Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 5, Issue. 10, October 2016, pg.123 – 129

Data Security Using DNA Cryptography

Mansi Rathi[1], Shreyas Bhaskare[2], Tejas Kale[3], Niral Shah[4], Naveen Vaswani[5]

[1]*Student, Computer Engineering Department & Mumbai University, Mumbai, India*

[2]*Student, Computer Engineering Department & Mumbai University, Mumbai, India*

[3]*Student, Computer Engineering Department & Mumbai University, Mumbai, India*

[4]*Student, Computer Engineering Department & Mumbai University, Mumbai, India*

[5]*Assistant Professor, Computer Engineering Department & Mumbai University, Mumbai, India*

[1] mansi.june5@gmail.com; [2] shreyasb008@gmail.com; [3] tejk1995@gmail.com; [4] niralshah2395@gmail.com; [5] vaswani.naveen@gmail.com

Abstract --- Throughout history, there has never been a change more dramatic than the one brought about by the ability of humans to communicate. Communication, which initially took place between people at a particular place gradually increased to communication across continents. With the advent of email, which revolutionized the communication field, communication between people across large distances became possible which once seemed a daunting task. This definitely had its share of cons as email could be intercepted by anyone while it is being transmitted from one place to other. To avert this interception, cryptography was evolved. Since its inception, cryptography has been used for various purposes ranging from storing data of a particular system in encrypted form, encrypting data before sending it through a communication medium etc. As this technique evolved the number of attacks and different ways of attacks also evolved. Cryptography has come a long way since it was first used and now a plethora of cryptographic techniques are available in order to encrypt data. In this paper, we present a complex cryptography algorithm which uses DNA sequence in order to encrypt the data. This technique will enable to encrypt the data in a very complex form and thus prove to be a very efficient algorithm with high accuracy. DNA has been widely used as one of the advanced forms to represent information. An efficient algorithm which uses DNA sequencing for cryptography has been discussed in this paper.

Keywords – Cryptography, DNA Cryptography, DNA Sequencing

I. INTRODUCTION

When Julius Caesar sent messages to his generals, he didn’t trust his messengers. So he replaced every A in his messages with a D, every B with an E, and so on through the alphabet [1]. One of the most important needs of human being is to communicate and share information selectively. This gave rise to the art and science of coding the messages and information such that only the authorized or intended people could access this information. Even if unauthorized people got hold of this information it should be such that they could not extract any information from it [2]. In earlier times techniques to encode the information were used by government, military people and during the war so that their information doesn't get leaked to their enemy. But with changing times, we have entered into the era of technology so the need and use of cryptography has also become prominent. Cryptography has led to the growth of e-commerce and internet. The exchange of sensitive information over wired and/or wireless Internet, such as bank transactions, credit card numbers and telecommunication services are already common practices [3].

Cryptography is an art and science of hiding information. It is the ability to send information between sender and receiver such that it prevents others from reading it. A message in its original form is known as plaintext. The encoded information is known as Cipher text. The process of generating cipher text from plain text is called encryption. This includes an algorithm and a secret value called key. Number of key depends on the level of security involved [4]. In Private Key cryptography a single key is used for both encrypting and decrypting. While in Public Key cryptography there are two keys- a public and a private key. Each user has both keys and while the private key must be kept secret the public key is publicly known [5]. Famous Symmetric key algorithms are AES, DES, FEAL, IDEA, BLOWFISH and famous Asymmetric key algorithms are RSA, Diffie-Hellman [6].

