A Hybrid DNA Algorithm for DES using Central Dogma of Molecular Biology (CDMB)

U.Noorul Hussain, T. Chithralekha
Department of Banking Technology
Pondicherry University
Puducherry, India

A.Naveen Raj, G.Sathish, A.Dharani
Department of Information Technology
Sri Manakula Vinayagar Engineering College
Puducherry, India

,

ABSTRACT
DNA cryptography is an emerging field of DNA computing research in information security world. Still this field is in theoretical stage and not in active practice. To overcome this lacuna the present study proposed a fast and secured hybrid algorithm using DES (Data Encryption Standard) with biological concept CDMB (Central Dogma of Molecular biology) developed and implemented in high level programming language. This edifice has enabled to do computations digitally and gave the high level of security, effectiveness and applicability.

Keywords
Cryptography, Data Encryption Standard, DNA Computing, DNA Cryptography, Central Dogma of Molecular Biology, Ribonucleic Acid.

1. INTRODUCTION
The Enormous parallelism, excellent energy efficiency and extraordinary information density required in DNA (Deoxyribonucleic Acid) Molecules being explored for computing, storage (G.Cui et.al, [13]) and cryptography. In this research area, new computers, storage and cryptography may be invented and this may lead to an innovative revolution in information security world.DNA Cryptography is a new born field of cryptography which utilizes DNA as an information carrier with the help of molecular biology techniques. The traditional cryptography made great progress in 20th century with the expansion of electronic technology and it is widely used at present.DNA cryptography got attention after DNA Computing was first proposed by L.Adleman in 1994, and now it has been the leading edge of cryptography.

The world filled with huge information, starting from much confidential to least one. This means, there is a need to secure them by any of the method that makes them safe in good hands. Cryptography is also meant to keep the information more secure from attacks. Dated during BC, King Julius Caesar, who used an alphabet with a left shift of three called "Caesar cipher" to carry the secret message to his soldier in command. Thereafter, arrives the substitution cipher, transposition cipher, etc. The invasion of computer made the computation easier and also the introduction of novel algorithms is difficult to break. The Cryptographic techniques are mostly of mathematical computation which take in message and convert them into ciphers with the keys. Increase in key size, increases the hacking time and recovery of plain text [6]. The residual part of this article consists of related works, proposed algorithm, analysis and conclusion.

2. RELATED WORKS
Kang Ning [8], proposed an algorithm for DNA Cryptography. This algorithm starts with the sender end, the sender generates key himself, which contains introns starting and pattern codes. Then, sender translates the plaintext into DNA Sequence using information conversion program which afterwards simulate the splicing, transcription and translation process of central dogma with the respective program. In these processes, the necessary padding is also implemented. Through these steps, the starting and pattern codes of the introns, the places of the introns, the removed spaced introns, and the codon-amino acids mapping of the protein are added into the key file, and the enciphered information is also created. These two files can be then transferred to receiver through different channels (enciphered file through public channel, and key file through secure channel). On the receiver end, receiver receives the enciphered information and key file from different channels, and then uses the key information in the key file to decipher the enciphered information. Receiver first performs the reverse translation, reverse transcription and reverse splicing process using the respective program, and the information stored in the key file. After these processes, receiver can get the DNA form of information, and receiver can then recover the plaintext using the recovery program. By these means, the receiver finally gets what the sender intended to tell him.

Advantages of this work are Protein form of information can be transferred through public channel, and the size of the protein form of information is generally smaller than that of the original information. The theoretical analysis shows that this method is powerful against brute force attacks. The experiments reveal that this method is very efficient and very robust in computation, storage and transmission. But it suffers with limitation as differential cryptanalysis may break the partial information of the cipher text.

Lu Mingxin et.al [11], proposed the DNA Symmetric Cryptosystem (DNASC). The DNA hybridization is used for Key generation and decryption processes. In Key Generation, encryption key is a set of probes; the decryption key is a set of corresponding complementary probes. The hybridization conditions can be used as a part of decryption key or not. The sender will select probes from an existing experiment as the encryption key. The decryption key is sent to intended receiver through a secure channel Decryption. In Decryption, The receiver uses his decryption key to hybridize the DNA chips (cipher text) and receives the hybridization signals. This process is not a mathematic computation, but a biological

exploratory process. This process signals with the serve of computer and then obtains plaintext.

The advantages are this system can be realized massively in a parallel way. Its cost will be cut really with the advance of DNA chip technology, and it will continue to be spread throughout commercial usage. This system also could thus be vulnerable in usage with the future DNA computer. However they have certain limitations such as the security of this system relies on certain computation problems that are believed to be good, but not proven. The described system is still far-off from being a perfect crypto system.

