ENCRYPTING INFORMATION USING DNA SEQUENCES WITH MATRIX ALGEBRA

Dr. K. Menaka
Assistant Professor
Department of Computer Science
Shrimati Indira Gandhi College, Tiruchirappalli – 2.
(E-mail: kmenakasigc@gmail.com)

***Abstract*—**Cryptography with DNA sequences is a newly emerging technique that helps for the secured transmission of data. With the help of the biological properties of the DNA sequences, it is possible to enhance the security of a message with minimum cost and reduced computational time. DNA cryptography shows how powerfully the cryptographic methods can be used to work along with DNA sequences. The main aim of DNA computing is to provide more security to the information being transferred with less time and space complexities. Matrix algebra is the most influential among mathematical tools that are available for the investigation of linear networks. This paper proposes a novel idea of using transposition proposition matrix along with DNA sequencing concept for the secure and effective transmission of data

Keywords— DNA cryptography, Matrix Algebra, Finite Difference, Information Encryption.

I. INTRODUCTION

Encrypting information is the order of the day for the purpose of transmitting data in a secure manner. Here, the original message is converted into an equivalent alternative by a suitable encoding mechanism. The encrypted message is then transmitted to the receiver. An encoding scheme by combining the important properties of the biological DNA (Deoxyribinuclei Acid) sequences could provide an effective stealth transmission of data. Thus the hidden data would be more secure that it could not be easily broken. In the proposed algorithm, the message to be encrypted is taken and converted into various representations with the framed complementary rules and the stealth message to be sent is encoded at the sender's side. At the receiver's side, the decoding of information is done and the original message is extracted out.

Many of the available cryptographic algorithms uses either the concept of mathematics or the concepts of physics [1]. Since DNA possesses higher information capacity, it is very well suited for computational purposes [2]. DNA based cryptography thus plays a major role and is over advantageous than other cryptographic methods. In this work, this has been taken into account and proposed a novel algorithm to improve the security and to increase the complexity of an encryption system.

Efficient techniques are needed to perform DNA computing rapidly. Numerical and statistical approaches could provide a better outcome for this challenge. The various functionalities of matrix algebra can also be effectively utilized for facing the above challenge. This paper thus proposes a novel idea which combines matrix algebra with DNA sequences and it is one of such attempts in the above direction. The results obtained show the compactness of this approach and scope for budding applications.

While transmitting information, information security is on the main focus. Over the years, various cryptographic methods were proposed and enormous mathematical computations have also been used for securely transferring the message. Effective encryption algorithms are required in order to make the data more confidential and to improve data security. DNA based encryption method is one of the recent techniques in cryptographic field in which the integrity if the message can be preserved. In this paper a DNA based cryptographic approach is proposed where the message integrity is also preserved. A DNA sequence is composed of four distinct letters, A, C, G and T. Each nucleotide contains a phosphate attached to a sugar molecule (deoxyribose) and one of four bases, guanine (G), cytosine (C), adenine (A) or thymine (T). DNA elements are found in numerous copies, in some cases thousands of copies.

This paper proposes a new DNA cryptographic technique based on DNA encoding and transition proposition matrix. The main idea is to convert the given message into DNA representations after having converted the message into binary and ASCII forms. Complementary rules are then applied and Transition Proposition (TP) matrix is formed. Now the message is in the form of a matrix. The symmetric encryption technique uses a key which is shared by the sender and the receiver. The proposed method uses symmetric key which is also represented in the form of a matrix, The XOR operation is performed between the partially encrypted message and the key. This approach thus systematically preprocesses the message and performs multiple stages of encryption and transmits to the destination.

The intended receiver then decrypts the message and converts into its original form. Any intruder who is trying to read the message will not be able to read it as the message is not in an understandable form to the eavesdroppers. The proposed algorithm is efficient in computation and is also very secure as it has been designed by following the principles of cryptography.

