning “to write”). Cryptography is the study of means of converting information from its normal comprehensible form into an incomprehensible format, rendering it unreadable without the secret knowledge. The process of converting information (plain text) by transforming it into unreadable format (cipher text) is known as encryption. Encryption techniques can be sometimes broken by cryptanalysis, also called as code breaking, although modern cryptographic techniques are virtually unbreakable. Cryptography encrypts the actual message that is being sent. This mechanism employs mathematical schemes and algorithms to scramble data into unreadable text. It can only be decoded or decrypted by the party that possesses the associated key[1] [7].

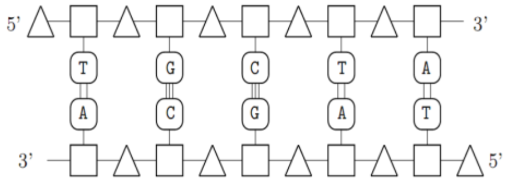
Steganography is derived from the Greek word “stegnos” (meaning “covered/secret”) and "graphein" (meaning "to write/draw") [1]. Steganography is the study of means of concealing the information in order to prevent hackers from detecting the presence of the secret information. The process of concealing the message in a cover without leaving a remarkable trace is known as Steganography. Steganography is the form of convert communication in which a secret message is camouflaged with a carrier data. Steganography masks the very presence of communication, making the true message not discernable to the observer. Cryptography and Steganography achieve the same goal using different means. Encryption encodes the data so that an unintended recipient cannot determine its intended meaning. Steganography in contrast attempts to prevent an unintended recipient from suspecting that the data is there [7]. The proposed method combines the two techniques (cryptography and steganography) to provide a very high degree of security for the data.

2-Related work

In 2003 , Muhalim Mohamed Amin et al. proposed system enhanced the LSB technique by randomly dispersing the bits of the message in the image and thus making it harder for unauthorized people to extract the original message[1]. Alan Siper et al. publish ,in 2005, The Rise of Steganography discuss Revelations about Steganography such as history, why it is used, how it works, techniques, counter-measures, risks, legal and ethical issues, and the future[8]. In DNA Shimannovsky et al. proposed an interesting and original arithmetic code. Pak Chung Wong and al presented new potential applications for DNA organic data memory. The advantages of these approaches in production keys were presented by Masanori Arita and Tanaka et al. in 2005. Recently, Nozomu Yachie et al. presented a very complete methodology, simple, flexible and robust of data storage based on sequence alignment of genomic DNA of living organism. D. Heider et al. publish, in 2007, a program called DNA Crypt whose use is centered on the patent protection of genetically modified organisms (GMOs). Shu-Hong Jiao and Robert Goutte publish, in 2009, Hiding data in DNA of living organisms[9].In 2011 Bibhash Roy et al. proposed system called Enhanced key Generation Scheme based Cryptography with DNA Logic[10]. In 2012 , Suman Chakraborty and Samir Kumar Bandyopadhyay Proposed two steps steganography approach Secret information is hidden within the Cover image and Cover image is hidden within DNA sequence[6].Mohammad Reza Najaf Torkaman et al. publish, in 2013, Innovative Approach to Improve Hybrid Cryptography by Using DNA Steganography[3].

3-DNA Steganography

Nowadays, scientists are working on a different kind of Steganography algorithms to ameliorate the security of the system. There are many Steganography algorithms to hide the secret data into the host carrier. DNA Steganography is one of the cutting edge techniques in this area. Basic concepts of the DNA Steganography are based on the properties of natural DNA sequences in the cell. In molecular biology, genetic information is stored in deoxyribonucleic acid which is known as DNA in the cells. DNA is made by four nucleotides which are Thymine (T), Cytosine (C), Guanine (G), and Adenine (A). These bases are linked by a backbone of DNA strands which are sugar components and phosphate groups. This backbone identifies the direction of the DNA strands [3].



Figure(1):The Backbone of DNA strand [4]

Each single strand is linked by hydrogen bond to make the DNA double strand. The standard situation of nucleotides allows to make a hydrogen bond between C and G; or A and T. This complementary standard rule is known as Watson-Crick base-pairing. C and G are bonded by triple hydrogen bond, although A and T is linked by double hydrogen bonds. This complementary concept is the fundamental issue in genetic activities that leads to double DNA strands are twisted together and make DNA double helix. The mixture of these basic nucleotides which are Thymine (T), Cytosine (C), Guanine (G), and Adenine (A) make the long polymer strands which able to make massive amount of combinations of DNA double helix that stores the every living features and properties of creatures such as human and mammal. There are several DNA data hiding approaches based on [3]. The DNA reference sequence can be selected from the different DNA database. One very important DNA database is EBI that provides fundamental genetic information [5].