3. BIOLOGICAL BACKGROUND OF THE STUDY

3.1 Nucleic Acid: DNA and RNA

The Deoxyribonucleic Acid (DNA) is the genetic content of the cell which carries the information from the parents, to their offspring. The double helix structure of DNA composed is of four nucleotides such as Adenine (A), Thymine (T), Guanine (G), and Cytosine (C). In this, Adenine always pair up with Thymine and Guanine pair up with Cytosine.

The RNA (Ribonucleic Acid) consists same as DNA except the Thymine replaced by Uracil (U). If the sequence of the DNA consists of GTAGA in one vertical strand then pairing will be as CATCT on the other strand which is the complementary rule.

3.2 Central Dogma of Molecular Biology

The DNA sequences are matched with the protein sequence called "Central Dogma of Molecular Biology". It is discovered by the Watson Crick who describes the process as Transcription and Translation. Transcription is the process in which the DNA strands is converted in RNA strands whereas the Translation converts the RNA strands into protein sequence. Based upon the building of DNA codon, the amino acid is made [8].

Combination of any three nucleotides of DNA makes the codon table and there is a corresponding protein for it. So by combination, out of 4 nucleotides 3 to be selected and hence 4x4x4=64 codon can be formed. So each codon can be applied for the amino acids as shown in table-I. As there is single amino acid name in contained more than one codon, the code for amino acid are suffixed by the alphabet in the amino acid name [8].

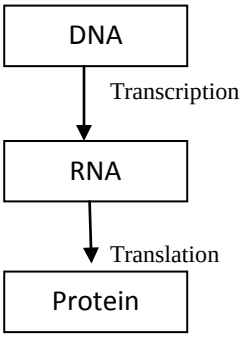

Fig. 1: Central Dogma of Molecular Biology process

The DNA which can accept the nucleotides and proteins, the way to convert the messages into proteins is by this Central Dogma of Molecular Biology, So as to make the computation easier to fix into the DNA.

Table 1. Codon table of DNA for Corresponding Amino acids

Nucleotide Codon	Amino Acid Name	Amino Acid code
GCT	Alenine	A
GCC		A1
GCA		A2
GCG		F
CGT	Arginine	F1
CGC		F2
CGA		W
CGG		U
AGA		U1
AGG		U2
ATT		U3
.	.	.
.	.	.
.	.	.
GTT	Valine	V
GTC		V1
GTA		V2
GTG		V3

4. ALGORITHM

The proposed system is the symmetric algorithm, both DES and Central Dogma of Molecular Biology (CDMB) uses the same key. As this hybrid algorithm involves the Data Encryption Standard (DES) [15], as usual it involves the permutation with Fiestal Function, with it the Central Dogma of Molecular Biology also explained below:

Step 1: Get the plaintext.

Step 2: By the initial permutation table the message to be encrypted is permuted.

Step 3: Then the 64 bit block is divided into equal halves of 32 bit left and right blocks. In this the right half is extended to 48 bits by expansion and XORed with keys and divided into 8 groups. The 6 bits of 8 blocks are input to the substitution box.

Step 4: The S-Box gives out the 4 bit output and the resulting 4 bits are grouped to give 32 bit block.

Step 5: Now this block will become the left half and the 3rd step left half is XORed with step 3 and will become the right half.

Step 6: To obtain the cipher text step 3 to 5 is repeated another 15 times. The output is the cipher text which is received. Based on the codon look up table, each letter in the DES output is replaced by DNA sequence.

Step 7: The DNA sequences are converted into RNA sequence by replacing the Thymine with Uracil.

Step 8: By the amino acid code RNA sequences are converted to protein sequence.

Step 9: As this is the symmetric cryptosystem, same key used for encryption and decryption so the reverse of the encryption are applied to decrypt the plaintext

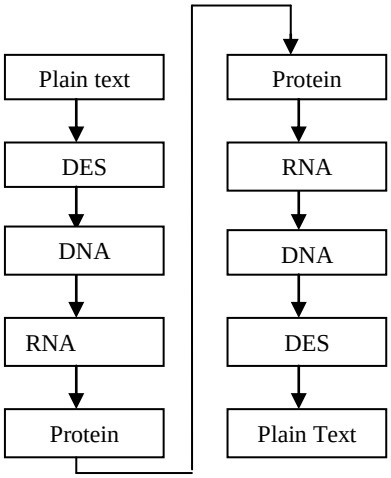

Fig 2: Process of encryption & decryption with DES and CDMB

5. IMPLEMENTATION AND ANALYSIS

In this section, implementation and the analysis parts are discussed. The comparison of the proposed algorithm with the existing algorithm is made the analysis results in Table III shows that our algorithm is effective than existing algorithm particularly in security, and the computational speed shown in figure 3. There is no need of using compression technique because the protein form of information is very smaller than the large DNA sequence.

