A Novel DNA based Encrypted Text Compression

D. Prabhu
Assistant Professor,
Dept Of Information
Technology Mailam
Engineering
College,Tamilnadu,India

M.Adimoolam
Assistant Professor,
Dept Of Information
Technology Christ College of
Engineering
&Technology,Puducherry,
India

P.Saravannan
Assistant Professor,
Dept Of Information
Technology Mailam
Engineering
College,Tamilnadu,India

ABSTRACT

Lack of trust and secrecy for communications laid foundation to a new area of security. Information security concerns with the protecting information and information systems from unauthorized access, use, disclosure, disruption, modification or destruction. Hence encapsulation of the original message with some kind of encryption techniques provides security. Most of the previous encryption techniques concentrated only on the secrecy of the message and not the size of the data being sent. Hence the quality of service and bandwidth becomes a bottleneck here. There is a problem of how to provide secrecy and also efficient bandwidth usage using special compression techniques which can compress an encrypted text. Our work addresses this problem by providing effective encryption and compression using concepts of polymerase chain reaction from DNA digital coding.

General Terms

Cryptography, Data Compression

Keywords

DNA Digital Coding, polymerase chain reaction, DNA amplification

1. INTRODUCTION

Cryptography made its roots in the world with the rise of lack of trust and necessity for secrecy in communications. Even after establishment of cryptography there were problems to choose a right encryption technique since plenty of options were available. Encryption technique was broadly classified into symmetric and asymmetric key cryptography. Both the sender and receiver receive used the same key for encryption and decryption this was the only method publicly known till June 1976[a].The asymmetric or public key cryptography uses two pairs of keys (public key, private key).These concepts of cryptography were applied to information systems which emerged as a new field as information security. Cryptography addressed confidentiality and integrity goals of information security. Kerchoff principle stated that "A cryptosystem should be secure even if everything about the system, except the key, is public knowledge"[2].

1.1 DNA Cryptography

DNA cryptography is a new born cryptographic field emerged with the research of DNA computing, in which DNA is used as information carrier and the modern biological technology is used as implementation tool.[2] In a pioneer study, Adleman[1] demonstrated the first DNA computing, which used DNA to solve a directed Hamiltonian path problem .It marked the beginning of a new stage in the era of information .This approach has been extended by Lipton[2] to solve another NP-complete problem which is the satisfaction problem .In the following researches, scientists found that the vast parallelism, exceptional energy efficiency and extra ordinary information density are inherent in DNA molecules. DNA computing provided, a parallel processing capability with molecular level, introducing a new data structure and calculating method. It can simultaneously attack different parts of the computing problem put forward challenges and opportunities to traditional information security technology .For example in 1995,boneh et al[6]. demonstrated an approach to break the Data Encryption Standard (DES) by using DNA computing methods. In 1999 Clelland[10] et al. achieved an approach to Steganography by hiding secret message encoded as DNA standards among a multitude of random DNA.DNA is a new born cryptographic field emerged with the research of DNA computing. In which DNA is used as information carrier, this modern biological technology is used as an implementation tool. The vast parallelism and extra ordinary information density inherent in DNA molecules are explored for cryptographic purpose such as encryption, authentication, and signature and so on.

The research of DNA cryptography is still at the initial stage and there are many problem solved .The new born DNA cryptography is for from mature both in theory and realization and this might be the reason why only few example of DNA cryptography where proposed .On the other hand current DNA technology is beta stage of laboratory exploration and focus on experiments ,why there is not a general theory about applying DNA molecule in cryptography .Some key technologies in DNA research such as polymerase chain reaction (PCR).

DNA synthesis and DNA digital coding have only been developed and well accepted in recent years. PCR is a fast DNA amplification technology based Watson–crick complementary model, but it would still be extremely difficult to amplify the message –encoded sequence without knowing the correct two primer pairs. DNA cryptography should be implemented by using modern biological technique as tools and biological hard problems as main security basis to fully exert the special advantages .As described above, the correct two primer pairs would be used as the key by applying the special function of primers to PCR amplification. On the other hand the security of

traditional cryptography is based on difficult mathematic problem which is mature both in theory and realization. There are many effective cryptosystem of traditional cryptography such as DES, AES, RSA are implemented widely some even have been extended (DDES & TDES). There by DNA cryptography does not absolutely repulse traditional cryptography and it is possible to construct hybrid cryptography

By studying the analysis of encryption technique, several algorithms are proposed on the basis of traditional encryption technique like DES, RSA, and AES etc. Most of our encryption are based on the basis of number system conversion .An analysis of algorithm of an encryption scheme using DNA technology, the message text will be converted into DNA digital code format, then PCR amplification is implemented there by creating large number of primes generation result large value of DNA data of the form (A, T, G, C). Data Compression is applied to get effective compression ratio of the encrypted text.

