



## International Journal of Engineering, Business and Enterprise Applications (IJEBA)

[www.iasir.net](http://www.iasir.net)

### Integration of DNA Cryptography for Complex Biological Interactions

Sanjeev Dhawan<sup>1</sup>, Alisha Saini<sup>2</sup>

<sup>1</sup>Faculty of Computer Science & Engineering, <sup>2</sup>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:** The DNA cryptography is a new and promising direction in cryptography research. DNA can be used in cryptography for the purpose of storing and transmitting the information, as well as for computation. Due to the very high storage capacity of DNA, this field is becoming extremely promising. DNA is the gene information which conceals the information of all living beings. Though the DNA cryptography has its own application in the area of vast information storage, immense parallel processing, low energy consumption which has been proposed and proved by the researchers. Although in its primal stage, DNA cryptography is exposed to be very effective. Currently, it is in the development phase and requires a lot of work and research to reach a mature stage. However, the use of the DNA in cryptography has high technical laboratory requirements and computational limitations, as well as the labor-intensive extrapolation means so far. These make the proficient use of DNA cryptography complicated in the security world now. Therefore, more theoretical study should be done before its real applications.

**Keywords:** Cryptography, Decryption, Encryption, DNA, DNA Computing, DNA Cryptography, Number conversion

#### I. Introduction

The word cryptography comes out from ancient Greek. It is a mishmash of two words: (a) *krypto* means "hidden" and (b) *grafo* means to "write". So the literal meaning of cryptography is "hidden writing". It is the ancient science of encoding messages so that only the sender and receiver can understand them. Cryptography is the science which uses mathematics to encrypt and decrypt the data. Cryptography enables us to store sensitive information and transmit it across insecure networks (like the Internet) so that it is kept unreadable by anyone except the intended recipient [1]. Cryptography is an art of writing in secret code and is an ancient skill. The first documented use of cryptography, in writing dates back to approximately 1900 B.C. Some experts argued that cryptography appeared unexpectedly sometime after writing was invented, with applications varying from diplomatic missives to war-time battle plans. There is no surprise that new forms of cryptography came soon after the general development of computer communications. In data and telecommunications, cryptography is essential while communicating over any unreliable medium, which comprise just about any network, particularly the internet [2]. Within the framework of any application-to-application communication, there are some precise security requirements, which include:

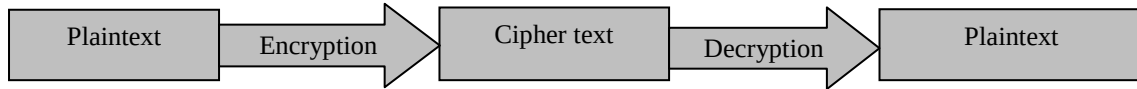
- **Authentication:** The process of proving one's identity. (The primary forms of host-to host authentication on the internet today are name-based or address-based, both of which are notoriously weak.)
- **Privacy/confidentiality:** Ensuring that no one can read the message except the intended receiver.
- **Integrity:** Assuring the receiver that the received message is original one. It has not been altered in any way from the original.
- **Non-repudiation:** A mechanism to prove that the sender really sent this message [2].

#### A. Cryptographic Scenario

Data that can be read and understood without any special measures is known as plaintext or cleartext. The method of disguising plaintext in such a way as to hide its substance is known as encryption. Encrypting plaintext results in unreadable data called ciphertext. We use encryption to ensure that the information is hidden from anyone for whom it is not intended, and even those who can see the encrypted data. The process of reverting ciphertext to its original plaintext is known as decryption. General diagram of cryptography is shown below in fig 1. The typical scenario in cryptography is that Alice (sender) wants to send some message secretly to Bob (receiver). The message which is to be sent is in the ordinary language understood by all i.e. the plaintext. The process of converting plaintext into a form which cannot be understood without having special information is called encryption. The unreadable form is known as cipher text and the special knowledge for encryption is known as

the encryption key. The conversion of cipher text again into plaintext with a special knowledge is known as decryption, whereas special knowledge for decryption is known as the decryption key. Only the receiver has this special knowledge and only receiver can decrypt a cipher text with this knowledge called the decryption key [1].

