Volume 3, Issue 3, March 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

DNA Based Cryptographic Techniques: A Review

Kritika Gupta*
Department of Computer Science and Engineering, PEC university of Technology.

Shailendra Singh
Department of Computer Science and Engineering, PEC University of Technology.

Abstract— Today, information has become very important resource and so is its security. Many traditional mathematical algorithms have been developed for encrypting the information or data for security purposes but they have limitations. DNA encryption, on the other hand, is much more effective as it has got much more storage and computing capabilities. DNA encryption comes from DNA computing, initiated with the idea of "computing using DNA not on DNA". This approach was further extended to solve other problems. A lot of work have been done in the area and many researchers have done encryption based on different techniques like-DNA digital encoding, Polymerase Chain Reaction (PCR), DNA synthesis, Electrophoresis etc.

Keywords— PCR, OTP, Steganography, Message Integrity, Authentication.

I. INTRODUCTION

Deoxyribonucleic acid, DNA is made up of two twisted strands composed of four bases, Adenine (A), Cytosine (C), Thymine (T) and Guanine (G). The four bases represent the genetic code. A sequence of DNA contains these four nucleic acid bases where 'A' and 'T' are complementary, and 'G' and 'C' are complementary. DNA cryptography involves encrypting or encoding the data using DNA computing techniques through which scientists can synthesize, amplify, isolate, digest and sequence DNA strands very easily. The extraordinary parallelism, storing and computing capabilities of DNA are also available for some other cryptographic purposes like- authentication, signature, etc.

In 1994, Adleman used DNA computing to solve an instance of the directed Hamiltonian path problem [1] and that was further used by Lipton to solve SAT problems[2]. Because of powerful computing capabilities, DNA was again used by Ashish Gehnani. This time they did encryption using DNA. Ashish Gehani with Thomas LaBean, and John ReifThis[3], in 2000, presented an initial investigation into the use of DNA-based information security and discussed two classes of methods: (i) DNA cryptography methods based on DNA one-time pads, and (ii) DNA steganography methods. They again mentioned detail procedures for two DNA one-time-pad encryption schemes: (i) a substitution method using libraries of distinct pads; and (ii) an XOR scheme utilizing molecular computation and indexed, random key strings. These methods can be applied either for the encryption of natural DNA or for artificial DNA encoding binary data [3].

II. ENCRYPTION BASED ON DNA MANIPULATING TECHNIQUES

DNA cryptography utilizes biological methods for encryption and decryption. Among those, Polymerase Chain Reaction (PCR) and DNA chip technology are the most prominent cryptographic techniques. All these techniques are described below.

PCR (Polymerase Chain Reaction) is a fast DNA amplification technology based on Watson–crick [4] complementary model. The purpose of designing PCR is to increase the amount of DNA, as it is very difficult to deal with small amount of DNA strands. PCR can select small strands of DNA and amplifies them. For performing PCR, one should know the sequence of DNA to be amplified and design the right primer for it, where primer is a sequence containing few numbers of nucleotides complimentary to the specific sequence of DNA which is to be amplified.

DNA digital Encoding, we often need to encode the DNA sequence in some kind of readable form. In the field of information science, the most basic encoding method is binary encoding in which we represent each nucleotide bases by any combination of 0 and 1. DNA chip technology is used to manipulate biological data. Tens of thousands, even millions of DNA probes are arranged in a square area less than 1 square inch on glass or silicon matrix. These chips like silicon chips can be used to handle and store the data in the form of DNA sequences. Input and output of the DNA data can be moved to conventional binary storage media by DNA chip arrays.

Ashish Gehnai, Thomas H. Labean and John H. Reif present a DNA cryptosystem for 2D images using DNA chip and randomly assembled one time pad. They did encryption and decryption of input and output data in the form of 2D images recorded on the microscopic array of a DNA chip. Again an encryption scheme have been designed by Guangzhao Cui, Limin Qin, Yanfeng Wang and Xuncai Zhang[5] based on DNA technologies for DNA encryption. They preprocess the data first to get completely different ciphertext for same plaintext to prevent attack from a possible word as PCR primers. Secondly, the DNA digital encoding technique has been applied to the ciphertext. After coding sender synthesizes the secret-message DNA sequence which is flanked by forward and reverse PCR primers, each 20-mer oligo nucleotides long. Thus, the secret-message DNA sequence is prepared and at last sender generates a certain number of dummies and

puts the sequence among them. Once the data gets in encrypted form and reached to receiver's side the same procedure, in reverse, is followed to decrypt it.

Similarly many algorithms have been developed and some of them like-BDEA (Bi-serial DNA encryption Algorithm) [6] used mathematical operations with DNA manipulating techniques for encrypting data. Traditional encryption algorithms like-RSA, DES, AES etc. are based on strong mathematical operations and are, up to some extent difficult to break. So, DNA cryptography came with a belief that they will create a bridge between existing and new technology and will together form a hybrid cryptosystem.

