International Association of Scientific Innovation and Research (IASIR)
*(An Association Unifying the Sciences, Engineering, and Applied Research)*

ISSN (Print): 2279-0047
ISSN (Online): 2279-0055

**International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)**

**www.iasir.net**

# Secure Data Transmission Techniques Based on DNA Cryptography

Sanjeev Dhawan[1], Alisha Saini[2]
[1]Faculty of Computer Science & Engineering, [2]Research Scholar
Department of Computer Science & Engineering, University Institute of Engineering and Technology,
Kurukshetra University, Kurukshetra-136 119, Haryana, INDIA.
E-mail (s): rsdhawan@rediffmail.com, alisha191@rediffmail.com

***Abstract***: *DNA cryptography is a new promising direction in cryptography research that emerged with the evolution in DNA computing field. DNA can be used not only to store and transmit the information, but also to perform computation. The extensive parallelism and extraordinary information density inbuilt in this molecule are exploited for cryptographic purposes. The difficulties that are identified in DNA cryptography are the absence of theoretical basis and practical methodologies which can readily be implemented in the field of security. As DNA cryptography is in the development phase and requires a lot of work and research to reach a mature stage. DNA as a means of cryptography has high technical laboratory requirements and computational limitations. The potentiality of DNA on computing can open up further computation methods and challenges.*

***Keywords***: *Cryptography, DNA, DNA computing, DNA Encryption, DNA cryptography*

## I. Introduction

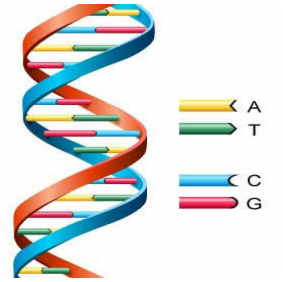
In a pioneering study, Adleman [1] sets the steps for the biological computing research, which used DNA to solve the complex mathematical problems. It marked the commencement of a new stage in the era of information. In the various researches, scientists find that the huge parallelism, extraordinary energy efficiency and extraordinary information density are inbuilt in DNA molecules [2], [3]. DNA computing provides a parallel processing ability with molecular level and involves the study of biology, chemistry, mathematics and computer science. It can simultaneously work on different parts of the computing problem that put forward challenges and opportunities to traditional information security technology [4], [5]. As an example, Boneh *et al.* [4] put forward an idea to break the Data Encryption Standard (DES) by using DNA computing methods. DNA cryptography [6], [7] is a new born cryptographic field that emerged with the research of DNA computing, in which DNA serves as information carrier and the modern biological technology is used as implementation tool. The extreme parallel processing and extraordinary information density inbuilt in DNA molecules are explored for cryptographic purposes such as encryption, authentication, signature and so on [8]. The study on DNA Cryptography is used to solve heavy combinatorial problems using its property of parallelism on huge storage. There are many problems that can be solved using DNA cryptography, the research of DNA Cryptography is still at the initial stage. The new born DNA cryptography [9], [10] is far from mature both in theory and realization, and this might be the reason that why only few examples of DNA cryptography were proposed. There is not any general theory about applying DNA molecules in cryptography [11], [12] but still current DNA technology is still in a period of laboratory exploration and focuses on experiments. Few of the key technologies in DNA research which have only been developed and well accepted in recent years [13] are Polymerase Chain Reaction (PCR), DNA synthesis, and DNA digital coding. PCR is a fast DNA amplification technology based on Watson-Crick complementary. But it is tremendously difficult to amplify the message-encoded sequence without knowing the proper two primer pairs i.e. the forward and reverse primer. The implementation in DNA cryptography can be performed by using modern biological techniques as tools and biological hard problems as main security basis to fully utilize the special advantages. As mentioned above, the two primer pairs (reverse and forward) could work as a key by applying the special function of primers to PCR amplification. On the other hand, if we consider traditional cryptography, its security is based on difficult mathematic problems which are mature both in theory and realization. There are many efficient cryptosystems of traditional cryptography such as DES, AES, RSA were invented. Thereby, DNA cryptography does not absolutely resist traditional cryptography and it is possible to construct hybrid cryptography of them.

