International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 10828-10833
© Research India Publications. http://www.ripublication.com

# Secure Symmetric Encryption Scheme Using Genetic Algorithm

**Cimi Thomas M**[1]

*Research Scholar, Department Of Computer Science, Karpagam Academy of Higher Education,*
*Pollachi Main Road,Eachanari Post, Coimbatore, Tamilnadu, India.*

*Orcid Id:0000-0002-8605-7440*

**Dr. S.Sheeja**[2]

*Associate Professor and Head, Department of Computer Applications,*
*Karpagam Academy of Higher Education, Pollachi Main Road,Eachanari Post, Coimbatore, Tamilnadu, India.*

## Abstract

Ensuring the security of sensitive and confidential data during transmission and storage is one of the greatest challenges faced by digital world. The most common technique used to protect data is encryption. Different encryption algorithms are available and they are used in security protocols to provide confidentiality, integrity and authentication. In spite of having strong encryption algorithms and security protocols, security breaches are still happening and reported frequently. Secure encryption schemes with new features different from traditional methods are needed to protect digital information. In this study a symmetric key encryption scheme based on DNA sequences is proposed. The proposed algorithm is a feistel cipher with many rounds and the round specific sub keys are generated by applying the principles of genetic algorithm. The study has proved that genetic algorithm is a viable method for selecting best fit keys for developing strong encryption schemes.

**Keywords:** DNA cryptography, Encryption, Decryption, Genetic Algorithm, Feistel structure.

## INTRODUCTION

The affordability of computing devices and easy access to World Wide Web has led to the development of wide range of web applications. These web applications require the transmission of personal and critical information over internet. There is tremendous increase in the number of online financial transactions owing to the popularity of E-commerce. Credit card information and user bank account details transmitted over network should be protected from unauthorized people. Ensuring the confidentiality and integrity of various databases is also challenging. DNA databases which has DNA profiles of people were initially used for criminal investigations. But now a days public genomic databases are available which are widely used to carry out research studies on genetic diseases and genetic genealogy. Security of these databases is a real concern as the DNA information can be misused in various ways. Encryption can be used to protect data during transmission and storage. Encryption algorithms can be classified as symmetric key encryption and asymmetric key encryption. Symmetric key encryption uses the same key for encryption and decryption. A number of symmetric key encryption algorithms are in existence and they are used in various security protocols and products. The advancement in computing and in the area of cryptanalysis demand the development or enhancement of existing encryption schemes. Researchers are trying to develop new cryptographic techniques which can withstand threats posed by quantum computers. In this study a symmetric key encryption scheme based on DNA sequences is proposed. The proposed algorithm has a feistel structure with multiple rounds of permutations and substitutions. The round specific sub keys for each round are generated by applying the principles of genetic algorithm.

DNA cryptography is a new promising field which has attracted researchers in recent years. DNA (Deoxyribo Nucleic Acid) contains our unique genetic code and is made up of monomers called deoxyribo nucleotides. Each DNA molecule has two long strands of nucleotides and each nucleotide is made of deoxyribose sugar, phosphate group and a nitrogenous base. Nitrogenous bases are A (Adenine), G (guanine), C (Cytosine) and T (Thymine). A bonds with T and G bonds with C. These nucleotides appear in random order in each DNA molecule and the order of these nucleotides determine DNA's genetic code. The strength of any cryptographic scheme is directly related to the randomness offered by them. Researchers have shown that randomness found in the nucleotide sequences can be efficiently used to create strong and reliable encryption schemes. The use of real DNA for developing crypto systems is not feasible currently due to sophisticated lab requirements.

Genetic algorithm is an optimization method based on Darwin's theory of natural selection. Genetic algorithm repeatedly modifies a population of individual solutions. The initial population is generated depending on the problem under consideration and the fitness of each individual in the population is evaluated using a fitness function. Individuals are selected based on their fitness value and genetic operators are applied on them to create children. Children form the next generation and will have more fitness than their parents. This research work has used the principles of genetic algorithm to create strong sub keys.

The rest of the paper is organized as, first Section provides a review of symmetric ciphers with feistel structure and current works on DNA cryptography. Second Section describes the proposed algorithm. Results and security analysis are discussed in third section.

