DNA BASED CRYPTOGRAPHIC SYSTEMS: DESIGN, ANALYSIS AND EVALUATION

DOÇ. DR. ESRA ŞATIR

Genel Yayın Yönetmeni / Editor in Chief • C. Cansın Selin Temana

Kapak & İç Tasarım / Cover & Interior Design • Serüven Yayınevi

Birinci Basım / First Edition • © Aralık 2024

ISBN • 978-625-5552-16-7

Serüven Yayınevi / Serüven Publishing
Türkiye Adres / Turkey Address: Kızılay Mah. Fevzi Çakmak 1. Sokak, Ümit Apt No: 22/A Çankaya/ANKARA
Telefon / Phone: 05437675765
web: www.seruvenyayinevi.com
e-mail: seruvenyayinevi@gmail.com

Baskı & Cilt / Printing & Volume
Sertifika / Certificate No: 47083

DNA BASED CRYPTOGRAPHIC SYSTEMS: DESIGN, ANALYSIS AND EVALUATION

DOÇ. DR. ESRA ŞATIR[1]

1 Esra ŞATIR, Doç. Dr., Düzce Üniversitesi, Bilgisayar Mühendisliği, Konuralp, Düzce; 0000-0003-1793-2472

CONTENTS

1. INTRODUCTION

Nowadays, information security is major issue in all type of communication. Because of the fast development of technology, information is essential for each organization. Information security is the area where the information is protected from the uncertified access, use, modification or destruction of data [1]. Digital communication has become an important part of information systems especially in recent years. The communication in information systems must be safeguarded when information leakage or unexpected attacks occurs by an bserver. Cryptographic techniques are being widely used to secretly transmit the data. DNA cryptography is a new cryptographic field which employs DNA with the information carrier and computation. DNA cryptography has attracted too much attention because of the significant storage capacity of DNA [2]. For a concrete view, the main digital information carriers are provided in Table 1, mentioning their capacities, access times, and lifetimes.

Table 1. *Digital Information Carriers [3].*

Carrier Medium	**Carrier capacity**	**Time to access**	**Service life**
SSD (Slid State Drive)	≈ 1 TB	ms	≈ 10 years
CD, DVD, Blue Ray disks	≈ 128 GB	ms	≈ 10s years
HDD (Hard Disk Drive)	≈ 10 TB	10 ms	≈ 10 years
Magnetic Tapes	≈ 100 TB	minutes	≈ 15-30 years
DNA	≈ hundreds of EB	10s hours	centuries

Deoxyribonucleic Acid (DNA) is the biological medium for storing and transmitting the genetic material for all organisms on the earth. Its power is hidden in its impressive storage capacity. It is known that an ounce of DNA can store 30,000 terabytes of memory with the lifetime up to 1 million years. Besides, DNA has the ability of replicating and assembling itself through the process of evolution without any intervention after only an initial setup. This inexpensive computing power and efficiency of DNA renders it as an indispensable medium for transmission of digital information [4].

L.M. Adleman introduced the concept of DNA computing in 1994 by using the natural abilities of DNA in computer science. Adleman used DNA as a medium of data storage and parallel computation. In his work, he solved the Hamiltonian Path problem via only DNA and molecular operations. Since then, DNA computing has become a popular area yielding DNA cryptography as an important branch in the field. General information security process in the scope of DNA can be separated into three general methodologies. *First* of them is natural DNA cryptography where cryptographic algorithms are applied to the synthesized DNA strands in a test tube. Here, DNA chemical processes are applied to DNA strands to generate the encrypted data. *Second* one, Pseudo-DNA cryptography employs theoretical models with no biological material. Here, these theoretical modelling techniques are applied to binary data. Pseudo-DNA cryptography method starts with the message that is translated into binary strings and then transformed into pseudo-DNA strands. Pseudo- DNA operations are then applied to the pseudo-DNA strands to increase the security

of existing algorithms. Finally, the resultant pseudo-DNA strands are translated to binary strings and sent through the communication channel. *Third* one, DNA steganography is used to conceal information in DNA. Steganography means "secret writing". When messages are hidden within DNA strands, it is hard to detect and decode since there are tens of millions of possibilities to sift through [4].

Researchers are showing a growing interest in using DNA as a medium for cryptographic purposes. DNA based cryptography exhibits several benefits when compared to conventional cryptographic methods. These are high storage capacity, low error rate, and resistance to environmental factors. However, using DNA for cryptography suffers some important difficulties like the high cost of synthesis and sequencing, the requirement for special equipment and expertise and finally, the potential for errors during encoding or decoding [4].

This work presents the ongoing process of DNA cryptography including its theoretical basis, requirements and the problems in the field by comparing the contemporary studies in the literature. In section 2, the details about DNA and DNA coding issues have been provided. In section 3, DNA computing process, the traditional cryptographic approaches and DNA cryptography including the main requirements and the state of the art techniques have been explained. The metrics used for evaluation has been mentioned in section 4. Finally general outcomes of DNA cryptography have been presented in Section 5.

2. THEORETICAL BACKGROUND

In this section, we aim to present information about DNA structure and the components for DNA coding.

2.1. DNA Structure

The unit element of DNA is the molecule that is called nucleotide. A nucleotide consists of these elements; nitrogenous base, phosphate group and sugar. The sugar which is used to build the nucleotide is called as deoxyribose, that's why deoxyribo exists as a prefix in DNA abbreviation. The structural components of DNA has been indicated in branches in Figure 1.

