# DNA Cryptography

Subjects: Computer Science, Information Systems
Contributor: kareem ahmed , Dina Ibrahim , Mohammed Masaod , Abd Elmgeid Amin Ali

Encryption, also known as cryptography, is the most widely used approach to achieve confidentiality and verify integrity. It is also the main way to secure governments, social data and personal transactions, since all of these entities require a great amount of protection against known/unknown attacks and threats.

DNA cryptography image encryption chaotic

## 1. Introduction

Information security is one of the most important fields that has been gaining more attention for data protection, privacy preservation and data leakage prevention. The process of securing information systems includes monitoring and analyzing all data flows and behaviors of different components that comprise the system, such as people that interact with the system; hardware, such as network infrastructure, routers and domain name servers; and software configurations and keys. It is a necessity to ensure that the system is fully functioning while protecting its components from known and unknown threats. The basic requirement for any security system is to achieve the following requirements: integrity, confidentiality and availability. Other requirements also include accountability and authentication. Encryption, also known as cryptography, is the most widely used approach to achieve confidentiality and verify integrity. It is also the main way to secure governments, social data and personal transactions, since all of these entities require a great amount of protection against known/unknown attacks and threats. In this sense, cryptography is the most fundamental building block in the design of secure information systems. As a result, it has attracted the main interest of many researchers, allowing thousands of competing research articles to offer new ways for fast and robust encryption.

Cryptography is divided into two broad categories: symmetric cryptography and asymmetric cryptography. Symmetric cryptography uses a single key for both encryption and decryption, also known as single-key cryptography. Some examples are the Advanced Encryption Standard (AES), Data Encryption Standard (DES), 3DES and Rivest Cipher 4 (RC4). Asymmetric cryptography, also known as public-key cryptography, uses one key for encryption and another key for decryption. Some examples of public key cryptography are the Rivest–Shamir–Adleman (RSA) algorithm, Elliptic Curve Cryptography (ECC) and the Diffie–Hellman key agreement protocol.

Because of the wide spread of wireless networks, LANs, Bluetooth, and other forms of communication media, digital images should be protected to maintain privacy and copyrights. Images may be vulnerable to different types of attacks during transmission or exchange between authenticated users. Some common attacks related to networks are network denial, masquerading and eavesdropping. For any communication medium to be trusted between parties, it should satisfy reliability, anonymity, confidentiality, integrity and availability. When exchanging digital images, common ways for securing these images are cryptography and steganography. Cryptography not only guarantees confidentiality, but it can also address other issues such as data integrity, authentication and non-repudiation. With cryptography, information can be sent securely, and only the authenticated receiver can retrieve this information, thus avoiding the compromising or leakage of private data.

The RGB model is a widely used model for representing and storing digital color images. In this model, the image is represented by three independent channels. Each channel is a 2D matrix. The first matrix represents the red color, the second matrix represents the green color and the third matrix represents the blue color. The three matrices are of the same width and height. Each pixel in the color image is represented by three values that are the three corresponding values on the three matrices. Each value can range from 0 to 255. In this way, the RGB model can represent more than 16 million distinct colors. Many recent scientific papers have shown that combining DNA cryptography and chaotic sequences leads to very efficient and promising results with respect to nonlinearity, randomness and resistance to common attacks. Arthi et al. [1] presented an encryption technique based on 4D Lorenz hyper-chaos and DNA encoding. First, an integer wavelet transform IWT is applied to produce approximation and detail bands of the input image. Then, the LL band is permuted using chaotic sequences, and then they are encoded using DNA. Finally, an operation called DNA-XOR is performed to produce the encrypted image. Paul [2] presented a color image encryption technique that uses a 2D hyperchaotic map to scramble the initial image and then uses a logistic-tent map to produce a cover image. Then, SHA-2 is applied to this cover image to produce a mask image, and the two images are encoded using DNA, followed by many diffusions to produce cipher image.

## 2. DNA Cryptography

DNA cryptography is considered one of the most appealing directions in cryptography. Huge amounts of parallelism and exceptional data thickness make the attributes of DNA suitable for being used in cryptography applications, such as encryption, confirmation and digital signatures [3]. DNA cryptography is a broad category in cryptography, which involves improved encoding and a great amount of randomness. DNA cryptography uses some encoding techniques to convert binary streams into DNA sequences, and then it uses different concepts related to DNA, such as nucleotides, proteins, amino acids and complementary rules to perform encryption, decryption and key generation. DNA steganography is another interesting field, in which secret data are embedded into the biological structure of natural DNA. First, the secret data are transformed into a DNA sequence. Then, this sequence is merged with a real DNA pattern in such a way; therefore, the output of this operation is a natural DNA sequence. DNA is available in the NCBI database (National Center for Biotechnology Information). DNA cryptography is always used along with DNA steganography to further protect sensitive data. In this case, data is first encrypted and then hidden inside a DNA sequence. DNA cryptography depends on DNA calculations, as the motive for its use is to simulate natural tasks occurring in DNA sequences.