II. LITERATURE REVIEW

With the help of symmetric and asymmetric keys, encryption and decryption of data can be done in DNA computing. By combining the DNA computing with traditional cryptography, hybrid security feature could be achieved [3]. Bibhash Roy et al. proposed several methods on DNA sequencing based encryption and decryption process [4] [5] [6]. Representing the given plaintext (message) into its corresponding ASCII equivalent is the common step done by many researchers while performing DNA computation. Hossain [7] adopted this concept and also performed several rounds with the very essence of cryptographic substitution and transposition techniques. Image encryption has also becoming a familiar approach in the area cryptography.

Chen [8] adopted this method in his work with OTP (one time padding) concept for encrypting a message. The author of this paper, Menaka [9] proposed a scheme for message encryption using the properties of DNA sequences in which complementary and indexing rules have been applied for hiding the message. The author has also proposed a methodology [10] for enhancing the information encryption which combines the features of DNA sequences and the Data Encryption Standard (DES). Indexing technique has been applied over the complementary DNA sequences in an approach by Mohammed Reza Abbasv, Pourva Nikfard et al. [11].

Text hiding was adopted by Amal Khalifa and Ahmed Atito et al.[12] in their work in which complementary rules are applied to the text which is encrypted with amino acids and DNA sequences.

III. DNA / DNA Cryptography

In human body each cell contains a nucleus which characterizes all the physical and behavioral features of human body. They are bundled into chromosomes. A DNA is in the form of a double helix which is made up of two strands in which each strand can have either a Purine or a Pyramidine base. Adenine (A) and Guanine (G) are Purine bases and Thymine (T) and Cytosine (C) are Pyrimidines bases. In a double helix DNA the two strands are joined together and the bases are bonded each other by hydrogen bonds: A with T and C with G, which is called the complementary pairs of DNA strands. Hence, DNA is made of these four characters i.e. <AGCT>.

DNA cryptography which is considered as a division of biological science has large data storage capacity. It stores information of living organisms. The DNA information is unique in all living organisms. The combination of cryptography with molecular neology helps for the secure data transmission. DNA cryptography skill is needed in information security to guard and hide data. DNA molecules have massive parallelism and huge information capacity which proves that they can be used for encrypting messages. The difference between conventional cryptography and DNA cryptography is that the later uses key sequences in the form of DNA sequences like ATGCCAGT [13]. The ciphertext produced during encryption process is also in DNA sequence format and it will be converted back into original plaintext during the decryption stage. The information (i.e. the plaintext) after proceeding through the proposed algorithm in this work produced a new form of encoded message.

IV. PROPOSED METHODOLOGY

The encryption scheme of the proposed method starts by accepting the plaintext from the user. As a first step, all the characters in the obtained plaintext are converted into their equivalent ASCII representation. The converted ASCII form of information is then converted into binary representation. Each and every two digits of the binary sequence is taken and are converted into DNA form of data as per the following table (Table – I).

Table I – DNA codes

Bits	DNA code
00	A
01	C
10	G
11	T

The basic idea behind Table – I is as follows: every bits of the binary code is taken and analyzed. If it is 00, then it is represented as the DNA codon A (Adenine), if it is 01, then the equivalent DNA representation is C (Cytosine), for 01, it is G (Guanine) and for 11, the DNA representation is T (Thymine). The above process is used to convert all the binary information into its corresponding DNA representation. The ultimate goal of this method is to scramble the original data. The next stage of the encryption process is to apply complementary rules. As discussed earlier in this paper, the nucleotides are divided into two classes Purines R= {A, G} and Pyrimidines Y= {C,T}. Based on their chemical properties, they can also be grouped into Amino Group N= {A,C} and Keto Group K= {G,T}. The division can also be made based on the strength of the hydrogen bonds as strong H- bonds and weak H-bonds where S= {G,C} and W= {A,T} as per the suggestions of Ing-an He et al. [14]. Purine and Pyrimidine are two of the building blocks of nucleic acids and they are as shown in Fig. 1.