In this paper we have used Byte Rotation Encryption Algorithm (BREA). The first step of this algorithm is to break the plain text into blocks of 16 bytes each. Each block is represented as 2D array. The second step is to apply the byte rotation on rows and columns to encrypt the text [7].We have used a symmetric key and we would encrypt it using DNA sequence. In this each letter of the key is converted into different combinations of the 4 bases that make up the DNA. DNA strands contain 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). One of the most emerging techniques in the world of cryptography is DNA cryptography which works on the concepts of DNA computing. Using the biological structure of DNA a new technique for securing data was introduced called DNA computing/biological computing. Advantages of DNA computing include- Speed, Minimal Power Requirements, and Minimal Storage Requirements [8]. A gram of DNA contains about 1021 DNA bases, or about 108 tera-bytes. Hence, a few grams of DNA may have the potential of storing all the data stored in the world [9]. Hiding data in the form of DNA is called DNA cryptography. It is a subject of study about how to use DNA as an information carrier and it uses modern biotechnology as a measure to transfer ciphertext into plaintext [10]. Multiple DNA algorithms that has been studied and researched are Symmetric and Asymmetric Key crypto System using DNA, DNA Steganography Systems, Triple stage DNA Cryptography [8]. So we have used a combination of BREA and DNA sequence in our algorithm to make the encryption assured.

II. RELATED WORK

After DNA was cited as the most advanced form of information representation, many new algorithms were developed and proposed by the researchers in order to ensure data security. This section highlights some of the algorithms which used nucleotide sequence in order to encrypt the digital data.

One of the research suggested using Bi-Serial DNA Encryption Algorithm in which the text message is converted into hexadecimal code and binary code. Now this message is split in two parts out of which one is used as key and the other is used as message and adding to it, XOR operation is also performed in order to increase the compression factor. DNA based coded message is received after applying DNA digital coding and then the PCR amplification is implemented by using two prime pair as key and compression is performed for variable length of data [11]. This algorithm definitely increases the security of the encryption method, but it also increases the computational complexity as it uses two prime numbers using PCR amplification.

Another algorithm which was proposed by Shreyas Chavan was the encryption of plaintext using DNA Cryptography and Binary One Time Pad (OTP) Scheme. This algorithm basically uses two keys; which are used for the encryption on the sender side and decryption on the receiver side. One of the keys is a random string of nucleotides forming a DNA Sequence and the length of this key depends on the length of the plaintext. The second key is Binary Sequence that is used for OTP. The length of binary key is twice the length of the DNA sequence key [12].

Kang Ning put forward the idea of securing data by a method called as Pseudo DNA Cryptography Method. The security approach adopted in this method is of converting the data or text into protein according to the genetic code table and the key will be send to the receiver via a secure channel. Even though the theoretical analysis of this method may be powerful, the length of cipher text is much higher than the plaintext and the partial information that exists after encryption can be compromised easily [13].

Another algorithm proposed on similar lines was that of Kritika Gupta and Shailendra Singh. The algorithm which they proposed comprised of converting the plaintext into its ASCII code which was then converted into binary form. Now these binary values are encoded in DNA sequences. Following that, a DNA sequence is selected as a key and grouped in blocks of 8 characters each. Based on the character positions in the key, a table is created and with the help of table and key, data gets converted in the encrypted form [14].

III. PROPOSED METHODOLOGY

The algorithm which is basically used here is Symmetric Key Block Cipher Algorithm. The key points of this algorithm are discussed below:

1. Each block size in this algorithm is taken as 16 bytes.
2. The size of the key matrix is also the same that is 16 bytes.
3. The values of the Key Matrix are randomly generated and these values range from 0 to 127.
4. In this algorithm, the concept of poly alphabetic substitution is followed.
5. This algorithm also makes use of the Byte-Rotation technique [7].
6. The encryption of the key is done with the help of DNA Sequencing.