## LITERATURE REVIEW

Horst Feistel proposed an ideal block cipher in 1970 which had several rounds of substitutions and permutations. The structure of the cipher proposed by Feistel became popular and was named as feistel structure or feistel network [1].Many significant block ciphers currently in use are based on Feistel cipher structure. The encryption algorithm of a feistel cipher consists of multiple rounds of processing of the plaintext, each round consisting of a substitution step followed by a permutation step. A cryptographic system based on Feistel cipher structure uses the same algorithm for both encryption and decryption.

Data Encryption Standard (DES) [2] is the first popular feistel cipher with 16 rounds of processing. DES has a block size of 64bits and key size of 56 bits. Though DES with its small key size is not considered secure any more, DES is used as a benchmark in research studies because of strong security features. Advanced Encryption Standard [3] is a symmetric block cipher with proven security. AES has a block size of 128 bits and key size of 128,192 and 256 bits depending on the number of rounds. Matrix Array Symmetric Key Encryption (MASK) [4] is a symmetric block cipher with 128 bits block size and 128 bits key size which uses simple matrix and array manipulation and is faster than AES algorithm. A number of other symmetric key encryption algorithms are present in literature with their own merits and demerits.

Computing on DNA was first done by Adleman in 1994 .He used DNA computing to solve directed Hamiltonian path problem [5].Boneh [6] used DNA computing to break DES. Gehani [7] developed the first cryptosystem using real DNA sequences. It was a onetime pad scheme utilizing the high storage capacity of DNA molecules. To develop encryption schemes using DNA sequences, binary data has to be converted to DNA form. Heider[8] has given different DNA coding rules. A pseudo DNA cryptographic method motivated by the central dogma of molecular biology was proposed by the author in [9] .A new iterative encryption technique combining DNA sequences and genetic algorithm is proposed in [10]. Security enhancement of Vigenere cipher using DNA sequences was done by the authors in [11]. In [12] authors have proposed an image encryption algorithm using DNA sequences, the concepts of Genetic Algorithm is applied to select the best cipher image. In [13] authors have used genetic algorithm to generate best fit keys for encryption scheme.

## PROPOSED ALGORITHM

The proposed symmetric encryption scheme handles binary data in DNA form. It is a block cipher with block size of 128bits or 64 nucleotides. The encryption scheme has a feistel structure with 16 rounds.

### A. Generation of master key and sub keys

The proposed scheme has a master key and 16 sub keys. Master key is a combination of accession number of DNA string generated from public database and current system time .Each sub key is a DNA string of 32 nucleotides. Sub keys for each round are derived by applying the principles of genetic algorithm. The first step in genetic algorithm is the creation of initial population. Here individuals in the initial population are nucleotide sequences of length 32, extracted from the downloaded DNA string using system time as the random seed. The fitness of each individual is calculated using the formula $E=-\sum P(x)*\log_2 P(x)$ ,where x is the base (A,T,C&G) in the sequence and P(x) is the probability of its occurrence [14]. Individuals are selected based on their fitness value and one point cross over and mutation are applied to create individuals for next generation. The process is repeated for 3 generations .The effective use of genetic algorithm in deriving best fit sub keys is explained in our previous work [15].

### B. Encryption Algorithm

The steps in the proposed encryption algorithm are described below.

Step 1: Convert plaintext to equivalent binary form.

Step 2 : Take bits two at a time and encode it in to DNA form using DNA encoding rule. The rule used is ”00” is converted to A, “11” to T, ”01” to C and “10” to G.

Step 3: Take plaintext in blocks of 128 bits (64 nucleotides), divide it in to two halves to form left half and right half.

Step 4: Perform data dependent rotation on left half and right half (Initial Permutation).

Step 5: Perform DNA addition between left half and round specific key.

Step 6: Perform DNA XOR between right half and round specific key.

Step 7: Interchange left half and right half.

Step 8: Repeat steps 4, 5,6,7 sixteen times. In each round use specific sub key.

Step 9: Output from 16th round gives the cipher text .

Step 10 : Repeat step 3 to 8 for each block until the entire plain text is encrypted.

The XOR, Addition and Subtraction of DNA sequences are performed using the following tables.

