Liquid Steganography-Secret Message as Living Cells Using DNA Computing

Vanathi P[1], Rashmika NRS[2], Poorani S[3], Vaishnavi R[4], Vedha A[5]
[1]*Assistant Professor, Department of Information Technology, Adhiyamaan College of Engineering*
[2, 3, 4, 5] *UG Scholars, Department of Information Technology, Adhiyamaan College of Engineering*

Abstract- Over the decades, DNA-primarily based totally steganography has emerged as a promising area for securing touchy statistics transmitted over an untrusted channel. Cryptography and Steganography are two methods used for secure information transmission. The process of data hiding involves using a carrier to conceal data such as images, video, and audio. Leaks of secret messages (SM) can have disastrous consequences, which has led to the development of numerous technologies for safeguarding SM. Steganography has been used for centuries and aims to keep SMs hidden. Throughout history, various types of steganography have existed, such as steganography ink, acrostic, wax-covered boards, and even messages written in hair by ancient Greeks. With the introduction of image, audio, and video technologies, steganography has evolved, and researchers have recently turned to using DNA to transmit and protect SMs. This project proposes Liquid Steganography, a scheme that uses a DNA-based steganography technique called data hiding using double DNA sequences to hide SMs in living cells. The cell can be stored and transmitted in liquid, paper, or plate form, disguising it as a drink, cosmetics, or another item. The key table and key pool are distributed in advance to both the sender and receiver in Liquid Steganography. The proposed system comprises two algorithms: substitution and recovery. The Substitution algorithm goes through several steps to encrypt and hide messages within a DNA sequence, while the Recovery Algorithm is used to retrieve the original secret message from the DNA sequence. This project is a preliminary investigation into a new model for hiding messages in living cells, and with DNA steganography being a new field, there are many opportunities for improving existing steganography and steganalysis approaches.

Keywords: Substitution technique, Recovery Algorithm, Playfair cipher technique.

1 INTRODUCTION

Due to the rapid advancement of technology, the importance of information security and confidentiality has become increasingly crucial. Therefore, high levels of security are required as a critical feature for successful networks. In various applications, such as annotation, ownership protection, copyrighting, authentication, and military, powerful data protection is essential, leading to a steady increase in research into data hiding techniques. Data hiding involves using a carrier to conceal data such as images, videos, and audios. In the proposed method, a DNA reference sequence is used as the carrier for data hiding.

1.1 DNA

DNA is a prolonged molecule that incorporates the specific genetic code for every individual. It holds the instructions for creating the proteins necessary for our bodies to function. The structure of DNA is that of a double helix, which is formed by two intertwined strands that appear twisted. Each strand of DNA is composed of a long sequence of nucleotides, which are individual units made up of

- a phosphate molecule
- a five-carbon sugar molecule called deoxyribose
- a region rich in nitrogen.

The four different nitrogen-containing regions are referred to as bases:

- adenine (A)
- cytosine (C)
- guanine (G)
- thymine (T).

The order of these four bases makes up the genetic code, which contains the instructions for life. The two strands of DNA are held together by bonds to form a ladder-like structure. A always pairs with T,

and G always pairs with C to form the "rungs" of the ladder. The sugar and phosphate groups contribute to the length of the ladder.

1.2 DNA Cryptography

Cryptography refers to the scientific discipline that deals with encoding information to conceal messages. DNA Cryptography is a rapidly evolving technology, which involves concealing data in the form of DNA sequences. DNA computing provides faster processing speed with minimal storage and power requirements. DNA has a memory storage density of approximately 1 bit per cubic nanometer, whereas conventional storage media requires 10^12 cubic nanometers per bit. During the computation process, no power is necessary. One gram of DNA contains 10^21 DNA bases, equivalent to 10^8 terabytes of data. Therefore, it has the potential to store all of the world's data in just a few milligrams. The nitrogenous base units of DNA are best encoded using four symbols:

A (0) – 00
C (1) – 01
G (2) – 10
T (3) – 11

The bases A and G form pairs, while T and C form pairs.

2 RELATED WORK

In the literature review section, we provided a brief overview of all relevant models and the closest competitor to our proposed study. Specifically, we focused on the most recent research papers published in the previous two years for this study.

1. Shyamasree C M and Sheena Anees proposed the three-level DNA-based Audio Steganography method [1].

The proposed method involves three levels of DNA-based encryption. In the first level, the Playfair algorithm based on DNA is utilized. In the second level, the secret message is hidden within a randomly generated DNA sequence. For the third level, DNA is embedded within an audio file. To convert the raw data of the secret file into a DNA sequence, DNA digital coding is applied using a binary coding scheme. The four nucleotides are encoded as A (00), C (01), G (10), and T (11). The resulting codons, which consist of three nucleotides, correspond to 20 amino acids. The Playfair encryption algorithm is then used to encrypt the amino acid sequence, which is hidden within a randomly generated DNA sequence using a two-by-two complementary rule. Finally, the embedded DNA sequence is hidden within the audio file using the Least Significant Bit (LSB) modification technique.