The flowchart representation of the encryption process is given as:

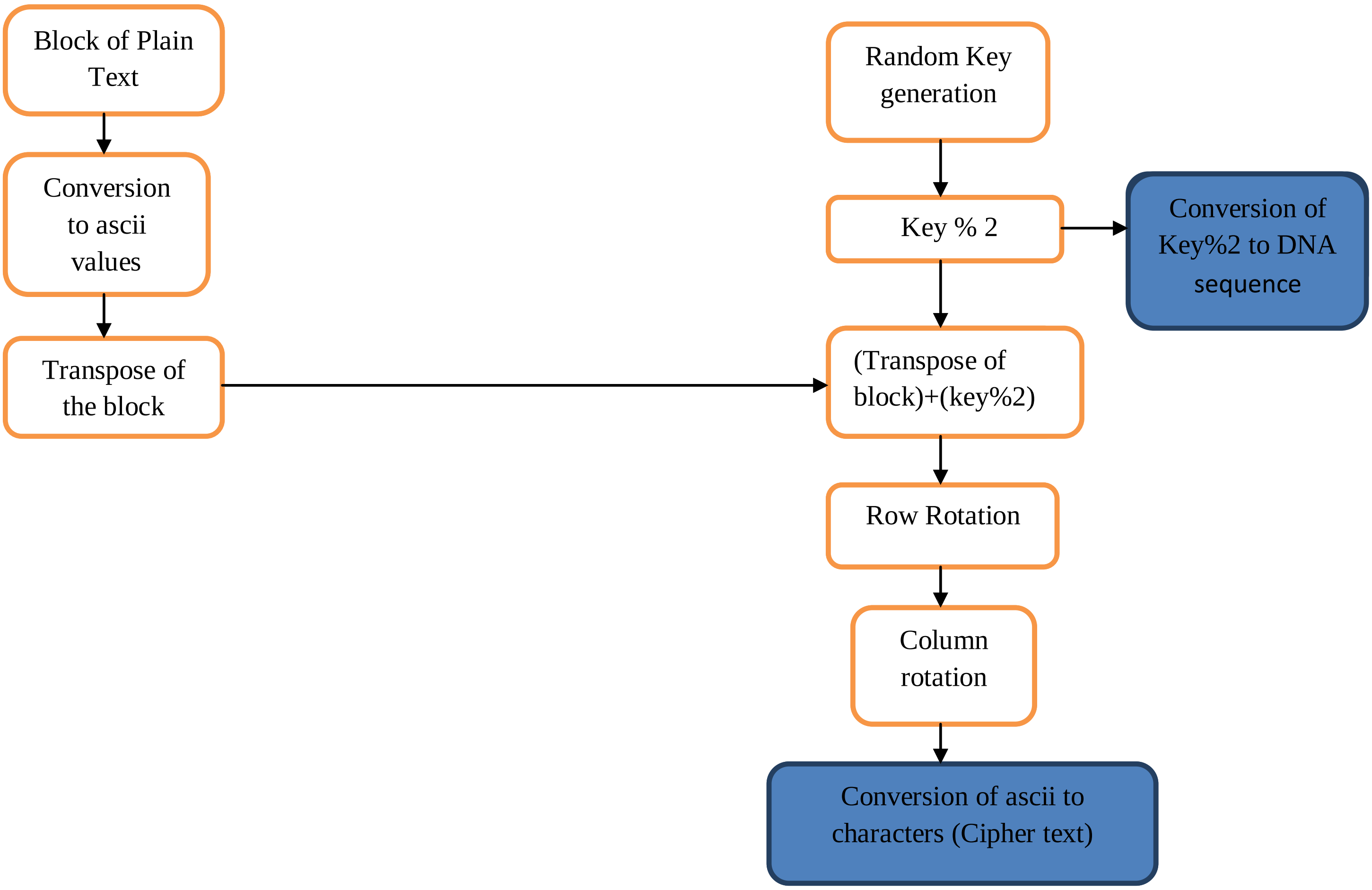

Block of Plain Text
Conversion to ascii values
Transpose of the block
Random Key generation
Key % 2
Conversion of Key%2 to DNA sequence
(Transpose of block)+(key%2)
Row Rotation
Column rotation
Conversion of ascii to characters (Cipher text)

Figure: 1- Encryption Process

The steps to be followed for implementing this algorithm along with example are as follows:

1.Plain Text to Cipher Text conversion:

1.1. The plaintext is partitioned into fixed-length blocks of size 16 bytes each. These blocks are represented by a matrix M. For example- Let the plain text be 'VANHUESENSUITING'.