**Table 1:** XOR operation for DNA Sequence

| XOR | A | C | T | G |
|---|---|---|---|---|
| A | A | C | T | G |
| C | C | A | G | T |
| T | T | G | A | C |
| G | G | T | C | A |

**Table 2 :** Addition operation for DNA Sequence

| ADD | A | C | G | T |
|---|---|---|---|---|
| A | A | C | G | T |
| C | C | G | T | A |
| G | G | T | A | C |
| T | T | A | C | G |

**Table 3:** Subtraction operation for DNA Sequence

| SUB | A | C | G | T |
|---|---|---|---|---|
| A | A | C | G | T |
| C | T | A | C | G |
| G | G | T | A | C |
| T | C | G | T | A |

In the proposed algorithm, processing is done on both halves of the data which is a slight deviation from the feistel structure. This is done to increase the security of the cipher.

### C. Decryption Algorithm

Step 1: Generate Round specific sub keys from master keys.

Step 2: Take cipher text in blocks of 128 bits (64 nucleotides), divide it in to two halves to form left half and right half.

Step 3: Interchange the two halves.

Step 4: Perform DNA subtraction between left half and Round specific key.

Step 5: Perform DNA XOR between right half and Round specific key.

Step 6: Perform rotation on left half and right half.

Step 7: Repeat steps 3,4,5,6 sixteen times .In each round use the keys in the reverse order.

Step 8: Repeat steps 2, 3,4,5,6, 7 until the entire cipher text is decrypted

Step 9: Decode nucleotide sequence to binary data.

Step 10: Convert binary data in to plaintext.

## RESULTS AND SECURITY ANALYSIS

The proposed method is implemented in Matlab R2015a. Public DNA database used in this work is European Bioinformatics Institute's (EBI's) European Nucleotide Archive which provides a huge collection of nucleotide sequences.

Implementation results show that using genetic algorithm strong sub keys can be generated from master key. Cross over operator increases the fitness of individuals (sub keys) and fitness of the population increases with each generation. In the current implementation, number of generation is fixed as 3. Table 4 shows individuals in three generations and their fitness value.

**Table 4:** Population sample and the fitness value.

| **Generation –I Population with Fitness value** | **Generation-II Population with Fitness value** | **Generation-III Population with Fitness value** |
|---|---|---|
| 1.9837<br>ACTAGCCCGGCCGTGTATGTTTTTGAATGAAC | 1.9943<br>ATGCCTCATGTAGTCCCACGGATGCTTATGAA | 1.9943<br>TCAGGACGCGTGCAAGTTTCCGCTAACACGTT |
| 1.9716<br>TCAGGACGCGTGGGGCTCCGCTAACACGTTAA | 1.9837<br>ACGTACCTGTCACAGGTATGAATGAAGCTACG | 1.9837<br>ACTAGCCCGGCCGTGTATGTTTTTGAATGAAC |
| 1.9469<br>GCAGCCTGTGTGGCTCCTACGCAAAAGCGGAC | 1.9576<br>GGCAGAGAGGTAGACGCGCCCCAATCACTTTG | 1.9685<br>TGTAGCCTAGCAGACGCGCCCCAATCACTTTG |
| 1.9424<br>CGTGTAGCACAGCAGCAATCACTTACTAACAT | 1.9424<br>ACGTACCTGTCAGCTCCTACGCAAAAGCGGAC | 1.9576<br>GGCAGAGAGGTAGACGCGCCCCAATCACTTTG |

| | | |
|---|---|---|
| 1.9252<br>AGATCAACTCCC<br>AGTGGGGGTCCA<br>CCGGCGTC | 1.9363<br>AGATCAACTCCC<br>CGTGGCTTGACC<br>GACGCTGC | 1.9576<br>TGTAGCCTAGCAC<br>CAAATGTCAGAAG<br>GGCCGG |
| 1.9252<br>TACTAGCCAGGC<br>CGTGGCTTGACC<br>GACGCTGC | 1.9288<br>TAGTCCCACGCA<br>CCAAATGTCAGA<br>AGGGCCGG | 1.9427<br>CGTGTAGCACAGT<br>ATCACGTTAACCC<br>CCCCTG |
| 1.9193<br>GGCAGAGAGGTA<br>GTCCCCCGCCGC<br>CTAAATCA | 1.8992<br>TGTAGCCTAGCA<br>AACTCCCCCGCC<br>CGCTAAAT | 1.9427<br>ACGTACCTGTCAG<br>ACGCGCCCCAATC<br>ACTTTG |

Since current system time is used in the generation of sub keys, the sub keys are time dependent. Millisecond difference will create entirely different sub keys. As a result the plain text encrypted at different points of time will produce entirely different cipher texts. The cipher text is kept in DNA form in the current implementation. Table 5 shows the time dependency of the proposed algorithm.