## II. DNA

Deoxyribo nucleic acid (DNA) is a nucleic acid that contains the genetic instructions used for the growth and functioning of all living organisms. It is a collection of the most complex organic molecules. The substance is found in every cell of the organism and is essential for the identity of any living being. The main responsibility of DNA molecules is the long-term storage of information. DNA is often compared to a set of blueprints, like a

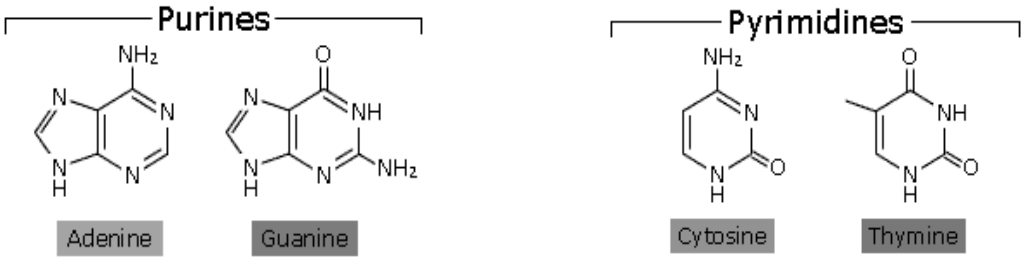
recipe or a code. Since it contains the instructions required to construct other components of cells such as proteins and RNA molecules. The DNA segments that hold this genetic information are called genes, but other DNA sequences have structural purposes or are involved in modifying the use of this genetic information. Just like a string of binary data is encoded with ones and zeros, a strand of DNA is encoded with four bases, represented by letters A (Adenine), T (Thymine), C (Cytosine) and G (Guanine). The information in DNA is stored as a code made up of these four chemical bases as shown in figure 1 below. The bases (nucleotides) are spaced every 0.34 nanometers along the DNA molecule, giving it a remarkable data density of nearly 18Mbits per inch. These nucleotides will only come together in such a way that A always pairs with T and C always pairs with G.

**Figure 1 Structure of DNA Molecule [24].**




The combination of the bases results in purines (combination of Adenine and Guanine) and pyrimidines (combination of Cytosine and Thymine) as shown in figure 2 below. The two strands of a DNA molecule are antiparallel where each strand runs in an opposite direction [23].This complementarily makes DNA a unique data structure for computation and can be exploited in many ways.

**Figure 2 Combinations of Bases Forming Purines & Pyrimidines [24].**




DNA is the basic storage medium for all living cells. The main function of DNA is to absorb and transmit the data of life for billions years. Roughly, around 10 trillions of DNA molecules could fit into a space, the size of a marbles. Since all these molecules can process data concurrently. Theoretically, we can calculate 10 trillions times simultaneously in a small space at one time. This section discusses the basic of DNA , similarity of DNA between the humans and other species and how the traits are inherited from one generation to the next generation.

- *Throughout the body - in cells*

  Our body is made up of about 50 and 100 trillion cells (a trillion is a thousand billion, or a thousand, thousand million). These cells are structured into tissues, such as skin, muscle, and bone. Each cell contains all of the organism's genetic instructions stored as DNA. Though, each cell uses only the instructions from a part of DNA. For example, a muscle cell utilizes the DNA that specifies the muscle apparatus, whereas a nerve cell utilizes DNA that specifies the nervous system. This all is like, if each cell reads only that part of a book of instructions that it requires.

- *Within the cell - in chromosomes*

  Each very long DNA molecule is firmly wound and packaged as a chromosome. Humans have two sets of 23 chromosomes in every cell one set is inherited from each parent. Thus a human cell as a result contains 46 of these chromosomal DNA molecules.

- *Within each chromosome - in genes*

  Each DNA molecule that forms a chromosome can be viewed as a set of small DNA sequences. These are divided into DNA function called genes, each of which guides the creation of one particular component of an organism. A set of human chromosomes consists of one copy of each of the

approximately 30,000 genes in the human "genome" - the term denotes the complete genetic instructions for an organism [25].

### *A. Biological Operations on DNA*

The biological model of DNA cryptography is the main activity carried out by the cell in our body which is integrated in the field of DNA computing. In this section, we will briefly discuss basic operations that are adopted from bio-chemistry operations as a computing tool in this field.