2. DIFFICULT BIOLOGICAL PROBLEM USED IN THIS SCHEME

DNA is the germ plasm of all life styles. In a double helix DNA string, two strands are complementary in terms of sequences, that is A to T and C to G according to Watson-crick rules, which is one of the greatest scientific discoveries.

The modern cryptography is based on difficult mathematic problem such as the Non-Deterministic Polynomial Time Complete (NP-C) Problem. Quantum cryptography is based on Heinsberg uncertain principle, which can also be regarded as a difficult biological problem. The difficult biological problem referred as "It extremely difficult to amplify the message encoded sequence without knowing the correct PCR two primer pairs "Polymerase Chain Reaction(PCR) is fast DNA amplification technology base complementary and is one of the most invention in modern biology .Two complementary oligonucleotide primers are annealed to double-standard target DNA strands ,and necessary target DNA strands and the necessary target DNA can be amplified after a serial of polymerase enzyme.. PCR is a very sensitive method and a single DNA target molecule can be amplified into 10^6 after 20 cycles in theory, this can done within short time. It is a special function in PCR amplification that having correct primer pairs. It would extremely difficult to amplify the message encoded sequence without knowing the correct the primer pairs. So believe that this biological problem is difficult and will last a relatively long time.

2.1 Exclusive Disjunction Operation

XOR operation is applied over the message and key to increase the repetition of normalized binary bit of 1's and 0's, there by gaining high compression factor. Moreover exclusive disjunction is computationally inexpensive

Truth Table		
P	Q	OUTPUT
1	1	1
0	1	0
1	0	0
0	0	1

Table1: XOR operation

2.2 DNA Digital Coding Technology

In the information science , the most fundamental coding method is Binary Digital Coding, which is anything can be encoded by two state 0 or 1 and a combination of o and 1,there are four kind of bases, which are ADENINE(A) and THYMINE(T) or CYTOSINE(C) and GUANINE(G) in DNA sequence. The simplest coding pattern to encode nucleotide bases (A,T,G,C) is by means four digits: 0(00),1(01),2(10),3(11),there are possibly 4!=24 possible pattern by encoding format like (0123/CTAG).
Pros:

1. To decrease redundancy of information coding and increase the efficiency compared to traditional encoding methods.

2. By using the technology of DNA digital coding, the traditional encryption method such as DES or RSA could be used to preprocess the plaintext.
3. The digital coding of DNA sequence convenient for mathematical operation and logical operation.

3. PROPOSED SYSTEM

In our system, get the text based message from the user and convert the message into Hexadecimal code and Binary code. Message is split into parts, one is used as message end other one is used as key, the XOR operation is performed for the purposed of high compression factor.
DNA digital coding is applied over the message and get the DNA base coded message, then PCR amplification is implemented by using two prime pair as key and compression is performed for variable length of data.

Though by various mode of operation happen in serial fashion and provide double layer (Bi-Layer)security and therefore it is called as Bi-serial DNA Encryption Algorithm(BDEA).
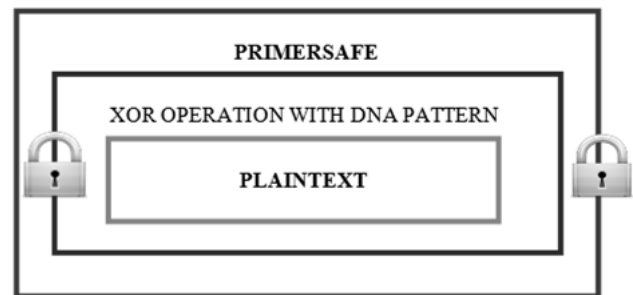

Fig1: Double Layer Security

3.1 Encryption

A plain text is converted into Hexadecimal code and then converted into Binary code using decimal convertor. XOR operation can be performed using the code as Binary code , in that one part is used as a message(kA) and another part used as key(kB), using the Binary code XOR operation to be performed and the XORed output is taken for DNA digital coding to pre-process the binary text into DNA digitalcode and PCR amplification is performed by using two primer pairs. Using the primer pairs, we can generated the maximum populated data depend upon the primer pairs and then high compression algorithm is used to compress the variable -length data to minimize the memory size.

The kA and kB size to be same, but key values are different from each other, after decryption process combining the correct keys only. we get the original data, otherwise there is a possibility of missing some part of data.