M =

V	A	N	H
U	E	S	E
N	S	U	I
T	I	N	G

1.2. The letters of alphabet are assigned numerical values based on their ASCII values. If the plaintext is smaller than the block size, then the empty spaces after the complete plain text are replaced by '.' (ASCII value: 46) till the length of the plaintext becomes equal to block size.

M =

86	65	78	72
69	85	83	69
78	83	85	73
84	73	78	71

1.3. Calculate the transpose of the matrix M denoted by Mt.

Mt =

86	69	78	84
65	85	83	73
78	83	85	78
72	69	73	71

1.4. The random key is generated using some random function; the values in the key matrix are generated in the range similar to that of the range of the characters that can be encrypted by this algorithm.

K=[k1,k2,......................., k16] K = Random (0, 127, 16).

K =

108	101	112	52
124	65	79	45
76	60	31	97
75	39	15	09

1.5. Find the matrix Kt using- Kt = K mod 2.

Kt =

0	1	0	0
0	1	1	1
0	0	1	1
1	1	1	1

*(In order to enhance the security, this key Kt is converted into DNA sequence using DNA sequencing/ciphering technique before it is stored or transmitted. The conversion of key (Kt) to DNA sequence (Ke) is explained in section **2.Key Conversion**.)

1.6. Add the matrices Mt and Kt and the resultant matrix is denoted by Mk.

Mk =

86	70	78	84
65	86	84	74
78	83	86	79
73	70	74	72

1.7. Rotate all the rows horizontally of Mk matrix to right such that rotate four bytes from the first row, rotate three bytes from the second row, rotate two bytes from third row and rotate one byte from fourth row. The resultant matrix is denoted by Mr. After rotating the rows the matrix will be-

Mr =

86	70	78	84
86	84	74	65
86	79	78	83
72	73	70	74

1.8. Rotate all the rows vertically downwards of Mr matrix such that rotate four bytes from first column, rotate three bytes from second column, rotate two bytes from third column and rotate one byte from fourth column. Denote the resultant matrix by Mc. After rotating the columns the matrix will be-

Mc =

86	84	78	74
86	79	70	84
86	73	78	65
72	70	74	83

1.9. This matrix values containing the ASCII values are then converted back to their respective symbols or characters to be stored or used further. The resultant matrix is denoted by C.

C =

V	T	N	J
V	O	F	T
V	I	N	A
H	F	J	S

C= 'VTNJVOFTVINAHFJS'

2. Key Conversion:

The key to be stored for decryption of the data is converted using two bytes at a time. The key matrix is Kt.

Kt =

0	1	0	0
0	1	1	1
0	0	1	1
1	1	1	1

The generated key (Kt) is encrypted using the DNA sequences which will be further stored and used for decryption purpose. The conversion of the key is done by considering two bytes at a time, and uses some basic permutation logic. As there are four DNA strands namely A (Adenine), C (Cytosine), G (Guanine) and T (Thymine) there are 256 combinations possible. Since the key is a sequence of pairs of either **00** or **01** or **10** or **11**; these pairs can be substituted by **A or C or G or T** as per the following table**.** The sequence of strands used is taken to be alphabetic and not as per their literal combination.