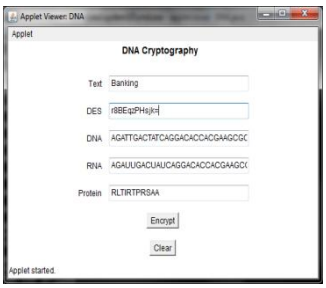

Fig 3: Implementation of Encryption

Thus, the presently proposed algorithm achieves a high-level of security due to the biological complex problem and cryptographic operations.

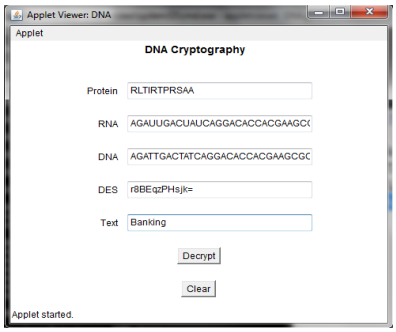

Fig 4: Implementation of Decryption

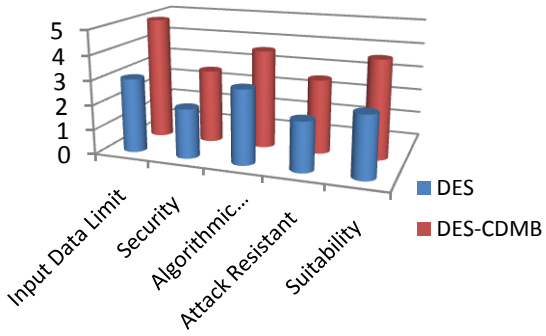

Fig 5: Result analysis of DES-CDMB

6. CONCLUSION

The objective this paper is to provide a stronger cryptosystem which will be used in the both the process of the DES and Central Dogma of Molecular Biology for the encryption and decryption by multiple cryptosystem for its higher security than the DES. The advantages of this system are unlimited length of the input data and the computation speed and length of the key. The properties of the length of the keys are proportional to hacking time and hence well utilized by this system for its security.

7. REFERENCES

[1] Morford, L, "A Theoretical Application of Selectable Markers in Bacterial Episomes for a DNA Cryptosystem", Journal of Theoretical Biology, p. 100-102, 2011.

[2] U.Noorul Hussain and T.Chithralekha, "Review of DNA Cryptology ", CiiT International Journal of Networking and Communication Engineering, Vol 3, No 13, October 2011, Pg.No 843-849.

[3] Niels Ferguson, Bruce Schneier, Tadayoshi Kohno, "Cryptography Engineering: Design Principles and Practical Applications" , 1st edition, 2010

[4] Ranbir Soram, Memeta Khomdram."Biometric DNA and ECDLP-based Personal Authentication System: A Superior Posse of Security". January 2010.

[5] Souhila Sadeg , Mohamed Gougache, Nabil Mansouri, " An Encryption algorithm inspired from DNA", IEEE, p 344 – 349, 2010.

[6] Shihua Zhou, Qiang Zhang, Xiaopeng Wei, "Image Encryption Algorithm Based on DNA Sequences for the Big Image", International Conference on Multimedia Information Networking and Security,884 - 888 ,2010

[7] Dr.G. Zayaraz, Dr. V. Vijayalakshmi and D. Jagadiswary. "Securing Biometric Authentication Using DNA Sequence and Naccache Stern Knapsack Cryptosystem". July 2009.

[8] Kang Ning. "A Pseudo DNA Cryptography Method". March 2009

[9] Xing Wang, Qiang Zhang, "DNA computing-based cryptography", Bio-Inspired Computing, p 1 - 3, 2009

[10] Xiutang Geng, Linqiang Pan, Jin xu, "A DNA sticker algorithm for bit substituion in a block cipher", "Journal of Parallel and Distributed Computing", September, 2008

[11] L. MingXin, L. XueJia, X. GuoZhen, and Q. Lei, "Symmetric-key cryptosystem with DNA technology," Science in China Series F: InformationSciences, Springer Verlag, Germany, vol. 50, no. 3, pp. 324–333, 2007.

[12] Sherif T. Amin, Magdy Saeb, Salah El-Gindi, "A DNA-based Implementation of YAEA Encryption Algorithm," IASTED International Conference on Computational Intelligence (CI 2006), 2006.

[13] G. Cui, Y. Liu, and X. Zhang, "New direction of data storage: DNA molecular storage technology," Computer Engineering and Application, vol. 42, no. 26, pp. 29–32, 2006

[14] Tanaka K, Okamoto A, Saito I., "Public-Key system using DNA as a one-way function for key distribution", "Biosystem",February 2005.

[15] Coppersmith, Don, "The data encryption standard and its strength against attacks", "IBM Journal of Research and Development", 243–250, (1994)