JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MODERN HIGHLY ENHANCED DNA BASED STEGANOGRAPHY APPROACH TOWARDS INFORMATION SECUTIRY

[1]Ms. Sumi M, [2]Subitha P
[1]Assistant Professor, [2]MCA Scholar
[1]Department of MCA
[1]Nehru College of Engineering and Research Centre, Pampady, India

Abstract: The world today is concerned with the security of digital data that is being transmitted. Digital data security is achieved through a variety of techniques. Cryptography and steganography are the two most popular techniques used in the field of cyber security, in which mathematics and logic are applied to design strong encryption methods. The characteristics of biological sequences make them suitable for use in digital data security procedures. DNA computing has been identified as a possible technology in the fields of Cryptography, Steganography and Authentication. DNA computing is a relatively new form of computing that, instead of using silicon-based technology, utilizes the abilities of the DNA molecule and biochemistry. It is making use of DNA characteristics for extremely parallel computation. DNA cryptography make use of DNA molecules as information carriers. DNA provides the most reliable personal identification. In order to construct concealed messages and avoid hacking, DNA steganography methodology was developed to create hidden messages in variable areas (single nucleotide polymorphisms) of the genome. This paper offers a brief explanation of the DNA-based data security method.

IndexTerms - Digital data security, cryptography, Steganography, authentication, DNA computing, silicon-based technology, biochemistry, DNA Cryptography, DNA steganography

I. INTRODUCTION

In the present era of e-business and e-commerce, information security is becoming increasingly crucial, requiring a high level of security and more powerful data protection methods to secure the confidentiality, integrity, and availability of shared information as well as of transmitted data. The potential of thieves getting private data and the transfer capabilities worries data transmission professionals a lot. Steganography and cryptography are the two techniques that are most frequently utilized in the field of cyber security because they integrate logic and mathematics to produce effective encryption systems. Due to the increasing use of the internet, the significance of these fields has substantially expanded in the current period.

A promising method for authentication, steganography, and cryptography is DNA computing. A relatively recent form of computer called DNA computing uses the DNA molecule and biology as opposed to silicon-based technologies. For massively parallel computation, DNA's unique properties are used. Parallel search has the potential to solve enormous issues if the proper architecture and DNA are used.

DNA molecules that have the ability to store, process, and transmit data are utilized as information carriers in the emerging subject of DNA cryptography. In order to make the signals concealed in different areas of the genome, DNA steganography methodology was developed (single nucleotide polymorphisms). Deoxyribonucleic acid (DNA) offers the most reliable personal identification method out of all the biometric technologies now in use [1]. These DNA concepts ensure non-vulnerable data transmission and provide new hope for unbreakable algorithms.

1.1 CRYPTOGRAPHY

Data encryption and decryption over open networks are two topics covered by cryptography. Confidential information is secured and protected via cryptography, a procedure that scrambles and transforms the information into an unreadable format. Before being sent to the recipients, the sender's secret message is encrypted in cryptography using a secret key and an encryption method. The recipients decrypt the message using the secret key and the proper decryption technique. An unauthorized user won't be able to extract the secret message without the secret key. Cryptography and cryptanalysis work together. Analysis and attempts to compromise the security systems put forth by the cryptography field are the objectives of cryptanalysis. In other words, the degree to which a cryptographic system is susceptible to cryptanalysis determines its level of strength. Several technologies (such RSA, ECC, etc.) have been developed to achieve a high level of security. Electrical engineering, computer science, and mathematics are all used in modern cryptography. Designing cryptographic algorithms around computational hardness presumptions that are thought to be challenging for an adversary to crack [2] is a very scientific approach.

