Analysis on DNA based Cryptography to Secure Data Transmission

S.Jeevidha
Dept. of CSE
Pondicherry University
Pondicherry, India

Dr.M.S.Saleem Basha
Asst Professor, Dept. of CSE
Pondicherry University
Pondicherry, India

Dr.P.Dhavachelvan
Professor, Dept. of CSE
Pondicherry University
Pondicherry, India

ABSTRACT
The biological research in the field of information technology paves the exploitation of storing capabilities, parallelism and also in conservative cryptography which enhances the security features for data transmission. DNA is the gene information which encodes information of all living beings. Though the DNA computing has its application in the field of huge information storage, massive parallel processing, low energy consumption which have been proposed and proved by the researchers and soon the molecular computer can replace the existing silicon computer and it exploits the world smallest computer. The combination of DNA molecules can be interpreted as a result to give a solution to a specific problem. The DNA strands can be replicated 500 times per second with greater accuracy. It can also be used in the field of cryptography based upon the vast parallelism which is used to break the existing cryptographic approach. This paper analysis an existing approach to the DNA computing method and DNA based cryptographic approach which provides the clear idea and limitations of existing research works.

Keywords
DNA, DNA Computing, DNA Cryptography

1. INTRODUCTION
In the year 1994, Adleman[1] sets the step for the bio-computing research which introduced the idea of DNA to solve the complex mathematical problem and also he concluded that DNA has computational latent. He got this idea from the book "Molecular biology of the gene" which was written by James Watson who discovered the structure of DNA in 1953. His idea is to solve the unsolvable problems in the computer field using chemistry by conservative computers which requires a vast amount of computation. The idea of DNA computing involves the study of mathematics, biology, and chemistry and computer science. The biological researchers proved that DNA is similar to computer (i.e.) Logic gates are used in the silicon based computer which is used to transmit signals from the binary codes. He solved the directed HPP (Hamiltonian Path problem) with seven vertices in graph which the molecules are encoded in a sequence and the computation is performed by biochemical operations. Lipton [2] solved another NP-complete, the satisfaction problem which can also be called as searching algorithm .This study ascertained how problems analogous to Boolean formulas are solved using extremely parallel [12] processing method. Such a method makes use the facility of DNA sequences that hybridize specifically to their complementary sequence. In 1953, Watson [3] discovered the structure of DNA is actually a double helix; The first DNA computer was developed with a concept of "DNA strands are useful to encode a information". He solved the problem in the molecular biology laboratory, in which he did experiment using Hybridization, Gel separation and PCR (Polymerase Chain Reaction) sequencing by handling DNA sequences in a test tube. Clelland [4] (1999) have demonstrated an approach to steganography by hiding secret messages encoded as DNA strands among a multitude of random DNA. He used substitution cipher for encoding a plaintext where a unique based triplet is assigned to each alphabet, numeral and characters. The DNA offers efficient parallel [12] molecular computation and huge storage which has been proved by the researchers is used in the various applications by solving the issues such as expansive and time consuming problems. The amplification technology of DNA is PCR (Polymerase Chain Reaction). It is tremendously difficult to amplify the message encoded sequence without knowing the proper forward primers and reverse primers. The primers can be used as the key to encrypt data. The study on DNA computing is used to solve heavy combinatorial problems by its parallel and huge storing capabilities. The emergence of DNA computers with higher speed can be used as a substitute for the existing silicon computer. According to molecular biologist DNA is a data storing device which performs parallel computations.Molecular biologists have a long thought that "DNA as an information storage device". The cells in our human body store this information with an impressive array of computing machinery which the researchers exploited to carry out a few of our own calculations. DNA computing may not be fast but it is massively parallel. With the right kind of setup, it has the potential to solve huge mathematical problems. It is an astonishing fact that computation using DNA represents a threat to various powerful cryptographic techniques. DNA computing is a new-fangled computational pattern that employs molecular manipulation to solve computational problems which also explores natural processes as computational models. In 1994, Adleman [1] in the lab of molecular science with the Department of Computer Science, and University of Southern California surprised the scientific community by using the tools of molecular biology to solve a different computational problem. The main idea was the encoding of data in DNA strands and the use of tools from

molecular biology to execute computational operations. The instruction in computer are carried out by logic gates which adapt binary code moving from the computer to perform operations, The silicon computer use logic gates which interpret input signal from silicon transistors and output signal to perform complex operations.