2. DNA sequencing has been proposed by Bama R, Deivanai S, and Priyadharshini K to ensure secure data authorization, storage, and transmission [2].

DNA sequencing for an electronic medical record system has the potential to revolutionize the way medical records are stored and accessed. By using DNA as a carrier for patient data, the system can provide secure and instant access to medical records, while also ensuring the privacy and confidentiality of patient information. The substitution method, which uses a binary coding scheme and complementary pair rule, is a powerful tool for encoding and decoding information in DNA. By keeping these schemes secret between the sender and receiver, the proposed DNA sequencing scheme can provide a high level of security and efficiency. Overall, the use of DNA sequencing for an electronic medical record system has many potential benefits, including increased security, efficiency, and accessibility of medical records. As generation continues to advance, innovative use of DNA in numerous fields can be clearly seen.

3. Pratik Pathak, Arup Kr. Chattopadhyay, and Amitava Nag proposed a location-based steganography technique [3].

The use of randomness to embed secret message bits in the audio file provides an additional layer of security, making it harder for unauthorized individuals to detect and extract the hidden information. The algorithm's complexity being O(n) means that the time required to embed the secret message bits increases linearly with the length of the message. Therefore, for longer messages, it may take more time to embed the secret bits. However, this scheme is still considered efficient as it is much faster than many other steganography techniques. Additionally, the high audio quality and lossless recovery mean that the embedded message can be retrieved without any degradation in the audio quality. Overall, this scheme provides a robust and secure method for hiding information within audio files.

4. The robust substitution technique was developed by Rohit Tanwar, Bhasker Sharma, and Sona Malhotra to implement audio steganography [4].
Using deeper layer bits for embedding helps in improving the resistance against intentional attacks because deeper layer bits are less likely to be noticed by attackers. Additionally, changing other bits willingly helps in reducing the distortion in the cover media caused by embedding the secret data, thereby improving the robustness of the technique. By addressing these two issues, the proposed technique can achieve a higher level of security and robustness.
5. K. Menaka proposed the indexing technique to hide the secret message inside the randomly generated DNA sequence [5].
The complementary rules based on Purine and Pyrimidines, based on Amino and Keto groups, and based on Strong and Weak H-bonds are utilized in this paper to encrypt the message. The message is first converted to a DNA sequence using digital coding patterns, and then the message index position in the fake DNA sequence is applied to each letter of the converted sequence. This paper suggests that DNA sequences have many properties that can be used for encryption, highlighting the potential of DNA cryptography as a secure and efficient method for data encryption.

3 METHODOLOGY

Using living cells as a medium for concealing secret messages adds an extra layer of security and makes it even more difficult for unauthorized individuals to access the hidden information. The researchers are proposing a method for using DNA as a stego medium to hide secret messages within living cells. They are using a clustered regularly interspaced short technique to hide the messages in different locations in the cells each time to make it more difficult to detect. The message is encoded into DNA code and then hidden in the genome using PMs. The receiver then uses PCR to extract the message and verify its authenticity using a hash value. The use of PCR primers and sequencing to extract the hidden message is a clever way to retrieve the information without damaging the living cells. This method could have many potential applications, such as in biotechnology or even espionage.

4 SYSTEM ARCHITECTURE

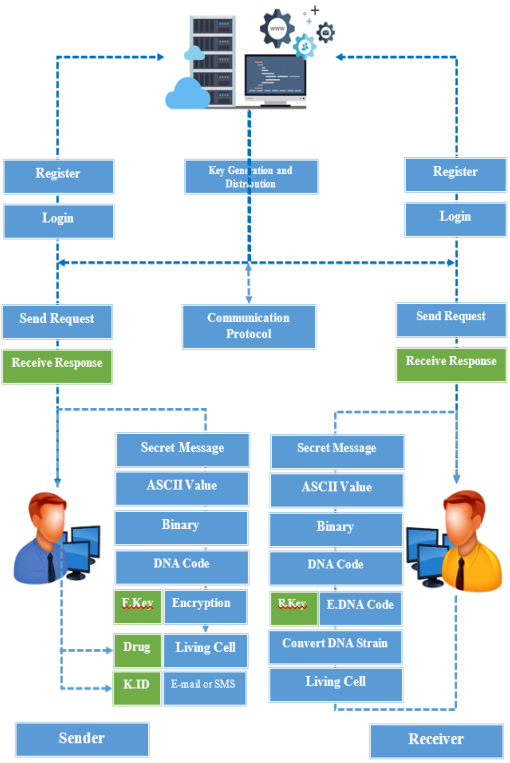


Figure 1 System architecture

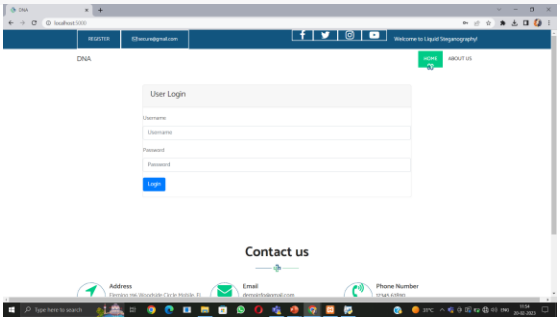
5 IMPLEMENTATION

The proposed work has five stages of implementation.
1. Web App for Liquid Steganography
2. User Management
3. Production and Distribution of Keys
4. Secret Communication Protocol
   4.1. Encryption of DNA
   4.2. Decryption of DNA
5. Transmission of Data