Deoxyribonucleic acid (DNA) (https://www.msdmanuals.com/home last accessed on 7 June 2022) is a complex chemical compound found in all organisms and is of great importance because it contains all necessary information to generate all kinds of proteins in a cell. DNA consists of two linked strands that wrap around each other and is called a double helix. DNA is composed up of four base units called nucleotides: adenine (A), thymine (T), guanine (G) and cytosine (C). The nucleotides in the first strand are connected with the nucleotides in the second strand to form chemical bonds, called base pairs. Adenine (A) is connected only to thymine, and vice versa. Guanine (G) is connected only to cytosine (C), and vice versa. A gene is a short segment of DNA, and it contains specific genetic information because it contains the code for specific protein that performs some function in one or more types of cells in the human body. There is another important concept called a chromosome, which is a structure existing inside a cell’s nucleus, and chromosomes contain genes. A chromosome may contain hundreds or thousands of genes.

As mentioned before, genes contain DNA. DNA is a blueprint (code) which is responsible for synthesizing a protein. Proteins are the basic building blocks for skin, muscles and other parts of the human body, but the importance of proteins is that they are responsible for making enzymes, which are responsible for all chemical processes and reactions inside the human body. The body is controlled by thousands of enzymes. Therefore, the amount and type of proteins control the functions and structures of the whole body. In this sense, the synthesis of proteins is the most important operation that affects all aspects regarding the structures and functions of the human body.

Proteins are long chains of amino acids connected to each other. The instructions for synthesizing proteins are stored in DNA, where nucleotides inside DNA are arranged in groups. Each group contains three consecutive nucleotides. Each group is also called a codon, and a codon is a group of three adjacent nucleotides. Each codon is considered an instruction to add a specific amino acid to the protein chain. For example, the GCT codon is an instruction to add the amino acid alanine. Therefore, the sequence of amino acids inside a protein chain is specified by the order of the codons in the gene for that protein. The process of turning DNA into a protein is caried out by two subprocesses called transcription and translation. In transcription, the instructions in DNA are copied (transcribed) to ribonucleic acid (RNA). RNA is similar to a strand of DNA, but thymine (T) is replaced by uracil (U). RNA typically contains codons of three bases, similar to DNA. **Table 1** shows the binary representations of mRNA codons. In translation, the mRNA that is obtained from DNA is translated into an amino acid [4]. The main differences between DNA and RNA are that the sugar in DNA is deoxyribose, whereas the sugar in RNA is ribose. Moreover, RNA uses uracil (U) instead of thymine (T). RNA codons are codons that are used during a process called transcription, but each mRNA molecule acquires its sequence of nucleotides by transcription from the corresponding gene of DNA.

**Table 1.** Binary representations of mRNA Codons.

| | A | U | C | G | |
|---|---|---|---|---|---|
| A [00] | AAA [000000] | AUA [000100] | ACA [001000] | AGA [001100] | A |
| | AAU [000001] | AUU [000101] | ACU [001001] | AGU [001101] | U |
| | AAC [000010] | AUC [000110] | ACC [001010] | AGC [001110] | C |

| | A | U | C | G | |
|---|---|---|---|---|---|
| | AAG [000011] | AUG [000100] | ACG [001011] | AGG [001111] | G |
| U [01] | UAA [010000] | UUA [010100] | UCA [011000] | UGA [011100] | A |
| | UAU [010001] | UUU [010101] | UCU [011001] | UGU [011101] | U |
| | UAC [010010] | UUC [010110] | UCC [011010] | UGC [011110] | C |
| | UAG [010011] | UUG [010111] | UCG [011011] | UGG [011111] | G |
| C [10] | CAA [100000] | CUA [100100] | CCA [101000] | CGA [101100] | A |
| | CAU [100001] | CUU [100101] | CCU [101001] | CGU [101101] | U |
| | CAC [100010] | CUC [100110] | CCC [101010] | CGC [101110] | C |
| | CAG [100011] | CUG [100111] | CCG [101011] | CGG [101111] | G |
| G [11] | GAA [110000] | GUA [110100] | GCA [111000] | GGA [111100] | A |
| | GAU [110001] | GUU [110101] | GCU [111001] | GGU [111101] | U |
| | GAC [110010] | GUC [110110] | GCC [111010] | GGC [111110] | C |
| | GAG [110011] | GUG [110111] | GCG [111011] | GGG [111111] | G |

This process occurs in the nucleus. In this process, information in DNA is copied into the mRNA. This process is conducted in three stages. (1) Initiation: This phase occurs when the enzyme RNA polymer is associated with DNA in a place called a promoter. Consequently, the enzyme is able to read DNA in one part and form the RNA strand. (2) Elongation: At this stage, the RNA strand elongates by adding nucleotides. The RNA polymer reads a DNA part and forms mRNA using a complementary base, and mRNA uses uracil (U) instead of thymine (T). (3) Termination: This is the last stage of transcription and occurs when the mRNA is completed and separated from the DNA.