3.2 Key Generation and Transmission

Diffie–Hellman key exchange (D–H) is a cryptographic protocol that allows two parties that have no prior knowledge of each other to jointly establish a shared secret key over an insecure communications channel. This key can then be used to encrypt subsequent communications using a symmetric key cipher.

Generation of keys

1. Primer1.

2. Primer 2.

3. Hexadecimal key**.**

Three keys are send to the receiver using key agreement proposed by Diffie–Hellman.
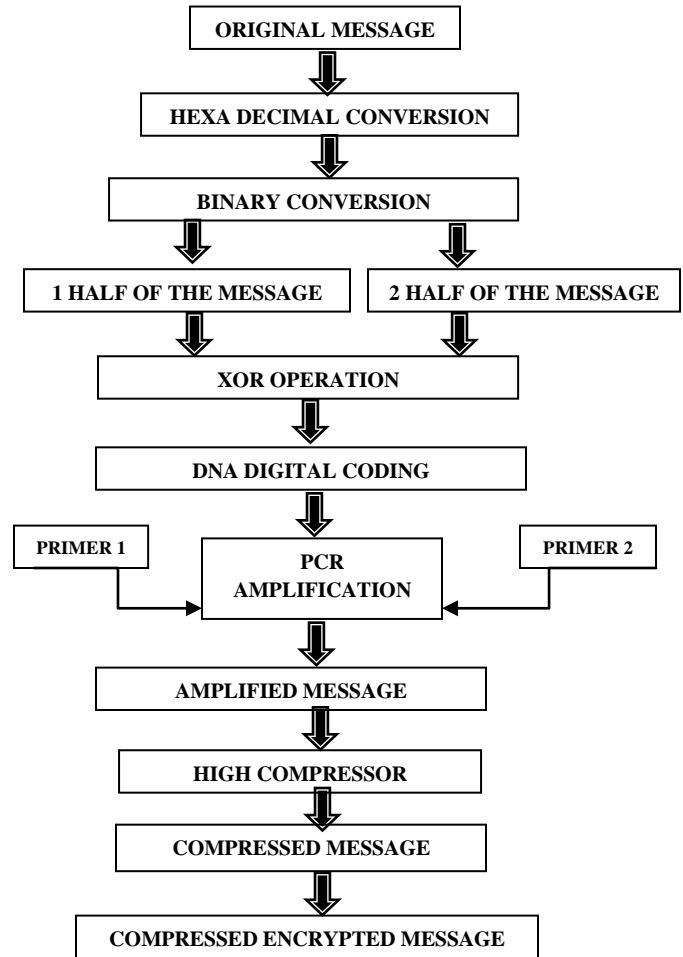

Fig2: Encryption

3.3 DECRYPTION

From the receiver side, get the encrypted data, by using high decompression algorithm to recover compressed data, then using the correct two primer pairs to retrieve the DNA digital coding.

DNA digital coding is converted into binary code , then XOR operation is performed by using the binary code and key is given by user, then combining the key and XORed output value to retrieve large Binary code and converted into Hexadecimal code. Hexadecimal code is converted into normal plain text by using Decimal Convertor.

In case anyone of the key is wrong, then there is possibi1lity missing of data or improper form of data. So, there is chance of maintaining more secure of data.

3.4 SECURITY

Regardless of many differences between DNA and traditional cryptography, they both satisfy the same characteristic of cryptography. The security requirements should based depend only on the secrecy of decryption key. In this scheme encryption scheme as a key kA, kB and pair of PCR primer pairs.

The security of the encryption scheme come from two levels:

The first level is biological difficult problems its extremely difficult to recover the DNA digital code without knowing the correct two primer pairs. When the adversary gets the sending sample and tries pick out the message encoded sequence without knowing the correct PCR primer pairs would take 10^{23} kinds of sequences. So only when both primer pairs are correct, then only retrieve the DNA digital code by amplification. On the other hand, if one the primer pair is wrong then it cause biological pollution, that leads to corruption of data.

PCR amplification require two primer pairs, to amplify the message encoded sequence , so we can use as one of the prime pairs as private key and another one as public key. It provides security unless there is development in DNA sciences , in case the DNA layer can be broken , the second layer is present to save the message from intruders as one part of message as key. So if an intruder break the DNA layer , the intruder can't get the full message because one part of message is kept as key. By combing the key with amplified binary message and key only, we recover original text else uncompleted message or polluted message. So,the biological and traditional method provides double safe guard to the message.

In case, intruder is hack our code, then we change the bit pattern in the DNA digital coding and also binary pattern, leads to Secure.