**Table 5:** Cipher text produced at different Points of time

| **Plaintext** | **Time** | **Cipher text** |
|---|---|---|
| Research creates knowledge | 22:54: 48.411 | TGGATGGGAATGGGGCGG<br>TCTAATGTTATGTCCATTT<br>CGGCTGGGAGGCCCCCTA<br>GGACTGGAGGGCGGCGGG<br>TCGTTGCTACCGTGTTCTA<br>TCTCGGTCGTGGAGAAAT<br>GATACTCAGAATCCCTGA |
| Research creates knowledge | 22:54: 50.521 | TGCACGTATCACGAGTCTG<br>ATAGCACGAAGAACTTCC<br>TTAAGTCAGTAGAGGACG<br>ATGAAGAAAGGGGACTAC<br>GGCTCGTACAAGTACCTG<br>AACAAGCTTACTACTCTGC<br>TGACGAAGAGCCAGCCGG |
| Research creates knowledge | 22:54: 51.833 | AAGCTAGCGAGTACTCGC<br>CTGGGTCTACAATATTCAT<br>GCTGGTGTAACCTTACGAT<br>CCGACATCCACTGTGCATA<br>TCATCTTTTTCATACACAT<br>TAACCGGACTTTCACTATT<br>GTGCCGCTGGGGCCT |

Security of a feistel cipher depends on block size, key size, number of rounds, sub key generation algorithm and round functions used. The block size of the proposed scheme is 128 bits, key size is160 bits and there are 16 rounds. Sub key generation is based on the principles of genetic algorithm which guarantees the creation of strong keys. The round function includes substitution and permutation. Both halves of the data are modified which increases the security. All these features together makes the proposed scheme reliable and secure.

Security of an encryption scheme can be analyzed by evaluating its resistance against known attacks. Brute force attack on keys is a form of attack where the attacker searches all possible keys until the correct key is found. The proposed method can withstand Brute force attack as there are $2^{160}$ possible keys and exhaustive key search is not feasible. The nature of master key further makes cryptanalysis difficult. The first part of the master key is the accession number of nucleotide sequence. The public database has a collection of millions of nucleotide sequences and brute force analysis to find a specific nucleotide sequence is highly impossible. System time to the accuracy of milliseconds is used and this 9 digit key has 864* $10^5$ values. Trying out 864* $10^5$ possible values on millions of DNA string in nucleotide database is required to guess the master key.16 sub keys are used ,each with a size of 32 nucleotides .$4^{32}$ combinations are possible for a single sub key which make brute force attack infeasible. Strong keys are the strength of the proposed scheme.

The strong keys helped in the generation of cipher text with high degree of randomness. The randomness of the cipher text is measured by estimating the entropy using Shannon's entropy [14] and the following results are obtained.

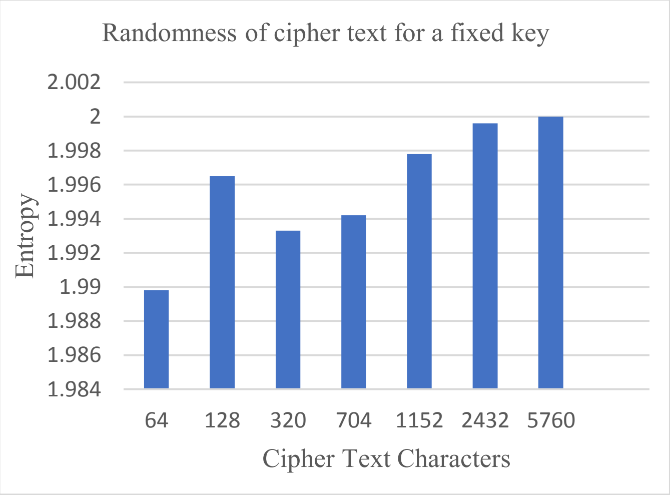

**Fig 1:** Entropy of cipher text for a fixed key

As the cipher text is in DNA form, the maximum entropy possible is 2 and this happens when each nucleotide has an equally likely chance of occurrence. Result shows that cipher text exhibit high degree of randomness.

Analysis of the proposed scheme shows that there is no correlation between cipher text bits and plain text bits and also between cipher text bits and key bits. The algorithm exhibits strong key avalanche and strong data avalanche properties. To demonstrate this plaintext 'Multiprogramming' is encrypted using a selected key and the output produced in each round is noted. Using the same key, another plain text with one bit difference is encrypted and outputs are recorded. The number of bit differences that occurred in each round for the two plaintexts are shown in the table below.

