APTIKOM Journal on Computer Science and Information Technologies
Vol. 5, No. 2, 2020, pp. 214~220
ISSN: 2528-2417
DOI: 10.34306/APTIKOM.J.CSIT.39

Transmission of Data in secure manner with DNA Sequence

Ravinder Paspula [1], K. Chiranjeevi [2], S. Laxman Kumar [3], NV Krishna Rao [4]
Departement of Computer Science and Engineering, JNTU-Hyderabad, Telangana, Hyderbad, India.
e.mail : ravindra.paspula@gmail.com

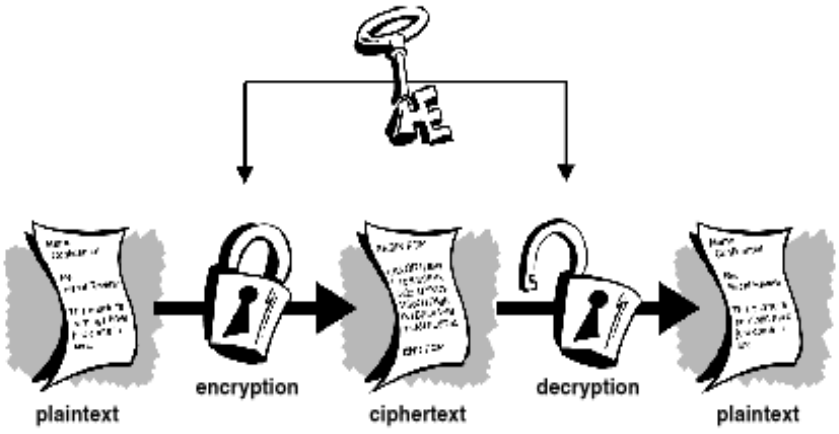
Abstract

A new-promising technology called DNA-Cryptography is emerged in the area of DNA- Computing field.DNA useful for store, sending the data and also to perform computation. Even it is under primal level, DNA-Based molecular Cryptography system is shows extremely efficient. This technique offers a unique cipher-text generation process and a new key generation practice. This method used to implement a procedure which includes two stages. First stage generates a session key and encryption key and it uses cipher block chaining mode-CBC, the secrete number(s) and incorporate plain-text M into level-1 cipher-text. The last stage converts the level-1 cipher-text into level-2 cipher-text (s). The level-2 cipher-text is again transformed into human made DNA-sequence (S') and is given to the receiver along with many other DNA-sequence. By this process it will become a more complicated for un-authorized user to gain original information. The receiver will apply the process to identify the human made DNA sequence with M hidden in it and extract the original message M.

Keywords*: Encryption, Decryption, DNA Sequence, Cipher Text, DNA cryptography, Cipher Block Chining Mode.*

1. Introduction

Cryptography is an indispensable tool for protecting information in computer systems. While cryptography is the science of securing data, cryptanalysis is the science of analyzing and breaking secure communication. Cryptography contains few aspects of providing privacy, authentication, digital signatures, electronic money, and other applications.



In the field of information technology the message that is transmitted through internet is being read by other people is very high. A message in a form that is easily readable by humans. This is the original message that has to be secured. The process of converting real text into unknown form is known as encryption. Encrypting real-text results in unreadable form called cipher-text. The method of converting cipher-text to its original real-text is called decryption.

2. Introduction to DNA Cryptography

A new technique for securing data was introduced using the biological structure of DNA called DNA Computing (aka molecular computing or biological computing). It was invented by Leonard Max Adelman in the year 1994, for solving the complex problems such as directed Hamilton path problem, NP-complete problem similar to The Travelling Salesman problem. Adelman is also known as the 'A' in the RSA algorithm – an algorithm that in some circles has become the de facto standard for industrial-strength encryption of data sent over the Web. The technique later on extended by various researchers for encrypting and reducing the storage size of data that made the data transmission over the network faster and secured. The concept of using DNA computing in the fields of cryptography and Steganography has been identified as a possible technology that may bring forward a new hope for *unbreakable algorithms*. Strands of DNA are long polymers of millions of linked nucleotides. These nucleotides consist of one of four nitrogen bases, a five carbon sugar and a phosphate group. The nucleotides that make up these polymers are named after the nitrogen base that it consists of; *Adenine (A), Cytosine (C), Guanine (G) and Thymine (T).*

