# Survey of different DNA Cryptography based algorithms

**Nikita Parab[1], Ashwin Nirantar[2]**

[1,2] *UG student, Department of Computer Engineering, PES Modern College of Engineering, Pune, Maharashtra*

---

**Abstract –***Cryptography is concerned with converting a plain text into cipher text which is storing and transmitting data in a particular form so that only those who are intended can read and process it. Due to the challenges in traditional cryptography, a new technique which makes use of the power of DNA is emerging to make data more secure. This technology deals with studying how to make DNA carry information. Modern biotechnology is used as a measure to transfer cipher text into plain text. Every technology often has its own sets of challenges and DNA cryptography is no exception. It requires high tech bio molecular laboratories and naturally, high computational complexity. This paper surveys the field of DNA cryptography, the algorithms which deal with DNA cryptography and the advantages and challenges associated with each of these algorithms. For anyone who is interested in this field, this paper can be a starting point into knowing what research has currently been done on DNA cryptography.*

***Key Words*: Cryptography, DNA Cryptography, Encryption, Decryption.**

## 1.1 Introduction

Cryptography and molecular biology were initially considered irrelevant, but now have begun to start working together more closely, because of the in-depth study of modern bio technology and DNA computing. DNA cryptography and information science was born after research in the field of DNA computing field by Adleman. Biological problems are the base for DNA Cryptography.

1) With very large scale of parallelism, the computing speed of DNA chains could reach 1 billion times per second.
2) DNA molecule has large capacity and can be used as a carrier of data. One trillion bits of binary data is possible to be stored in one cubic decimeter of a DNA solution.
3) Another key factor is low power consumption. The power consumption of a DNA molecular computer is only equal to one-billionth of a traditional computer.

In this paper, we do not present any new research results. The contribution comes from combining existing algorithms gathered from many sources and publications.

## 1.2 Definition

One way of defining DNA cryptography is hiding data in terms of DNA sequence. In theory, in addition to having the same computing power as that of a modern computer, a DNA computer will have a potency and function which these computers will not be able to match. As stated before, a DNA molecule has large capacity, DNA chains have large parallelism and a DNA molecular computer has low power consumption.

DNA cryptography does not completely repulse traditional cryptography and it is possible to construct hybrid cryptography.

## 2. Technology

Biotechnology, is closely associated with DNA cryptography and plays an important role in this field. Some of the DNA biotechnology and software of the field of DNA are:

1. Gel Electrophoresis
2. DNA chip technology
3. PCR technology
4. DNA code
5. DNA fragment stitching software is the DNA Baser Sequencer Assembler. It is used for splicing DNA fragments, we need to prepare some DNA fragments for splicing before using the software.

The following algorithms are proposed which make use of DNA cryptography in order to make communications more secure.

## 3.1 Bidirectional DNA Encryption Algorithm

Modern cryptography is based on a difficult mathematical problem, the NP-complete problem, quantum cryptography is based on Heisenberg's uncertainty principle, which is also a difficult biological problem.

Similarly, this scheme [1] also makes use of a difficult biological problem, which is stated as "It is extremely difficult to amplify the message encoded sequence without knowing the correct PCR two primer pairs". PCR, or Polymerase Chain Reaction is a fast DNA amplification technology in which two complementary oligonucleotide primers are annealed to double-standard target DNA strands, then necessary target DNA strands and the necessary target DNA can be amplified after a serial of polymerase enzyme. Being very sensitive, a single DNA target molecule can be amplified into 10^6 after 20 cycles in theory, which can be done within short time. It would be

extremely difficult to amplify the message encoded sequence without knowing the correct primer pairs.

This scheme makes use of DNA digital coding, which is an advancement over the traditional binary digital coding which makes use of 1's and 0's. In a DNA sequence, there are four bases, Adenine (A), Thymine (T), Cytosine (C) and Guanine (G). A simple coding pattern to encode nucleotide bases is by means of four digits, 00 (0), 01 (1), 10 (2), 11(3). A, T, G, C stand for 0, 1, 2, 3 respectively.

The advantages of using DNA digital coding are redundancy of information coding is reduced and efficiency is increased. Traditional methods such as DES or RSA could be used for preprocessing the plain text and this method is convenient for mathematical and logical operations.

In the proposed system [1], a text message is received from the user, which is converted into Hexadecimal and Binary code. The message is split into parts, one is used as a message and the other is used as a key. For the purpose of high compression factor, the XOR operation is used.