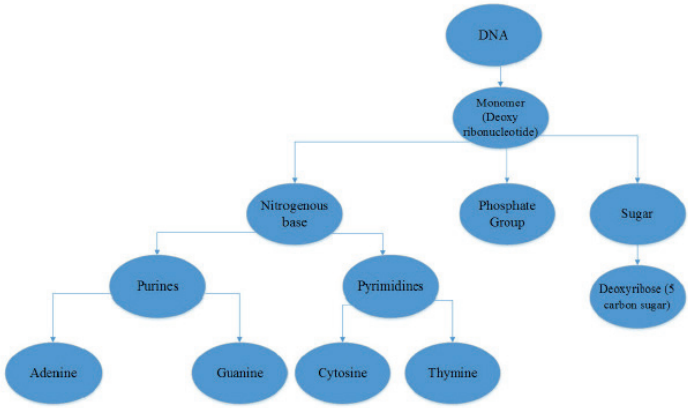

Figure 1. *Structural elements of DNA [4].*

Sugar contains an atom that has five carbons and numbered 1′ to 5′. In DNA structure, the phosphate group attaches to the 5′ carbon, whereas

the nitrogenous base attaches to the 1′ carbon. Hydroxyl group (OH) attaches to the 3′ carbon of the sugar. DNA contains four different molecules called nucleotides. Every nucleotide is defined by a nitrogenous base: Adenine (A),

Guanine (G), Cytosine (C), and Thymine (T). A and G belong to the nitrogenous bases called purine category, while C and T belong to the pyrimidine category [4].

DNA is hereditary material which carries the genetic information for all organisms. It contains two antiparallel biopolymer strands wrapped around each other. This forms a double helix structure as shown in Figure 2 [5].

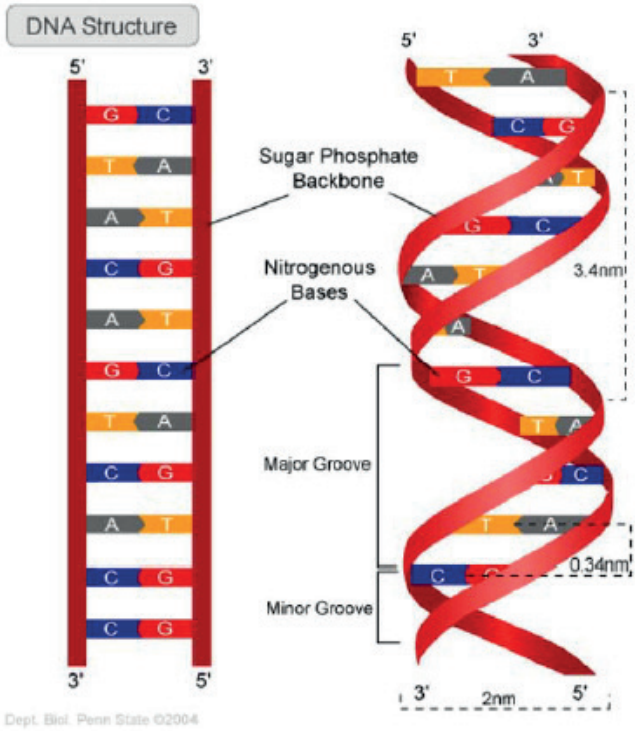

Figure 2. *Double-helix structure of DNA [5].*

In DNA, Adenine always pairs with base Tymine whereas Guanine always pairs with base Cytosine. This is known as Watson-Crick's complementarity principle and formed the discovery of spiralling staircase structure of double-stranded DNA known as 'double helix.' The double-stranded DNA can be expressed as:

5′ – ACG – 3′

3′ – TGC – 5′

From this representation, it is obvious that a single-stranded DNA sequence pairs with another sequence in the opposite direction [4].

2.2. DNA Coding Basics

In natural DNA cryptography, biological materials in laboratories are employed by implementing molecular operations on natural or artificial DNA strands. Biological operations of natural DNA cryptography can be applied in a computer-simulated environment. In this sub section, the concerning concepts have been explained as a basis for DNA cryptography issue.

2.2.1. DNA synthesis and sequencing

An implementation medium is the essential part for DNA cryptography processes. DNA synthesis (reading) is the process of forming artificial DNA strands with the use of a special machine known as synthesizer. The synthesizer combines DNA bases by considering the given input to generate millions of synthetic DNA sequences. The produced artificial DNA sequence is then used for experiments. Moreover, it can be used as a medium with the purpose of storage. Therefore, artificial DNA strands are the main materials for DNA computing and DNA cryptography [4].

There is an increasing demand for synthetic DNA in many sectors or research and commercial activities. There will be significant potential in

engineering biology, therapy, data storage and nanotechnology, if DNA can be provided at scale and low cost. By basing on the successes in next generation sequencing and gene editing technologies, DNA synthesis has already been an emerging industry. However, the synthesis of >200 bp sequences is too expensive. Therefore, it is a research area to developed alternative technologies with the purpose of overcoming the limitations in DNA synthesis as effectively as DNA sequencing (reading). The chronological development of DNA synthesis by considering the amount has been indicated in Figure 3 [6].

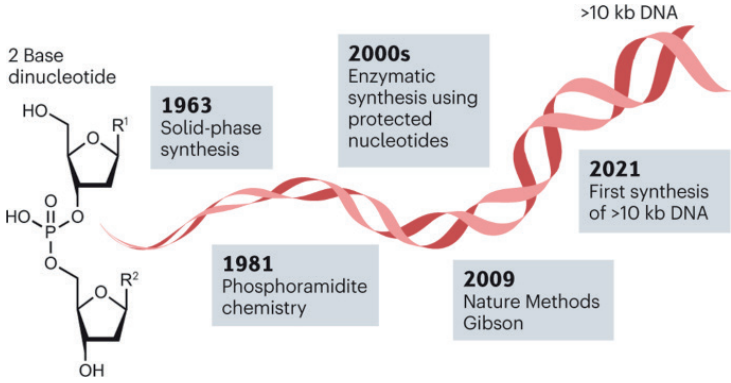

Figure 3. Chronological *developments of DNA synthesis technologies by considering the amount [6].*

The first wave in DNA technologies is following the sequencing of the human genome in the early 2000s. This ability of DNA reading is so close to surpass Moore's law. According to the Moore's law, the number of transistors doubles every 2 years, as indicated in

Figure. 4a.The second wave in this field is provided by novel technologies like de novo DNA synthesis and CRISPR gene editing. These technologies gave the ability to 'edit and write' DNA. These developments encouraged the researchers to implement the read, edit and write abilities of DNA for products such as vaccines, data storage and drug delivery devices or genome engineering to generate organisms with useful properties, such as heat-resistant plants. This significant advance of applying DNA technologies brings the need for synthesis on a big scale, to provide industry challenges including health, environmental and computer science and etc. [6].