A genetic code can be defined as the relationship between the nucleotide sequences of genes and the amino acid sequences of proteins, which is determined by the rules of translation. It consists of three-letter ‘words’ called codons, formed from a sequence of three nucleotides (e.g., UCA or GAC). Each codon specifies one amino acid in a protein. It can be expressed as either a DNA codon or an RNA codon. The process of protein production or translation occurs after the transcription process. **Table 2** shows all possible codons and their corresponding amino acids. There are 20 different amino acids that can be used in protein synthesis. The protein-building process is stopped when the process arrives at the three “stop” codons (UAG, UAA, UGA) [5].

**Table 2.** Sixty-four different codons and their corresponding amino acid [4].

| | | U | | C | | A | | G | |
|---|---|---|---|---|---|---|---|---|---|
| U | U | UUU | Phe | UCU | Ser | UAU | Tyr | UGU | Cys |

| | | U | | C | | A | | G | |
|---|---|---|---|---|---|---|---|---|---|
| | C | UUC | | UCC | | UAC | | UGC | |
| | A | UUA | Leu | UCA | | UAA | Stop | UGA | Stop |
| | G | UUG | | UCG | | UAG | | UGG | Trp |
| C | U | CUU | Leu | CCU | Pro | CAU | His | CGU | Arg |
| | C | CUC | | CCC | | CAC | | CGC | |
| | A | CUA | | CCA | | CAA | Gln | CGA | |
| | G | CUG | | CCG | | CAG | | CGG | |
| A | U | AUU | Ile | ACU | Thr | AAU | Asn | AGU | Ser |
| | C | AUC | | ACC | | AAC | | AGC | |
| | A | AUA | | ACA | | AAA | Lys | AGA | Arg |
| | G | AUG | Met | ACG | | AAG | | AGG | |
| G | U | GUU | Val | GCU | Ala | GAU | Asp | GGU | Gly |
| | C | GUC | | GCC | | GAC | | GGC | |
| | A | GUA | | GCA | | GAA | Glu | GGA | |
| | G | GUG | | GCG | | GAG | | GGG | |

The rest of this section presents one of the most used rules in DNA cryptography, which is the DNA complementary base pairing rule. As mentioned before, DNA consists of two strands of nucleotides called a double helix. The nucleotides are cytosine (C), guanine (G), adenine (A) or thymine (T). They are bound by hydrogen bonds. For example, A with T includes two hydrogen bonds, and C with G includes three hydrogen bonds. This chemical structure of the bases is called the Watson–Crick complementary base pairing principle. Arranging these rules helps to calculate and store information. The encoding rule is the following: (A-00), (C-01), (G-10) and (T-11), and the complementary numbers of them are (00-11), (01-10), (10-01) and (11-00). There are eight cases of combinations that satisfy complementary base pairing, as shown in **Table 3** [6].

**Table 3.** DNA encoding rules.

| Rule | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 00 | T | T | G | C | G | C | A | A |
| 01 | G | C | T | T | A | A | G | C |
| 10 | C | G | A | A | T | T | C | G |
| 11 | A | A | C | G | C | G | T | T |

## References

1. Arthi, G.; Thanikaiselvan, V.; Amirtharajan, R. 4D Hyperchaotic map and DNA encoding combined image encryption for secure communication. Multimedia Tools Appl. 2022, 81, 15859–15878.
2. Paul, L.; Gracias, C.; Desai, A.; Thanikaiselvan, V.; Shanthini, S.; Rengarajan, A. A novel colour image encryption scheme using dynamic DNA coding, chaotic maps, and SHA-2. Multimed. Tools Appl. 2022.
3. Babae, M. A novel text and image encryption method based on chaos theory and DNA computing. Nat. Comput. 2013, 107, 101–107.
4. Ahmed, K.; El-Henawy, I. Increasing robustness of data encryption standard by integrating DNA cryptography. Int. J. Comput. Appl. 2017, 39, 91–105.
5. Chen, J. A DNA-based biomolecular cryptography design. In Proceedings of the 2003 International Symposium on Circuits and Systems, ISCAS'03, Bangkok, Thailand, 25–28 May 2003; Volume 3.
6. Liu, H.; Wang, X.; Kadir, A. Image encryption using DNA complementary rule and chaotic maps. Appl. Soft Comput. 2012, 12, 1457–1466.