**Table 6:** Data Avalanche with a fixed key

| Rounds | No: of bits changed |
|---|---|
| 1 | 37 |
| 2 | 66 |
| 3 | 64 |
| 4 | 62 |
| 5 | 59 |
| 6 | 62 |
| 7 | 68 |
| 8 | 67 |
| 9 | 67 |
| 10 | 60 |
| 11 | 63 |
| 12 | 67 |
| 13 | 63 |
| 14 | 60 |
| 15 | 63 |
| 16 | 63 |

Results show that the algorithm exhibits strong avalanche property. The result is compared with the data avalanche exhibited by AES and modified AES discussed in the literature [16] and is shown in the table7.

**Table 7:** Comparison of Data Avalanche with existing algorithms

| Encryption Algorithm | Average change in no: of bits of cipher text for 1 bit change in plain text |
|---|---|
| AES | 46 |
| Modified AES | 50 |
| Proposed Algorithm | 62 |

## CONCLUSION

DNA cryptography is a new promising field in information security. Though working with real DNA is not feasible currently, the randomness offered by the nucleotides in a DNA string can be used to create secure encryption scheme. In this study, a new symmetric encryption algorithm is proposed which is a feistel cipher based on DNA sequences. Genetic algorithm is used to generate sub keys for each round and study has shown that the application of genetic algorithm helps in selecting best fit keys. The proposed encryption algorithm is not faster than popular symmetric key algorithms but has strong avalanche property and can withstand security attacks. The round functions used will be improved to increase the encryption speed in the future work.

## REFERENCES

[1] W. Stallings, Cryptography and Network, Security, Principles and Practices, Fourth Edition,Prentice Hall,November 2005

[2] https://csrc.nist.gov/csrc/media/publications/fips/46/3/archive/1999-10-25/documents/fips46-3.pdf

[3] https://csrc.nist.gov/csrc/media/publications/fips/197/final/documents/fips-197.pdf

[4] A.J Paul, P.Varghese,P. Mythilli, "Matrix Array Symmetric Key Encryption", Journal of Computer Society of India, Vol.37,Issue No.1,Jan-March 2007,pp 48-53.

[5] L.M. Adleman,Molecular Computation of Solutions to Combinatorial Problems",SienceVol 266(5187),Nov 1994,pp 1021-1024.

[6] D.Boneh, C Dunworth, R.J Lipton,JSgall , On the Computational Power of DNA, Discrete Applied Mathematics,1996,71(1):p:79-94.

[7] A.Gehani ,TLaBean, JReif, DNA Based Cryptography, Aspects of Molecular Computing,2004,Springer p-167-188.

[8] D.Heider, A Barnekow, DNA based watermarks using the DNA –Crypt Algorithm,BMC bioinformatics,2007,8(1),p(176).

[9] K.Ning,"Pseudo DNA Cryptography Method", http://arxiv.org/abs/0903-2693,2009.

[10] H.M Mousa, "DNA-Genetic Encryption Technique", International Journal of Computer Network and Information Security,2016,7,1-9.

[11] M.Najaftorkaman, N. S Kazazi,"A method to encrypt information with DNA based cryptography", International Journal of Cyber-Security and Digital Forensics(IJCSDF),4(3):417-426,2015.

[12] R.Enayatifar,A.Abdullah,I.FauziIsnin,"Chaos-based image encryption using a hybrid Genetic Algorithm and a DNA sequence",Optics and Lasers in Engineering 56(2014)83-93.

[13] S.Jawaid,A.Jamal,"Generating the best fit key in

cryptography using Genetic Algorithm",International Journal of Computer Applications,Vol98,No.20 July 2014.

[14] C. Shannon, "Communication Theory of Secrecy Systems", Bell Systems Technical Journal,28,656-715, 1949

[15] C.Thomas, S.Sheeja,"DNA based Feistel Cipher with Sub keys selected using Genetic Algorithm", Journal of advanced research in dynamical and control systems,13-special issue, September 2017,p 153-158.

[16] AJ.Paul, A.Saju, R.N.Lekshmi,"Data Based Transposition to Enhance Data Avalanche and Differential Data Propogation in Advanced Encryption Standard" International Journal of Computer Applications,vol67,No:12, April 2013.