The structure of DNA was first understood in the second half of 1900s. Over four decades, relevant steps were taken to establish the underpinning chemistry for the stepwise synthesis of DNA, nucleotide by nucleotide as mentioned in Figure 4b. Chemical methods were then developed to reliably provide short DNA strands containing <200mer nucleotide called oligonucleotides. After these valuable steps, these methods were optimized for automatic synthesizers, which became mandatory tools for gene engineering and sequencing. These developments in enzymatic and hybrid approaches to generate DNA is longer and more complex than oligonucleotides. Accordingly, companies have commercialized these approaches, offering services ranging from custom synthesis to benchtop DNA printers, making DNA synthesis accessible to non-expert users. This made the DNA technologies reachable for many researchers. In recent years, thousands of DNA nucleotides have been produced, by showing that, the gap between the abilities to read and write DNA may close in the near future [6].

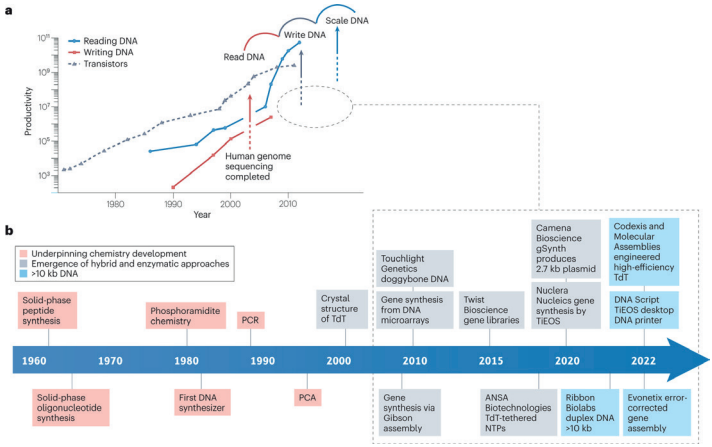

Figure 4. *The progress in DNA write and read technologies.*

a. *Productivity of DNA reading and DNA writing by considering Moore's Law.*

b. *Timeline of milestones in DNA synthesis technologies [6].*

2.2.2. DNA length

To scale and understand the results of performed experiments in DNA cryptography processes, the length of DNA molecules has to be measured. DNA length indicates the total number of used nucleotides in formation process. For example, if a strand has 20 nucleotides, its length is written as 20 mer. This indicates that the DNA strand contains 20 molecules which can bind with others. For instance, if a double-stranded DNA contains 20 base pairs, its length is 20 bp [4].

2.2.3. DNA coding

In a DNA sequence, there are four nucleic acid bases: Adenine (A), Cytosine(C), Guanine (G), and Thymine (T). A complements with T, whereas G complements with C. In the literature it has seen that the binary values 00, 01, 10, and 11 are used to represent T, G, C and A, respectively as indicated in Table 2. Therefore, it is obvious that 00 complements with 11 like T and A, whereas 01 complements with 10 like G and C [7].

Table 2. *Bit-base translation Table [8].*

Bits in Computer Systems	Oligos in DNA molecule
00	T
01	G
10	C
11	A

Complement of a single base can be defined as follows:

T = complement of (A)

A = complement of (T)

G= complement of (C)

C = complement of (G)

Like in DNA encoding and decoding processes, the algebraic operations on DNA sequences (addition or subtraction) are implemented in binary base. In encryption and decryption, the addition and subtraction operations stays reciprocal. Besides, the inverse operation of XOR does not change. Accordingly, the addition and XOR operations have been provided in Table 3 [7].

Table 3. *Addition and XOR operations on DNA sequences [7].*

+	A	C	G	T	XOR	A	C	G	T
A	A	C	G	T	A	A	C	G	T
C	C	G	T	A	C	C	A	T	G
G	G	T	A	C	G	G	T	A	C
T	T	A	C	G	T	T	G	C	A

3. DNA İN CRYPTOGRAHY

In this section the role of DNA in cryptography has been explained by investigating DNA computing, traditional and DNA cryptography by including the main requirements of a DNA cryptographic system and comparison of the state of the art techniques.

3.1. Computing with DNA

DNA computing is a recently applied method for solving challenging computational problems. It is firstly proposed by Adleman in 1994 to encode and solve Hamiltonian path problem (HPP). Afterwards this new, DNA computing, field has been concerned by many researchers around the world [9].

In order to explain the details of DNA computing, it will be useful to investigate the model that is developed to solve the Hamiltonian path problem. The purpose of Hamilton path problem is to detect a path beginning at *vin*, and ending at *vout*. Here, every other vertex is entered exactly once on a directed graph. A random oligonulecotide DNA sequence with the length of 20- mer was produced for each vertex *i*. The process for solving the HPP can be explained by the following items:

1. Produce random paths through the graph.

2. Keep the paths beginning with *vin* and ending with *vout*.

3. If the graph has n vertices, then keep the paths that enter exactly n vertices.

4. Keep the paths that enter all of the vertices of the graph at least once.

5. If any paths remain, say "yes", otherwise say "no" [9]

The graph structure for HPP has been depicted in Figure 5.

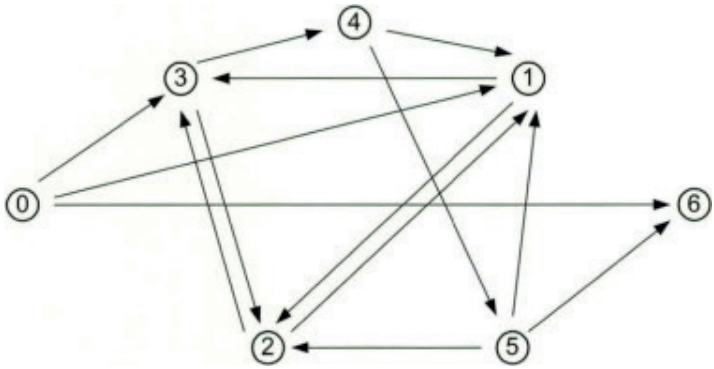

Figure 5. *The Directed graph for Adleman's experiment [10].*

To solve HPP, Adleman encoded DNA sequences for all possible answers. Besides he removed all unnecessary solutions which do not meet the requirements via a series of restrictive conditions. Then, Adleman solved the HPP. When compared the traditional computing algorithm, it can be seen that DNA computing provides a huge parallelism and high amount of density. Therefore, DNA cryptography benefits from DNA computing by employing some biological technologies of DNA computation process [9].