**Figure 1 General block diagram of cryptography.**



### B. Types of Encryption Schemes

The two main types of Encryption schemes are: (i) Secret key cryptography or symmetric encryption, and (ii) Public key cryptography or asymmetric encryption.

- **Symmetric Encryption:** It is also referred as conventional encryption. Symmetric encryption is a form of cryptosystem in which encryption and decryption are performed using the same key. Symmetric cryptography is susceptible to plain text attacks and linear cryptanalysis which mean that they are hackable and is simple to decode.
- **Asymmetric Encryption:** Asymmetric algorithms use two keys, one to encrypt the data and the other one to decrypt. These inter-dependent keys are generated together. One is labeled as the Public key and is distributed freely. The other one is labeled the Private Key and must be kept hidden. Often referred to as Public/Private Key Encryption. These systems can provide a number of different functions which depends on how they are used. The most common usage of asymmetric encryption is to send messages with a guarantee of confidentiality.

### C. Types of Cryptography

There are mainly three sub fields or prominent branches of cryptography named as: (i) Modern Cryptography, (ii) Quantum Cryptography, and (iii) DNA Cryptography. These three fields depend upon different difficult problems concerning to different disciplines for which there is no known solution until now.

- The modern cryptography is based upon the difficult mathematical problems such as prime factorization, elliptic curve problem, for which there is no known solution found so far.
- Quantum cryptography which is also relatively a new field is based upon the Heisenberg uncertainty principle of Physics.
- DNA cryptography depends upon the difficult biological processes concerning to the field of DNA technology [10] such as Polymerase Chain Reaction (PCR) for a sequence without knowing the correct two primer pairs and another is extracting information from the DNA chip without having the knowledge about the sequences present in different spots of DNA chip [1].

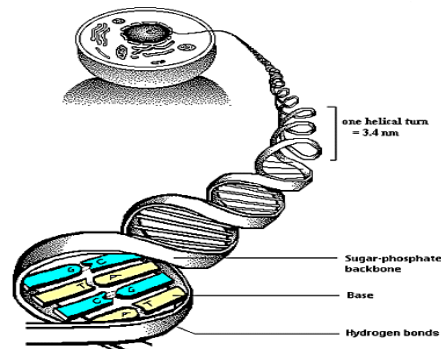
In this paper, it has tried to find out the basics of DNA Cryptography that what is DNA and DNA cryptography. Later on the comparison between the traditional and DNA Cryptography on the basis of various attributes is given. And lastly how DNA computational logic can be used in cryptography for encrypting, storing and transmitting the information i.e. the art of cryptography security to make anyone message unreadable by encoding it with the help of DNA is considered.

## II. DNA

Within the cells of any organism is a substance called Deoxyribonucleic Acid (DNA), which is a double-stranded helix of nucleotides, carries the genetic information of a cell. The data density of DNA is impressive. 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) [3]. A DNA molecule is found in every cellular organism as a storage medium for genetic information. It is a polymer constructed from monomers called nucleotides distinguished by their chemical group of bases that attached to them. DNA consists of three main components: sugar, phosphate and base. There are four different bases: adenine, guanine, cytosine and thymine, abbreviated as A, G, C and T respectively. Since nucleotides differ only in terms of their bases, abbreviations of bases are used to identify them. In mathematical modeling it can be illustrated as  $X = \{A, C, G, T\}$ . Figure 2 illustrates the basic DNA structure. It consists of four distinct bases: Adenine (A), Thymine (T), Guanine (G) and Cytosine (C). A will always bind with T and G always with C that represent Watson-Crick complementary law. Every single nucleotide is linked together end-to-end to form DNA strands. A short single-stranded polynucleotide chain, generally less than 30 nucleotides long, is known an oligonucleotide (or shortly oligo). The DNA sequence has a polarity i.e. a sequence of DNA is different from its reverse. The two dissimilar ends of the DNA sequence are known under the name of the 5 - end and the 3 - end, respectively. The most important feature of DNA is the ability to bind together between two single strands called as Watson-Crick complementarity.

