International Journal of Mathematics and
Computer Science, **14**(2019), no. 4, 953–964

Two layer Encryption Schemes for Symmetric algorithms using DNA Sequences

Siddaramappa Vantigaru[1], Ramesh Krishnarajanagar Basavaraj[2]

[1]Research Scholar
Department of CS&E
R. V. College of Engineering
VTU Bangalore, India

[2] Department of EIE
R. V. College of Engineering
VTU Bangalore, India

email: siddavmk@gmail.com

(Received July 12, 2019, Accepted August 23, 2019)

Abstract

Information security and confidentiality has become important because of the rapid growth of the internet. Advanced information technology and communication devices allowed a person unauthorized access to sensitive data rapidly. There are different techniques used to protect data from an intruder like cryptography and steganography. In this paper, the authors proposed DNA steganography as one layer and symmetric cryptography of DES and AES second layer with parallel programming to increase efficiency of algorithms. The DNA based steganography and Cryptography to protect data from an interceptor and proposed parallel AES and DNA steganography takes less time for encryption(a 20-30 percent time reduction). The proposed algorithms protect information in two fold methods. DNA XOR operation has more secure resistance compared to Binary XOR operation because of the 8 possible combinations for DNA Steganography operation.

Key words and phrases: DNA Steganography, ADECS, DNAXOR,EXON and INTRON.
AMS (MOS) Subject Classifications: 68P25, 94A60.
ISSN 1814-0432, 2019, http://ijmcs.future-in-tech.net

1 Introduction

With increasing progress in digital network, there is need to protect data from attackers on network. The DES was a symmetric algorithm designed by Horst Feistel in the early seventies at IBM and the AES was a symmetric algorithm developed by Vincent Rijmen and Joan Daemen in 1998. DNA is a chemical component present in all living cells. DNA is made up of purines and pyrimidines G,C,T and A to function in living cells. In 1953 James D and Crick identified details how DNA was organized in living cells and got a Nobel Prize in 1962. Watson and Crick identified nucleotide pairs for structure of DNA [1]. Steganography can be thought as hiding a secret message within another message. DNA steganography is hiding a message in DNA sequences.

The NCBI contains a huge amount of data centers and different database resources related to biological makeup. The NCBI database contains different file formats and links to other resources like Medical databases, location of gene and history of related data. It also contains different algorithms related to analysis of data and comparisons. The web server performs tasks and displays results by using remote computation power and resources for a particular parameter. All of the resources can be accessed through the NCBI home page [2]. Figure 1 shows an architecture of how DNA are present in a cell [14].

In this research paper we propose a two layer encryption algorithm. The first layer algorithm implemented is based on DNA cryptography which converts a message into a DNA sequence consisting of A,T,G and C nucleotides letters, A-adenine, T-thymine, G-guanine and C-cytosine. Like computers which deal only with 1's and 0's, a human corresponds to these four letters only. The first layer converts a message into middle cipher text which becomes input for the second layer traditional symmetric algorithms AES and 3DES. In the second layer we have implemented parallel concepts to reduce time taken for encryption and decryption steps. The second layer produces final cipher text using open mp parallel algorithm.