- *Synthesis*: Synthesis is a process of designing and reorganizing the information in DNA sequence form. In DNA computing, designing and synthesizing information in the DNA sequence form is an important process where a slight off beam design might leads to wrong result.
- *Ligation:* DNA ligation is a process of combining two single linear DNA fragments together. More exclusively, DNA ligation involves creating a phosphodiester bond between 3 -hydroxyl of one nucleotide and the 5 -phosphate of another.
- *Hybridizatio*n: It is a process of combining the complementary single-stranded nucleic acids into a single molecule. Nucleotides will bind to their complement under the normal conditions, so two entirely complementary strands will attach to one other at the end of the process.
- *Polymerase Chain Reaction (PCR)*: PCR is a process that rapidly amplifies the amount of specific molecules of DNA in a given solution using primer extension by a polymerase. DNA polymerases execute several functions including the repair and duplication of DNA. Each cycle of the reaction doubles the quantity of the molecules, giving an exponential growth in the number of operations.
- *Gel Electrophoresis*: Gel electrophoresis is a technique to sort DNA strands based on their length or weight through a gel such as agarose gel, in electrical field based on the fact that DNA is negative charge. Larger or longer strands will travel slowly into positive charge and after some time, the strands spread to different bands according their size of length [14].

## III. Background Work

Boneh *et al.* [4] started his pioneer work on breaking Data Encryption Standard which leads to more difficulties and much cost for experiment. Kang [16] explained the pseudo encryption methodology based upon Gehani *et al.* [12] work. He explained the encryption i.e. the conversion of plain text into DNA sequences and how these sequences are converted to the spliced form of data and protein form of data by cutting the introns according to the specified pattern. It is then translated to mRNA form of data and mRNA is converted into protein form of data. Lastly the protein form of data is sent through the secure channel. Cui *et al.*[9] proposed the encryption scheme by using the PCR, the DNA digital coding. The original message is converted to hexadecimal code and then to binary. The binary digits are then transformed into DNA sequence and it is considered as a DNA template. The forward primer is chosen to perform PCR, which changes the DNA sequence. The original message now is entirely different from the obtained sequence. The reverse primer is used to convert the PCR DNA to the original message and it is transformed into binary and which is further converted into plaintext message. It has both biological difficulty as well as mathematical difficulty. By the use of symmetric key block cipher and biochemical methods such as transcription and translation. Guangzhou Cui [15] proposed an encryption scheme in which a plaintext of message is converted into 4*4 matrixes and initial permutation is performed. With the help of generated key the XOR operation is performed. A matrix can be transposed and a secret key is generated which is given to the DNA module and permutation is performed to produce the resultant cipher text. Borda [18] proposed a secret writing method using the concept of one time pad in DNA. Using XOR and chromosome indexing the message is converted into binary in which each bit is encoded with nucleotides and primers are added. With the help of short oligonucleotides sequences, a long DNA sequence can be generated. The delimitation of DNA segment can be done using shortening the length.

By using molecular biology concept, the new cryptographic method [16] has been used to simulate the DNA biological operations .The sender translates mRNA form of data into protein according to genetic code table. The sender knows the starting codes and pattern codes. The cut out DNA introns are translated into mRNA form of data. The key is sent to the receiver in a secure channel to recover DNA form of information. By applying symmetric key cryptosystem with DNA technology [19], 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 then embedded in DNA chip with the most difficult DNA microarray technology to attain information security in a cryptosystem. Kang Ning [16] proposed that the sender encodes the message in the original DNA sequence using DNA coding which allows this to be DNA transcription and DNA Translation. The resulting protein will be like a public key which can be sent to the receiver over a public channel. Meanwhile, the sender sends to the receiver, a secret key which consists of the message. It needs to reassemble the DNA at the receiving end such as the location of the non coding regions that need to be reinserted. This type of cryptography is useful against powerful attacks. The procedure of encryption and decryption requires an informational message to be

transmitted. In an approved encryption and decryption algorithms, the plaintext message data is encoded in DNA strands using the publicly known alphabet of short oligonucleotides sequences. Using these short oligonucleotides sequences, a long DNA sequence can be generated. By using DNA hybridization, the delimitation of DNA segment can be done using short length. The message is converted to binary form in which each bit is encoded with nucleotides and is encapsulated with primers [18].