Sequence	**Nucleotide used**
00	A
01	C
10	G
11	T

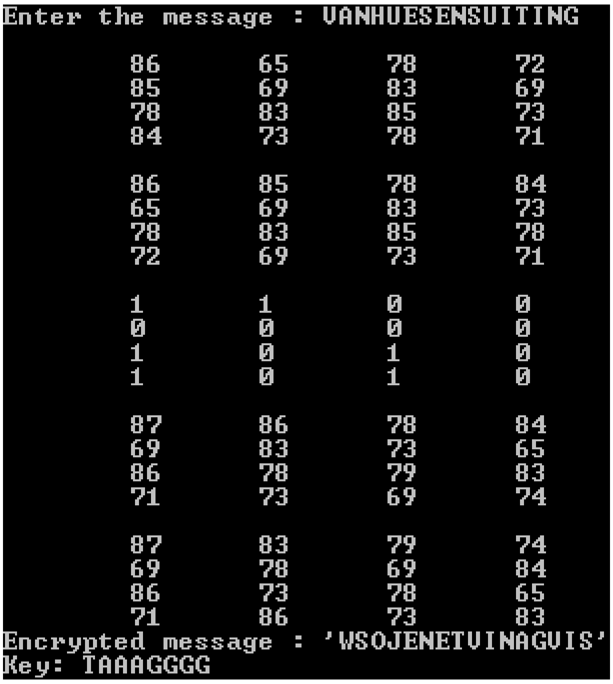
So our encrypted key would be Ke**- CACTATTT**

Decryption Process:

The decryption is just the reverse of the encryption process. In this process, firstly the DNA encrypted key must be obtained to convert it to the normal key matrix which is to be used for decryption purpose by subtracting it from the ASCII values of the cipher text. After that, the reverse column and row rotation is to be performed respectively and finally the transpose of that matrix is to be taken. This matrix thus gives us the original plain text.

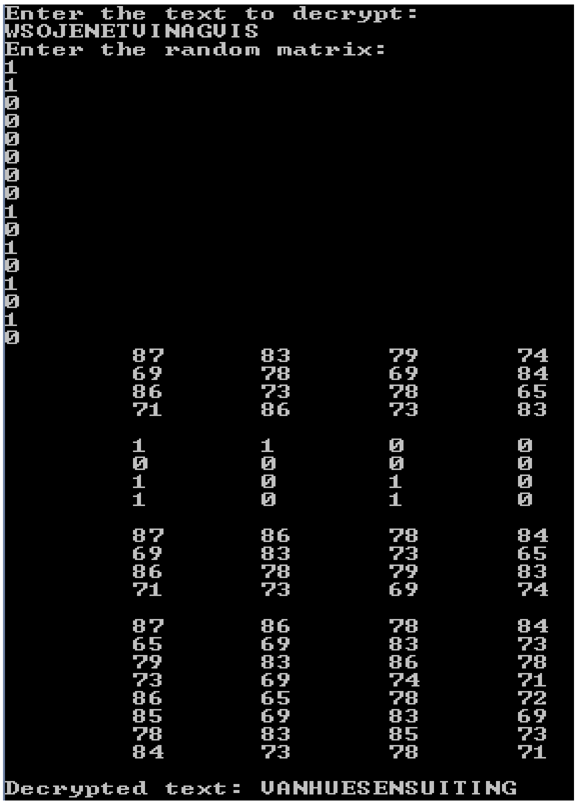
IV. RESULTS

In this section, the results which are obtained after applying the algorithm on a plaintext are displayed. The following results are obtained:



```
Enter the message : VANHUESENSUITING

        86        65        78        72
        85        69        83        69
        78        83        85        73
        84        73        78        71

        86        85        78        84
        65        69        83        73
        78        83        85        78
        72        69        73        71

        1         1         0         0
        0         0         0         0
        1         0         1         0
        1         0         1         0

        87        86        78        84
        69        83        73        65
        86        78        79        83
        71        73        69        74

        87        83        79        74
        69        78        69        84
        86        73        78        65
        71        86        73        83
Encrypted message : 'WSOJENETVINAGVIS'
Key: TAAAGGGG
```

Figure 2 Encryption



```
Enter the text to decrypt:
WSOJENETVINAGVIS
Enter the random matrix:
1
1
0
0
0
0
0
0
1
0
1
0
1
0
1
0
           87        83        79        74
           69        78        69        84
           86        73        78        65
           71        86        73        83

           1         1         0         0
           0         0         0         0
           1         0         1         0
           1         0         1         0

           87        86        78        84
           69        83        73        65
           86        78        79        83
           71        73        69        74

           87        86        78        84
           65        69        83        73
           79        83        86        78
           73        69        74        71
           86        65        78        72
           85        69        83        69
           78        83        85        73
           84        73        78        71

Decrypted text: VANHUESENSUITING
```

Figure 3 Decryption

In figure 2, the plaintext that has been used is ‘VANHUESENSUITING’. After applying the algorithm, we get the encrypted message as ‘WSOJENETVINAGVIS’ and the key generated is TTGGAGAC.