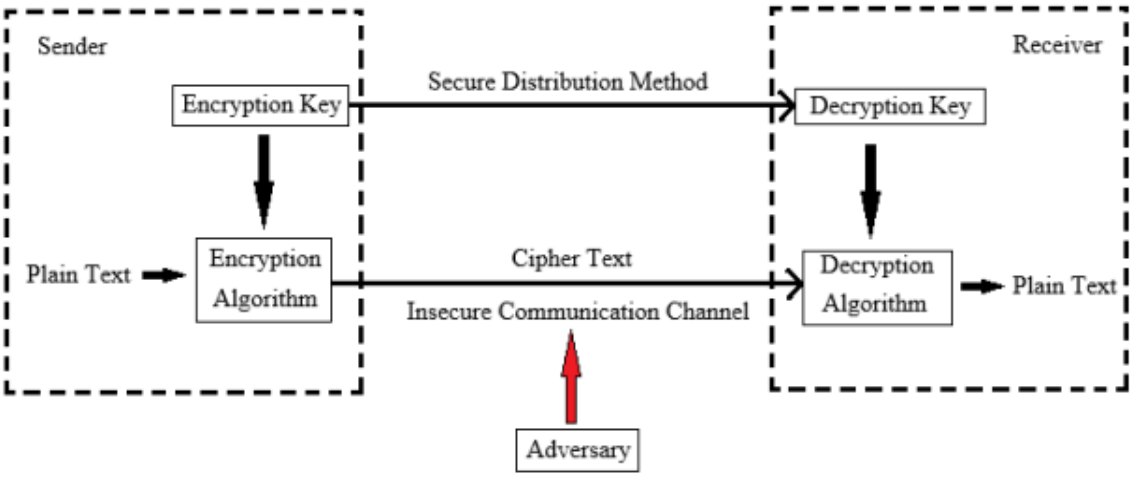

Fig1: cryptosystem

1.2 STEGANOGRAPHY

A confidential communication can be hidden by using steganography within the cover media. Greek terms steganos, which means hidden, and graphein, which means to write, are the roots of steganography. A thorough explanation of the steganographic procedure:

Cover medium + Hidden data + Keystego = Stegomedium

The file containing the data, or hidden data, is the cover medium. Another key called keystego can also be used to encrypt the generated file. The stego medium is the last file that will be transferred. The covering file's medium may, as is typical, include an audio or image file. Several steganography techniques have been introduced, depending on the cover material. Among these, text steganography, audio steganography, video steganography, and image steganography are the most popular [3]. The ability to conceal using the steganography methods mentioned above is quite constrained. With DNA steganography, a cover medium's ability to conceal information is improved. Although cryptography seeks to make data unreadable by a third party, steganography aims to hide the data from a third party. The cover file (X) and secret message (M) are fed into the steganographic encoder. The secret message is embedded in a cover file by the Steganographic Encoder function, f(X, M, K). The Stego Object that results from this will look exactly like your cover file. Coding has finally been finished. To obtain the hidden message, a steganographic decoder is fed a stego object.

