Data Security in DNA Sequence Using Random Function and Binary Arithmetic Operations

Siddaramappa V

Asst. Professor, VTU, India

***Abstract*-** The main target of this research paper is to propose an algorithm to implement data security using encryption and decryption method in binary sequence of original text message. The binary sequence of message is converted into DNA sequence, which consists of nucleotide letter A,T,G,C., then by using random function, which generates the sequence of random numbers for each nucleotide. The generated random numbers are used to cryptography the ASCII values of binary DNA sequence of original text message, with binary addition and subtraction method. Hence, this paper aims to describe and review about the data security by using the random function in DNA sequence.

***Index Terms*-** Binary Sequence; DNA sequence; Random number generator function; data security, Binary addition and Binary subtraction.

I. INTRODUCTION

In network it is very important to protect the data. In order to achieve this there are several technologies being used. One of the secured way to protect the data is encryption and decryption method. Because of increasing the number of Internet users, data security has become very important task of sending and receiving the data.

Therefore, now a days it has become more significant to secure the data. Before employing biological properties of DNA sequence, the popular symmetric cryptography algorithms include the old Data Encryption Standard (DES) and the new advanced Encryption Standard (AES)[1], argue that the AES system suffers from an obvious weakness, the key must be known to both parties. Thus the problem of confidential communication reduces to that of how to distribute these keys securely.

The new class in cryptosystem is an asymmetric cryptography algorithm. In Asymmetric cryptography algorithm a different key is used for encryption and decryption. So, a user has a pair of keys like public and private key. Cryptography aims to construct schemes or protocols that can still accomplish certain tasks even in the presence of an adversary. A basic task in cryptography is to enable users to communicate securely over an insecure channel in a way that guarantees their transmission is private and authenticated. Providing privacy and authenticity remains a central goal for cryptographic protocols. Cryptography's aim is to construct schemes or protocols that can still accomplish certain tasks even in the presence of an adversary. A basic task in cryptography is to enable users to communicate securely over an insecure channel in a way that guarantees their transmissions are guaranteed. Providing privacy and authenticity remains a central goal for cryptographic protocols, but the field has expanded to encompass many others, including e-voting, digital coins, and secure auctions.[2]

Two different kinds of genetic material exist, deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) present in a cell. DNA consists of 4 types of nucleotides like, Adenine (A), Thymine (T), Cytosine(C), and Guanine (G). James D.Watson and Francis Crick were the two co-discoverers of the structure of DNA in 1953. They used x-ray diffraction data collected by Rosalind Franklin and proposed the double helix or spiral staircase structure of the DNA molecule. Their article, Molecular Structure of Nucleic Acida: AStrcuter for Deoxyribose Nucleic Acid is celebrated for its treatment of the B form of DNA (B-DNA), and as the source of Watson –Crick Base pairing of nucleotides. They were, with Maurice Wilkins, awarded the Nobel Prize in Physiology or Medicine in 1962.[3]

Binary sequence is a sequence, which contains 0's and 1's. As CPU understands only machine level language, so processing with binary numbers is efficient and processing speed will be high.

A random function is a type of random element in which a single outcome is selected from some family of functions, where the family consists some class of all maps from the domain to the co domain. For example the class may be restricted to all continuous functions or to all step function. The values determined by a random function is evaluated at different points from the same realization would not generally be statistically independent but, depending on the model, values determined at the same or different points from different realizations might well be treated as independent.[4].