The DNA base coded message is obtained by applying DNA digital coding over the message, after which PCR amplification is implemented which makes use of two primer pair as the key. For variable length data, compression is performed. Various modes of operation happen in serial fashion, and double layer security is provided, hence this is called as Bi-serial DNA Encryption Algorithm.

The Diffie-Hellman (DH) key exchange protocol is used to establish a shared secret key over an insecure communications channel for two parties that have no prior knowledge of each other. This key can then be used for encryption of subsequent communications using a symmetric key cipher.

For decryption, the encrypted data is obtained from the receiver by using a high decompression algorithm compressed data is recovered, then the correct two primer pairs are used to retrieve the DNA digital coding. This is converted into binary code, then XOR is performed and key is given by the user. Combining the key and XOR-ed output, large Binary code is retrieved which is converted into Hexadecimal code. This is further converted into normal plain text by using a decimal converter.

If any of the key is wrong, there is a chance of missing data or improper form of data. So secure data can be maintained.


**Fig -1**: Encryption

### 3.2 DNA Cryptography using Quantum key exchange

Through a public channel, exchanging a key securely is difficult in secret key cryptography. By Quantum Cryptography [2], these shortcomings can be overcome. The main advantage of using QC is that it can be used for authentication, and distribution of random keys or information between parties can be permitted by establishing a quantum channel instead of an ordinary channel. DNA based algorithm can be used for message encryption. The following laws of physics are the base for Quantum Cryptography.

#### 1) No cloning theorem

"It is impossible to create a copy of an arbitrary unknown quantum state". Due to which replay attacks are prevented. The attacks where the attacker simply sends a text data which was previously sent by some other user for reproducing the effect. The key will be detected which will result in termination.

#### 2) Heisenberg principle.

"It is in general impossible to measure, for instance, both the location and speed of a quantum object with perfect accuracy". The information is sent in the form of photons representing 0s and 1s by the QC systems. If an attempt is

made to eavesdrop, it is able to detect tampering by unintentionally altering the photons being transmitted. The beam of photons doesn't encode the actual secret message, it contains only an encryption key .If any part of the key is intercepted, the communication parties detect the altered photons and can remove or delete that part of the key. Once they've transmitted enough photons, the shared key is used to encrypt the message .This key can be sent over public communication lines but the photon key has to arrive reliably at its destination.

However, quantum cryptography is not completely unbreakable if the systems are not built correctly. But if you build the systems correctly, no hacker will be able to hack the system.

**DNA coding Technology**

The DNA coding converts the input alphabet into DNA which is then converted into a triplet code. The Input message that has to be encrypted contains characters which on processing generates a triplet code. The generated code contains combination of three bases for each character.

| | | | |
|---|---|---|---|
| A=CGA | K=AAG | U=CTG | 0=ACT |
| B=CCA | L=TGC | V=CCT | 1=ACC |
| C=GTT | M=TCC | W=CCG | 2=TAG |
| D=TTG | N=TCT | X=CTA | 3=GCA |
| E=GGC | O=GGA | Y=AAA | 4=GAG |
| F=GGT | P=GTG | Z=CTT | 5=AGA |
| G=TTT | Q=AAC | =ATA | 6=TTA |
| H=CGC | R=TCA | ,=GAT | 7=ACA |
| I=ATG | S=ACG | .=GAT | 8=AGG |
| J-AGT | T=TTC | ;=GCT | 9=GCG |

**Fig -2**: Triplets for DNA coding

The proposed system contains a secure message transfer protocol consisting of the BB84 protocol, authentication, secure key exchange, a DNA based algorithm and AES encryption. All of these are explained in short:

1. **BB84 Protocol**: BB84 stands for Bennett – Brassard 1984. By using this protocol, A can send a private key to B. A begins with two string of bits, 'a' and 'b', each n bit long. Then A encodes these two strings as a string of n qubits.

$$|\psi\rangle = \bigotimes_{i=1}^{n} |\psi_{a_i b_i}\rangle.$$

$a_i$ and $b_i$ are the $i^{th}$ bits of $a$ and $b$, respectively. Together, $a_i b_i$ give us an index into the following four qubit states:

2. **Authentication**

There are different types of authentication such as, User Authentication (UA), Message Authentication (MA) and so on. The goal of MA is to provide the communication parties with a means to make sure that received messages originated from the other participant. In particular, MA allows the communication parties to send each other messages in such a way that any modification of them can be detected with very high probability.

