A Hybrid Cryptographic Algorithm Based on Matrix Algebra using DNA sequences and S-Boxes

[1]Dr. K. Menaka, [2]M. Vishalini

[1]Associate Professor, Dept. of Computer Science, Shrimati Indira Gandhi College, Tiruchirappalli-2

[2]Research Scholar, Department of Computer Science, Shrimati Indira Gandhi College, Tiruchirappalli-2

Abstract-Cryptography using DNA sequences is an emerging area which has facilitated a massive impact in the field of information security. The DNA sequences possess many significant properties which now help the cryptographic algorithms for secure data transmission. This kind of approach in using the properties of DNA sequences for hiding the message is a promising approach in this era. In traditional cryptography, S-Box (also called Substitution Box) is a basic component for performing substitutions in symmetric key algorithms. This paper thus proposes a novel hybrid algorithm which uses the properties of DNA sequences with some concepts of Hill Cipher and S-Boxes combined with the power of matrix algebra for the secure transmission of data.

Keywords- *Data Hiding; Encryption; DNA Sequences; Hill cipher; S-Boxes; Secure Transmission.*

I. INTRODUCTION

In recent days, information has become an important resource and information security is considered to be the most important during transmission of message. The three basic essentials for information security are Confidentiality, Data Integrity and Authenticity. To provide these basics and to hide information during transmission, cryptographic algorithms are generally applied. Cryptographic algorithms hide the message and protect them from public unauthorized access. Data sent by the sender can be secured during transmission by using cryptographic techniques. For securing the data during transmission, there are many cryptographic methods available. DNA (Deoxyribinucleic Acid) cryptography is now emerging very fast and it possess high information density [1]. This very important feature of DNA makes us to use it cryptography. Messages from the sender are thus being encrypted and are being hidden from the intruders and the original receiver only can decrypt the message. The hidden message is thus more secure that it could not be broken easily. In the proposed method, the message to be sent is taken and converted into encrypted form in the sender's side. The original message can be retrieved by decoding this message at the receiver's side.

Conventional cryptography has many algorithms which help for secure transmission of data. One of them is Hill Cipher which is a polygraphic substitution cipher based on the concept of linear algebra. It was invented by Lester S. Hill in 1929. It uses a scheme in which each letter in the English alphabets is represented as numbers like, A=0, B=1, …*Z*=25. The message to be encrypted is first represented in the equivalent decimal form and then represented as matrix. This matrix representation of message is then multiplied with the key matrix against modulo 26. For decryption of message, each block is multiplied by the inverse of the key matrix used for encryption. Another interesting concept in traditional cryptography is the S-box (Substitution Box). It is a basic component in symmetric key algorithms which performs substitution. An S-box takes some number of input bits 'n' and converts them into some other number of output bits 'm' where both 'm' and 'n' need not be of equal length. An $m \times n$ S-box can be implemented as a lookup table with 2^m words of *n* bits each. For example, it can take a 6-bit input, in which the output is found by selecting the row using the first and last bits and the column is found using the middle four bits. DNA computing helps to improve the performance by using many efficient techniques. The challenge of sending message in a secure manner could be achieved by combining any numerical or statistical method with DNA computing. This could also be achieved by effectively using the concept of Matrix Algebra.

This paper thus proposes a novel idea which combines the some of the concepts of the traditional algorithms like Hill Cipher & DES, Matrix Algebra and DNA sequencing. The main idea in this work is to convert the given message of plain text into a ciphertext in the form of DNA sequences. To achieve this, the algorithm takes several steps during the stage of encryption to make strong and effective ciphertext.

II. LITERATURE REVIEW

In the literature, it is stated that encryption of data during transmission could be achieved impressively by means of DNA computing. Also by combining the traditional cryptography with DNA computing, hybrid security features can be obtained [2]. Chowdhury, S et al. [3] proposed a cryptographic method which uses Hill Cipher for the basic encryption of their message. Adinarayana Reddy K et al. [4] proposed a method which modifies the Hill cipher based on circulate matrices and performs security analysis. Andysah Putera Utama Siahaan [5] proposed a method in which

Genetic Algorithm is used to determine the key for the process of encryption and decryption of message.

The genetic algorithm possesses an evaluation function which gives a key that fits the composition will be effectively obtained. Bibhash Roy et al. proposed several methods on DNA sequencing based encryption and decryption process [6] [7] [8]. Mohammad Reza Abbasy, et al. proposed [9] a data hiding scheme in which data was efficiently encoded and decoded using the properties of DNA sequence. Abhishek Majumdar and Meenakshi Sharma used DNA sequences in the field of cryptography to achieve a high level of encryption for the message being transmitted[10].