## IV. DNA Cryptography

DNA cryptography is a new promising direction in cryptography research that emerged with the evolution of DNA computing field. DNA can be used not only to store and transmit the information, but also to perform computation. The extensive parallelism and extraordinary information density inbuilt in this molecule are exploited for cryptographic purposes. Several DNA based algorithms are proposed for encryption, authentication and so on. In this paper, the research conducted by a number of authors related to the discipline of DNA Cryptography is taken into consideration and it has tried to find out the basics of DNA Cryptography that how DNA cryptography field emerged and how DNA computational logic can be used in cryptography for encrypting, storing and transmitting the information i.e. the art of cryptography security to make anyone message unreadable by encoding it. It has been shown that how DNA cryptography uses DNA as the computational tool with a number of molecular techniques to manipulate it along with various algorithms for encryption.

### *A. Traditional Cryptography and DNA Cryptography*

- *Development*
  Traditional cryptography can be traced back to Caesar cipher 2000 years ago or even earlier. Related theory is almost sound. All the practical ciphers could be seen as traditional ones. DNA cryptography is having history from less than last two decades, the theory basis is under research and the application costs very much.
- *Security*
  Only computational security can be achieved for traditional cryptographic schemes except for the one-time pad, that is to say, an adversary with infinite power of computation can break them theoretically. For the DNA cryptography, the main security basis is the restriction of biological techniques, which has nothing to do with the computing power and immunizes DNA cryptographic schemes against attacks using quantum computers. Nevertheless, the problem as to what is the extent this kind of security and how long it can be maintained it is still under exploration.
- *Application*
  Traditional cryptosystems are the most convenient of which the computation can be executed by electronic, as well as DNA computers, the data can be transmitted by wire, fiber, wireless channel and even by a messenger, and the storage can be CDs, magnetic medium, DNA and other storage medium. Using the traditional cryptography we can realize purposes as public and private key encryption, identity authentication and digital signature. Under the current level of techniques, only by physical ways can the cipher text of DNA cryptography be transmitted. Due to the vast parallelism, exceptional energy efficiency and extraordinary information density inherent in DNA molecules, DNA cryptography can have special advantages in some cryptographic purposes, such as secure data storage, authentication, digital signature, steganography, and so on. DNA can even be used to produce unforgettable contract, cash ticket and identification card.

Researches of the two kinds of cryptography are still in progress, and many problems remains to be solved especially for DNA cryptography, this making it hard to predict the future. But from the above discussions we think it is likely that they exist and develop conjunctively and complement each other rather than one of them falls into disuse thoroughly. Furthermore, the comparative analysis between different cryptographic schemes using DNA technology has clearly explained in the table 1.

### *B. Advantages of DNA Cryptography*

After going through DNA Cryptography various advantages of using DNA along with cryptography came to be known which are described as follows:

- The biggest advantage of cryptography is its secure nature. Although, it never needs to be transmitted or exposed to anyone.
- Moreover, encrypting it along the DNA sequence makes it more secure. One gram of DNA contains 10^21 DNA bases = 10^8 tera-bytes of data. A few grams of DNA can hold all the data stored in world.
- Since DNA is used for encryption, Signature authorization is not needed. DNA replaces the cause of Digital signatures and digital timestamps.
- Can work in a massively parallel fashion: DNA is modified biochemically by a variety of enzymes, which are minute protein machines that read and process DNA according to nature's design. There is a

wide variety and number of these "operational" proteins, which manipulate DNA on the molecular level. For example, there are enzymes that are used to cut DNA and enzymes used to paste it back together. Just like a CPU has a basic suite of operations like addition, bit-shifting, logical operators (AND, OR, NOT NOR), etc. that allow it to execute even the most complex calculations, DNA has cutting, copying, pasting, repairing, and many other capabilities. In the test tube, enzymes do not function consecutively, working on one DNA at a time. Many copies of the enzyme can work on many DNA molecules at the same time.