4-Complementary rules

The complimentary DNA sequence means the sequence on the other strand of DNA directly opposite your specified sequence.

There are six major Complementary rules for each letter of DNA sequence. For all letter x, C(x), C(C(x)), C(C(C(x))) is not equal [6] :

1. A→T , T→C , C→G , G→A
2. A→T , T→G , G→C , C→A
3. A→C , C→T , T→G , G→A
4. A→C , C→G , G→T , T→A
5. A→G , G→T , T→C , C→A
6. A→G , G→C , C→T , T→A

5- Proposed method

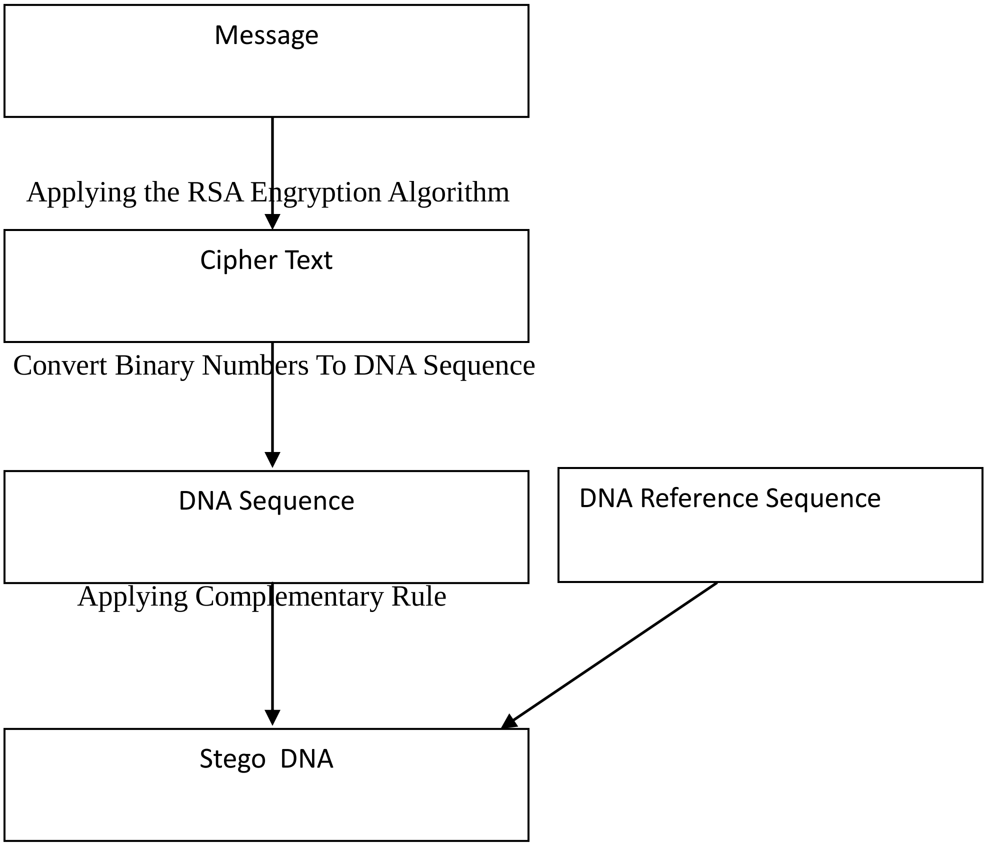
The structure of proposed algorithm is combinational of concepts of cryptography and steganography. first applying the RSA encryption algorithm on secret message and encrypt it. This algorithm is based on the difficulty of factorizing large numbers that have 2 and only 2 factors (Prime numbers). The system works on a public and private key system. The public key is made available to everyone. With this key a user can encrypt data but cannot decrypt it, the only person who can decrypt it is the one who possesses the private key. It is theoretically possible but extremely difficult to generate the private key from the public key, this makes the RSA algorithm a very popular choice in data encryption. Then message can convert from binary to DNA sequence. This paper is based on table following rules. 00 is converted to A, 01 converted to C, 10 converted to G, 11 converted to T(table(1))

Table (1): Binary Representation Of Nucleotides

Nucleotide	Binary Form
A	00
C	01
G	10
T	11

Then will hide the DNA sequence in a known DNA sequence by using complementary. sender extracts reference sequence from EBI database which is accessible publicly .The complementary DNA sequence means the sequence on the other strand of DNA directly opposite your specified sequence .In this paper was used the following complementary law:

1-(A A)→A
2-(A T)→A
3- (T T)→T
4-(T A)→T
5-(C C) →C
6-(C G)→C
7-(G G) →G
8-(G C) →G


Message
Applying the RSA Engryption Algorithm
Cipher Text
Convert Binary Numbers To DNA Sequence
DNA Sequence
DNA Reference Sequence
Applying Complementary Rule
Stego DNA

Figure(2) :embedding secret message

Application : we explain the proposed method

Message ='hi'

Cipher text=117 209

Binary numbers=0111010111010001

```
DNA  sequence=             C     T    C  C       T      C     A      C       (from table (1)we have)
                           |     |    |  |       |      |     |      |
DNA Reference Sequence= G  G  T  T G  G     G  A  A     G     T   A A G A G A A
```

Stego DNA= CGTTCCGTACAAACAGAA

Index= 13568101114

Now, embedding stage is finally completed. Then, sender sends 1,3,5,6,8,10,11,14 to the receiver with the stego DNA. In the next section, the receiver will apply the extracting phase for extracting the original message.

Explain the extraction method:

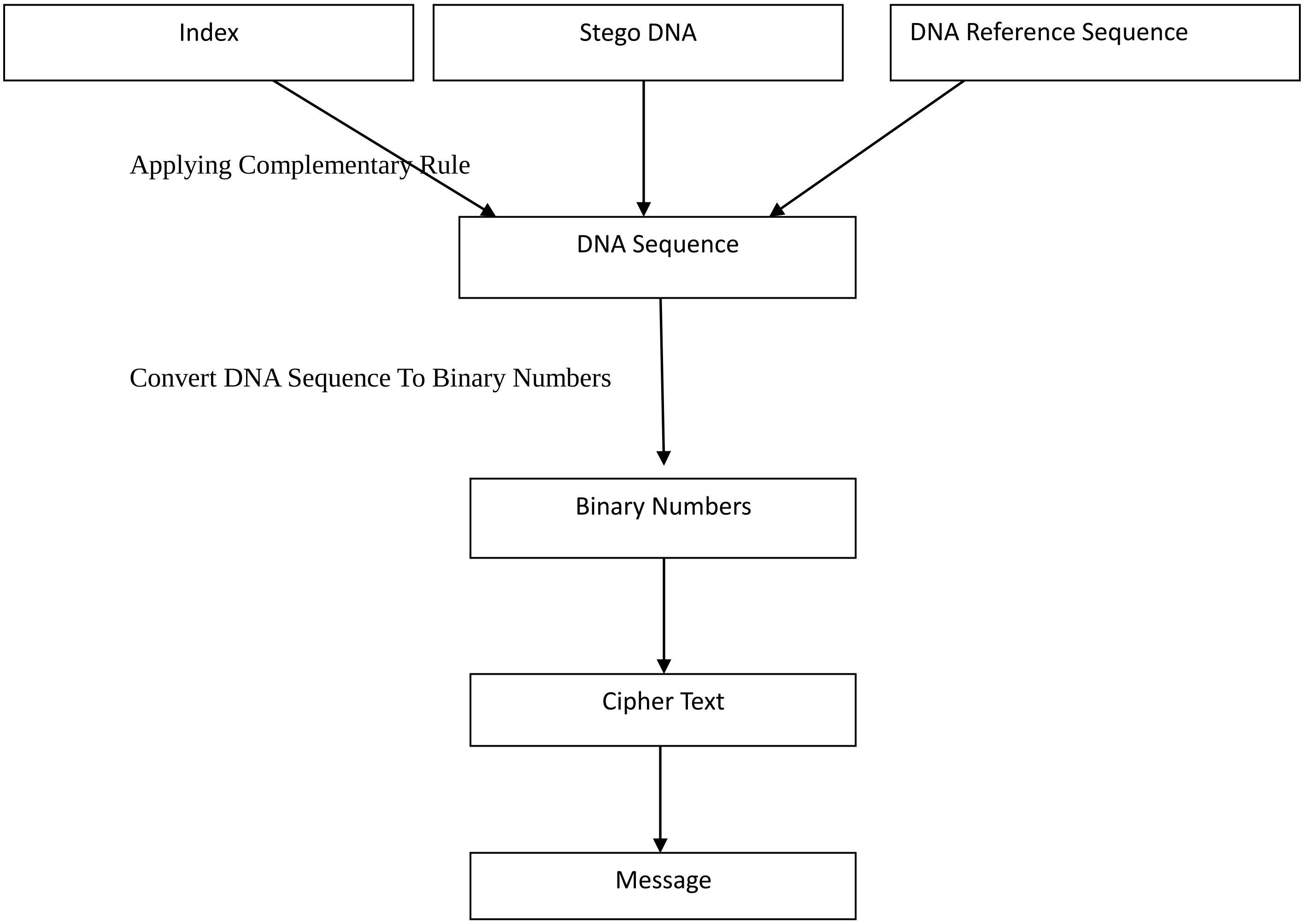
```
Index         =      1    3     5  6     8    10  11     14

Stego DNA=          C G  T T   C  C G  T  A   C   A AA  C  AGAA
                    |    |     |  |    |      |   |     |
                    |    |     |  |    |      |   |     |
DNA Reference = G G  T T  G  G G  A  A  G    T AA G  AGAA

complementary rule
                    ↓    ↓     ↓  ↓    ↓      ↓    ↓     ↓

DNA  sequence=     C    T    C  C    T     C    A     C
```