2. DNA COMPUTING

The computation using DNA sequences is called DNA Computing. The information can be encoded in a DNA sequence and it can be computed. Due to the vast parallelism of DNA computing, DNA itself acts as a microprocessor in which the silicon processors are replaced in future. As DNA it stores huge amount of information and perform immense parallel computation, it can process large number of problems in a fraction of second. Each DNA molecule or group of molecules could be treated as a processor that can store billion times larger data than our personal computer perform parallel computation at the same time. DNA and RNA are alluring media to store data which exceeds the storage capacity of electronic, physical and optical media. A single molecule can also extend to store large memory and build a storage device which is much times faster than a traditional computer. It can be used to search possible solutions simultaneously which is represented by DNA strands. By using the parallelism [1] concept, it can be used in the field of encryption and decryption by generating public and private keys.DNA can be embedded with computer chip and it can be used as a bio chip. DNA computation requires only less energy which can be automated, as it is a nature of DNA strands. The DNA computer hardware and software can all be made up of bimolecular where it does not require power efficiency. Adleman[1] assigned a graph of seven cities to different sequences of DNA which is 20 molecules long, then dropped the sequences into a heat of millions more sequences of DNA which is naturally bonded with the cities. It generated a random path of thousand sequences as like the computer which can be used to break the codes by filter random numbers.

3. DNA CRYPTOGRAPHY

According to the Watson [3], The DNA strands can be useful to encode information. In encryption method ,the data has been developed through DNA.As the DNA cryptography is the emerging topic only few algorithms have been proposed and it is highly far away from the real time implementation [11].Even though the DNA cryptography is effective but it is not as much effective than traditional cryptography, But it can be combined with existing cryptography to give hybrid security[12].This encryption methodology been used to hide the original data which also includes the set of methodologies to store and encrypt the message in terms of DNA sequences. The advantage of DNA encryption methodology is that it addresses the set of extended ASCII through which all kind of digitized information can be encrypted. The DNA sequences can be synthesized and merged with denatured DNA or dummy DNA strands of DNA database.

3. BIOLOGICAL STUDY

3.1 DNA

DNA is Deoxyribo nucleic acid, and it is the family of molecules which is referred as nucleic acid with two strands of sugar phosphate backbone.

Fig1: Four Nitrogen Bases

It is the substance from which our genes are made; it has the capacity to perform calculations many times quicker than any other computer existing in this world. It contains the genetic instruction needed to construct other cells like RNA and proteins. The complex structure of the living body consists of human parts which are the result of applying simple operations to the initial information encoded in a DNA sequence called genes. Likewise the complicated mathematical operation is made up of simple addition and subtraction .The major advantage of DNA which contains a four bases, Adenine (A), Thiamine (T), Guanine (G) and Cytosine(C).According to Watson Crick [3] complement condition, in a double helix DNA, The oligonucleotides or DNA strands with the four bases bind together A is double bonded with T, C is triple bonded with G. These oligonucleotides combine together in a anti parallel way with respect to the chemically distinct ends 5' and 3' of the DNA molecule which is represented in fig(1). This complementary property making DNA a unique data structure for computation can be exploited in many ways. For example, the chemical reaction such as Hybridization, Ligitation is useful in solving mathematical problems. In the process of hybridization the two complementary single strands are combined together to form a double strand by the chemical reaction ,In the process of ligitation ,the two double DNA strands are combined together to form a double strand. The simplest coding patterns, To encode the 4 nucleotide bases (A, T, G, C) is by the digital coding of DNA sequence. It is very convenient in mathematical and logical operation which gives the impact on the creation of DNA bio –computer. The computation investigated experimentally influenced the way to the upcoming researchers to develop this computing paradigm in theoretical and practical way. The difficulty raised by the Polymerase Chain Reaction is the DNA structure design, assembly of DNA sequence, self repairing and concatenation of DNA sequences with PCR.DNA enzymes simply make mistakes, cutting where they shouldn't be inserting a T for a G.DNA can also be damaged by thermal energy and UV energy from the sun. If the error occurs in one of the strands of double stranded DNA, repair enzymes can restore the proper DNA sequence by using the complement strand as a reference The chemical reaction such as Hybridization, Ligitation is useful in solving mathematical problems. (i.e.) In the process of hybridization the two complementary single strands are combined together to form a double strand by the chemical reaction , In the process of ligitation ,the two double DNA strands are combined together to form a double strand. The simplest coding patterns to encode the 4 nucleotide bases(A, T, G, C) is by the digital coding of DNA sequence which is very convenient in mathematical and logical operation which gives the impact on the creation of DNA bio –computer.