As an instance, Lee et al. proposed a DNA computing experiment for the HPP which is presented in graph in Figure 6a. Here, the directed graph has 8 nodes and 14 edges with two Hamiltonian paths. Node 0 and 7 were designate as the start node s and the destination node t, respectively. They encoded the nodes with randomly generated oligonucleotides, each of size I8-mer as shown in Figure 6b. In Adleman ' s DNA computing, firstly the graph in terms of a set of oligonucleotides is encoded and then the following well known genetic engineering procedures are impended: (1) DNA synthesis, (2) annealing and ligation, (3) Polymerase Chain Reaction (PCR) amplification, (4) DNA extraction, and (5) DNA detection. By employing the procedures (1) through (3), DNA computing generates a large number of DNA molecules, each of which corresponds to a path from the start node to the destination, including the ones for Hamiltonian paths, if any. Procedures (4) and (5) test if such molecules exist [10].

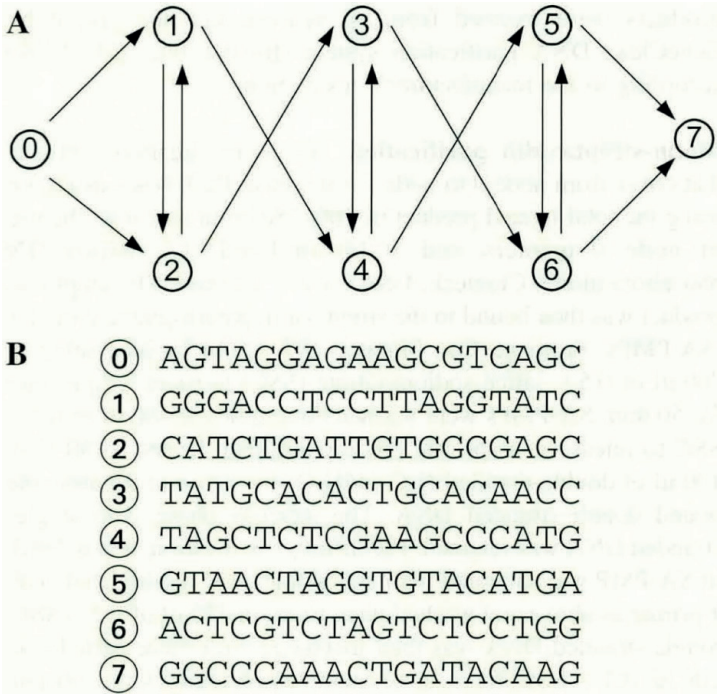

Figure 6a. *The directed graph of Lee et al. for HPP.* **6b.** *Encoding each node in (a) via oligonucleotides [10]*

3.2. Traditional cryptography

Cryptography has an essential place in data protection. Its aim is to ensure the security and the integrity of the message and the sender authenticity. Since most people interact electronically every day by e-mails, e-commerce, cellular phones, etc., protecting data by employing the cryptographic techniques has gained an extreme attention. The increase in

the amount of data transfer has raised the need on cryptographic algorithms and authentication by users [11]. The general flow of cryptography has been provided in Figure 7.

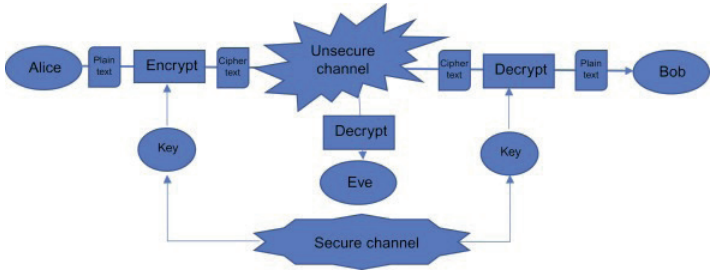

Figure 7. *The general flow of cryptography [12].*

In Figure 7, Alice and Bob can encrypt the data while sending and decrypt the data receiving. This would avoid the illegitimate user Eve from decoding the data that is sent over the unsecure channel. However under ideal conditions, Alice and Bob should keep the algorithm secret. Thus, Eve cannot decode it. But keeping the algorithm secret is not sensible and not practical. Moreover, making the algorithm public renders it resilient against attackers. Using an algorithm which is kept secret, is never recommended. The security of an ideal cryptographic process should be based on keeping the key secret instead of the algorithm, to prevent Eve from decrypting the message. Here, it is obvious that the solution is a pre-shared secret key between Alice and Bob, from which Eve is unaware of [12].

Cryptography consists of the following two main categories according to the employed key:

1. Symmetric key cryptography: In this type of cryptography, both the sender and the recipient employ the same secret key for encryption and decryption phases like Advanced Encryption Standard (AES) and Data Encryption Standard (DES) [11]. This secret key is assumed to be shared between by the user and the recipient via a secure communication channel. The sender creates and uses the secret key to encrypt the message (to produce the cipher text from the plain text) and then sends it to the recipient. When recipient obtains the key, he/she uses it to decrypt the message (to produce the plain text from the cipher text). Here, the security of the cryptographic process relies on keeping this key secret. General flow of the symmetric key cryptography has been provided in Figure 8.