In figure 3, the cipher text used is the same one which is received after encryption in figure 1 which is ‘WSOJENETVINAGVIS’. After applying the decryption algorithm, we get the encrypted message as ‘VANHUESENSUITING’ and the key generated is TAAAGGGG thus proving the correctness of the proposed algorithm.

V. FUTURE WORK

This algorithm can be further enhanced by implementing the DNA sequencing to the cipher text also. This will enable the cipher text to get double encrypted. The areas in which this algorithm can be used are military purposes as any country’s data is very important and confidential. It can be also used for banking applications where it can be used to encrypt the vital data of the customer such as the account number or pin or password.

VI. CONCLUSIONS

In this paper, we have devised a new cryptographic technique. This new cryptographic technique uses DNA sequencing to encrypt the data. The algorithm presented in this paper has been tried and the results are also presented. The results depict the effectiveness of the algorithm presented and is an indication that it can be used in various applications.

REFERENCES

[1] ”*The Basics of Cryptography-Fisher College of Business*”. [Online] Available: https://fisher.osu.edu/~muhanna.1/pdf/crypto.pdf

[2] “*Cryptography Just for Beginners*”. [Online] Available: https://www.tutorialspoint.com/cryptography/cryptography_tutorial.pdf

[3] Rodriguez-Henriguez, F.; Saqib, N.A; Diaz Perez, A; Koc, C.K. “*Cryptographic Algorithms on Reconfigurable Hardware*”

[4] “Introduction to Cryptography”. [Online] Available: http://www.ggu.ac.in/download/Class-Note14/public%20key13.02.14.pdf]

[5] Yaman Akdeniz, "*Cryptography & Encryption*" August 1996, Cyber-Rights & Cyber-Liberties (UK) Available: http://www.leeds.ac.uk/law/pgs/yaman/cryptog.htm.

[6] Professor Guevara Noubir, Northeastern University, “*Fundamentals of Cryptography: Algorithms and Security Services*”.

[7] Sunita Bhati, Anita Bhati, S.K.Sharma, “*A New Approach towards Encryption Schemes: Byte-Rotation Encryption Algorithm*”, Proceedings of the World Congress on Engineering and Computer Science 2012. Vol II WCECS 2012, October 24-26, 2012, San Francisco, USA.

[8] “*The Future of Data Security: DNA Cryptography and Cryptosystems*” [Online]. Available: http://securityaffairs.co/wordpress/33879/security/dna-cryptography.html

[9] Ashish Gehani, Thomas LaBean, John Reif, Department of Computer Science, Duke University, “*DNA-Based Cryptography*”.

[10] Yunpeng Zhang* and Liu He Bochen Fu College of Software and Microelectronics, Northwestern Polytechnical University, Xi’an, China, “*Research on DNA Cryptography*”.

[11] D.Prabhu, M.Adimoolam, “*Bi-serial DNA Encryption Algorithm*” [Online]. Available: https://pdfs.semanticscholar.org/1754/f0eb5852500598a70af4002e186cd2f3c6ce.pdf

[12] Shreyas Chavan, “*DNA Cryptography Based on DNA Hybridization and One Time pad scheme*”, International Journal of Engineering Research & Technology, Volume 2 Issue 10, October-2013.

[13] Kang Ning, “*A Pseudo DNA Cryptography Method*”, **arXiv:0903.2693** [cs.CR]**,** Cornell University Library, March-2009.

[14] Kritika Gupta, Shailendra Singh, “*DNA Based Cryptographic Techniques: A Review*”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3 Issue 3, March 2013.