2.1 Advantages of DNA computing

1. **Speed** – Conventional computers can perform approximately 100 MIPS (millions of instruction per second). Combining DNA strands as demonstrated by Adelman, made computations equivalent to 10^9 or better, arguably over 100 times faster than the fastest computer.
2. **Minimal Storage Requirements** – DNA stores memory at a density of about 1 bit per cubic nanometer where conventional storage media requires 10^12 cubic nanometers to store 1 bit.
3. **Minimal Power Requirements** – There is no power required for DNA computing while the computation is taking place. The chemical bonds that are the building blocks of DNA happen without any outside power source. There is no comparison to the power requirements of conventional computers.

Cryptographic technique in which each letter of the alphabet is converted into a different combination of the four bases that makes up the human deoxyribonucleic acid (DNA).Complementarily is achieved by distinct interactions between nucleobases adenine, thymine guanine and cytosine. Adenine and guanine are purines, while thymine, cytosine and uracil are pyrimidines. Purines are larger than pyrimidines. Both types of molecules complement each other and can only base pair with the opposing type of nucleobase. In nucleic acid, nucleobases are held together by hydrogen bonding, which only works efficiently between adenine and thymine and between guanine and cytosine. The base complement A=T shares two hydrogen bonds, while the base pair G≡C has three hydrogen bonds. All other configurations between nucleobases would hinder double helix formation. DNA strands are oriented in opposite directions, they are said to be anti-parallel.

–CTAG‖,–	**CATG‖,**	**–GTAC‖,**	**–GATC‖,**	**–TCGA‖,**	**–TGCA‖,**	**–ACGT‖**	**–AGCT‖.**

Six complementary rules-3 used for DNA cryptography:

(AC-CG-GT-TA)
(AG-GT-TC-CA)
(AG)-GC-CT-TA)
(AT-TC-CG-GA)
(AT-TG-GC-GA)
(AC-CT-TG-GA)

For instance any one of them like :(GC) (CT) (TA) (AG) is applied to projected methodology

3. Problem Statement

Data Security is a challenging issue of data communications today that touches many areas including secure communication channel, strong data encryption technique and trusted third party to maintain the database. The rapid development in information technology, the secure transmission of confidential data herewith gets a great deal of attention. The conventional methods of encryption can only maintain the data security. The information could be accessed by the unauthorized user for malicious purpose. Therefore, it is necessary to apply effective encryption/decryption methods to enhance data security Strong cryptography or cryptographically strong is general terms applied to cryptographic systems or components that are considered highly resistant to cryptanalysis. Demonstrating the resistance of any cryptographic scheme to attack is a complex matter, requiring extensive testing and reviews, preferably in a public forum. Good algorithms and protocols are required and good system design and implementation is needed as well. Present system uses private key cryptography for internet banking application or websites. Private-key methods are efficient and difficult to break. However, one major drawback is that the key must be exchanged between the sender and recipient beforehand, raising the issue of how to protect the secrecy of the key. When the President of the United States exchanges launch codes with a nuclear weapons site under his command, the key is accompanied by a team of armed couriers. Banks likewise use high security in transferring their keys between branches. These types of key exchanges are not practical, however, for e-commerce between, say, amazon.com and a casual web surfer.

4. Literature Survey

Plaintext encryption using DNA sequence is a new technique which is under initial stage and trying to provide secure communication over the internet. This technique is deemed to be securing that would be very difficult for unauthorized users to break the hidden plaintext and extract the original message. Only concern receiver can decrypt and receive the original message.The following listed authors have done tremendous research on DNA based message encryption and hiding techniques reported recently. Cheng Guo et.al, proposed a method in which letters of original message are replaced with the other letters on a particular location set by the algorithm. This algorithm converts a given letter with a selected letter outlined by the complementary rule. For example, if a complementary rule is outline as (AC-CG-GT-TA), then the result of θ(G) are going to be T, and result of θ(T) are going to be A.