III. 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>.

Adleman [1] inspected that these four characters (A,T,C and G) could be used for effective computation and from that inspection, computing in DNA established to commence.

DNA computing uses DNA molecules for encodes genetic information for all living organisms. It is considered to be a highly interdisciplinary study and one of the fastest emerging areas in the fields of both Computer Science and Biology.

Knowledge of DNA cryptography is needed today in the area of information security to the purpose of effective data hiding during the transmission of data. By making use of DNA cryptography, strong data hiding techniques could be devised in order to send the data in a secure manner. This combined approach consists of key generation, encryption and decryption processes [11]. Conventional cryptographic algorithms do not provide much security as the DNA cryptography provides.

Combination of conventional cryptography with molecular biology helps for the secure transmission of data. The skill in DNA cryptography is needed for guarding and hiding the data during communication. DNA cryptography differs from conventional cryptography in the sense that the former uses key sequences in the form of DNA sequences like ACGTAGCT. The resulting ciphertext will also be in the form of DNA sequences which is then converted into original plaintext during decryption process.

IV. PROPOSED METHODOLOGY

The proposed methodology provides an encryption scheme which is a novel idea in the area of DNA cryptography. This scheme is a combination of the powerful matrix algebra, some biological properties of molecular sequences and the concept of S-box used in conventional cryptography. It starts by taking the plaintext and converting each and every letter of the plaintext into its decimal equivalent. It is then represented in the form of a matrix.

The key to be used in the encryption process (key1) is also taken in the form of a matrix and it is constructed randomly during every run. This key is a symmetric key which will be shared by the sender to the receiver in a secure channel as in traditional cryptography. The key matrix (KM1) is taken and matrix multiplication against modulo 26 is performed between the decimal form of plaintext in matrix form and the key matrix (KM1). The obtained matrix is termed as Partial Cipher Text (PCT1). Row shifting is performed in the key matrix (KM1) as in Advanced Encryption Standard algorithm of the traditional cryptography and it is taken for the second round of encryption. This key is then multiplied with PCT1 against modulo 26 to obtain Second Partial Cipher Text (PCT2). PCT1 and PCT2 are then multiplied against modulo 26 to obtain the Third Partial Cipher Text (PCT3). The above stages help for confusing the plaintext with more stages of encryption and to make the message more secure and unreadable.

At the final stage of encryption, an S-box is constructed with 15 rows and 4 columns which consist of only unique numerals. The numbers finally obtained in the PCT3 are located in the S-box and the corresponding row and column of the S-box is obtained. The obtained row & column of the S-box is represented in binary form. Each and every two bits of the binary sequence is taken and they are converted into equivalent DNA representation as per Table – I.

Table. I: DNA Based Coding

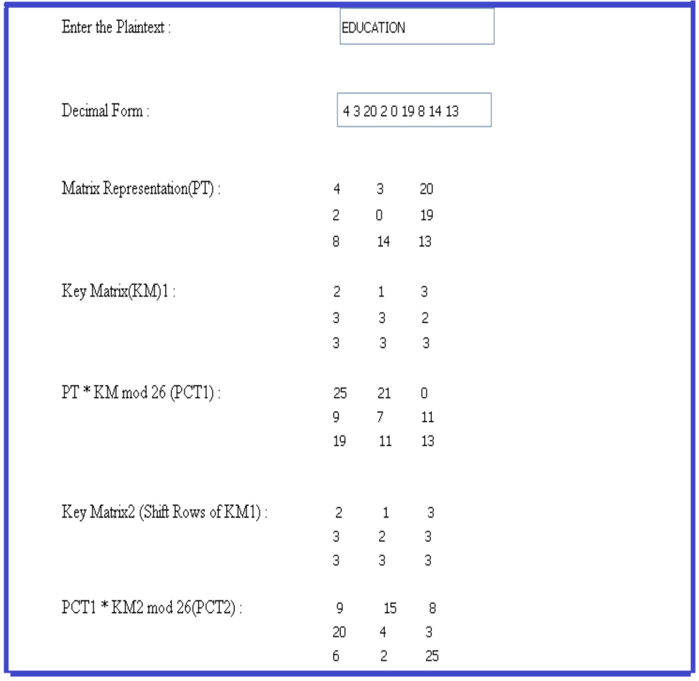
Bits	**DNA Codes**
00	A
01	C
10	G
11	T

The results obtained show the firmness of this method and scope for potential applications in this area.