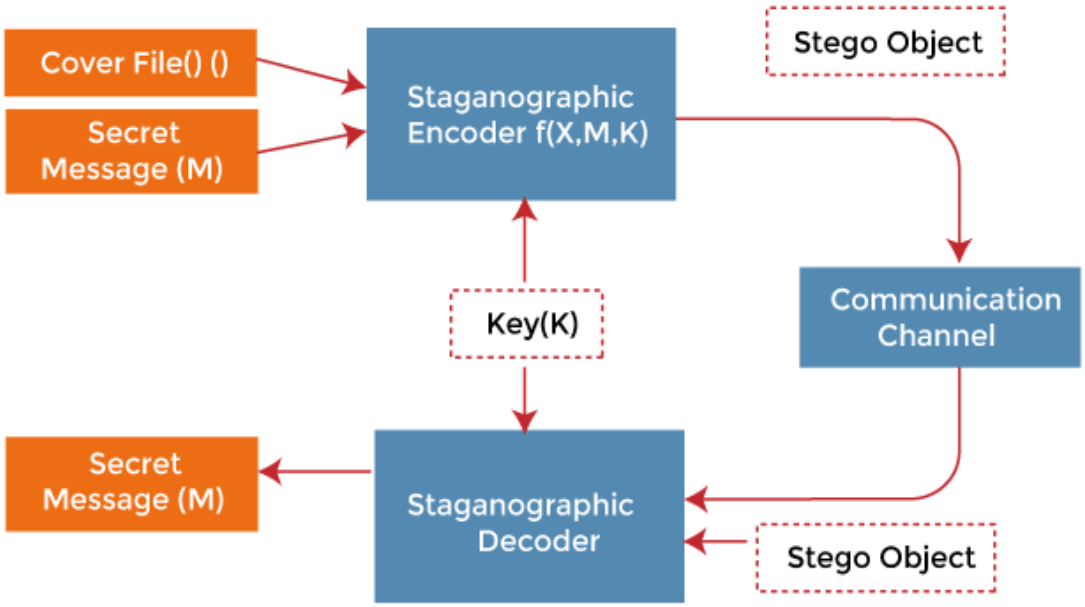

Fig2: steganography

II. BIOLOGICAL BACKGROUND

2.1 DNA STRUCTURE

The genetic material found in the cells of all living things is known as deoxyribonucleic acid, or DNA, and it contains the genetic instructions needed for survival, reproduction, and evolution. A nucleotide composed of a phosphate group, a sugar group (deoxyribose), and a nitrogen base serves as an information carrier in fully developed living things. DNA is a double helix-shaped molecule made up of sugar and phosphate groups that is divided into two long, anti-parallel (opposite polarity) strands. The linkages between the bases connect the two strands. Adenine (A), thymine (T), guanine (G), and cytosine (C) are the four different types of nitrogen bases (C) [4]. Each DNA strand has two distinct endpoints that determine its polarity: the 3'end and the 5'end. In the chemical structure of DNA, two linear sequences of bases form a specific bond. This bond follows the complementarity principle: adenine always bonds with thymine via double hydrogen bonds (A-T) and vice versa (T-A), and cytosine always bonds with guanine via triple hydrogen bonds (C-G) (G-C). Because a codon is composed of three adjacent nucleotides, there are 4^3 = 64 possible codon combinations. The shape and function of the resulting protein in living organisms are determined by the configuration of these combinations. DNA stores all of the extensive and intricate information about an individual using only the letters A, C, T, and G.

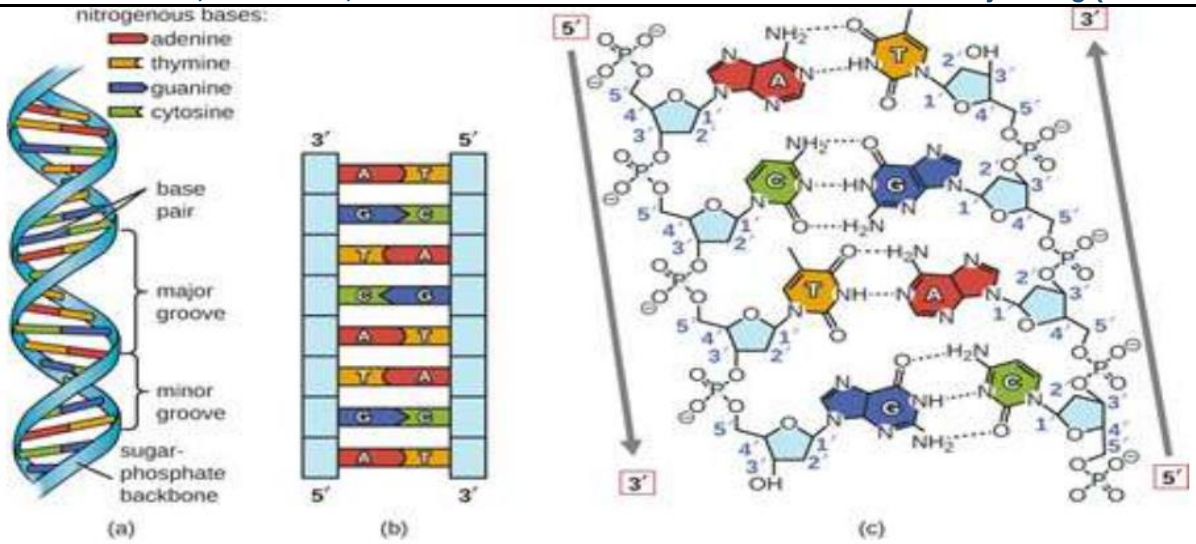

Fig3: basic structure of DNA

2.2 DNA COMPUTING

In place of conventional silicon-based computer technologies, DNA computing makes use of molecular biology, biochemistry, and other biological processes [5]. It acknowledges that biomolecules are the basic building blocks of electronic devices. DNA computing or molecular Computing makes use of DNA's natural combinational capabilities for extremely parallel processing. This implies that you can try every solution to a given problem until you randomly calculate the best one. DNA can perform this type of computation considerably more quickly than a normal computer, where significant parallelism would call for a lot of hardware, not just additional DNA.