The National Center for Biotechnology Information (NCBI) provides data analysis, retrieval and resources that operate on the data in GenBank and a variety of other biological data is available through NCBI's Web site. NCBI data retrieval resources includes Entrez, PubMed, Locus Link and the Taxonomy Browser. Data analysis resources includes BLAST, Electronic PCR, OrfFinder, RefSeq, UniGene, Database of Single Nucleotide Polymorphisms (dbSNP), Human Genome Sequencing pages, GeneMap'99, Davis Human–Mouse Homology Map, Cancer Chromosome Aberration Project (CCAP) pages, Entrez Genomes, Clusters of Orthologous Groups (COGs) database, Retroviral Genotyping Tools, Cancer Genome Anatomy Project (CGAP) pages, SAGE map, Online Mendelian Inheritance in Man (OMIM) and the Molecular Modeling Database (MMDB). Augmenting many of the Web applications are custom implementations of the BLAST program optimized to search specialized data sets. All of the resources can be accessed through the NCBI home page at: http://www.ncbi.nlm.nih.gov [5].

Cryptography is one of the most active area of research in computer science. It survives only where efforts to reduce computational complexity have failed, because the intractability of various problems keeps unwanted intruders at bay. Predicting its future, however, is difficult. Researchers are constantly devising new cryptosystems that are often based on new, untested intractability assumptions. For every cipher that a cryptanalyst breaks, two more seem to sprout up in its place. Despite the fact that revolutionary discoveries in algorithmic might render entire classes of cryptosystems obsolete overnight, the field likely will continue to survive due to its breadth and diversity alone[7].

DNA SEQUENCE AND RANDOM NUMBERS

A living organism consists of different organs. Organs are made up of group of tissues and cells. Each group of cells performs different functions based on structure and origins in the body.

Each cell consists of cell membrane, cell wall, mitochondria, nucleus etc., each nucleus consists of nucleolus, which consists of DNA's. Deoxyribonucleic acid (DNA) consists of Purines are Adenine (A) and Guanine(G).Pyrimidines are Thymine(T) and Cytosine(C). Fig.1. shows the structure of double stranded Helix DNA sequence.

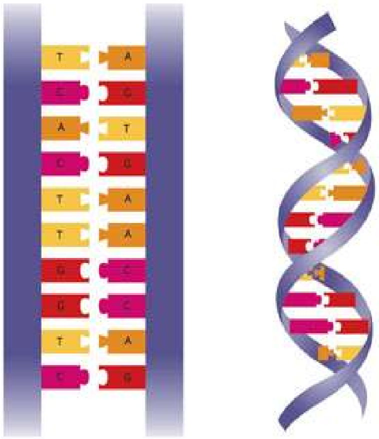

Fig.1: Structure of double stranded DNA.

The detection of the distortions of the sequence when the host sequence changed to some degrees. That was the best spot to start the wholly detection of the secret data . By advent of biological aspects of DNA sequences to the computing areas, new data hiding methods have been proposed by researchers, based on DNA sequences. The key portion of their work is, utilizing biological characteristics of DNA sequences. In order to convert binary data into amino acids as a DNA sequence, the base pairing rules must be used. Synthesizing nucleotides in real environment (biology) is done in constant rules [6].

The table I give the binary values for DNA sequence.

TABLE I
NUCLEOTIDE SEQUENCE FOR BINARY SEQUENCE

Binary Sequence		Nucleotide
0	0	A
0	1	T
1	0	C
1	1	G

The table II, gives the ASCII values for A,T,G,C and its binary equivalent representation.

TABLE II
ASCII VALUES FOR NUCLEOTIDE

Nucleotide	ASCII values	Binary Representation
A	65	1000001
T	84	1010100
G	71	1000111
C	67	1000011

Random number generator

A random function is a type of random element in which a single outcome is selected from some family of functions, where the family consists of some class of all maps from the domain the co domain.

The sequence of random numbers are generated using the function rand()and modulus(%) operator, within the range of 50 numbers.

Encryption and Decryption

Encryption method is done by Binary addition of ASCII values of ATGC and random key generated. Decryption method is done by binary subtraction of encrypted binary numbers with random key values.

Binary addition and Binary subtraction

Binary addition and binary subtraction performs operations only on 0's and 1's based on rule [8].

Table III is used to perform binary addition between the binary representation of ASCII values of DNA sequence and the binary representation of random numbers.