V. RESULTS AND DISCUSSION

This work possesses several phases of conversion in various formats during the encryption stage for sending the message from the sender to the receiver. The key to be shared between the sender and the receiver is first randomly generated and for making partial cipher text (PCT1), this key is used. During the second stage of encryption, row shifting is done for the key and this helps for the generation of second partial ciphertext (PCT2). The use of two keys during encryption helps to make a strongly encoded partial cipher and hence this adds additional intricacy and obscurity to the algorithm. The proposed method has been implemented with ASP. NET as a front end tool with Intel Pentium Dual Core Processor.

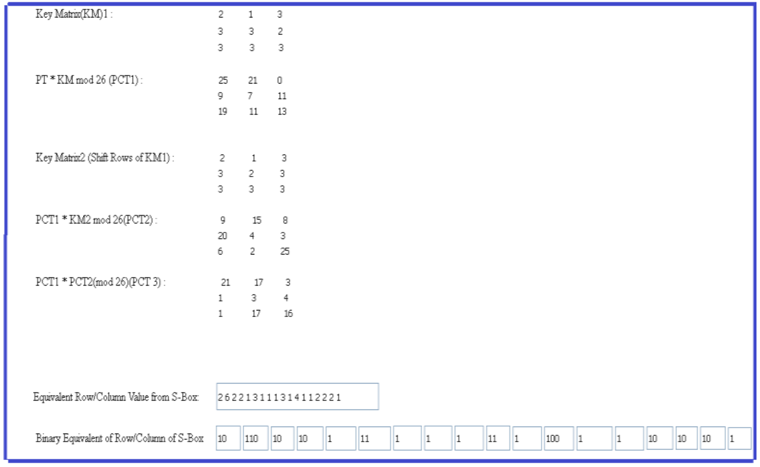
The following are the screenshots obtained during the encryption process. Fig. 1 represents the initial stage of encryption process. This shows the conversion of the plaintext (PT) into its corresponding decimal equivalent and its representation in the form of a matrix. The randomly generated key matrix (KM1) is then multiplied with it and PCT1 is obtained in this stage. PCT1 is then multiplied with key matrix (KM2) which is obtained after performing row shifting operation in KM1 and this brings out PCT2.



```
Enter the Plaintext :                     EDUCATION

Decimal Form :                            4 3 20 2 0 19 8 14 13

Matrix Representation(PT) :               4     3     20
                                          2     0     19
                                          8     14    13

Key Matrix(KM)1 :                         2     1     3
                                          3     3     2
                                          3     3     3

PT * KM mod 26 (PCT1) :                   25    21    0
                                          9     7     11
                                          19    11    13

Key Matrix2 (Shift Rows of KM1) :         2     1     3
                                          3     2     3
                                          3     3     3

PCT1 * KM2 mod 26(PCT2) :                 9     15    8
                                          20    4     3
                                          6     2     25
```

Fig..1:Initial Stage of Encryption Process

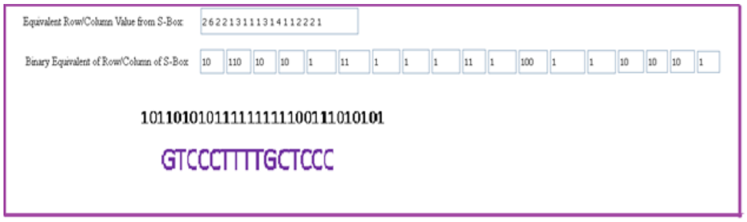
In the third stage of encryption process, the obtained PCT2 is multiplied with PCT1 to get the third partial ciphertext (PCT3). The constructed S-box with 15 rows and 4 columns takes the method to its final stage of the encryption. Each and every cell value in the obtained PCT3 (which is in the form of a matrix) is taken and this value is located in the S-box. The row and column of the S-box which has got the matched value is then taken. The obtained row, column of the S-box is represented in binary form as in the following Fig.2.



```
Key Matrix(KM)1 :                                  2     1     3
                                                   3     3     2
                                                   3     3     3

PT * KM mod 26 (PCT1) :                            25    21    0
                                                   9     7     11
                                                   19    11    13

Key Matrix2 (Shift Rows of KM1) :                  2     1     3
                                                   3     2     3
                                                   3     3     3

PCT1 * KM2 mod 26(PCT2) :                          9     15    8
                                                   20    4     3
                                                   6     2     25

PCT1 * PCT2(mod 26)(PCT 3) :                       21    17    3
                                                   1     3     4
                                                   1     17    16

Equivalent Row/Column Value from S-Box:            2 6 2 2 1 3 1 1 1 3 1 4 1 1 2 2 2 1

Binary Equivalent of Row/Column of S-Box           10 | 110 | 10 | 10 | 1 | 11 | 1 | 1 | 1 | 11 | 1 | 100 | 1 | 1 | 10 | 10 | 10 | 1
```