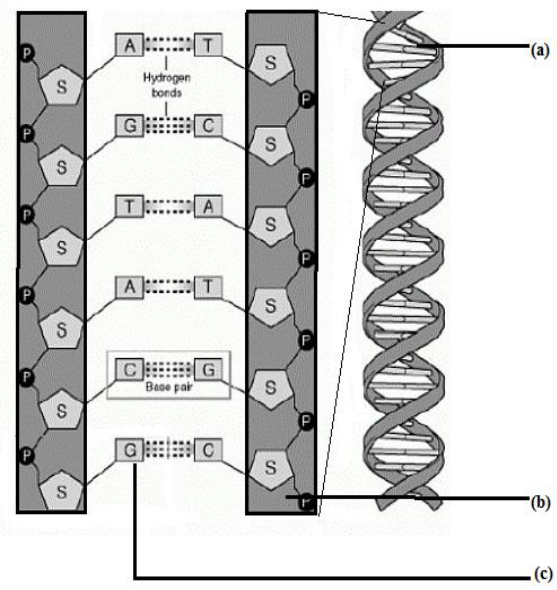

Fig 2: DNA Molecular Structure

a) Double Helix Structure b) Sugar Phosphate Backbone c) Base pair of oligonucleotides

Instead of using electrical pulses, The DNA uses chemical properties to perform these operations such as Synthesis, Separation, Merging, Extraction, Annealing, Amplification, Cutting, Ligitation and Detection. It is possible to anneal and separate the two antiparallel and complementary DNA strands because the hydrogen bonding between two complimentary sequences is weaker than the other one. The known DNA sequence can be taken out from the test tube of DNA strands by denaturing the double stranded molecules, The probe can be attached to the filter and denatured molecules remains.DNA can act as logic gates with DNA sequence as input , By using the chemical operation the DNA sequence can perform the logical operation .In two ways the DNA computation can be performed one is in the liquid state other is in the solid state as DNA strands are capable of producing billion answers simultaneously.

3.2 BIOLOGICAL OPERATIONS

The biological concepts of DNA cryptography is the main activity of the cell in our body which is incorporated in the field of DNA computing.

3.2.1 HYBRIDIZATION:

The double stranded DNA is formed by combining single stranded DNA sequences, In this process A always pairs with T and G always pairs with C according to Watson crick complement condition Consider an example of two single stranded sequence 5'TGTGCGCG 3'and 5'GCGCACAC3'will combine to form a Double stranded sequence in a double helix structure.

5'TGTGCGCG 3'

3' ACACGCGC 5'

3.2.2 DENATURATION:

The double stranded DNA molecule can be heated up to 90 degree Celsius, so that it is resolved into two single stranded DNA.

3.2.3 LIGITATION:

If the DNA sequences are dropped in pipette the DNA sequences recombine with other Sequence by means of some enzyme. The oligonucleotides is heated to 95 degree centigrade and cooled to 20 degree Centigrade – 1 degree Centigrade per minute for hybridization. The reaction is then subjected to ligitation with another DNA sequence to produce new DNA sequences.