The first level is biological difficult problems its extremely difficult to recover the DNA digital code without knowing the correct two primer pairs. When the adversary gets the sending sample and tries pick out the message encoded sequence without knowing the correct PCR primer pairs would take 10^{23} kinds of sequences. So only when both primer pairs are correct, then only retrieve the DNA digital code by amplification. On the other hand, if one the primer pair is wrong then it cause biological pollution, that leads to corruption of data.

PCR amplification require two primer pairs, to amplify the message encoded sequence , so we can use as one of the prime pairs as private key and another one as public key. It provides

security unless there is development in DNA sciences , in case the DNA layer can be broken , then another layer is present to save the message from intruders as one part of message as key. So if an intruder break the DNA layer , the intruder can't get the full message because one part of message is kept as key. By combing the key with amplified binary message and key only, we recover original text else uncompleted message or polluted message. So,the biological and traditional method provides double safe guard to the message.

In case, intruder is hack our code, then we change the bit pattern in the DNA digital coding and also binary pattern, leads to Secure.

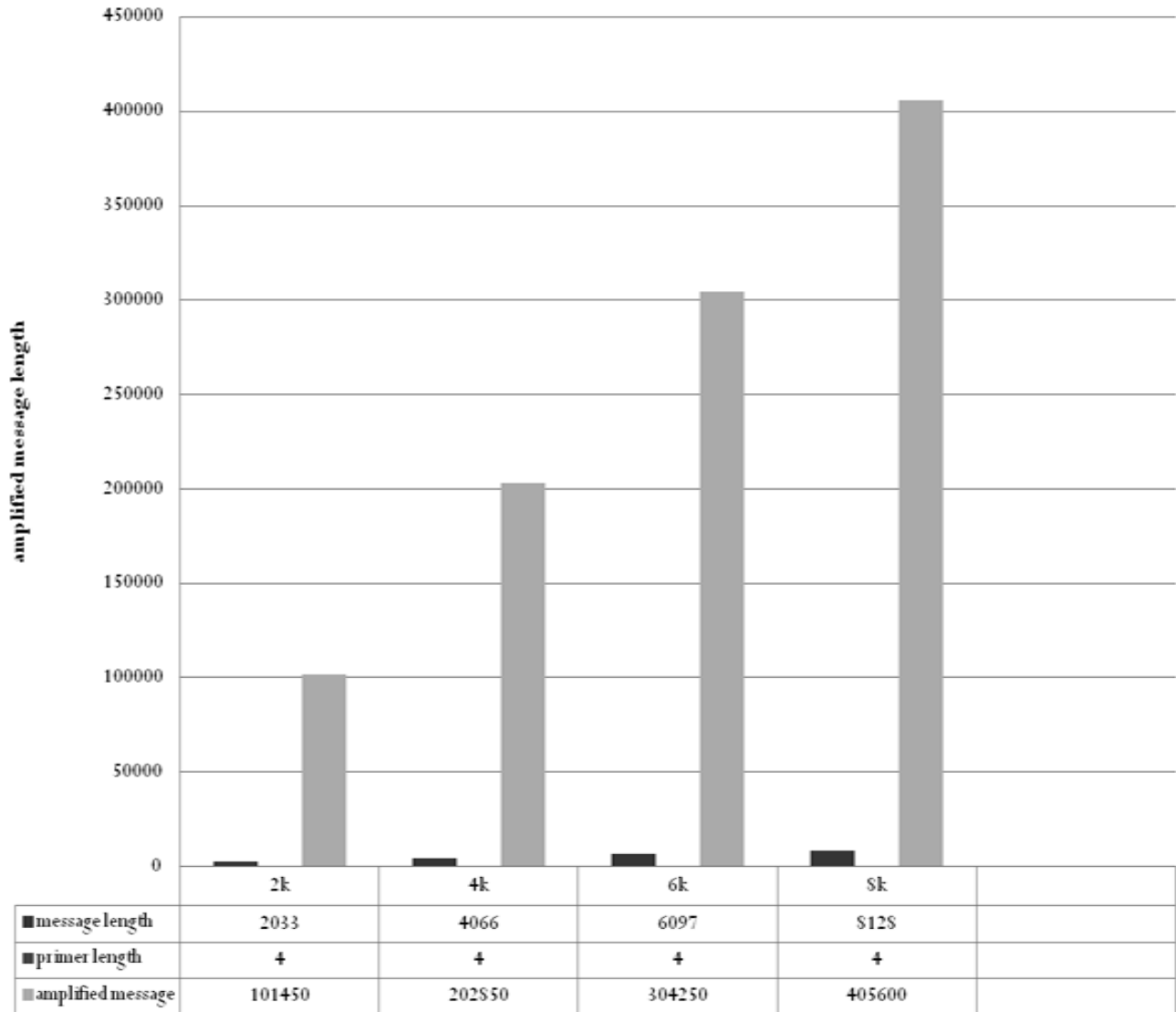## 3.5 System Analysis & PCR Amplification Results
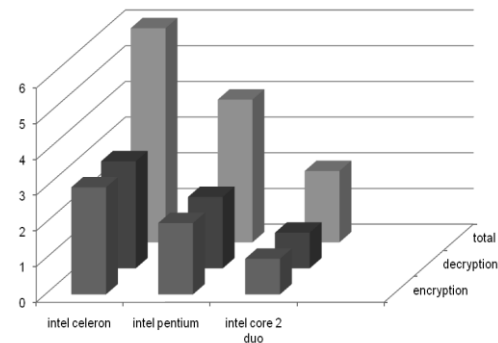
Graph1.Sytem Analysis