Fig..2: Second and Third stage of the Encryption Process

The final stage of the encryption is the representation of the binary format of the ciphertext in the form of DNA codons as in Table –I which will be completely in unreadable form to the intruders and this is shown in the following figure.



```
Equivalent Row/Column Value from S-Box:     2 6 2 2 1 3 1 1 1 3 1 4 1 1 2 2 2 1

Binary Equivalent of Row/Column of S-Box    10 | 110 | 10 | 10 | 1 | 11 | 1 | 1 | 1 | 11 | 1 | 100 | 1 | 1 | 10 | 10 | 10 | 1

        101101010111111111100111010101

          GTCCCTTTTGCTCCC
```

Fig.3:Final stage of the Encryption Process

From the above screenshots, it is observed that the algorithm is very strong since it goes through various stages during the progression of encryption. The plaintext taken here is "EDUCATION" which is converted into "GTCCCTTTTGCTCCC" finally. Also, for a particular message, the ciphertext to be produced will not be the same at all times because of the random generation of key matrix at every time. The main focus of this method is to use the combined power of DNA sequences with potential properties and the power of matrix algebra along with S-boxes of the traditional cryptography. The final step of making the partial ciphertext into DNA form of representation strengthens the algorithm. The above screenshots elucidate the generation of ciphertext in all stages while performing the encryption process. Decryption is the reverse of the encryption process.

VI. CONCLUSION

This paper elucidates that this hybrid algorithm of using the power of matrix algebra along with the DNA base codon and the concept of S-box in traditional cryptography is highly effective and it secures the message while it is transformed from the sender to the receiver. Since the algorithm possesses various stages and as it uses many hybrid concepts, it is significantly impossible for the eavesdroppers to break out the message that is being sent. Thus, with this approach the overall security of the message is enhanced during the encryption process and it helps for securely sending any confidential information.

REFERENCES

[1] Adleman, L. M.(1984). Molecular computation of solutions to combinatorial problems. Science, 266(5187), 1021-1024.

[2] Tushar Mandge Vijay Choudhar (2013,February). A DNA Encryption Technique Based on Matrix Manipulation and Secure key Generation Scheme. International Conference on Information Communication and Embedded Systems (ICICES), 2013.

[3] Chowdhury, S. I., Shohag, S. A., & Sahid, H. (2011). A Secured Message Transaction Approach by Dynamic Hill Cipher Generation and Digest Concatenation. *International Journal of Computer Applications*, *23*(9), 25-31.

[4] Adinarayana Reddy K, , Vishnuvardhan B, Madhuviswanatham,Krishna A. V. N. (2012). A Modified Hill Cipher Based on Circulant Matrices. Procedia Technology 4, Published by Elsevier Ltd, 114 – 118.

[5] Andysah Putera Utama Siahaan (2016). Genetic Algorithm in Hill Cipher Encryption. American International Journal of Research in Science, Technology, Engineering & Mathematics, 84-89.

[6] Bibhash Roy, Gautam Rakshit, Pratim Singha, Atanu Majumder, Debabrata Datta (2011, February). An improved symmetric Key cryptography with DNA Based strong Cipher, International Conference on Devices and Communication (ICDeCom), 2011.

[7] Bibhash Roy et al (2011 January). A DNA based Symmetric key Cryptography. ICSSA, 2011.

[8] Bibhash Roy, Gautam Rakshit, RitwikChakraborty (2011). An Enhanced key Generation Scheme based cryptography with DNA Logic. International Journal of Information and Communication Technology Research, 1(8), 370-374.

[9] Mohammad Reza Abbasy, Azizah Abdul Manaf, and M.A. Shahidan (2011 July). Data Hiding Method Based On DNA Basic Characteristics. International Conference on Digital Enterprise and Information Systems, 2011.

[10] Abhishek Majumdar, Meenakshi Sharma (2014). Enhanced Information Security using DNA Cryptographic Approach. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 4(2), 72-76.

[11] M.X. Lu (2007). Symmetric-key cryptosystem with DNA technology. Science in China Series F: Information Science, .50(3), 324-333.