Bonding between single strands occurs when two different bases attract with each other. In this complementarity, A will bond only with T and G only with C. Pair of (A, T) and (G, C) is therefore known as a complementary base pair. The two pairs of bases form hydrogen bonds between each other with two bonds between A and T and three between G and C.

**Figure 2 Basic structure of DNA molecule [4].**



A different number of hydrogen bonds between these two pairs give a different strength for them. In other words, a pair of G and C is stronger than a pair of A and T. It is one of the characteristics that are always considered in designing DNA strands for DNA computing. The classical double helix of DNA is formed as shown in figure 2 when two separate strands bond. Two requirements must be fulfilled for this to occur firstly, the strands must be complementary, and secondly they must have opposite polarities.

### III. DNA Cryptography

DNA Cryptography is a new field which has been emerged with the research of DNA Computing, in which the DNA is used as an information carrier and modern biological technology is used as an implementation tool. The vast parallelism along with extraordinary information and density inherent in DNA molecules are explored for the cryptographic purposes such as encryption, authentication, signature and so on. DNA can be used in cryptography for storing the information and transmitting that information. It can also be used for computation. Although in its early stage, DNA cryptography was shown to be very effective. Currently, several DNA computing algorithms has been proposed for some cryptography, cryptanalysis and steganography problems. These are very powerful in their respective areas. However, the employ of the DNA as a means of cryptography has very high technical laboratory requirements and computational limitations, as well as the labor intensive extrapolation means so far. These make the efficient use of DNA cryptography difficult in the world of security now. Therefore, more theoretical examination should be performed before its real applications. As some of the modern cryptography algorithms (such as DES, and more recently MD5) are broken, the new directions of the information security are being sought to protect the data. The idea of using DNA computing in the area of cryptography and steganography is a promising technology that may convey forward a new hope for leading or even unbreakable algorithms.

Adleman with his pioneering work [Adleman, 1994] [5], set the stage for the new field of bio-computing research. His main idea was to use actual chemistry of DNA to solve problems that are either impossible by conventional computers, or require an enormous amount of computation. By using DNA computing, the Data Encryption Standard (DES) cryptographic protocol can be broken [Boneh, *et. al.*, 1995] [6]. The one-time pad cryptography with DNA strands and the research on DNA steganography (hiding messages in DNA) are shown in [Gehani, *et. al.*, 2004] [7]. However, researchers in DNA cryptography are still looking at much more theory than practicality. The constraints of its high technical laboratory requirements and computational limitations, combined with the labor intensive extrapolation means. Thus prevent DNA computing from being of efficient use in today's security world. DNA cryptography has been ragged about much in the media as of late but whether or not this technology is appropriate for the future is still debatable. There has been a distinct lack of hard evidences that put forward to illustrate whether the technology is even feasible, much less appropriate in the foreseeable future. Ashish *et al.* [7] published a paper entitled 'DNA-based Cryptography' which puts an argument forward that the high level computational ability and incredibly compact information storage media of DNA computing has the possibility of DNA based cryptography based on one time pads. They argue that current practical applications of cryptographic systems based on one-time pads is limited to the confines of conventional electronic media whereas a small amount of DNA can suffice for a huge one time pad for use in public key infrastructure (PKI).

In table 1 below, a simple comparison is discussed between traditional and DNA cryptography. This table considers the attributes including storage medium which is used to store up the data in the technique, the storage capability of the storage medium used, the security level provided by both the techniques, time complexity i.e. the

time taken to process the technique and the constancy of the results for the particular technique. Traditional cryptography usually runs on computers over the network, so the storage mediums are the silicon chips of the computers. While DNA cryptography deals with the DNA strands which are manipulated by biological techniques. By considering DNA as the storage medium, it has got a huge storage capacity when compared to the equivalent amount of the silicon chips. This property of data makes DNA cryptography very attractive and beneficial field of research. When it comes to security of both the techniques, it can be seen that traditional Cryptography provides one fold security as it relies on only computational difficulties. DNA cryptography provides a two fold security by involving computational difficulties as well as the biological difficulties. The Time complexity of the efficient cryptographic algorithms is few seconds, whereas DNA cryptography that involve technologies like PCR and DNA chip can take hours to complete the entire process.