Graph 1 shows the time taken by system in various processors for encryption, decryption, both together as a complete set.

Metrics derived from the Graph 1 is increase computing power is directly proportional to reduced computing time

ANALYSIS OF PCR AMPLIFICATION

	2k	4k	6k	8k	
■ message length	2033	4066	6097	8128	
■ primer length	4	4	4	4	
■ amplified message	101450	202850	304250	405600	

Graph.2 Analysis of PCR Amplification

Graph 2 shows the amplified text size of various size of input messages with the primers.

Metrics derived from the Graph 2 is increase in message length is directly proportional to increase amplified message length keeping the primer constant.

Several other possibilities must be discovered on further development of the project leaving it for future enhancement

4. CONCLUSION
In this paper, we designed Bi-serial DNA encryption algorithm containing technologies of DNA synthesis, PCR amplification, DNA digital coding, XOR operation as well as traditional cryptography.

The intended PCR two primer pairs was used as the key of this scheme that not independently designed by the sender or receiver. This operation could increase the security of encryption method. On the other hand, the traditional encryption method and DNA Digital Coding are used to preprocess operation we can get completely different cipher text from the same plaintext, which can effectively prevent attack from possible word as PCR primers. The complexity of biological difficult problem and cryptography computing difficulties provide a double layer security. And the security analysis shows that encryption scheme has high confidential strength. More over the cost of encryption will be greatly with process of techniques. We have concentrated on the space and time

complexity during computation, in that way introducing compression technique to compress the encrypted data and to minimize the utilization of memory and time during computation .This will provide potential support to our proposed system.

5. ACKNOWLEDGMENTS

Our thanks to the experts who have reviewed our paper and given valuable suggestions .we also thank our students manicavel, jagadeesh and ragunath for their support in the constant development and contribution on this project. We also thank our college management for providing the necessary infrastructure and constant support

6. REFERENCES

[1] Guangzhao Cui, Limin Qin ,Yanfeng Wang and Xuncai Zhang,"An encryption scheme using DNA technology". BICTA 2008G.

[2] Z. Cui, "New Direction of Data Storage: DNA Molecular Storage Technology," Computer Engineering and Applications, vol. 42, pp.29–32, 2006

[3] J. Kawai and Y. Hayashizaki, DNA book. *Genome Res*, Vol. 13, pp.1488–1495, 2003

[4] T. Kamei, "DNA-containing inks and personal identification systemusing them without forgery," *Jpn. Kokai TokkyoKoho*, 2002, p.8.

[5] A. Leier, C. Richter and W. Banzhaf, "Cryptography with DNA binarystrands," *Biosystems*, vol. 57, pp. 13–22, 2000

[6] C. T. Celland, V. Risca and Bancroft C. "Hiding messages in DNA microdots," *Nature*, vol. 399, pp. 533–534, 1999.

[7] L. Kari, "DNA Computing: Arrival of Biological Mathematics," *TheMathematical Tntelligencer*, vol. 19, pp. 9–22, 1997.

[8] R. J. Lipton, "Using DNA to solve NP-complete problems," *Science*,vol. 268, pp. 542–545, 1995

[9] L. M. Adleman, "Molecular computation of solutions to combinationalproblems," *Science*, vol. 266, pp. 1021–1024, 1994.

[10]Whitfield Diffie and Martin Hellman, "New Directions in Cryptography", IEEE Transactions on Information Theory, vol. IT-22, Nov. 1976, pp: 644–654.

[11] Auguste Kerckhoffs, "La cryptographie militaire", Journal des sciences militaires, vol. IX, pp. 5–83, Jan. 1883, pp. 161–191, Feb. available online