III. ENCRYPTION BASED ON DNA SEQUENCES

Here some basic steps, to encrypt or to hide the data into DNA sequences are followed. At first the binary data, text or image, is transformed into its ASCII code. Then the code again gets converted into binary form and from there the whole string gets substituted into a DNA sequence. Afterwards the encryption algorithm is applied to resulted DNA sequence and a key is used in the process. The key used is a random DNA sequence which can be generated by random key generator or be picked up from DNA sequence database like-GENEBANK. The ciphertext is then transferred to the receiver along with the key. So the receiver apply the decryption algorithm on the received cipher DNA sequence and the key, to get the plain DNA sequence which again gets converted into binary form and then to the actual data.

Leier[7] designed an encryption scheme according to which a binary encoded plaintext message strand is mixed with some dummy strands to fulfill information hiding based on DNA binary sequences. An index-based symmetric DNA encryption algorithm has been proposed by Zhang Yunpeng and Wang Zhong[8] in 2011 4th International Congress on Image and Signal Processing Conference. The algorithm converts each character into ASCII then to the DNA sequence. A special DNA sequence is selected as the encryption index and the key is generated for the whole procedure. Afterwards they applied some mathematical operation to the encryption process and send to the receiver's end where the whole process is repeated but in reverse order for decryption.

Likewise some more algorithms have been proposed for encryption based on DNA sequences. Deepak Kumar under the guidance of Dr. Shailendra Singh [9] has proposed a method for secret writing using DNA sequences. He hide the plain DNA sequence in other DNA sequence with the help of a key which gets transmitted along with the cipher DNA sequence to the receiver.

IV. PROPOSED Algorithm

Here an encryption algorithm based on OTP (one-time-pad) has been proposed using DNA sequence. So far we have come across many ideas which involve data encryption using traditional mathematical operations and/or data manipulating DNA techniques. But once an encryption algorithm has been applied and the data is transmitted on the transmission media: there's a possibility that the data, even if in the cipher form gets manipulated by any interceptor. To avoid that we check our data for any kind of manipulation at the receiver's end.

A. Steps for the encryption

1) The plaintext is needed to be represented into its ASCII code first.
2) ASCII code is then gets converted into binary to get the data in the form of 0's and 1's.
3) Now the binary values are encoded in DNA sequences using table (Table 1) of binary to nucleotide conversion where each of the four bases is represented by combination of 0 and 1.
4) After, a DNA sequence is selected as a key and grouped in the blocks of 8 characters each.
5) Then a table is created based on the positions of each character in the key sequence.
6) Based on that table and the randomly selected DNA sequence data gets converted into encrypted form(Figure 1).
7) The encrypted sequence with the key is send to the receivers end.

Table 1: Nucleotide to Binary Conversion

Nucleotide	Binary Form
A	00
C	01
G	10
T	11

B. Steps for Decryption

The cipher sequence along with the key are received here and the decryption algorithm is applied to find the actual DNA sequence hidden in the cipher DNA sequence.

1) The DNA sequence is then gets decoded into binary using the same conversion table.
2) Binary is converted into ASCII and ASCII to the actual text.

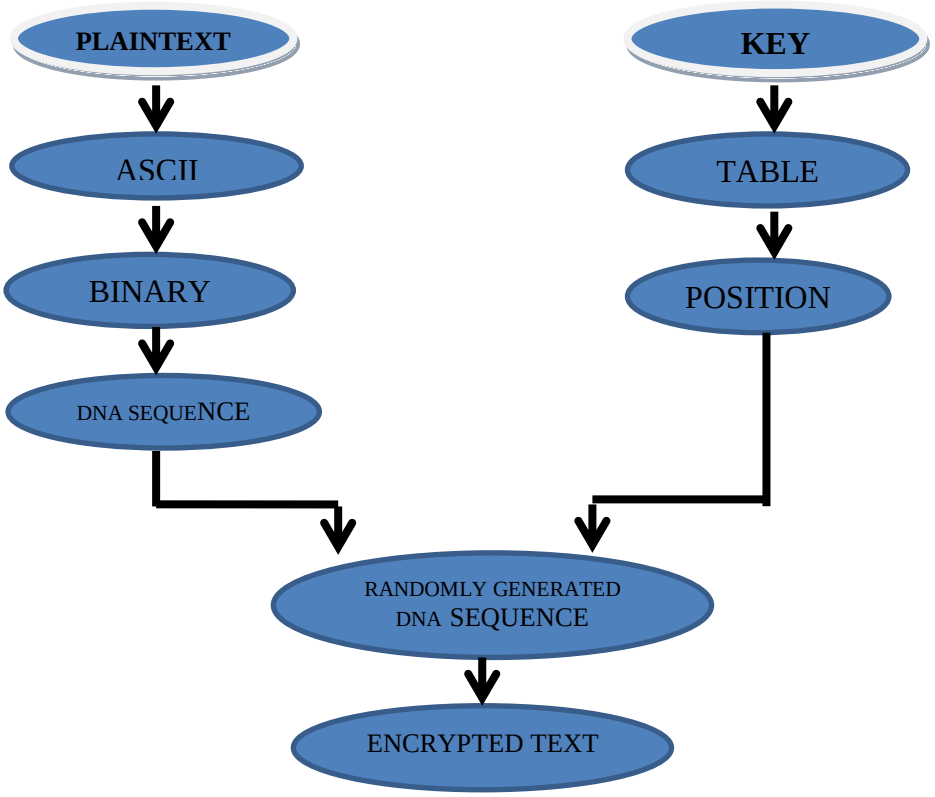

Figure 1.