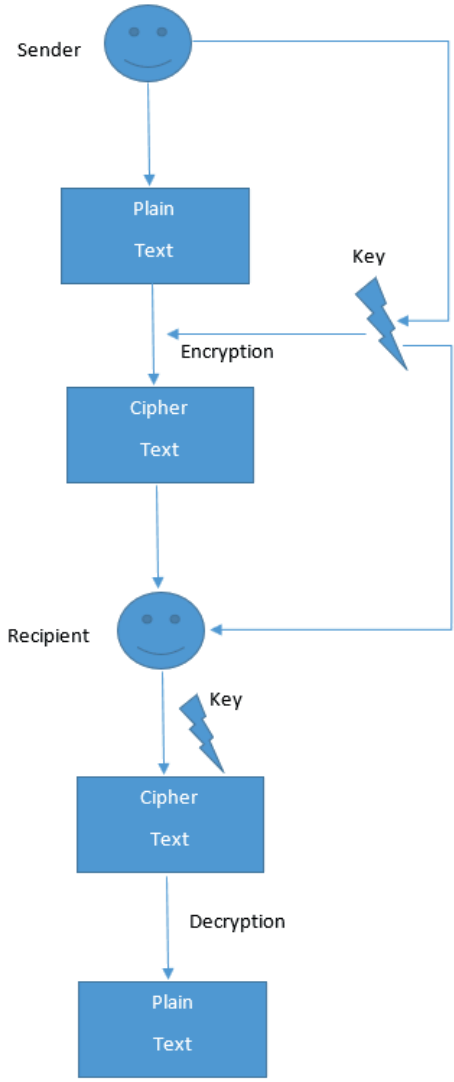

Figure 8. *The general flow of symmetric key cryptography*

2. Asymmetric key cryptography: This type of cryptography contains two keys. The first one is public key and the second one is private key. Both thsee keys work in pairs of matched public and private keys like the Rivest, Shamir and Adleman (RSA) algorithm. [11]. The sender generates the public and private keys via a defined mathematical process. Here, the public key is transferred between the sender and the recipient. However, the matched private key with this public key is used to encrypt the message (to produce the cipher text from the plain text). When the recipient obtains the public key, he/she uses it to derive the private key to decrypt the message (to produce the plain text from the cipher text). Here, the security of the cryptographic process relies on keeping this private key secret. General flow of the asymmetric key cryptography has been provided in Figure 9.

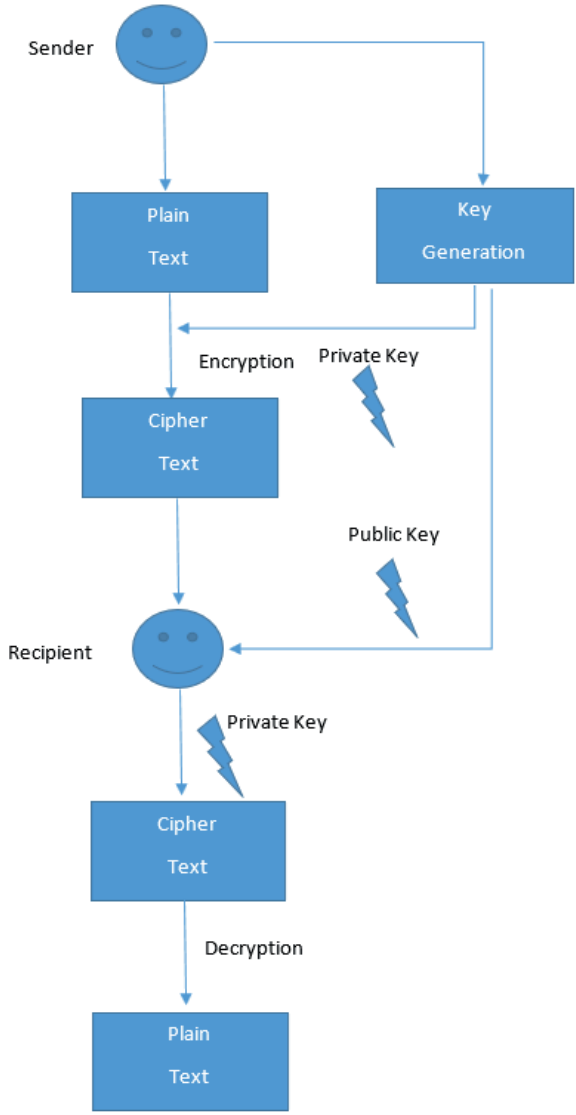

Figure 9. *The general flow of asymmetric key cryptography*

The goals of cryptography can be listed as follows:

a. Confidentiality: Data transmission between two computers must be accessed by only an authorized user.

b. Authentication: Data transmission between two computers must be accessed by only an authorized user.

c. Integrity: The transmitted information can be modified by only the authorized party

d. Non-Repudiation: Ensures the message that the sender or the receiver should be able to deny the transmission.

e. Access Control: The information during transmission can only be accessed by only the authorized people [11]

3.3. DNA Cryptography

In this subsection, pseudo-DNA cryptography has been considered since it employs theoretical models with no biological material, as suitable with the computer systems. Here, these theoretical modelling techniques are applied to binary data. Pseudo-DNA cryptography method starts with the message that is translated into binary strings and then transformed into pseudo-DNA strands. Pseudo- DNA operations are then applied to the pseudo-DNA strands to increase the security of existing algorithms. Finally, the resultant pseudo-DNA strands are translated to binary strings and sent through the communication channel [13].

In DNA cryptography, DNA bases, namely A, T, C and G, are used to encrypt and store the data unlike the traditional approaches. In traditional approaches, binary bits 0 and 1 are used used to encrypt and store the data. A combination of DNA bases is known as DNA sequence. Here, the sender selects one of the combinations of DNA bases for encrypting the data to improve security. The data is encrypted and converted to a DNA sequence in this way [13].

For example, let the encoding rule for the DNA bases is as follows:

00 → A, 01 → T, 10 → C and 11 → G.

Let' s encrypt 'H'. If the binary ASCII value of 'H' corresponds to 01001000. It is obvious that, H is represented as 'TACA' by using the DNA encoding rule given above [13].

The binary data, namely the plain text, can be converted to oligos by this way. Then pseudo DNA operations are applied on this DNA strands to enhance the algorithm. This is the process on which the researchers focus to achieve significant results in terms of the measurements metrics in the literature. After this process, the resultant DNA base strand, namely the cipher text, is converted to binary data by applying the inverse of the given procedure above. The workflow of pseudo DNA cryptography has been provided in Figure 10.