3.2.4 POLYMERASE CHAIN REACTION:

It is one of the ways to amplify the DNA sequences; the small amount of double stranded DNA which is to be complementary to each other is taken with the great amount of Primers. Primers are nothing but the small amount DNA fragment in single stranded DNA in the test tube which is subjected to heat at 90 degree Celsius. The reaction is used for amplifying the amount of DNA fragment which is performed in cycles. It is converted to single stranded DNA. By using primers it is possible to generate billions of different DNA sequence. An enzyme builds complementary strands with the base of short primers.

The following are the steps in the reaction of PCR

STEP1: DENATURATION:

The Double stranded DNA is subjected to get separated in to single stranded DNA at 90 degree Celsius heat.

STEP2: HYBRIDIZATION OF PRIMERS:

The amount of primers is much greater than the amount of amplified DNA at (40-60) degree Celsius.

STEP3: ELONGATION:

The polymerase enzyme adds nucleotides to the strand of short primer on base of original DNA strand. The DNA strands between primers amplified. It is done approximately at 72 degree Celsius.

3.2.5 GEL ELECTROPHOSIS:

Based on the constraint, the DNA molecules are negatively charged. The length of the DNA molecules can be obtained. DNA fragments in certain condition move constantly in the electromagnetic field, the speed of electromagnetic field is inversely proportional to logarithmic of their weight. The longer sequence will go slower than the shorter one, after a period of time depending upon the length of the fragment DNA sequences can be separated by using the special chemical amalgam to identify the sequence.

4. BACKGROUND WORK OF DNA CRYPTOGRAPHY

Boneh [5] started his pioneer work on breaking Data Encryption Standard which leads to more difficulties and much cost for experiment. The Kang [9] explained the pseudo encryption methodology based upon A.Gehani[7] work. He explained the encryption i.e. the plain text is converted to DNA sequences and these sequences are converted to the spliced form of data and protein form of data by cutting the introns according to the specified pattern and it is translated to mRNA form of data and mRNA is converted into protein form of data. The protein form of data is sent through the secure channel. Chen [8] proposed sticker based DNA cryptography method by using number of enzymes with memory strands and sticker strand which may annealed together. The stickers connected to memory strand are memory complex. During sticker computation the region is

identified exactly by one bit position and one Boolean variable in which the memory strands are N bases length and K overlapping regions. If the sticker is attached to the memory strand and the Bit is equal to one and if no sticker is attached to the region, it is equal to zero which has the logical true or false. Identical memory strands of longer length will be represented as a set of bit. The bit strings corresponds to a unique organization of memory strand Adleman [10] provided the description of the breaking DES by using the gram of DNA even in the presence of large errors. G.Z cui [12] has proposed the encryption scheme by using the PCR, the DNA digital coding. The original message is converted to hexadecimal code and again it is converted into binary. The binary digits are transformed into DNA sequence and it is considered as a DNA template. The forward primer is chosen to perform PCR. The DNA sequence is changed. The original message is entirely different from the obtained sequence. The reverse primer is used to convert the PCR DNA to original message and it is transformed into binary and which is then converted into plaintext message. It has both biological difficulty as well as mathematical difficulty. It is difficult for the adversary to identify the original message. By the use of symmetric key block cipher and biochemical methods such as transcription and translation. G.C [6] proposed an encryption scheme. A plaintext of message is converted into 4*4 matrixes and initial permutation is performed. The XOR operation is performed with the generated key .A matrix can be transposed and a secret key is generated which is given to the DNA module and the permutation is performed to produce the cipher text. Monica Borda[15] proposed a secret writing method using DNA with the concept of one time pad. Using XOR and chromosome indexing the message is converted into binary in which each bit is encoded with nucleotides and encapsulated with primers. A long DNA sequence can be generated using short oligonucleotides sequences. The delimitation of DNA segment can be done using shortening the length. By using molecular biology concept, the new cryptographic method [9] has been used to simulate the DNA biological operations .The sender knows the starting codes and pattern codes which records the introns places a. The cut out DNA introns are translated into mRNA form of data. Sender again translates mRNA form of data into protein according to genetic code table. The key is sent to the receiver in a secure channel to recover DNA form of information. By applying contemporary bio technology [17] hooked on the field of cryptography, Encryption and Decryption keys are created by DNA probes. Encryption is done by DNA fabrication, Decryption is done by DNA hybridization. The Cipher Text is embedded in DNA chip with the most difficult DNA microarray technology is to attain information security in a cryptosystem. Kang Ning [9] idea is that the sender encodes her message in the original DNA sequence which allows this to be DNA transcription and DNATransalation. The resulting protein is like a public key which can be sent to the receiver in a public channel. Meanwhile, the sender sends to the receiver, a secret key which consists of the message. It needs to reassemble the DNA such as the location of the non coding regions that need to be reinserted. This type of cryptography is used to secure powerful attacks and also it is very difficult to generate keys .The process of encryption and decryption requires an informational message to be transmitted.In an agreed encryption and decryption algorithms, The plaintext message data is encoded in DNA strands using the publicly known alphabet of short oligonucleotides sequences.