C. The Key

The proposed method is based on OTP (one-time-pad) encryption where the key is secret, is random, as large as or greater than the plaintext and never reused in whole process. For large plaintext, a large key space is needed which is pretty impossible in traditional algorithm. But DNA has enormous storage capacity and it could be used for storing large amount of data. The key is first grouped into a blocks of 8 characters and each block is used once to encrypt the data.

In the whole process what if our data, at the time of transmission, gets stolen or manipulated by any interceptor? Decryption need keys and algorithms but alterations don't. So anyone can manipulate the ciphertext and can disturb the message integrity, the assurance that data received are exactly as sent by an authorized entity or contain no modification, insertion, deletion, or replay. Again there are situations when we need to make sure that the communicating entity is the one that it claims to be and there we authenticate the data.

V. Steganography Based On Dna Sequences

Steganography means covered writing in which the actual data appears to be something else. The main advantage of steganography over cryptography is that the message does not attract attention to it as it remains hidden in some other data. A few stenographic algorithms have been proposed for data, images, audios etc. Suman Chakraborty with the guidance of Prof. Samir K Bandyopadhya[13] design the algorithm called 3 level image steganography based on DNA sequences for the same. At first level the message gets embedded into the cover image then the stego-image is converted into DNA sequence and finally the compression of the sequence takes place. Similarly Amal Khalifa and Ahmad Atito[14] hide the message in two levels. Initially the DNA-based Playfair cipher is applied for encryption of the message and then a substitution method is used to hide the encrypted message into some other DNA. Likewise they, Suman Chakraborty, Samir K Bandyopadhya and Sudipta Roy [15] came up with another scheme for image steganography based on Sudoku matrix.

VI. Conclusion And Future Scope

Here an attempt is made for the message encryption along with the idea of adding authentication and message integrity. Encryption can be applied before or after authentication to maintain data confidentiality and data integrity so that no one other than the intended receiver can read or modify the data. Yes, the research of DNA cryptosystem lacks practical implementation but the future of this area looks very promising, seeing as DNA is a medium for ultra-compact information storage.

References

[1] L. Adleman, *Molecular Computation of Solutions to Combinatorial Problems*, Science, vol. 266, pp. 1021-1024, 1994.

[2] R. J. Lipton, *Using DNA to solve NP complete problems Science*, Vol. 268, pp. 542-545, 1995.

[3] A. Gehani, T. LaBean, and J. Reif, *DNA-based Cryptography*, Lecture notes in Computer Science, Springer, 2004.

[4] J. D Watson, F. H. C. Crick, *A structure for deoxyribose nucleic acid*, Nature, vol. 25, pp. 737-738, 1953.

[5] G. Cui, L. Qin, Y. Wang, and X. Zhang, *An encryption scheme using DNA technology*, in IEEE 3rd International conference on Bio-Inspired Computing.

[6] D.Prabhu, M.Adimoolam Lecturer, *Bi-serial DNA Encryption Algorithm (BDEA)*.

[7] A. Leier, C. Richter, W. Banzhaf, H.Rauhe, *Cryptography with DNA binary stands*, BioSystem, vol. 57, pp. 9-22, 1997

[8] Zhang Yunpeng, Wang Zhong, *Index-Based Symmetric DNA Encryption Algorithm*, School of Software and Microelectronics Northwestern Polytechnical University Xi'an, Shannxi, 710072, P.R.China.

[9] Deepak Kumar, Shailendra Singh, *Secret Data Writing Using DNA Sequences*, Department of Computer Science & Engineering PEC University of Technology Chandigarh, India

[10] William Stallings. Cryptography and Network Security, Principles and Practices, Forth Edition, Prarson Education, 2008.

[11] G. Z. Cui, L. M. Qin , Y. F. Wang and X. C. Zhang, *Information Security Based on DNA Computing*, 2007 IEEE International Workshop on Anti-Counterfeiting Security, Identification, 2007, pp. 288-291.

[12] Beenish Anam, Kazi Sakib, Md. Alamgir Hossain, Keshav Dahal, *Review on the Advancements of DNA Cryptography*, School of Computing The Bradford University West Yorkshire, UK

[13] Suman Chakraborty, Prof. Samir K Bandyopadhyay,*An approach of 3-level image staganography using dna sequence and 3×3 matrix pixel pair differencing algorithm*, Kolkata, India,

[14] Amal Khalifa, Ahmed Atito, *High-Capacity DNA-based Steganography*. Faculty of Computer Science, EGYPT.

[15] Suman Chakraborty, Sudipta Roy and Prof. Samir K. Bandyopadhyay, *High-Capacity DNA-based Steganography*, Kolkata, India.