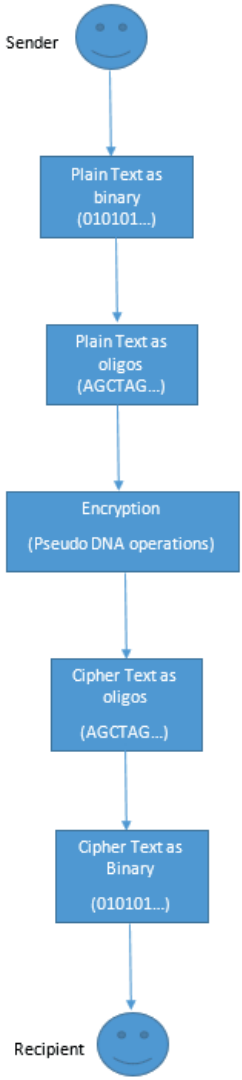

Figure 10. *The workflow of pseudo DNA cryptography.*

On the other hand, the most well-known DNA technologies can be used in DNA cryptography for encryption, decryption and key generation as presented in Table 4 [1].

Table 4. *DNA computing technologies for DNA cryptography [1].*

DNA computing technology	Key points
Polymerase chain reaction (PCR)	In this method, it contains two primer-oligonucleotides to increase the DNA pattern in every cycle This method is used in key generation, encryption, and decryption.
One the Pad (OTP)	OTP is randomly chosen secret DNA sequence but occurs only once. This method is used for encryption, key generation, and decryption.
DNA decoding	In this method, the plaintext is first converted into ASCII code, and then into binary code. After this decoded into DNA pattern This method is used in a preprocessing step.
DNA chip	Biochip has many sports conceal on a solid surface. These sports contain DNA pattern which is used to find expression of the gene.

3.4. Requirements for DNA cryptography

There are some requirements that every DNA cryptographic algorithm should fulfil. They can be been explained as follows:

1. DNA Encoding of Complete character set fulfilment – This requirement means encoding an alphabet including all necessary symbols like uppercase, lowercase letters, numbers and special

characters. Here, it should be ensured to encode all the characters of the plaintext into DNA sequence. On the other hand, the generation of this encoding table should be based on a well-defined procedure that can be repeated any number of times to generate the encoding table automatically.

2. Dynamic Encoding Table Generation - In order to provide a remarkable level of security, different encoding tables should be generated for periodic intervals or for every interaction session between the sender and recipient. This will yield to produce different DNA sequences for each element in the character set.

3. Unique sequence for encoding of every character of plaintext to DNA sequence – The encoding process of plaintext into DNA sequence should be unique for every element of the character set in every session between sender and receiver

4. Robustness of encoding – While DNA encoding of the plaintext, a robust encoding scheme which is very difficult to decipher should be provided. This will render the algorithm resilient against attacks.

5. Biological Process Simulation - DNA encryption and decryption algorithm should be based on biological processes. These biological processes are simulated to adapt the encoding to digital computing environment. Here, a complete algorithm that is completely based on simulation of difficult biological processes is required. This will render the algorithm resilient since modern cryptographic algorithms have been broken via DNA cryptography.

6. Dynamicity of encryption process – This last requirement can be explained as different cipher texts should be produced by the same plaintext [14].

A concrete example for requirement 1; "DNA

Encoding of Complete character set fulfilment" has been provided in Figure 11 [15].

DNA Base Sequence				
ACAT-a	ACTC– q	CAAT– G	ATAA– W	ATTA- ,
ACTG- b	ACCG– r	CATG– H	ATTT– X	ATCC- .
ACCC- c	TCTC– s	CACG– I	ATCG– Y	TTTA- ?
ACGA– d	TCCC– t	CAGT– J	ATGC– Z	TTCG- /
TCAT– e	CCTT– u	GAAG- K	TTAA– 0	CTTC- :
TCTG– f	CCCC– v	GATA– L	TTTT– 1	CTCG- ;
TCCG– g	GCTA– w	GACG– M	TTCC– 2	GTTC– "
TCGT– h	GCCC– x	GAGG– N	TTGG- 3	GTCC– '
CCAG– i	AAAA– y	AATA– O	CTAT– 4	AGAG- {
CCTA– j	AATT– z	AACG– P	CTTG– 5	AGTA– [
CCCG– k	AACC–A	TATC– Q	CTCC– 6	AGCG- }
CCGG– l	AAGG- B	TACG– R	CTGA– 7	AGGG-]
GCAA– m	TAAT– C	CATC– S	GTAT– 8	TGAA- |
GCTT– n	TATG– D	CACC– T	GTTG– 9	TGTT- \
GCCG– o	TACC– E	GATT– U	GTCG– <	TGCG- +
GCGC– p	TAGA– F	GACC– V	GTGT- >	TGGC- =
CGAA--	CGCC-)	GGAT- *	GGCC- ^	AGTT- $
CGTT-–	CGGG– (	GGTG- &	GGGA-%	AGCC- #
TGTA -@	TGCC- ¡	CGTA- ˜	CGCG- ‘	GGCG- £

Figure 11. *Example of encoding a full alphabet via DNA sequences [15].*

In table 5, comparison of the state of the art techniques by considering each mentioned requirement above, has been presented.

Table 5. *Investigation results of the existing DNA encryption algorithms in terms of the given requirements [16].*

Authors	Character Set Fullfillment	Dynamic Encoding Table	Unique Sequence for Encoding	Encoding Robustness	Biological Process Simulation	Dynamicity of Encryption
G. Cui et al. [17]	X	X	X	X	*	*
Q. Zhang et al. [18]	X	X	X	X	*	X
S. Sadeg et al. [19]	X	X	X	X	√	X
S.T. Amin et al. [20]	X	X	X	X	√	X
O. Tornea & M.E. Borda [21]	X	X	X	X	*	X
M. Sabry et al. [22]	X	X	X	X	*	*
X. Wang & Q. Zhang [23]	X	X	X	X	X	X
A. Akanksha et al. [24]	√	X	X	X	X	X
K. Ning [25]	X	X	X	X	√	X
N. H. UbaidurRahman et al. [14]	√	√	√	√	√	√
E. Şatır & O. Kendirli [16]	√	X	√	√	√	√

X - Minimum. √ - Acceptable. * - Partially

4. EVALUOTIONAL PARAMETERS

The design of an encryption/decryption algorithm should be complex enough to resist against security attacks. The best way to reach such kind of complexity is to work towards scalability since this will ultimately lead to large-scale of complexity. The main idea to increase the complexity of a system by augmenting its size is to achieve the desired security which will require significant efforts to attack the system successfully. These desired properties can be satisfied by DNA cryptography since it provides huge parallelism and storage capacity, simultaneously. The power of DNA encryption does not only rely on the molecules or encoding. It relies on the positions where we want to save the data to protect it from attacks for a longer time [15].