A, T, C, and G, four distinct molecules, serve as the genetic coding for DNA. When these four "bits" are linked together, they have a huge storage capacity. These four molecules were combined in a test tube and spontaneously self-assembled into DNA strands. Each DNA strand may stand in for a distinct travel path for the salesman if some combination of these molecules represented a city and a flight path. This calculation would take place simultaneously as the DNA strands assembled themselves in parallel.

2.3 ORIGIN OF DNA COMPUTING

In 1994, Leonard Adelman was the first to suggest utilizing DNA to solve difficult mathematical problems. Adelman held a position at Southern California University as a computer scientist. The famous directed Hamilton Path problem, often known as the "travelling salesman" problem in mathematics and computer science, was addressed in an essay he published in the science magazine, revealing how DNA may be utilized to address the issue. The objective of the task is to determine the shortest path between a set of cities that passes through each just once. The goal was to find a path from start to finish that only passes through all of the points exactly once. For conventional (serial logic) computers, this problem is difficult because each path must be tried one at a time. Adleman demonstrated that DNA can be assembled in such a way that a test tube full of DNA building blocks could assemble itself to encode every conceivable path in the travelling salesman problem at the same time [6].

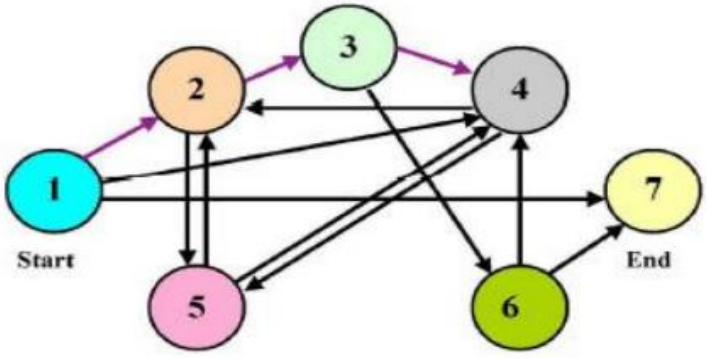

Fig4: graph in Adleman's experiment

The following algorithm solves the problem:

1. Generate random paths through the graph.
2. Only pathways that start with Vin and end with Vend should be considered.
3. If the graph has n vertices, then keep only those path that enter exactly n vertices.
4. Keep only those paths that enter all of the vertices of the graph atleast once.
5. Say "YES" if any paths are still open, "NO" otherwise.

To implement step 1, each vertex of the graph was encoded into a random 20-nucleotide strand. Then the product of step 1 was amplified by polymerase chain reaction (PCR). Step 3 was implemented using technique called gel-electrophoresis. By employing an iterative technique known as affinity purification, step 4 was completed. The existence of a molecule that encodes a Hamiltonian path was then verified.

2.4 ADVANTAGES OF DNA COMPUTING

1. Speed: As Adleman demonstrated, DNA strands performed computations equivalent to 9, 10, or better, arguably more than 100 times faster than the fastest computer. A conventional computer can perform approximately 100 MIPS.

2. Minimal storage requirements: Unlike traditional storage media, which requires 12 10 cubic nanometers to store a single bit, DNA can store memory at a density of about 1 bit per cubic nanometer.
3. Minimal power requirements: When the calculation is running, no external electrical power is required for DNA computing. The chemical bonds in DNA form naturally without the assistance of any external energy source.
4. Dense information storage: One gram of DNA can store approximately 1x10^14 MB of data.
5. Parallel computing: The most difficult problems can be solved by Massively parallel DNA computers can quickly and efficiently answer even the most challenging tasks in a matter of weeks.

III. CRYPTOGRAPHY BASED ON DNA COMPUTING

3.1 DNA RULES

DNA cryptography is more advanced, Because of the high processing power of DNA molecules. It makes use of DNA pairs as information carriers. DNA Cryptology combines cryptology and modern biotechnology. DNA can be used in cryptography for data storage and transmission, as well as computation. DNA cryptography utilizes the power of DNA to hide the transmitting data. As a result, DNA Cryptography can be defined as the concealment of data in terms of DNA Sequence.00, 01, 10 and 11 are encoded by 4 bases A, C, G, and T. Of the 4! = 24 types of coding rules, only 8 rules satisfy the Watson-Crick complementary rule. The rule that used for encrypting the data during the encryption process must be used for decrypting the data in the decryption side.