**Table I Comparison Between Traditional and DNA Cryptography**

| Attributes      | Traditional Cryptography              | DNA Cryptography                                                     |
|-----------------|---------------------------------------|----------------------------------------------------------------------|
| Storage Medium  | Computer/Silicon Chips                | DNA Strands                                                          |
| Storage Ability | 1 gm silicon chip contains 16 MB      | 1gm DNA contains $10^{21}$ DNA bases that carries 10 <sup>8</sup> TB |
| Security Level  | One Fold                              | Two Fold                                                             |
| Time Complexity | $\geq$ few seconds                    | $\geq$ few hours                                                     |
| Constancy       | Depends on implementation environment | Depends on present environmental conditions                          |

If we consider the constancy of results of cryptographic techniques, it refers that encryption and decryption always gives the same results. If the constancy of the results of encryption and decryption provided is analyzed, it can be seen that traditional cryptographic algorithms depends on the implementation conditions [8]. Implementation conditions consist of the platform and the language limitations which are used to encode the algorithm. Whereas the constancy of DNA cryptography is very much dependant on the environmental conditions such as temperature, pH.

#### IV. DNA Encryption Decryption Technique

This approach presents the way in which DNA binary strands can be used for cryptography, which encrypts the data by hiding information, in order to provide rapid encryption and decryption. It is shown that DNA cryptography based on DNA binary strands is secure under the supposition that an interceptor has the same technological prospective as sender and receiver of encrypted messages.

##### A. Encryption

It is a Cryptographic technique in which each letter of the alphabet is changed into a different combination of the four bases that make up the human DNA. A portion of DNA spelling out the message to be encrypted is synthesized, and the strand is slipped into a normal fragment of human DNA of similar length. As only a single DNA strand in about 30 billion will enclose the message. The recognition of even the existence of the encrypted message is most unlikely.

As an example, it is shown here how to change the word "CRYPTO" into DNA-ready code. Firstly the ASCII table converts each of the individual letter of the document to be encrypt into numerical value (C=67, R=82, Y=89, P=80, T=84, O=79). The generated ASCII value is in Base 10 form. Further base-10 to base-4 conversion is applied which will convert the decimal form of output to the quaternary form. This will use the real numbers 0, 1, 2 and 3 for the representation of the generated ASCII values (67=1003, 82=1102, 89=1121, 80=1100, 84=1110, and 79=1033). Finally, numbers with 0, 1, 2, and 3 can be changed into their DNA base equivalents and are replaced with A, T, C, and G respectively. As an example CRYPTO will becomes TAAGTTACTTCTTTAATTTATAGG. DNA strands aren't long enough to store complicated information like a book or a photograph, so the best available solution is to fragment the data into little pieces and spread it among the different cells. To make it work, the researchers have to work on a system that allows the fragments to identified and eventually put back in the right order. So they created a three-part structure for the DNA: header, message, and checksum. The header is an 8-base-long string that is divided into four levels of classifying the information - zone, region, area and district, which helps each fragment to be put back in the right order. After the

message carries the real usable data, the checksum provides a replication of the original header, which is helpful in controlling for the minor mutations to the bacteria. Now the information has been encrypted and placed in lots of different cells of bacteria [9].

## **B Decryption**

To retrieve the data on the other end, the Decrypter would take the DNA and run it through what's known as next-generation high-throughput sequencing, or NGS. This particular type of sequencing, analyses and compares multiple copies of the same sequence and then uses majority-voting to check out the correct base, if parts of the data have decayed. After this the compression algorithms could be reversed to restore the raw data into its original form. The last step would be rearranging the fragments back together in the accurate order so that the DNA strands could be decoded back into useful data to get the exact result. This is where we go from just data storage to the data encryption. The person trying to read the data will need the formula that will tell the right order of the headers and checksums - without that formula, the data will remain meaningless [9].