While Watson and Crick worked on the problem of the structure of the DNA molecule, Erwin Chargaff, an Austrian biochemist who emigrated to the U.S. during the Nazi era, was studying the actual composition of DNA. It was known that DNA consisted of A, G, T, C, but what Chargaff found was that the proportion of each of these bases in DNA was not random. In fact, across different animal species (from sea urchin to salmon, and so on), he found that the number of as in the DNA was always equal to the number of Ts in the DNA, and the number of Gs was equal to the number of Cs. This rule, %A = %T and %G = %C, is known as Chargaff Parity Rule 1 and played a key role in Watson's and Crick's understanding of the structure of DNA. Boneh et al and Adleman - et al, proposed a model to break a Data Encryption Standard as an alternative approach for data hiding technology. Mohammad Reza Abbasy et al. proposed an information hiding methodology in which message was encoded and decoded efficiently by using the DNA sequence.

5. Proposed System

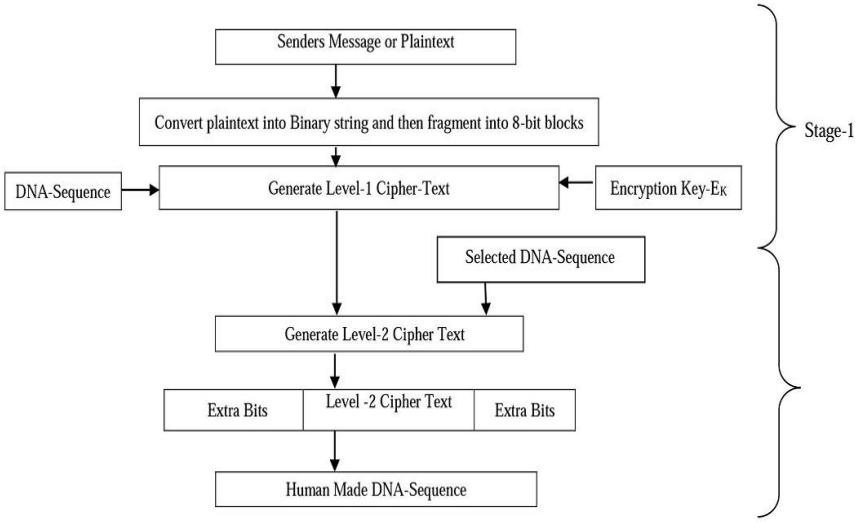
The security of information flowing over the internet is crucial nowadays. With the growth of information technology the number of threats a user is supposed to face with grew exponentially. In order to provide good security and reliable data transmission, a new method of encryption process is proposed here. This method uses DNA sequence which composed of four bases- adenine-A, cytosine -C, guanine-G and thymine-T. It is the arrangement of the bases in a sequence, for instance like ATCCAGTT, that determines the encoded gene. The natural sequence pattern with complementary coding can be used to shield the message.

Coding DNA Bases	Decimal Numbers	Corresponding Binary digits
A	0	00
C	1	01
G	2	10
T	3	11

5.1 Research Method

The proposed algorithm contains two stages. Stage-1 uses encryption key -K_E and **C**ipher **B**lock **C**haining mode CBC- is a mode of operation for a block cipher. The Encryption key is calculated based on a number generated randomly at source side. After that convert a plaintext into 8-bit binary blocks and then every block of plain-text and

encryption key-K_E will be send it to CBC mode. The outcome of one block of CBC will be used as the input (key) for the next block. In stag-1 each and every block will generates intermediate form of cipher-text also called as level-1 cipher-text. Each block of intermediate form of cipher-text is converted into DNA sequence by applying DNA encoding scheme. And then apply DNA complementary rule on generated DNA sequence and convert that DNA sequence into binary format this is called as level-1 cipher-text. In Stage-2, sender has to choose a selectively reference DNA-Sequence randomly from publicly available DNA- sequences. The selected DNA-sequence will be act as a one of the key for the encryption stage-2. Receiver must have the information about the selected and used DNA-sequence. Then this selected DNA-sequence is converted into binary string using binary coding scheme. After that converts a binary string into K-bit blocks. Both blocks of stage-1 and stage-2 will be combined together in such a way that each block of the stage-1 is to be appended in front of the each k-bit block of stage-2. When the length of the k-bit block of stage-2 is less than the length of the stage-1 block, the stage-2 k-bit block will be repeated that is level-2 cipher-text. And then change the level-2 cipher-text into a faked-DNA- sequence also referred as human-made DNA-sequence using binary coding scheme of DNA.



Proposed Algorithm Flowchart

7. Algorithm-1 for implementing Encryption

Steps for Encryption (Stage-1)

Step 1: The plaintext is used under the form of ASCII code (in decimal format).