3. **Secure key exchange**

BB84 protocol is used to implement the secure key exchange module. The generated random stream of bits consists of 0's and 1's, which represent a stream of photons. To get n random bits, the function is called n times. If the 'random' function returns a value less than or equal to 0.5, then that particular basis is considered '+' else it is considered 'x'.

4. **DNA based algorithm**

By using DNA based encryption process a DNA coded sequence is generated, and then this sequence is given to the AES algorithm and used as the key. This key will be in the form of triplet codes.

5. **AES Encryption**

The Advanced Encryption Standard, or AES, is used for encrypting and decrypting the sequence. Hence the security of the overall cryptographic process is increased.

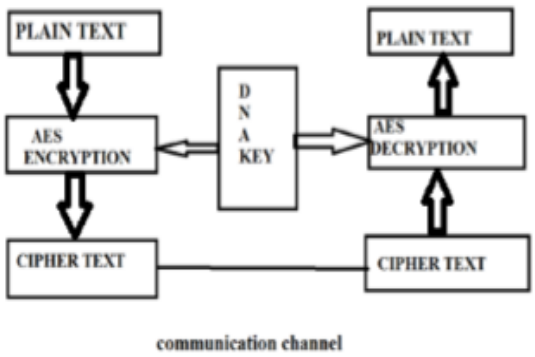


**Fig -3**: AES Encryption

### 3.3 Implementation of DNA Cryptography in Cloud Computing

Cloud computing is the latest technology in the field of distributed computing. The cloud employs encryption techniques to secure the data that will be used or stored in the cloud that provides various online and on-demand services for data storage, network services, platform services and etc. Without any delay of information exchange, sensitive data can be protected. Many organizations are unenthusiastic to use cloud services due to data security issues as the data resides on the cloud services provider's servers.

These issues have been attempted to solve in the past by using digital signature with DH key exchange and AES Encryption algorithm, by combining RSA, digital signature and Kerberos Authentication.

In this paper [3], the Bi-serial DNA encryption algorithm is used to provide security in cloud computing applications. Some additions are made to the previously stated Bi-serial DNA Encryption algorithm, they are:

1. Key combination is used for added security. A 72-bit key is generated, by adding 8-bit ATGC to the 64-bit key values obtained by key combinations. By using Diffie Hellman algorithm, the ATGC is sent to the receiver side, and the key value will be changed randomly every time.
2. Cloud computing deals with text which may not necessarily be English, so the plaintext is converted into Unicode, then its ASCII value is converted into Hexadecimal, and the rest encryption process is the same as stated in 3.1.

### 3.4 Secure Medical Image Encryption based on Intensity level using Chao's theory and DNA Cryptography

The confidently, reliability, security in storage and transmission of digital image are the primary concerns in many applications. In telemedicine the diagnosis and treatment is based on patient information in the form electronic medical images, which is transmitted from a different location using telecommunication. The physicians' diagnosis is based on electronic medical image. Because of open source the quality of images is affected by noises and intruders, which may causes erratic problems. The digital medical image requires zero tolerance to these noises. These images are very large in size and contain confidential data. Hence the compressing of the medical image is not possible, which may lead to the issues like the speed of the transmissions, cost of storage, reliable and robust to store, ensuring the security of the sensitive data. The security of the digital medical image has become more essential and also must fulfill the requirements like integrity, reliability and confidentiality.

Using DNA techniques for digital medical encryption is still in the premature stage. In this paper [4], for the basic operation they have used the Chen's hyper chaotic map and The Lorenz chaos system. The Chen's hyper chaotic map has spatiotemporal complexity and mixture property. The sequences are very complex and complicated to predict and analyze. So, it is suitable to enhance the security of medical image encryption. The Lorenz chaos system is high dimensional chaotic map and it is very complex. The chaotic sequence generated using this is more unpredictable and hence it provides high security which is required for digital medical images.

The novel approach for digital medical image encryption is performed by using Chao's theory and DNA encoding. In the proposed model first, based on pixel values two grayscale images are generated from the input digital medical image. The grayscale images can be odd pixel value images or even pixel value images. These grayscale images are generated from the input digital medical image, after which they are transformed into 8-bit binary images separately.