Table1: DNA rules satisfying Watson-Crick complementary rule

Rules	Rule1	Rule2	Rule3	Rule4	Rule5	Rule6	Rule7	Rule8
00	A	A	G	G	T	T	C	C
01	C	G	A	T	C	G	A	T
10	G	C	T	A	G	C	T	A
11	T	T	C	C	A	A	G	G

3.2 DNA OPERATIONS

Once the data is encoded, all the information is available as combinations of A, C, G and T. During the encryption process, DNA addition is performed using the rules given in the table 2. During the decryption process, DNA subtraction is performed using the rules that are given in the table 3. A second method for encryption is using DNA XOR operation. In this case, for both the encryption and decryption process, DNA XOR operation is used.

Table2: DNA addition

++	C=00	T=01	A=10	G=11
C=00	C	T	A	G
T=01	T	A	G	C
A=10	A	G	C	T
G=11	G	C	T	A

Table3: DNA subtraction

--	C=00	T=01	A=10	G=11
C=00	C	G	A	T
T=01	T	C	G	A
A=10	A	T	C	G
G=11	G	A	T	C

Table4: DNA xor

XOR	C=00	T=01	A=10	G=11
C=00	C	T	A	G
T=01	T	C	G	A
A=10	A	G	C	T
G=11	G	A	T	C

3.3 TRADITIONAL CRYPTOGRAPHY V/S DNA CRYPTOGRAPHY

Security is represented by encryption or ciphering text, which is the process of converting plain text into encrypted, non-recordable text. The purpose of cryptography is to conceal information. Cryptography must be used to ensure the security and confidentiality of sensitive information. Traditional cryptography secures and stores data using complex mathematical procedures. A DNA cryptosystem emerges to avoid this complexity. The DNA cryptosystem is a novel approach to data encryption based on DNA sequence [7].

Table5: comparison between current cryptography and DNA cryptography

Basis	Traditional Cryptography	DNA Cryptography
Ideal System	Silicon chip based	DNA chip based
Information Storage	Silicon computer chips	DNA strands
Storage Capacity	1 gram silicon chip contain 16 Megabytes.	1 gram DNA chip contain 10^8 Terabytes.
Processing Time	Slow	Fast
Performance Dependency	Implementation and system configuration	Environmental conditions

IV. OBJECTIVE

1. Apply the DNA steganography methodology to hide secret messages in a genome.
2. Preserving the original functionality of the reference DNA
3. Using a double hiding layer technique that ensures more security
4. Lowering the cracking probability
5. Lowering the probability of knowing the data inside the DNA sequence by an adversary

V. METHODOLOGY

The analysis of digital data security based on DNA computing is the method adopted in this research. The highly improved DNA steganography approach conceals data inside a DNA sequence using many layers of encryption, offering higher levels of concealment and greater security [6].

5.1 DNA STEGANOGRAPHY

In DNA steganography, the molecular sequence of the DNA is used as a cover medium. The fundamental idea behind DNA steganography is to use a random DNA sequence as the cover medium, encrypt the secret message inside of it, and then deliver the altered DNA sequence to the intended receiver. The secret message is extracted from the changed DNA sequence by the receiver using the appropriate decryption technique.

Table6: DNA digital coding

DNA Nucleotide	Decimal	Binary
A	0	00
C	1	01
G	2	10
T	3	11

5.2 ENCRPTION

The algorithm employs two unique keys. The final letter in the message (M) is XORed with the character that comes before it in the message (M), and so on. The first key (K1), which is an integer in the range of 0 to 255, is used to perform this operation. As a result, the message is encrypted using the first key. The DNA sequence is divided into identical-length segments using the second key (K2), which is produced at random. At the start of each segment, the generated cypher characters are inserted as binary bits one by one. Following that, Table 2 is used to translate the binary sequence into DNA bases. The second key should be a small number so that the DNA sequence is as short as possible while still concealing the secret message. The encryption process follows the given steps,

1. Divide M into characters, M = m1, m2, m3….,mn and convert each character to its 8-bit binary equivalent based on the ASCII standard.
2. K1 is formed by randomly generating a number between 0 and 255, and the key is then converted into an 8-bit binary sequence.
3. The final character in the message M is XORed with K1.
4. The result is XORed with the character preceding the last one in M, and the process is repeated until all of the characters have been converted and stored in A.
5. A protein sequence is created from the binary sequence A.
6. Table 6 is used to convert a random sample DNA sequence S into a binary bit sequence.
7. Create a random number, preferably one that is small, K2, and then divide the DNA sequence S into segments, with each segment's length being K2.
8. Insert the first binary value of A into the first DNA binary segment, and the second binary value of K1 should go into the second binary segment, and so on.
9. Create a fake DNA sequence by concatenating all the binary sequences and then converting it.