Cryptography is the procedure to create such kind of algorithms, whereas; cryptanalysis is the procedure in which the developers examine the cipher for its vulnerabilities and accordingly, improve the algorithm by giving insight for future directions. Randomness, avalanche effect, and entropy per bit are some of the desired properties to evaluate the cipher text [15].

4.1. Frequency (Mono Bit) Test

The frequency test calculates the number of 0's and 1's in the ciphertext. This process investigates whether the number of zeros and ones are equal. This is one of the desired properties of a ciphertext. Value 0.01 is the level of significance for this test. This

means that only 1 sample out of 100 will be rejected. Ideally, the final value should be "1", that means a perfect balance between 0 and 1 in the string. The following explanations should be considered for this test:

n: binary string length

ε: the index of bits in the binary string like $\varepsilon = \varepsilon_1, \varepsilon_2, \ldots \varepsilon_n$

S_{obs}*: the absolute value for simplified* X_i

$$X_i = 2\varepsilon - 1 = \pm 1 \quad (1)$$

$$S_n = X_1 + X_2 + \ldots\ldots + X_n \quad (2)$$

$$S_{obs} = \frac{S_n}{\sqrt{n}} \quad (3)$$

The final value for this criteris is calculated as follows:

$$p - value = erfc\left(\frac{S_{obs}}{\sqrt{2}}\right) \quad (4)$$

Here, *erfc* is a complementary error function. This test evaluates the p-value. If the computed p-value is less than 0.01, it means that the given sequence is not random. On the other hand, if p-value is more than 0.01, the string can be considered as a random string [15].

4.2. Avalanche Effect

Avalanche effect means that a small change in plain text or key yields a significant change in the ciphertext.

It's an extremely valuable property for cryptographic algorithms. The more the avalanche effect is, the better the algorithm will be. If the avalanche effect is greater than 50%, this indicates that the cipher is more random and less predictable for attackers. Avalanche effect is explained as follows [15]:

$$Avalanche\ Effect = \frac{Number\ of\ the\ flipped\ bits\ in\ cipher\ text}{Total\ number\ of\ bits\ in\ cipher\ text} \quad (5)$$

4.3. Entropy

According to Shannon's concept, entropy means the uncertainty in the cipher text bits. Uncertainty of the cipher text is determined by estimating the number of plaintext bits that can be recovered from mixed ciphertext to obtain the original message without any errors. On the other hand, entropy is the weighted average of optimal bit representation size, such as the average size of an encoded message. Entropy can be defined by using the following equation:

$$H(X) = -\sum_{x \in X}(\Pr[X] log_2(\Pr[X])) \quad (6)$$

Here, X corresponds to $X = \{0, 1\}$. If the string is considered in binary, the highest entropy value is 1 when the 0s and 1s are equally distributed [15].

4.4. Hamming Weight

This metric is used to evaluate the randomness of a binary sequence. If two strings of equal length have

different symbols at some positions, the total number of those positions is called hamming weight. The higher value of the hamming weight has, the better randomness the binary sequence will have. Hamming weight can be defined by the following equation [15]:

$$Hanning\ Weight = \frac{Total\ number\ of\ non-zero\ bits}{Length\ of\ the\ cipher\ text} \quad (7)$$

5. POTENTIAL PROBLEMS OF DNA TECHNIQUE

DNA-based cryptography offers several advantages like high storage capacity, low error rate, and resistance to environmental factors when compared to traditional cryptographic methods. However, using DNA for cryptography suffers some important difficulties like the high cost of synthesis and sequencing, the requirement for special equipment and expertise and finally, the potential for errors during encoding or decoding.

In spite of its massive parallelism and extreme computation abilities, the DNA computing is still resides in a theoretical stage rendering itself not feasible enough. Since DNA computing model employs molecular technique and applies experimental stage to resolve a certain problem, there is a lack of the necessary computing schemes for different variety of problems. This indicates that there is not a uniform DNA computing and coding model for every handled problem.

When the existing DNA computing mode is considered, the time complexity of DNA computing with respect to the space complexity doesn't increase with the computational complexity, notably. Therefore, DNA computing only transforms the time complexity into space complexity. Once the complication of problems exceed the physical limit of DNA sequence employed in the bio-chemical technique, DNA computing is still not enough to implement. For instance, Boneh spend nearly 4 months to construct $DES^{-1}(E)$ solution. However, the

quantities of cipher key of AES algorithm used by the US federal government is 21 times with respect to DES algorithm. Therefore, it will cost many years if construction of $AES^{-1}(E)$ solution according to the Boneh' s way is aimed. Therefore, it can be claimed that Boneh' s method can only break the symmetric system under 64 bits. Mathematical cryptography can increase the length of the cipher easily rendering the cryptography resilient against powerful attack using DNA computer. Namely, in terms of the existing DNA computing mode, although DNA computer improve the probability of the cipher to be broken by the people, it is disable to provide a real intimidation for the security of cryptography. DNA cipher is a useful supplement to the existing mathematical cipher and it is a good prior choice especially to the lower demand real-time encryption system.

DNA has much more potential in steganography and certification. With the rapid developments in modern biotechnological processes, the costs of biological experiments have reduced rendering them similar with the formal experimental operations. The further developments in biotechnology and the invention of a better DNA cipher design will provide a new orientation for the research in information security field. However, the security, feasibility, stability of DNA cryptography issues still need a deep and further research [9].