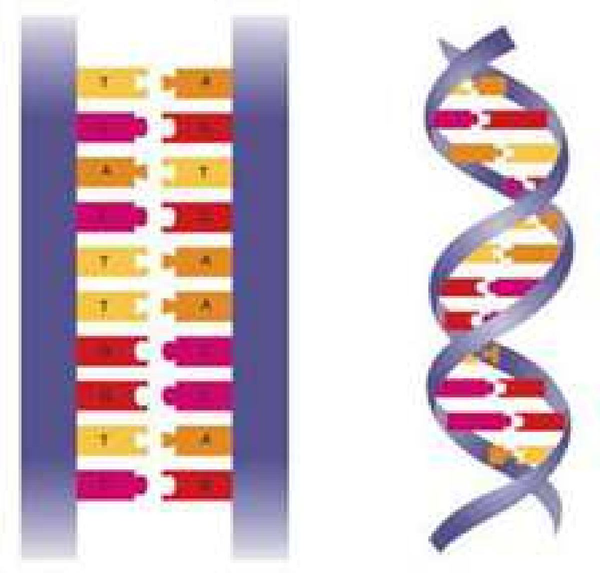

Figure 1: Structure of double stranded DNA

2 Literature survey

Nowadays the use of DNA is important in performing trillions of mathematical parallel operations with the use of small space, power and money [3]. The watermarking techniques have attracted towards digital circuits and found answers through digital computers, the use of DNA with the same kind of operations used for digital medium [4,5]. We have to discover nature of binary sequences which is difficult to identify. Plain text can be easily exposed fast because algorithms are already known to public. The public and private key algorithms of cryptography are used for encryption and authentication for binary sequences [6].

Chemically created DNA sequences are considered a very secure stage for digital medium. Due to rapid techniques in DNA during the period 1999-2013 from 23 character to 739 KB obtained successfully [7,8,9]. They have proposed that the error can be decreased using base 3 [9]. The cryptography of advanced algorithm hides data but it has some limitations in running time and memory [10].

One of the most common implementation techniques of encrypting the data that is converting the plain text to cipher text and decrypting the data is by using the single core system [11]. One of the most common implementations of encrypting the data that is converting the plain text into cipher text and decrypting the data uses the single core system. Here only one core is used irrespective of the size of the file which has to be encrypted or decrypted. This method will work slowly if the data file is big in size.

One of the conventional methods in enhancing the encryption and decryption of the plain text is by double encryption and double decryption which is proposed in [12]. The master thread executes all programs implemented by open mp as a single process. It executes as a single memory area until the first parallel programs are executed.

AES performs a blocker cipher of 128, 192, or 256 key expansion to encrypt and decrypt data blocks in different rounds. DNA storage have long duration and data storage in DNA in wet lab is becoming less expensive [13]. The two layer concept is proposed to encrypt the biometrics and palm prints based on convolution and cyclic two layer encoding methods to protect data[15]. Nowadays, online telemedicine communication protects data using a double chaotic layer encryption method algorithm proposed using symmetric logistic and chaotic media [16]. The two layer encryption algorithm proposed is based on Arnold map and hybrid 4d hyper to protect data and transmission of secure information over a network [17].

3 Preliminary Setup

3.1 Binary information is converted into DNA sequences from mapping 00,01,10,11$->$ A,G,C,T.

The human chromosome made up of Adenine(A), Guanine(G), Cytosine(C) and Thymine(T) is a string of nucleotides analogous to a computer executing only 0's and 1's. There are 8 different combinations possible. Here we have selected best 00 and its inverse is 11 so that in a DNA model A is always paired with T and G is paired with C in all living cells organism. Like binary, complementary DNA sequence have complementary rules. Table 1 describes how to convert nucleotides A,G,C,T into two bit binary representation using 8 different possible choices.

Example : DNA sequence
CCGAGCCACAT $< - >$1010010001101000100011

Table 1. Combination of DNA mapping into binary values

DNA	1	2	3	4	5	6	7	8
A	00	00	01	01	10	10	11	11
T	11	11	10	10	01	01	00	00
G	01	10	00	11	00	11	01	10
C	10	01	11	00	11	00	10	01

3.2 DNA sequence converted into complementary

A DNA sequence is made up of nucleotides A,G,C,T in human chromosome each nucleotide has complementary different nucleotide. In a computer system where 1 is complementary with 0 and 0 is complementary with 1, a DNA sequence nucleotides have complementary pairing. Table 2 describes different complementary pairing of nucleotides A,G,C,T. A complementary T,T complementary A, G complementary C,C complementary G,
Example : DNA sequence

Table 2. DNA XOR operation

xor	**A**	**G**	**C**	**T**
A	A	G	C	T
G	G	A	T	C
C	C	T	A	G
T	T	C	G	A

CCGAGCCACAT$< - >$GGCTCGGTGTA

3.3 Human genome

The human genome contains 3 billion base pairs and contains 23 chromosomes. Each chromosome has genes which contains Exons and Introns. Exons are coding sequences and Introns are non-coding sequences. Exon sequences can be used for DNA steganograohyor layer f 3 fold security. Table 3 describes genes of human chromosome number 12 and Chromosome X which consists of G3PD and DMD genes for glyceraldehydes convert glucose into energy molecules and dystrophin for skeleton connections. G3PD contains 4444 nucleotides A,G,C,T and DMD contains 2,220,381 nucleotides A,G,C,T.