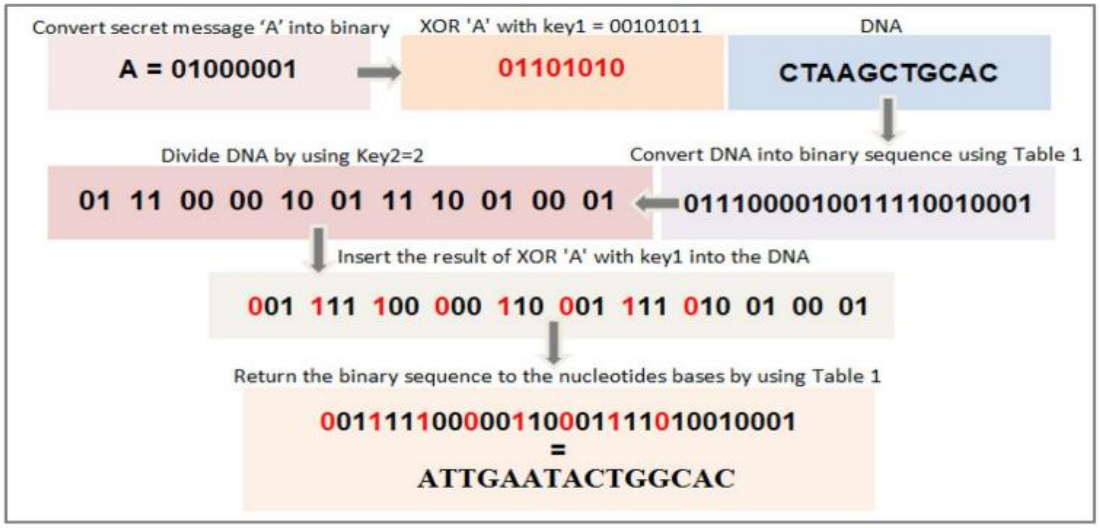

Fig5: encryption

5.3 DECRYPTION

K1 and K2 must be understood by the recipient in order to decrypt the message. Also, the sender must send the recipient the original DNA. On receiving the cyphertext from the sender side, the receiver performs the given steps,

1. Use Table 6 to transform the incoming fake DNA sequence into a binary sequence.
2. The binary DNA sequence will be divided into segments, each of which will have a size equal to K2 + 1.
3. Concatenate the first bit from each segment to create significant bits B
4. B's first 8 binary bits should be XORed with K1 before B's next 8 bits should be XORed with B's prior 8 bits, and so on.
5. Transform the DNA sequence's binary bits into ASCII text values.