Table 1. Comparison between different DNA cryptographic methods and DNA Technology

Cryptographic methods	DNA Technology Used	Concept
An Encryption Scheme Using DNA Technology	DNA digital coding PCR primers	A message is converted to DNA template in which primers are used as key to encode and decode the message [6].
An Encryption Algorithm Inspired From Dna.	Symmetric key block cipher algorithm Transcription(DNA-RNA) Translation(RNA-Protein)	A message is converted into matrix with initial permutation and XOR operation is performed with the key which is subjected to DNA module transcription and translation. [13].
A Pseudo DNA cryptography Method	Transcription Splicing Translation	Sender translates mRNA form of data into protein according to genetic code table. The key are send to the receiver in a secure channel [9].
Asymmetric Encryption and Signature method with DNA technology	DNA-PKC PUBLIC KEY PRIVATE KEY (Generated from primers)	An asymmetric method used to protect the data from tampering [18].
Symmetric Key Cryptosystem With Dna Technology	DNA fabrication DNA hybridization DNA chip DNA Microarray	Encryption and Decryption keys created by DNA probes. Encryption is done by DNA fabrication. Decryption is done by DNA hybridization and Cipher Text is embedded in DNA chip. Most difficult DNA microarray technology is to attain information security in a cryptosystem [16].
A DNA-based, Bimolecular Cryptography Design	Carbon Nano Tube Technology	A nano scale used to alter the message [17].

A long DNA sequence can be generated using short oligonucleotides sequences. The delimitation of DNA segment can be done using short length, this technique was done using DNA hybridization, and the message is converted to binary form which each bit is encoded with nucleotides and encapsulated with primers [15] .The encryption and decryption keys are formed by DNA probes and the cipher text is embedded in a DNA chip. The encryption of the message involves the fabrication of new DNA chip and the decryption is done by

using the hybridization of DNA [17].This paper analysis the comparison between different cryptographic schemes using DNA technology has clearly explained in the above Table (1).