Step 2: Stage-1 includes following

- Generate a common Encryption-key (E_K) (by using Random number Generator).
- Select Complementary-Rules.
- Select DNA Binary-coding Schemes.
- All are shared between sender and receiver prior to communication (16-bit).

Step 3: Compute Session Key (S_K).

- Compute Session Key based on secret-number-(s) and Random Number (Rn)

Step 4: Compute Encryption Key (E_K) for stage-1 Encryption

Step 5: Compute Level-1 Cipher-Text

- Generate: Level-1 cipher-text by Performing XOR Operation on Step 1 and Step 2.

Steps for Encryption (Stage-2)

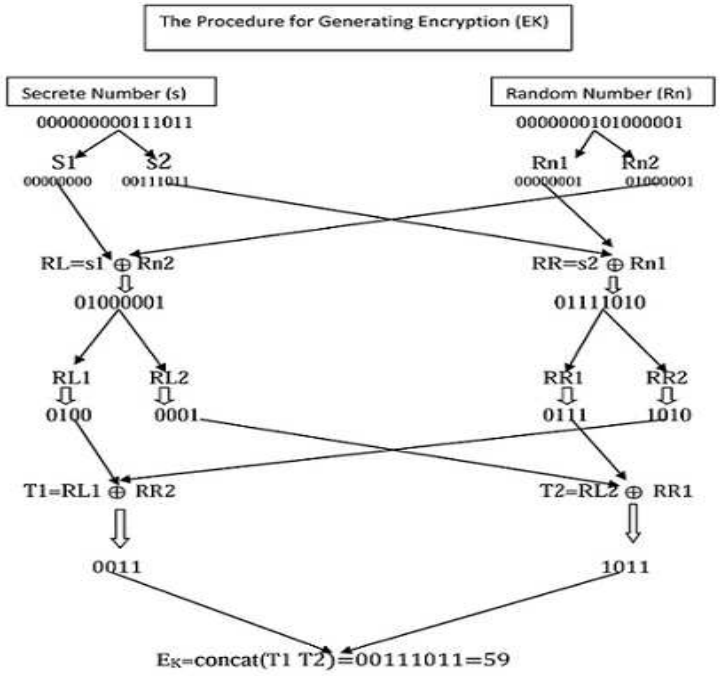
Step 1: Select DNA-sequence (S).

- Change the DNA-sequence *S* into a binary string by using the binary-coding-scheme.
- Fragments the binary string into K-bit blocks. (Select k randomly).

Step 2: Fragment the level-1 cipher-text into K-bits block.

Step 3: Encrypt the DNA-Sequence.

- Append the K-bit block of level-1 cipher-text once at a time, in-front of the of each K-bit block of S
- Concatenating the above segments to generate level-2 cipher-text.
- Use the binary-code-scheme to changes the level-2 cipher-text into new DNA-sequence.



7.1 Algorithm – 2 for Generating Session Key and Encryption Key

Find the secret-number(s)-(16-bits) is given to receiver before communication takes place.

Sending end Computations for Encryption Key - E_K:

Step 1: Sender will select a random number-(R 16- bits) by using random number generator

Step-2: Fragment the random number 'R' into 2 equal parts each of which having 8-bits- (R1 & R2).

Step-3: Fragment the shared secret-number (s) into 2 parts each of which having 8-bits (s1 & s2).

Step-4: Perform X-OR operation on s1 and R2 and s2 and R1. R1= s1 ⊕ R2, R2= s2 ⊕ R1

Step-5: Divide the R1 into RL1 and RL2 and R2 into RR1 and RR2 having 4-bit each.

Step-6: Perform 4-bit EX-OR Operation on RL1 and RL2 and between RL2 and RL1. T1=RL1 ⊕ RR2 T2=RL2 ⊕ RR1

Step-7: Generate Encryption key-E_K for stage-1 $_{by}$ performing concatenating operation on T1 and T2. E_K = concate (T1, T2).