Example:

Table 3. **Human genome genes example**

Gen Name	Chromosome no	Length of gene	number of Exons	Lenth of Exons	Lenth of Introns
G3PD GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE	12	4,444	9	1,425	3,019
DMD DYSTROPHIN	X	2,220,381	79	10,500	2,209,881

Table 4 describes keys generation for encryption of data from gene G3PD and DMD contains only exons because exons nucleotides A,G,C,T sequence is unique in a human genome DNA. Each exon has specific location in genome and sequences so that we have generated keys from those DNA sequence converted into binary from Table 1.

Table 4. Details of DNA Reference Sequence from gene G3PD for DNA Stegnography

Exons	DNA Reference Sequence (Keys for encryption of Data)
00001932108	GGGGGAAGTGGGGGCTGGGAAGGAACCACGGGCCCCCGCCCGAGGCCCAT GGGCCCCTCCTAGGCCTTTGCCTGAGCAGTCCGGTGTCACTA CCGCAGAGCCTCGAGGAGAAGTTCCCCAACTTTCCCGCCTCTCAGCCTTTGAAAGAAAGAAA GGGGAGGGGGCAGGCCGCGTGCAGCCGCGAGCGGTGCTGGGCT CCGGCTCCAATTCCCCATCTCAGTCGTTCCCAAAGTCCTCC TGTTTCATCCAAGCGTGTAAGGGTCCCCGTCCTTGACTCCCTAGT GTCCTGCTGCCCACAGTCCAGTCCTGGGAACCAGCACCGATCACC TCCCATCGGGCCAATCTCAGTCCCTTCCCCCCTACGTCGGGGCCC ACACGCTCGGTGCGTGCCCAGTTGAACCAGGCGGCTGCGG AAAAAAAAAAGCGGGGAGAAAGTAGGGCCCGGCTACTAGCGG TTTTACGGGCGCACGTAGCTCAGGCCTCAAGACCTTGGGCTGGG ACTGGCTGAGCCTGGCGGGAGGCGGGGTCCGAGTCACCGCCTGCC GCCGCGCCCCCGGTTTCTATAAATTGAGCCCGCAGCCTCCCGCTT CGCTCTCTGCTCCTCCTGTTCGACAGTCAGCCGCATCTTCTTTTGCGTCGCCAG
00003562276	CCGAGCCACATCGCTCAGACACCATGGGGAAGGTGAAGGTCGGAGTCAACGG
00003571091	ATTTGGTCGTATTGGGCGCCTGGTCACCAGGGCTGCTTTTAACTCTGGTAAAGTGGATAT TGTTGCCATCAATGACCCCTTCATTGACCTCAACTACATG
00003682485	GTTTACATGTTCCAATATGATTCCACCCATGGCAAATTCCATGGCACCGTCAAGGCTGAG AACGGGAAGCTTGTCATCAATGGAAATCCCATCACCATCTTCCAGGA
00003678358	GCGAGATCCCTCCAAAATCAAGTGGGGCGATGCTGGCGCTGAGTACGTCGTGGAGTCCAC TGGCGTCTTCACCACCATGGAGAAGGCTGGG
00003562150	GCTCATTTGCAGGGGGGAGCCAAAAGGGTCATCATCTCTGCCCCCTCTGCTGATGCCCCC ATGTTCGTCATGGGTGTGAACCATGAGAAGTATGACAACAGCCTCAAGATCATCAG
00003663529	CAATGCCTCCTGCACCACCAACTGCTTAGCACCCCTGGCCAAGGTCATCCATGACAACTT TGGTATCGTGGAAGGACTCATG
00003460425	ACCACAGTCCATGCCATCACTGCCACCCAGAAGACTGTGGATGGCCCCTCCGGGAAACTG TGGCGTGATGGCCGCGGGGCTCTCCAGAACATCATCCCTGCCTCTACTGGCGCTGCCAAG GCTGTGGGCAAGGTCATCCCTGAGCTGAACGGGAAGCTCACTGGCATGGCCTTCCGTGTC CCCACTGCCAACGTGTCAGTGGTGGACCTGACCTGCCGTCTAGAAAAACCTGCCAAATAT GATGACATCAAGAAGGTGGTGAAGCAGGCGTCGGAGGGCCCCCTCAAGGGCATCCTGGGC TACACTGAGCACCAGGTGGTCTCCTCTGACTTCAACAGCGACACCCACTCCTCCACCTTT GACGCTGGGGCTGGCATTGCCCTCAACGACCACTTTGTCAAGCTCATTTCCTG
00001902446	GTATGACAACGAATTTGGCTACAGCAACAGGGTGGTGGACCTCATGGCCCACATGGCCTC CAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACC CTCACTGCTGGGGAGTCCCTGCCACACTCAGTCCCCCACCACACTGAATCTCCCCTCCTC ACAGTTGCCATGTAGACCCCTTGAAGAGGGGAGGGGCCTAGGGAGCCGCACCTTGTCATG TACCATCAATAAAGTACCCTGTGCTCAACCA