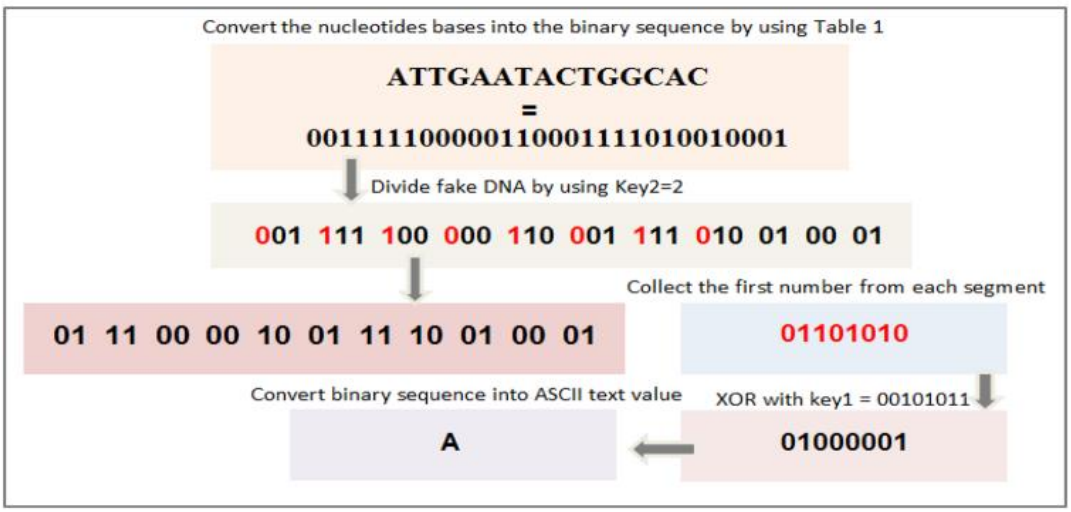

Fig6: decryption

VI. RESULT ANALYSIS

The size of the reference DNA is around 163 million. Hence, the likelihood of predicting the reference DNA sequence is 1/(1.63×10^8). The likelihood of correctly predicting the various binary coding (A, C, G, and T) combinations is 1/24. The likelihood of discovering the message and reference DNA sequence is 1/(n−1), where the number n denotes the number of bits in the faked DNA sequence. A random key is used to segment the message and DNA; As a result, the likelihood of correctly predicting the message's segmentation is 1/2^m-1 since the secret message is having a bit length of m. The likelihood of correctly predicting DNA segmentation is 1/2^(s-1), where the letter s denotes the number of bits in the reference DNA sequence. The XOR operation is used to encode data within the DNA sequence, and the probability of the XOR combination is calculated to be ½^8m. [8] As a result, the likelihood of discovering the message hidden in the DNA sequence is:

(1/1.63*10^8) * (1/24) * (1/n-1) * (1/2^m-1) * (1/2^s-1) * (1/2^8m)

VII. CONCLUSION

The research of data encryption in DNA sequences is a relatively new one. DNA computing is still in its early stages, and its applications are still not well understood. Information security professionals are always looking for uncrackable encryption to protect the data we communicate over the internet. On the other hand, it is envisaged that the uses of DNA authentication techniques would continue to expand as they have already demonstrated substantial potential in the market. The significantly improved DNA-based steganography technology includes a number of features. The extremely low probability that a third party will be aware of the data contained in the DNA sequence is one of the biggest advantages. It is challenging to separate the ciphertext from a lengthy list of DNA sequences.

Because of its immense parallelism and storage capacity, DNA has been intended to have benefits over its traditional counterparts in the fields of cryptography, security, and data encryption. With the development of biotechnology and the finding of a superior DNA encryption design, the study of DNA cryptography in information security will certainly advance.

DNA cryptography still needs to be studied more in order to reach the limits of cutting-edge technology. Even employing Quantum Computing, it would be possible to break the current record for the length of time it takes to decrypt data by combining DNA Cryptography with other conventional techniques. The complexity of a cypher key generated by hybridizing DNA cryptography and conventional methods would be much higher. The secret message is extracted from the changed DNA sequence by the receiver using the appropriate decryption technique.

REFERENCES

[1] Ofualagba Mamuyovwi Helen "DNA Computing Based on Information Security Technology", International Journal of Pure and Applied Science, Vol. 21 No.9 June, 2021.

[2] Mandrita Mondal, Kumar. S. Ray "Review on DNA Cryptography" 15 March 2019.

[3] Partha Saha, Lubna Yasmin Pinky, Mohammad Ashraful Islam, Papia Akter "Higher Payload Capacity in DNA Steganography using Balanced Tree Data Structure", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019.

[4] Haval I. Hussein, Wafaa M. Abdullah, "A Modified Table Lookup Substitution for Hiding Data in DNA", 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018.

[5] Omar Haitham Alhabeeb "A Review of Modern DNA-based Steganography Approaches" International Journal of Advanced Computer Science and Applications 06 November 2021 Vol. 12(No. 10):184-196 DOI:10.14569/IJACSA.2021.0121021.

[6] O.A. Al-Harbi, W.E. Alahmadi, and A.O. Aljahdali, "Security analysis of DNA based steganography techniques" SN Applied Sciences, 2 (2), 2020.

[7] Omar G. Abood, Shawkat Guirguis, "DNA Computing and Its Application to Information and Data Security Field: A Survey", International Journal of Academic Engineering Research (IJAER) ISSN: 2000-001X Vol. 3 Issue 1, January – 2019, P.

[8] A. Khalifa, "A secure steganographic channel using DNA sequence data and a bio-inspired XOR cipher," Information (Switzerland), 12 (6), 2021.