- Large storage: A gram of DNA contains about 10^21 DNA bases, or about 10^8 tera-bytes of data. Hence, a few grams of DNA have the capability of storing all the data stored in the world.
- The main goal of the research of DNA cryptography is exploring characteristics of DNA molecule and reaction, establishing corresponding theories, discovering possible development directions, searching for simple methods of realizing DNA cryptography, and laying the basis for future development.
- Input and output of the DNA data can be moved to conventional binary storage media by DNA chip arrays [22].

### *C. Limitations of DNA Cryptography*

Apart from advantages, DNA cryptography comprises of few disadvantages as listed below:

- Lack of the related theoretical basis.
- Difficult to realize and very expensive to apply.

**Table 1. Comparison between various cryptographic schemes using DNA technology.**

| Cryptographic schemes | 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 [15]. |
| 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 [16]. |
| An Encryption Algorithm Inspired From DNA. | Symmetric key block cipher Algorithm Transcription(DNA-RNA) Translation(RNAProtein) | 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 [17]. |
| A DNA-based, Bimolecular Cryptography Design | Carbon Nano Tube Technology | A nano scale used to alter the message [20]. |
| 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 [19]. |
| 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 [21]. |

## V. Conclusion

DNA cryptography is basically hiding of data in terms of DNA sequences. This is done by using various DNA technologies with the biological tools. Here in this paper with the summarization of DNA, basics of where DNA is found are discussed. Various biological operations that can be carried on DNA are explained. Further DNA cryptography and the biological work on DNA cryptography is taken into consideration. It is shown that how traditional cryptography differs from the emerging DNA cryptography. Few of the advantages along with the limitations of DNA Cryptography are mentioned. Later on, the existing DNA cryptographic techniques are discussed and a comparison between different cryptographic schemes using DNA technology is explained. The future work will consist of analyzing and comparing the performance of all the DNA cryptographic techniques based on secure data transmission processes.

## VI. References

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

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

[3] P. L. Cox J, "Long-term data storage in DNA," *Trends Biotechnol*ogy, vol. 19, pp. 247–250, 2001.

[4] D. Boneh, C. Dunworth and R. Lipton, "Breaking DES using a molecular computer," *American Mathematical Society*, pp. 37–65, 1995.

[5] L. M. Adleman, "On applying molecular computation to the date encryption strands in DNA based computers," in *Proc. of the 2ed Annu*., 1996, pp. 28–48.

[6] M. Amosa, G. Paun and G. Rozenbergd. "Topics in the theory of DNA computing," *Theoretical cience*, vol. 287, pp. 3–38, 2002.

[7] G. Z. Xiao, "New field of cryptography: DNA cryptography," *Chinese Science Bulletin*, vol. 51, pp. 1139–1144, 2006.

[8] S.V. Kartalopoulos, "DNA-inspired cryptographic method in optical communications," in *authentication and data mimicking Military Communications Conference*, 2005, pp. 774–779.

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

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

[11] M. X. Lu, "Symmetric-key cryptosystem with DNA technology," *Science in China Series F: Information Sciences*, vol. 3, pp. 324–333, 2007.

[12] A. Gehani, T. H. LaBean and J. H. Reif, "DNA-based cryptography," *DNA Based Computers V. Providence: American Mathematical society*, vol. 54, pp. 233–249, 2000.

[13] T. Kazuo, O. Akimitsu and S. Isao, "Public-key system using DNA as a one-way function for key distribution," *Biosystems*, vol. 81, pp. 25– 29, 2005.

[14] Rohani binti abu Bakar and Junzo Watada, *"DNA Computing and its applications",* Survey, Volume 2, Number 1, ICIC International 2008 ISSN 1881-803X -March 2008.

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

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

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

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

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

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

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

[22] www.worldofjoy.weebly.com/1/category/dna%20cryptography/1.html.

[23] Donald Nixon, "DNA and DNA Computing in Security Practices – Is the Future in Our Genes"**,** GSEC Assignment Version 1.3, SANS Institute 2000 – 2002.

[24] William Stallings, *"Cryptography and Network Security"*, Third Edition, Prentice Hall International -2003.

[25] www.koshland-science-museum.org/exhibitdna/intro02.jsp,Copyright2011 National Academy of Sciences.