Example: DNA Steganography and 3DES
GACCAAGCCTGCAAAAGCAAATTCAAAAAATGTCCGTAAGACTTAAAATCCACACAAGAA
TCACGAAGTAGTGCCCGAAGTCGTAGAAGAAGGCTTTTT
AGAAGATATGGCAGTCAGAAGACATTAAGGCTTCGTAGCGAGCACCCACCGAC
Start: GACCAAGCCTGC
Length: ATTCAAAA = 104
Checksum: AATGTCCG = 55008
Data:
TAAGACTTAAAATCCACACAAGAATCACGAAGTAGTGCCCGAA GTCGTAGAAGAAGGCTTTTTAGAAGATATGGCAGTCAGAAGACA TTAAGGCTTCGTAGCGA

4 Proposed Algorithms

4.1 Proposed 2 layer Symmetric Algorithm for AES

DNA Steganography provides high security because a DNA sequence of one human cannot match another person except twins. In a dry lab we utilized DNA as an information storage device like Disk, CD and generate keys from it.

Algorithm 1: In DNA steganography, the first step converts a message into Binary numbers while the second step converts Binary numbers into DNA sequences using table 1; this is consider as input for DNA steganography step 2 and DNA reference of human to produce cipher based indexing as output. These two steps consider as one layer for converting a message into index based DNA sequence as a middle cipher text.

Algorithm 2: The output of algorithm 1 as input for traditional AES and 3DES algorithms. We have proposed parallel concepts open mp to implement AES algorithms to convert middle cipher text into final cipher text of message.

Algorithm 1:
DNA steganography step 1
Input: information in terms of binary representation
Output: DNA sequence representation
Input: information in terms of binary representation

```
for i=0 to size of binary inf[maxsize-1]
for j=0+1 to size of binary inf[maxsize-2]
do i and j position to DNA sequence
map(00,11,01,10) -¿ DNA(A,T,G,C)
```

```
end for
end for
```

DNA steganography step 2

```
Input: DNA sequence of original data by step 1, DNA reference sequence of human
Output: array position element of DNA sequence present on DNA reference sequence POS[8319....69]
DNAO[ATGCAAGCAT........]
DNAREF[GTAATTGCGAGAGGGTATGC......AT]
for i=0 to size of DNAO[maxsize-1]
for j=0+1 to size of DNAO[maxsize-2]
for i=0 to size of DNAR [maxsize-1]
for j=0+1 to size of DNAR [maxsize-2]
compare length2 and obtain position of DNAO original sequence in reference sequence POS[8319....69]
end for
end for
end for
end for
```

The second layer is a cryptographic algorithm in this output of DNA steganography cipher is input for AES algorithm.