Table IV is used to form a binary subtraction between key generated by random numbers with encrypted binary sequence for the length of 8 bits.

TABLE III
RULES FOR BINARY ADDITION

A	B	A+B	Carry
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

TABLE IV
RULES FOR BINARY SUBTRACTION

A	B	A-B	Borrow
0	0	0	0
0	1	1	1
1	0	1	0
1	1	0	0

II. IMPLEMENTATION

The algorithm to implement data security in binary representation of DNA sequence is done by using the random function as well as using encryption and decryption algorithm, based on the method of binary addition and binary subtraction rule.

Algorithm for Encryption

```
Step 1: Convert text message into binary message.
Step 2: Convert binary message into DNA sequence
        using table 1.
Step 3: Generate a random numbers for each nucleotide
        of DNA sequence , using rand() function.
Step 4: Convert DNA sequence of step 2 into standard
        ASCII values, using table 2, then convert into
        Binary values.
Step 5: Convert random numbers of step 3 into binary
        numbers.
Step 6: Perform binary addition for step 4 & step 5.
Step 7: Repeat step 6 for entire length.
Step 8: gives out the complete binary encrypted message
        of original message.
```

Algorithm for Decryption

```
Step 1: Read the sequence of encrypted binary message
        with random numbers
Step 2: Apply Binary subtraction from key generated
        by random numbers with encrypted binary
        sequence for the length of 8 bits.
Step 3: Convert the result of step 2 into decimal number
        and convert to ATGC using table 2.
Step 4: Convert the result of step 3 into binary sequence
        using table 1.
Step 5: Results the original binary message.
```

III. DISCUSSION

Finding exact match of binary sequence is very difficult task. To break the data security, intruder has to know all possible combinations of binary representation of A,T,G and C Random number key value, Binary arithmetic operations. The probability is to find exact two bit binary message is $1/2^{30.}$

IV. CONCLUSION

One of the major problems in network is data security. Considering DNA characteristics and random keys, achieve new ideas in data security. This paper focus on the data security issues for providing a secure and effective encryption and decryption method by random number keys generation. Through this system model we can secure our data in a network.

REFERENCES

[1] Rothe J. (2002). Some facets of complexity theory and cryptography: A five-lecture tutorial. ACM Computing Surveys, 34(4).p. 504-549.

[2] Coron, J.S., “What is cryptography?”, IEEE Security & Privacy Journal, 12(8), 2006, p. 70-73.

[3] J. Watson, N. Hopkins, J. Roberts, J. Steitz, Molecular Biology of the Gene, fourth ed., Benjamin Cummings, Menlo Park, CA, (1987).

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

[5] A.L. Lehninger, D.L. Nelson, M.M. Cox, Principles of Biochemistry, Worth, New York, (2000)

[6] Mohammad Reza Abbasy, Bharanidharan Shanmugam., Enabling Data Hiding for Resource sharing in Cloud Computing Environments Based on DNA Sequences,(2011).

[7] Young, A., “The future of cryptography: Practice and theory”, IEEE IT Professional Journal, 2003, pp. 62-64.

[8] Torno, D.Exorand Technol., Orleans., France Parhami, B., Arithmetic operators based on the binary stored carry or borrow representation, on page(s): 1148-1152 7-10 Nov. 2010

AUTHORS

First Author – Siddaramappa V, Asst.Professor, VTU, India., Email: siddavmk@gmail.com

Siddaramappa V, received the BE degree in Computer Science and Engineering from the Visvesvaraya Technological University, (VTU) Karnataka, India, in 2007. He received the ME degree in Computer Science and Engineering, specialization in Bio-Informatics from Bangalore University, Karnataka, India, in 2010. He is working in City Engineering College, Department of CSE and ISE Bangalore, India as a Asst. Professor.

He is currently working on research in Computer Science and Engineering. His research interest includes Network Security, Neural Networks, Cryptography, Rule based Expert system for Medical Diagnosis Diseases, Bio-Informatics and Communications Protocol Design.