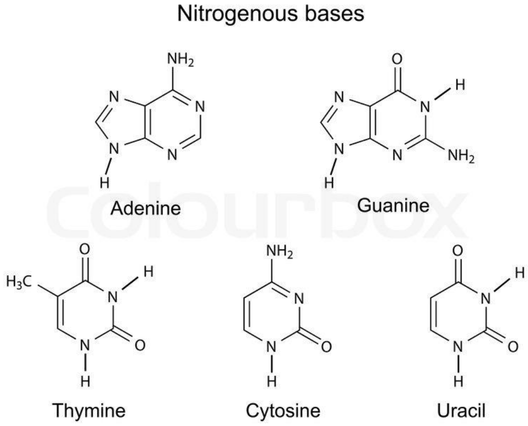


Figure 1 : Structure of Purine and Pyrimidine

The following are the complementary rules adopted in the proposed methodology:
Complementary Rule 1: (AG) (GA) (CT) (TC) → based on Purines and Pyrimidines
Complementary Rule 2 : (AC) (CA) (GT) (TG) → based on Amino and Keto groups
Complementary Rule 3 : (AT) (TA) (GC) (CG) → based on Strong and Weak bonds.

Thus, if the obtained DNA form of the message is GCGA CACC CATA CATA CATT, then after applying the complementary rule (Rule 1), the message becomes, ATAG TGTT TGCG TGCG TGCC.

The next stage is to find the transposition among the letters of the partially encrypted scrambles sequence. After finding the transitions, the Transition Proposition Matrix (4 x 4 TP matrix) is formed. The TP matrix for the above sequence will look like the following (Table II).

Table II – TP Matrix of the Scrambled message

	A	C	G	T
A	0	0	1	1
C	0	1	2	0
G	0	3	0	4
T	1	0	4	2

The algorithm proceeds by generating the key for further encryption of data. The key is generated randomly every time by using random number generator function (number between 0 to 9) and it is also stored in 4x4 matrix for further processing. The final stage of encryption is to perform XOR operation between the scrambled partially encrypted message and the key after converting them to 8- bit binary representation. Both the message and the key are in the form of matrix representation. Hence, each and every cell of the matrices is taken and the XOR operation is performed and the result is also obtained as a matrix. The following table shows the generated key in the form of matrix representation.

Table III – Key Matrix

4	0	1	2
1	4	3	1
0	6	2	1
5	3	1	0

After performing the XOR operation, the following matrix is obtained.

Table IV – Resultant Matrix

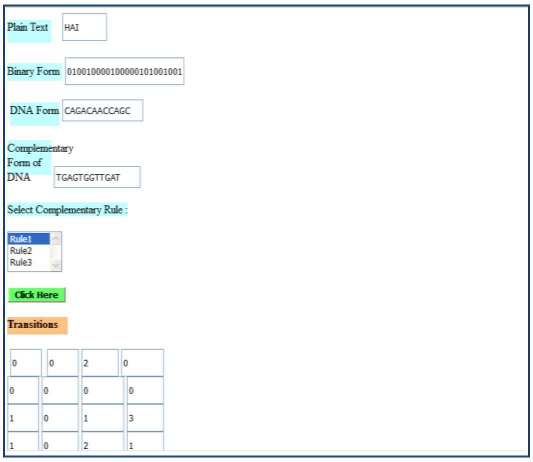
4	0	0	3
1	5	1	1
0	5	2	5
4	3	5	2

I.

V. RESULTS AND DISCUSSION

In this work, information which is to be sent from the sender to the receiver has gone through various phases of conversion in different formats during the process of encryption. Three different secret keys are used in this work which is formed based on the biological properties of the DNA sequences [9] on which complementary rules are formed. Any of these keys (any rule: Rule1, Rule2 or Rule3) can be chosen during the encryption process. Based on the choice of these rules, the very important phase of the encryption process i.e. complementary representation of the message is accomplished in this work. This adds additional complexity and difficulty to the algorithm. The algorithm has been implemented with ASP.NET as front end tool with Intel Pentium Dual Core Processor.