5. LIMITATIONS

Although the Adleman [1] proved the extraordinary parallelism in the Hamiltonian path of solving seven cities which required 7 days to work in the lab for a graph of 7 vertices, the much work of the HPP is of affinity separation called extraction which had to repeat for each vertex. If the Hamiltonian path problem is improved to 50 or more cities, then it would require an excess of DNA molecules in tons.DNA has its parallel processing capabilities which allow the DNA based computer to solve hard computational problems in a reasonable amount of time. Lipton [2] solved the error which occurred in the Adleman experiment by reducing the usage of DNA molecules .Lipton's SAT method needs a least amount of DNA molecules with the use of more than 70 variables. If it is increased to more than 100 variables, the requirement of DNA molecules will be increased to millions of kilograms. In biological systems, this facility for error correction means that the error rate can be quite low. For example, in DNA replication, there is one error for every 10^9 copied bases or in other words an error rate of 10^{-9} with the assumption of 1 base per square nanometer. Comparing to the 7 GB hard drive to a one base per square nanometer, it is 100,000 times smaller.Clelland (1999) have demonstrated an steganography approach by hiding secret messages encoded as DNA strands among a multitude of random DNA. He used substitution cipher for encoding a plaintext where a unique base triplet is assigned to each letter t, each numeral and some special characters. One time cryptography with DNA strands is the practical implementation of DNA cryptography .It is quiet hectic because of the computational constraints. The amount of information collected on the molecular biology of DNA over the last 40 years is almost over powering to extent. The data density of DNA is like stirring a string of binary data with 1s and 0s by exploiting the massive parallelism and huge storing capabilities. The DES cryptographic Protocol can be broken. Gehani [7] proposed one time pad which is much secure so that we cannot identify the random secret key. To identify the secret key, biochemical process such as transcription, PCR and translation are not required. G.C [6] Even Though the encryption scheme provides a double layer security it is entirely different from the original scheme. Even though the above DNA cryptography proves a promising future for securing data. It has to be checked with more test cases and implementation. The data which is encrypted may be secure .A single imperfection on the design of encryption scheme can allow successful attacks.

6. CONCLUSION

DNA cryptography is encrypting or hiding a data in terms of DNA sequences. This can be done using several DNA technologies with the biochemical methods. Traditionally the DNA cryptography is implemented using biological tools. This DNA cryptographic method can also be interpreted with other schemes in order to apply this technology in various fields. The performance of DNA cryptography can be tested in order to prove the efficiency of an algorithm.

7. REFERENCES

[1] Leonard M. Adleman "Molecular Computation of solution to combinatorial problems" Science, New Series, Vol. 266, No. 5187. pp. 1021-1024 Nov. 11, 1994

[2] R. J. Lipton," Using DNA to Solve NP-Complete problems," Science, vol. 268, pp. 542 545, 1995

[3] J. D. Watson, F. H. C. Crick, "A structure for de oxy ribose nucleic acid", Nature, vol. 25, pp. 737-738, 1953

[4] Taylor Clelland, "Hiding messages in DNA Microdots". Nature Magazine vol.399,June 1999

[5] D. Boneh, "Breaking DES using Molecular computer", American Mathematical Society, pp 37-65, 1995

[6] Guangzhou Cui "An Encryption scheme using DNA Technology",IEEE pg 37-42 ,2008

[7] A. Gehani, T. LaBean, and J. Reif, "DNA-Based Cryptography", Lecture Notes in Computer Science, Springer. 2004.

[8] Zhihua Chen. "Efficient DNA Sticker Algorithm for DES" pg 15-22.IEEE 2008.

[9] Ning Kang, A pseudo DNA cryptography Method, http://arxiv.org/abs/0903.2693 ,2009

[10] L.M Adleman "On Applying Molecular Computation to the Data Encryption Standard." Journal of Computational Biology, 6 (1). pp. 53-63. 1999

[11] G. Z. Cui, L. M. Qin, Y. F Wang and X. C. Zhang, "Information Security Technology Based on DNA Computing,"IEEE International Workshop on Anti-counterfeiting Security, , pp. 288–291, 2007

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

[13] Souhila Sadeg " An Encryption algorithm inspired from DNA" IEEE pp 344 - 349 November 2010

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

[15] Monica BORDA "DNA secret writing Techniques" IEEE conferences 2010

[16] LU MingXin, "Symmetric Key Cryptosystem With Dna Technology" Science China pp 324-333,June 2007

[17] J Chen "A DNA-based, Bimolecular Cryptography Design"ISCAS'03.Proceedings2003

[18] LAI XueJia, LU MingXin "Asymmetric encryption and signature method with DNA technology" Vol. 53 No. 3: 506–514 March 2010