Step-8: Session-key (S_K) computation for Stage-1

Calculate the S_K= s ⊕ Rn

Session-key computation

Divide's' by using 16 and convert the remainder into its equivalent hexadecimal form and keep it in 'S_K1' and Divide the result once again by 16, keeping hex form of the remainder in 'S_K2'. Do until the result is less than 16 (S_K1, S_K2, S_K3….. S_{Kn}).

t= s ⊕ Rn=123 ⊕321=39483 S_K1 =39483%16=11=B (hex form)
t=39483/16=2467 S_K2=2467%16=3=3 (hex form)
t=2467/16=154 S_K3=154%16=10=A (hex form)
t=154/16=8 S_K4=8=8 (hex form)
S_K=B3A8

Step-9: Make together all the 'S_Ks' in order to get theStage-1 session-key S_K.
Step-10: Send 'S_K' as Stage-1 session-key by using a secure communication channel along with Stage-2's session-key.
End

7.2 Receiving end Computations

Step-1: Computation of decryption key-D_K as follows:

Requirements:

1. Shared Secret-number-'s'
2. Session-key 'S_K'

Step-2: Separate all the digits of 'S_K' and then convert them into decimal form.

Ex: S_K = B3A8 (say), S_K1=B, S_K2=3, S_K3=A, S_K4=8

t = S_K-n
t = (t * 16) + S_K-n-1 continue up to S_K-1
Random number, Rn = t ⊕ s
Ex: t1= (8*16) +10=154
t2= (154*16) +3=2467
t3= (2467*16) +11=39483
Rn = (123 ⊕ 39483) =321

Step-3: Fragment the 'Rn' into 2 equal parts each of having 8-bit. (Rn1 and Rn2) and divide the shared secret number into 2 parts as s1 and s2 both of having 8-bit each.
Step-4: Perform X-OR operation on following:
R1= s1 ⊕ Rn2 R2= s2 ⊕ Rn1
Ex: R1=0 ⊕ 65 = 65 R2 = 123 ⊕ 01 = 122

Step-5: above two outputs R1 and R2 again sub-divided into 2 parts namely R11, R12 and R21, R22 having 4-bit each.
Step-6: Perform X-OR operation on following T1=R11 ⊕ R22 T2=R12 ⊕ R21. T1=4⊕ 7= 3 T2= 01 ⊕ 10 =11
Step-7: if we concatenate both T1 and T2 will give Decryption key-D_K of 8-bit for Stage-1. D_K= concate (T1, T2)
Decryption key, KD= concate (0011, 1011) = 00111011 = 59
End

8. Conclusion and Future Scope

We must identify that the DNA-sequence 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 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.

In this system, we use chaotic encryption for encryption systems dealing with plaintext. This encryption 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] Beenish Anam, et al- proposed a "A Secured Technique for Image Data Hiding", Communications in Computer and Information Science, June-2009, Vol. 29, pp. 151-159.

[2] L. M. Adleman- DNA computation of solution to combinatorial problems, Nov-1994.

[3] Snehal Javheri et al-Sinhgad Institute of Technology, Dept of Computer Engineering, University of Pune.

[4] A. Khalili, et al-proposed a "Toward Secure key distribution in Truly Ad-Hoc Networks,," in proceedings of the 2003 IEEE Symposium on Applications and the Internet -2013.

[5] Ashish Gehani, et al- DNA based cryptography, IEEE transaction on DIMACS DNA Based Computers -2000.

[6] Ashish Gehani, et al-"DNA-based Cryptography," Annual MIMACS Meeting on DNA Based Computers(DNA 5), MIT, June 1999.

[7] Ashish Gehani et al, "DNA-based cryptography", Lecture Notes in Computers Science, vol.2950,pp.167- 188,2004.

[8] Atul Kahate,"Computer and Network Security".

[9] Beenish Anam, Kazi Sakib, et al- "Review on the Advancements o DNA Cryptography"

[10] Bibhash Roy, Gautam Rakshit, Pratim Singha, et al- "An improved Symmetric key cryptography with DNA Based strong cipher"-*IEEE trans. ICDeCom-2011*, Feb' 24-25'2011, pp.1-5.

[11] Bibhash Roy et al, "A DNA based Symmetric key Cryptography," IEEE trans.

[12] Mohammadreza Najaftorkaman 1, Nazanin Sadat Kazazi 2- A Method to Encrypt Information with DNA-Based Cryptography.

[13] Anupriya Aggarwal et al- Secure Data Transmission Using DNA-Encryption.

[14] Radu Terec, Mircea-Florin Vaida., "DNA Security using Symmetric and Asymmetric Cryptography", 2011-ISSN 220-9085.