## **C Summery**

### **C.1. Encryption**

*Step-1: The binary data, message or the text is used under the form of ASCII code (converted in numerical format).*

*Step-2: Number conversion is then applied on the numerical format converted from ASCII code (e.g. Base 10 to base 4 conversions).*

*Step-3: This encoded (converted) message is then changed to DNA base equivalent.*

*Step-4: Then these digits are grouped into and substituted as A for 0, T for 1, C for 2, and G for 3.*

*Step-5: We then fit the primers on either side of this message. Primers will act as stoppers and detectors for the message. This has to be given to the receiver prior to the communication.*

*Step-6: This message is followed by our own DNA sequence followed by another stopper/primer.*

*Step-7: This message is then flanked by many sequences of DNA or by confining it to a microdot in the micro-array.*

*Step-8: If considered as a pseudo method: this sequence is transferred to the receiver through the Internet. Else the micro-array is sent physically (though time consuming).*

### **C.2. Decryption**

*Step-1: The Decrypter would take the DNA and run it through what's known as next-generation high-throughput sequencing (NGS).*

*Step-2: It analyzes and compares multiple copies of the same sequence and then uses majority-voting to check out the correct base.*

*Step-3: Compression algorithms could be reversed to restore the raw data into its original form.*

*Step-4: Snap the fragments back together in correct order so that the DNA strands could be decoded back into useful data to get the correct result.*

Input and output of the DNA data can be moved to conventional binary storage media by DNA chip array. At this point the binary data may be encoded in DNA strands with the help of an alphabet of short oligonucleotides sequences.

## **V. Conclusions**

The DNA cryptography is a relatively new cryptographic field of research evolved with the DNA computing. In this field, DNA is used as a message carrier. The high storage capability and parallel computability of DNA molecules are exploited for encryption, authentication, and authorization. In this paper, a brief introduction to the existing cryptography, DNA and DNA Cryptography have been discussed. With the summarization of cryptography and DNA cryptographic research, their comparison on the basis of various attributes and the algorithm for encryption decryption has been discussed. The future of this area looks very promising by integrating the basic nature and design aspects of DNA for encryption and further will try to implement DNA based methods in order to break cryptosystem.

## **VI. References**

- [1] Beenish Anam, Kazi Sakib, Md. Alamgir Hossain, Keshav Dahal, "Review on the Advancements of DNA cryptography"-2010.
- [2] Gary C. Kessler, "An Overview of Cryptography", Journal of *Digital Forensic Practice*, May 2011.
- [3] (2012) The scribd website. [Online]. Available:<http://www.scribd.com/doc/46249955/Paper-on-DNA-Computing>
- [4] (2012) Available:<http://www.accessexcellence.org/RC/VL/GG/structure.php>.
- [5] L. M. Adleman, "Molecular computation of solution to combinatorial problems", Science, vol. 266, pp. 1021-1024, November 1994.

- [6] D. Boneh, C. Dunworth, and R. Lipton, "Breaking DES using a molecular computer," in Proc. *DIMACS workshop on DNA computing*, 1995, pp. 37–65.
- [7] A. Gehani, T. LaBean, and J. Reif, "DNA-Based Cryptography", Lecture Notes in Computer Science, *Springer*, 2004.
- [8] A. A. Jabri, "A statistical decoding algorithm for general linear block codes," in 8th IMA International *Conference on Cryptography and Coding*, Cirencester, UK, Dec. 2001, pp. 1–8.
- [9] Schneier, B., "Description of a New Variable- Length Key, 64-Bit Block Cipher (Blowfish)", *Springer-Verlag*, and, Fast Software Encryption, Cambridge Security Workshop Proceedings, 1993.
- [10] G. Rozenberg and A. Salomaa, "*DNA computing: New ideas and paradigms*", Lecture Notes in Computer Science, Springer-Verlag, vol.7, pp.188-200 -2006.