Various Screen shots obtained during the encryption process are shown above. Fig. 2 shows the first among them in which the plaintext is first obtained which is the message to be encrypted. That message is first converted into ASCII form and then to binary and DNA forms. Then the sender has to choose which complementary rule to be used for further encryption of the data which is also be in the form of DNA sequence. The transitions among the sequences are then found and represented as Transition Proposition (TP) matrix. The TP matrix represents the transitions such as A→ A, A→C, A→G, A→T in its first row; C→A, C→C, C→G, C→T in its second row; G→A, G→C, G→G, G→T in its third row and T→A, T→C, T→G, T→T in its fourth (i.e. last row).


Plain Text: HAI

Binary Form: 010010000100000101001001

DNA Form: CAGACAACCAGC

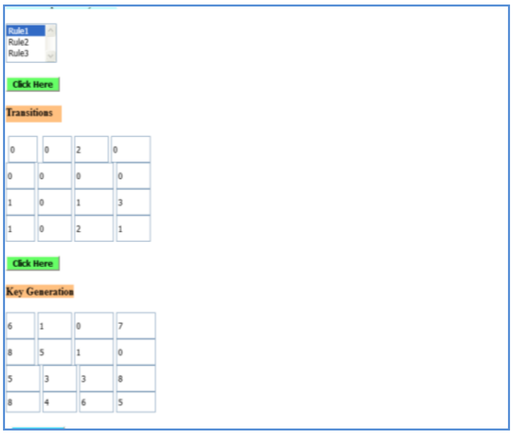
Complementary Form of DNA: TGAGTGGTTGAT

Select Complementary Rule : Rule1

Transitions

0	0	2	0
0	0	0	0
1	0	1	3
1	0	2	1

Figure 2: First Stage of Encryption Process



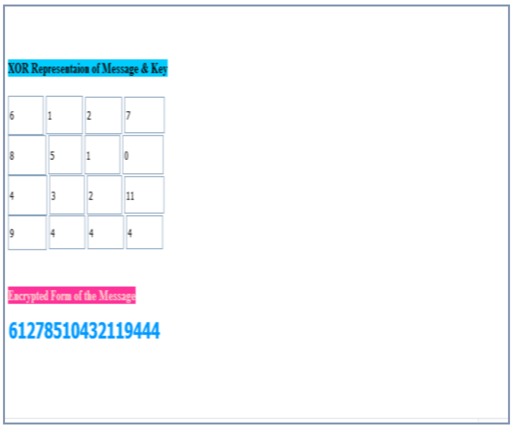
Transitions

0	0	2	0
0	0	0	0
1	0	1	3
1	0	2	1

Key Generation

6	1	0	7
8	5	1	0
5	3	3	8
8	4	6	5

Figure 3: Middle Stage of Encryption Process


XOR Representaion of Message & Key

6	1	2	7
8	5	1	0
4	3	2	11
9	4	4	4

Encrypted Form of the Message

61278510432119444

Figure 4: Final Stage of Encryption Process

Fig. 3 represents the screen shot for key generation. The key is generated randomly at run time. Since symmetric key encryption is followed in this work, the key is shared among the sender and the intended receiver. The final stage of the encryption process is to do XOR operation between the message (which is partially encrypted and in the TP – matrix) and the key (which is also in matrix form). Each and every cell of the matrices is taken and XOR operation is performed. From the screenshot of Fig.4, it is seen that the obtained ciphertext is “61278510432119444” for the plaintext “HAI”. The algorithm has performed several stages of encryption in order to obtain a form of data which will look entirely different from the original message. Also, for a particular message, the ciphertext to be produces will not be the same at all times since choosing of complementary rules may produce a different TP matrix and also the random generation of key will produce different key matrix at different times. The main focus of this method is to use the combined power of DNA sequences with potential properties and the power of matrix algebra. The Complementary rules used intermediately which are formed with the biological properties of the DNA sequences also strengthens the algorithm. For decrypting the message, the inverse process of the encryption is performed by the intended receiver.