DNA sequence=CTCCTCAC (from table(1) we have)

Binary numbers=0111010111010001

Cipher text=117 209

Message ='hi'



Figure(3): extracting original message

6-Security Issues

In terms of security, each intruder must be aware from the following information, correctly. Without this fundamental information, possibility of extracting original message is near to zero, scientifically. They are:

RSA Algorithm : the only person who can decrypt a message is the one who possesses the private key.

DNA reference sequence: there are 163 million DNA reference sequence on EBI database. Therefore, the likelihood of making a doing well conjecture by attacker is 1/24 .

Binary coding rule: as mentioned, the sender is free to select any equivalent binary form for every nucleotide. It means that, A can be ‘00’, ‘01’, ‘10’, or ‘11’; C can be ‘00’, and so on. In other words, all the binary coding rules are 4×3×2×1=24. So, the likelihood of making correct guess by attacker is 1/24.

Complementary pairing rule: like binary coding rule, there is 4×3×2×1=24 complementary alphabet among basic nucleotides. Therefore, the possibility of making successful attack is 1/24.

Eventually, the final probability of making a correct and successful guess by attacker is (1/(163*106)*(1/24)*(1/24)).

7-Conclusion

Basically, the purpose of cryptography and steganography is to provide a very high degree of security for the data . The secret message after it has been encrypted using the RSA algorithm hid in the DNA sequence known using complementary character. Considering DNA characteristics brings new ideas in data hiding. DNA sequences are potential to implement new data hiding techniques or even transforming previous schemes to new one. In this paper, a reference DNA sequence has been shared between sender and receiver. Not only this DNA reference sequence can be retrieved from EBI or NCBI [3] databases but it can also be simply selected from any database. Therefore, by considering any sort of database, there are 163 million targets to select it. Virtually, guessing the correct DNA sequence by attacker is unachievable .The crucial feature of the DNA sequences is visibility. Finding secret message in a DNA sequence is difficult because the visibility of the sequences is very low. Therefore, attacker cannot find out whether this sequence is a fake or not. In comparison with previous techniques, such as in images, not only implementing this method is not difficult but also it is formidable to detect, as well.

References

1-Muhalim Mohamed Amin, Subariah Ibrahim , Mazleena Salleh ," Information Hiding Using Steganography ", Universiti Teknologi Malaysia , 2003.

2- Jie Yang,Weiwei Lin, Taojian Lu., " An approach to prove confidentiality of cryptographic protocols with non-atomic keys", IEEE,2012.

3- Mohammad Reza Najaf Torkaman1, Nazanin Sadat Kazazi1, Azizallah Rouddini2 ," Innovative Approach to Improve Hybrid Cryptography by Using DNA Steganography ",(IJNCAA), 2012 .

4- Knudsen, L.R. ,"Block ciphers analysis", Aarhus University, 1994.

5- Shiu, H., Ng, K., Fnag, J. F., Lee, R., Huang, C.," Data hiding methods based upon DNA sequences", Information Sciences, 2010.

6- Suman Chakraborty1, Prof. Samir Kumar Bandyopadhyay," Two Stages Data-Image Steganography Using DNA Sequence",

International Journal of Engineering Research and Development, 2012.

7- Manoj Kumar Sharma, Dr. Amit Upadhyaya,Shalini Agarwal," Adaptive Steganographic Algorithm using Cryptographic Encryption RSA Algorithms ", Journal of Engineering, Computers & Applied Sciences ,2013 .

8- Alan Siper, Roger Farley and Craig Lombardo, "The Rise of Steganography", Proceedings of Student/Faculty Research Day, CSIS, Pace University, May 6th, 2005.

9- Shu-Hong Jiao, Robert Goutte," Hiding data in DNA of living organisms ",Vol.1, No.3, 181-184 (2009).

10- Bibhash Roy, Gautam Rakshit, Ritwik Chakraborty,

" Enhanced key Generation Scheme based Cryptography with DNA Logic", International Journal of Information and Communication Technology Research,2011.