The DNA sequence as A=01, T=10, G=11and C=00 is applied on 8-bit binary odd/even image and the DNA encoded odd/even matrices are obtained respectively. The Lorenz and Chen's chaotic sequences are generated using state variables and control parameters. Based on index of the sorted chaotic sequences, the pixels of the DNA encoded odd matrix and even matrix are scrambled.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| A | 00 | 00 | 01 | 01 | 10 | 10 | 11 | 11 |
| T | 11 | 11 | 10 | 10 | 01 | 01 | 00 | 00 |
| G | 01 | 10 | 00 | 11 | 00 | 11 | 01 | 10 |
| C | 10 | 01 | 11 | 00 | 11 | 00 | 10 | 01 |

Algorithm 1. Digital medical image encryption

Input: The input is the digital medical image as Ip(R, C)

Output: The output is the encrypted medical image as Op(R, C)

Step 1: Start

Step 2: The input digital medical image is represented as I(R, C),

Where R is the size of row and C is the size of column.

The input digital medical image is divided into odd and even images based on pixel values. Further, these images are converted into 8 bit binary odd image matrix and even image matrices.

Step 3: The binary images are revamped into DNA encoded odd and even matrices.

Step 4: Chen's hyper chaotic sequences X and Y are sorted in increasing order as X1 and Y1. The Lorenz chaotic sequences XX and YY are sorted in increasing order as XX1 and YY1.

Step 5: The index value of X1 and Y1 are used to scramble the pixels of the odd matrices and the index value of XX1 and YY1 are used to scramble the pixel of even matrices separately.

Step 6: ADD operation is used to perform addition of the two odd and even matrices.

Step 7: Transform the resultant matrix into binary using DNA decoding and into decimal to obtain encrypted image.

Step 8: Stop

The decryption is performed using inverse process of digital medical image encryption algorithm and in place of addition operation subtraction operation is used.

The performance analysis demonstrates that the proposed algorithm provides high security. There are 4 steps that they have done to check for Performance analysis

**1) Histogram Analysis**: the distribution of pixel value based on intensity

**2) Correlation Coefficient Analysis**: is measure of correlation between the contiguous pixels in the given images. The good encryption algorithm must have highly correlated adjacent pixels.

**3) NPCR and UACI**: two criterion used to measure the performance of image encryption methods against the differential attacks.

**4) MSE and PSNR**: two metrics used to check whether the distortion of noise or error effects the quality of the image.

In the proposed algorithm the primary values of state variables and the system parameters of the Chen's hyper chaotic map and Lorenz chaotic maps are used as secret key. The Chen's and Lorenz hyper chaotic system is highly sensitive to initial conditions of state variables and control parameters. If there is a slight modification then retrieving same input medical image from decryption process is not possible.

## 4. CONCLUSION

In this survey paper we have studied various algorithms based on DNA cryptography and the scope of DNA in the security of various kinds of data. It can concluded from the study of the methods that the proposed DNA Cryptography methods promise to be a better solution for implementation in secure networks. The research of DNA cryptography is still at an initial stage. It is far from mature, both in theory and realization. We discussed a concise outline about cryptography, quantum cryptography and DNA based cryptography and also about secure message transfer between two systems. Information about technologies used in DNA, such as PCR amplification.

We discussed the implementation of Bi-serial DNA encryption algorithm containing technologies of DNA synthesis, PCR amplification, DNA digital coding, XOR operation as well as traditional cryptography. Then we saw the extension to this BDEA Algorithm to non-English characters. Finally we saw the Secure Medical Image encryption using Chao's theory. It is also suitable for telecommunication applications. This method focused on generating matrices of the image based on pixel values and performing addition on those matrices for encryption.

Most importantly, DNA cryptography indicates that biological molecules can be used for cryptographic purposes and has irreplaceable properties.

## REFERENCES

[1] Prabhu, D, and M Adimoolam. "Bi-Serial DNA Encryption Algorithm (BDEA)." Journal(2011)

[2] Karthigaikumar, P. (n.d.). "Vlsi Implementation of DNA Cryptography Using Quantum Key Exchange."

[3] B, Prajapati Ashishkumar. 2016. "Implementation of DNA Cryptography in Cloud Computing and using socket."

[4] Prema T. Akkasaligar and Sumangala Biradar, "Secure Medical Image Encryption based on Intensity level using Chao's theory and DNA Cryptography", Computational Intelligence and Computing Research, IEEE Conference 2016.