VI. CONCLUSION

The unique properties of DBA sequences have been used in this work which helped for fortifying the encryption process of the message. The algorithm went through several steps and hence it is hard for an intruder to break the text. The power of matrix algebra also helped for further encryption of the message. Hence, any intruder who takes the ciphertext will not be able to reveal the true meaning of the message without the random key and without the knowledge of biological properties of DNA sequences and also without knowing matrix algebra concepts. Hence, in this work, the security of the encryption process has been enhanced using multiple stages of encryption.

REFERENCES

[1] Leuenberger, M. N., “Quantum computing in molecular magnets”, Vol . 410, 12 April 2001, Nature – International Journal of Science, pp.789-793, doi:10.1038/35071024..

[2] Adleman, L. M., “Molecular computation of solutions to combinatorial problems”, Vol. 266, Issue 5187, 11 Nov.1984, Science, pp. 1021-1024.

[3] Tushar Mandge Vijay Choudhary, “A DNA Encryption Technique Based on Matrix Manipulation and Secure key Generation Scheme” , ICICES Journal 2013.

[4] Bibhash Roy, Gautam Rakshit, Pratim Singha, Atanu Majumder, Debabrata Datta, ”An improved Symmetric Key cryptography with DNA Based strong Cipher”, ICDeCom 2011, Feb‟ 24 25‟‟2011, pp.1 5.

[5] Bibhash Roy et al, “A DNA based Symmetric key Cryptography”, ICSSA 2011, 24 - 25 Jan‟‟11.

[6] Bibhash Roy, Gautam Rakshit, Pratim Singha, Atanu Majumder, Debabrata Datta, An Enhanced key Generation Scheme based cryptography with DNA Logic”, IJICT 2010.

[7] Hossain, E. M., “A DNA cryptographic technique based on dynamic DNA sequence table” , 18-20 Dec. 2016, 19th International Conference on Computer and Information Technology (ICCIT), IEEE Xplore, Digital Library, pp. 270 – 275, DOI:10.1109/ICCITECHN.2016.7860208.

[8] Chen, J., “A DNA-based, bio molecular cryptography design. Circuits and Systems”, 25-28 May 2003, Proceedings of the International Symposium on Circuits and Systems, 2003. ISCAS '03, IEEE Xplore, Digital Library, pp. 822-825, DOI: 10.1109/ISCAS.2003.1205146

[9] Menaka, K, "Message Encryption Using DNA Sequences”, 27 Feb. – 1 March 2014, World Congress on Computing and Communication Technologies (WCCCT), IEEE Xplore Digital Library, pp. 182-184, DOI: 10.1109/WCCCT.2014.35.

[10] Menaka. K, Enhancing Information Encryption with Biomolecular sequences using NDES algorithm, International Journal of Advanced Research in Computer Science, **Volume 8, No. 9, November-December 2017.**

[11] Mohammad Reza Abbasy, Pourya Nikfard, Ali Ordi Mohammad Reza Najaf Torkaman. “DNA Base Data Hiding Algorithm”, Vol. 2, No. 1, Jan. 2012, International Journal on New Computer Architectures and their Applications, pp. 183-192.

[12] Amal Khalifa and Ahmed Atito, “High-Capacity DNA-based Steganography”, 14-16 May 2012 , 8th International Conference on informatics and Systems (INFOS2012) IEEE Xplore Digital Library, INSPEC Accession Number: 12864186.

[13] M.X. Lu, “Symmetric-key cryptosystem with DNA technology”, Vol.50, Issue 3, June 2007, Science in China Series F: Information Sciences, pp. 324-333.

[14] Ing-an He, Chun Li and Jun Wang; “Finding Protein Coding Genes in the Yeast Genome Based on the Characteristic Sequences”, *Internet Electronic Journal of Molecular Design”, Sep. 2005, Vol. 4, No. 9, Pp. 613-624.*