Algorithm 2: Open mp of AES algorithm

```
Input: DNA steganography cipher and Key for AES algorithm
Output: Cipher text with parallel programming
POS[8319....69]
Divide file containing DNA steganography cipher into n/2 size
Apply Parallel program concepts for n/2 file size
Merge output of two files into one file
```

4.2 Proposed 2 layer Symmetric Algorithm for 3DES

DNA steganography step 1

```
Input: information in terms of binary representation
Output: DNA sequence representation
Input: information in terms of binary representation
for i=0 to size of binary inf[maxsize-1]
for j=0+1 to size of binary inf[maxsize-2]
```

```
do i and j position to DNA sequence
map(00,11,01,10) -¿ DNA(A,T,G,C)
end for
end for
```

DNA steganography step 2

```
Input: DNA sequence of original data by step 1, DNA reference sequence of human
Output: XOR operation between DNA sequence original data and DNA reference sequence , DNACIPER[]
DNAO[ATGCAAGCAT........]
DNAREF[GTAATTGCGAGAGGGTATGC......AT]
for i=0 to size of binary inf[maxsize-1]
for j=0+1 to size of binary inf[maxsize-2]
do i and j position to DNA sequence
DNA(A,T,G,C) -¿ map(00,11,01,10)
end for
end for
DNACIPER[]: XOR between DNAO[]and DNAREF[]
```

Algorithm 2: 3DES algorithm

```
Input: DNA steganography cipher and Key for DES algorithm
Output: Cipher text DNACIPER[]
Apply DES program concepts with input key value and DNACIPER[]
```

5 Complexity Algorithm

The proposed algorithm has DNA steganography and Cryptography complexity. Based on DNA reference sequence and size of data. If size of data increases time and probability of finding correct sequence increases.
Probability to guess a single letter is : 1/(X1*X2*X4*X5)
X1=Binary representation of a single letter(ASCII)
X2= Binary combination rules 00,01,10,11
X3=Complementary pairing rules 4*3*2*1(A,T,G,C)
X4= no_of DNA reference sequences available present 163 million in ncbi
Probability to guess a single letter for DNA steganography is : 1/(8*8*24*163*10∧6)
Probability to Guess Key for Cryptography and Exact cipher text into correct DNA a letter
Total probability a guess a single letter : cryptography key * 1/(8*8*24*163*10∧6)

6 Results and Discussion

DNA of human beings contains 3 billion basepairs so that probability of each correct a single letter is 1/3billion. NCBI and EMBL contains both genomics and proteomics data in billions. The complexity of finding one correct letter is 1/444*1425*3019. The proposed algorithm for small files takes more time but large files takes less time and difficult to guess keys and cipher text because two times cipher text generated and keys are generated from DNA sequences. Figure 2 shows existing algorithm in blue color and proposed algorithm in dark red color. Existing algorithm takes more time to encryption and decryption process and proposed algorithm takes less time for large data and initially it takes more time for small files. From figure 2 clearly shows which is better in terms of time for large data to encryption and decryption process.