REFERENCES

1. Singh, S., Sharma, Y., A Review on DNA based Cryptography for Data hiding, *International Conference on Intelligent Sustainable Systems (ICISS)*, Palladam, India, 2019, pp. 282-285, doi: 10.1109/ISS1.2019.8908026.
2. Zhang, Y., Liu, X., Ma, Y. *et al.* (2017). An optimized DNA based encryption scheme with enforced secure key distribution, *Cluster Comput,* 20, 3119–3130, https://doi.org/10.1007/s10586-017-1009-y
3. Garafutdinov, R.R., Chemeris, D.A., Sakhabutdinova, A.R.; Kiryanova, O.Y.; Mikhaylenko, C.I., Chemeris, A.V. Chemeris, (2022). Encoding of non-biological information for its long-term storage in DNA, *Biosystems, 215*, 104664, https://doi.org/10.1016/j.biosystems.2022.104664
4. Mahjabin, T., Olteanu, A., Xiao, Y., Han,W., Li, T, Sun, W., (2023). A Survey on DNA-Based Cryptography and Steganography, in *IEEE Access*, 11, 116423-116451, doi: 10.1109/ACCESS.2023.3324875.
5. Monika, Upadhyaya, S., (2015). Secure Communication Using DNA Cryptography with Secure Socket Layer (SSL) Protocol in Wireless Sensor Networks, *Procedia Computer Science*, 70, 808-813, https://doi.org/10.1016/j.procs.2015.10.121.

6. Hoose, A., Vellacott, R., Storch, M. *et al.,* (2023). DNA synthesis technologies to close the gene writing gap, *Nat Rev Chem* , 7, 144–161, https://doi.org/10.1038/s41570-022-00456-9

7. Xue X, Zhou D, Zhou C., (2020). New insights into the existing image encryption algorithms based on DNA coding, *PLoS One*. 23; 15(10): e0241184. doi: 10.1371/journal.pone.0241184.

8. Hafeez, I., Khan, A., Qadir, A., (2014). DNA-L-CEB: A high-capacity and mutation-resistant DNA data-hiding approach by employing encryption, error correcting codes, and hybrid twofold and fourfold codon-based strategy for synonymous substitution in amino acids, *Med. Biol. Eng. Comput*., *52*, 945–961.

9. Cui, G., Qin, L., Wang Y., Zhang, X., Information Security Technology Based on DNA Computing, *International Workshop on Anti-Counterfeiting, Security and Identification (ASID)*, Xizmen, China, 2007, pp. 288-291, doi: 10.1109/IWASID.2007.373746.

10. Lee, C., M., Kim, S., W., Kim, S., M., Sohn, U., (1999). DNA Computing the Hamiltonian Path Problem, *Molecules and Cells*, 9(5): 464-469, https://doi.org/10.1016/S1016-8478(23)13571-0.

11. Elamir, M., M., Mabrouk, M., S. , Marzouk, S.,Y., (2022). Secure framework for IoT technology based on RSA and DNA cryptography, *Egypt J Med Hum Genet* 23, 116, https://doi.org/10.1186/s43042-022-00326-5

12. Gupta, R., Gupta, P., Singh, J., 14 - Security and Cryptography, Editor(s): Robert Oshana, Mark Kraeling, Software Engineering for Embedded Systems (Second Edition), Newnes, 2019, pp. 501-547, ISBN 9780128094488, https://doi.org/10.1016/B978-0-12-809448-8.00014-X.

13. Pavithran, P., Mathew, S., Namasudra, S., Lorenz, P., (2021). A novel cryptosystem based on DNA cryptography and randomly generated mealy machine, *Computers & Security*, 104, https://doi.org/10.1016/j.cose.2020.102160.

14. UbaidurRahman, N.,H., Balamurugan, C., Mariappan, R., (2015). A novel DNA computing based encryption and decryption algorithm. *Procedia Comput. Sci*. 46, 463–475. https://doi.org/10.1016/j.procs.2015.02.045.

15. Imdad, M., Ramli, S., N., Mahdin, H., (2021). Increasing Randomization of Ciphertext in DNA Cryptography, *International Journal of Advanced Computer Science and Applications(IJACSA)*, 12(10), http://dx.doi.org/10.14569/IJACSA.2021.0121047

16. Şatir E., Kendirli, O., (2022). A symmetric DNA encryption process with a biotechnical hardware, *J. King Saud Univ. - Sci.*, 34(3): 101838, doi: https://doi.org/10.1016/j.jksus.2022.101838.

17. Cui, G., Li, C., Li, H., Li, X., DNA computing and its application to information security field, *5th Int. Conf. Nat. Comput. ICNC 2009*, Tianjian, China, 2009, pp. 148–152, 2009, doi:

10.1109/ICNC.2009.27.

18. Zhang, Q., Guo, L.,Xue, X., Wei, X., An image encryption algorithm based on DNA sequence addition operation, *BIC-TA 2009 - Proceedings, 2009 4th Int. Conf. Bio-Inspired Comput. Theor. Appl.*, Beijing, China, 2009, pp. 1–5, doi: 10.1109/BICTA.2009.5338151.

19. Sadeg, S., Gougache, M., Mansouri, N., Drias, H., An encryption algorithm inspired from DNA, *2010 Int. Conf. Mach. Web Intell. ICMWI 2010 - Proc.*, 2010, pp. 344–349, doi: 10.1109/ICMWI.2010.5648076.

20. Amin, S., T., Saeb, M., El-Gindi, S., A DNA-based implementation of yaea encryption algorithm, *Proc. 2nd IASTED Int. Conf. Comput. Intell. CI 2006*, pp. 116–120, 2006.

21. Tornea, O., Borda, M.E. (2009). DNA Cryptographic Algorithms. In: Vlad, S., Ciupa, R.V., Nicu, A.I. (eds) International Conference on Advancements of Medicine and Health Care through Technology. IFMBE Proceedings, vol 26. pp. 223–226, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04292-8_49

22. Sabry, M., Hashem, M., Nazmy, T., (2012). Three Reversible Data Encoding Algorithms based on DNA and Amino Acids' Structure, *Int. J. Comput. Appl.*,54(8): 24–30, doi: 10.5120/8588-2339.

23. Wang, X., Zhang, Q., "DNA computing-based cryptography", *BIC-TA 2009 - Proceedings,*

2009 4th Int. Conf. Bio-Inspired Comput. Theor. Appl., 2009, pp. 67–69, doi: 10.1109/BICTA.2009.5338153.

24. Akanksha, A., Bhopale, A., Sharma, J., Meer Shizan, A., Gautam, D., (2012), Implementation of DNA algorithm for secure voice communication, *Int. J. Sci. Eng. Res.*, 3(6):1–5

25. K. Ning, (2009). "A Pseudo DNA Cryptography Method", *09032693*, Available at: http://arxiv.org/abs/0903.2693. (Last accessed: 22 December 2024)