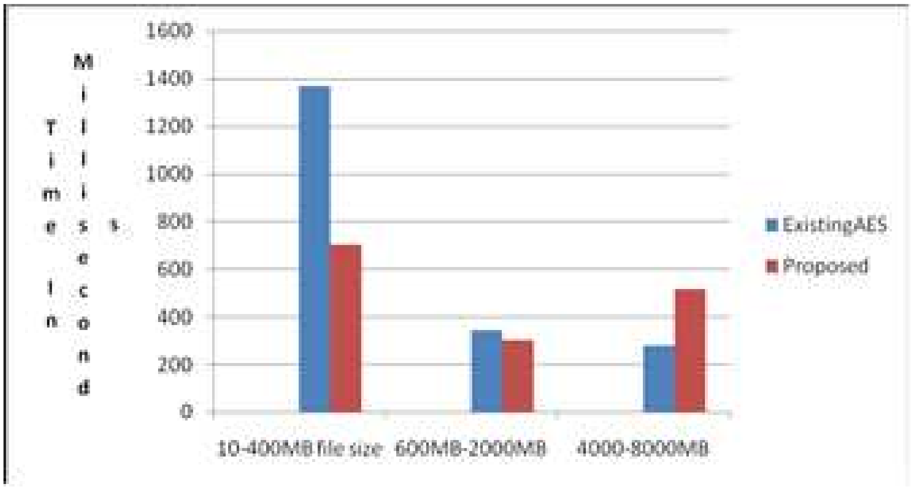

Figure 2: Comparison of Existing and Proposed Two layer AES algorithm

7 Conclusion

DNA XOR operation is reflexive and unique. The proposed encryption and decryption algorithms for data security and storing in DNA sequences are very secure because gene sequences are key values for DNA XOR operation. The proposed 2 layer DNA steganography and Cryptography algorithm are very secure and fast compared to existing AES algorithm and takes less time for large data and DNA steganography and 3DES provides security against cryptanalysis of frequency letters to find plaintext because only 4 letters and 8 different DNA XOR operations performed on plaintext.

8 Future Work

Secure storing big data in DNA sequence format and increasing the speed of accessing the data.

References

[1] J. Watson, N. Hopkins, J. Roberts, J. Steitz, Molecular Biology of the Gene, Fourth ed., Benjamin Cummings, Menlo Park, CA, 1987.

[2] National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/

[3] L. Adelman, Molecular computation of solutions to combinatorial problems, Science, **266**, Nov. 11, 1994.

[4] J. Lach, W. Mangione-Smith, M. Potkonjak, Fingerprinting Digital Circuits on Programmable Hardware, Information Hiding Workshop, Portland, Oregon, 1998.

[5] A. B. Kahng, J. Lach, W. H. Mangione-Smith, S. Mantik, I. L. Markov, M. Potkonjak, P. Tucker, H. Wang, G. Wolfe Watermarking Techniques for Intellectual Property Protection, DAC-98 35th ACM/IEEE DAC Design Automation Conference, San Francisco, CA, June 1998, 776–781.

[6] W. Diffie, M. E. Hellman, Privacy and Authentication: An Introduction to Cryptography, Proceedings of the IEEE, **67,** No. 3, March 1979.

[7] G. M. Church, Y. Gao, S. Kosuri, Next-generation digital information storage in DNA. Science, 337(6102):1628, 2012.

[8] C. T. Clelland, V. Risca, C. Bancroft. Hiding messages in DNA microdots. Nature, **399,** (1999), 533–534.

[9] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, E. Birney, Towards practical, high-capacity, lowmaintenance information storage in synthesized DNA, Nature, **494,** (2013), 77–80.

[10] J. Pichel, D. E. Singh, J. Carretero, Reordering algorithms for increasing locality on multicore processors. 10th IEEE International Conference on High Performance Computing and Communications, 2008, 123–130.

[11] Chih-Chung L, Shau-Yin Tseng, Integrated design of AES (Advanced Encryption Standard) encrypter and decrypter, in the The IEEE International Conference on Application-Specific Systems, Architectures and Processors, 2002, 277–285, 2002 doi:10.1109/ASAP.2002.1030726.

[12] Mart Abadi, Roger Needham, Prudent Engineering Practice for Cryptographic Protocols, IEEE Transactions on Software Engineering, **22,** no. 1, January 1996, 6-15.

[13] James Bornholt, Randolph Lopez, Douglas M. Carmean, ASPLOS ’16, April 2–6, 2016, Atlanta, GA, DOI: http://dx.doi.org/10.1145/2872362.2872397

[14] http://nitro.biosci.arizona.edu/courses/EEB195/Lecture04/Lecture04.html

[15] Hengjian Li, Jian Qiu, Jiwen Dong, Guang Feng, Biometrics encryption combining palmprint with two-layer error correction codes, Ninth International Conference on Digital Image Processing (ICDIP), Hong Kong 2017, doi: 10.1117/12.2281672

[16] M. A. Murillo-Escobar, L. Cardoza-Avendaño, R. M. López-Gutiérrez, C. Cruz-Hernández, A double chaotic layer encryption algorithm for clinical signals in telemedicine, Journal of Medical Systems, **41,** Issue 4, April 2017.

[17] Zhongpeng Wang, Fangni Chen, Weiwei Qiu, Shoufa Chen, Dongxiao Ren, A two layer chaotic encryption scheme of secure image transmission for DCT precoded OFDM-VLC transmission, Optics Communication, **410,** (2018), 94–101.