1. Web App for Liquid Steganography
In this module, we developed a web-based application that conceals sensitive information using DNA Steganography and allows for its secure transmission over any medium or channel. The Data Shielder applies DNA Steganography to embed the sensitive information into a randomly generated

DNA sequence. The embedded DNA sequence is then hidden within a cover image using LSB substitution. The resulting image is then sent back to the user through the web application. The user can then download the stego-image and transmit it over any medium or channel.

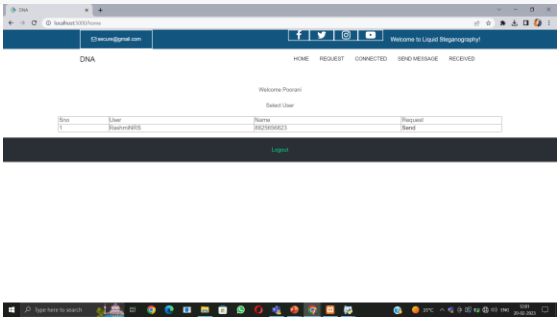
At the receiver end, the stego-image is uploaded to the web application, which then extracts the embedded DNA sequence using LSB substitution. The DNA sequence is decoded to reveal the original sensitive information. This method provides a secure and efficient way to conceal and transmit sensitive information over any channel.

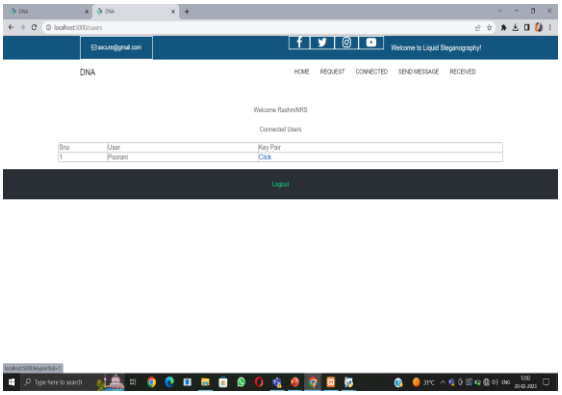


2. User management

The user management module in DNA steganography involves creating a system for authenticating users and managing user data for the web application. This would include user registration, user authentication, and password management. The module would also involve assigning different levels of access to users based on their roles, such as managers and personnel.

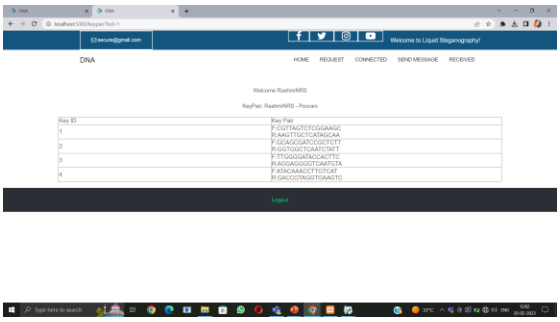
The user management module would be integrated with the web application to ensure that only authorized users are able to access and use the application. It would also enable the administrator to create, update, and remove users for the web application, as well as manage their access privileges.





3. Production and Distribution of Keys

While cryptography allows User1 and User2 to communicate privately, they still need a way to exchange the shared key so that only they know it and others do not. This is a difficult problem to solve in general. This shared key could be exchanged offline, for example, by sending a letter to each other, but this does not scale. They could also use public key cryptography: User1 could select the shared key and encrypt it with User2's public key, allowing only User2 to read it. This works, but we'd like to see if there are any other methods for securely exchanging keys. We will investigate two approaches: Diffie-Hellman Key Exchange and Key Exchange with a Key Distribution Center. There is a Key Distribution Center (KDC) in this module that both User1 (Client - C) and User2 (Server - S) trust and with which they share keys - this means that there are already KKDC, C and KKDC, S. Client sends a request to KDC with its own and a server's identities - C, S. The KDC generates the shared key KC, S and transmits it to the client EKKDC, C (S, KC, S) and the server EKKDC, S. (C, KC, S). A ticket is the name given to the message sent to the server. KDC can also provide this ticket to the client for use in their first communication with the server. The client must send the identities of the client and server to KDC so that KDC knows which keys to use to encrypt the replies. The response sends the shared key KC, S to both parties. It must be encrypted, or else someone on the network could snoop on it. Only the client and server can read the responses because they are encrypted. Why are the identities of the other parties included in the messages? They must determine who this shared key applies to and who the other party holding the same key is.

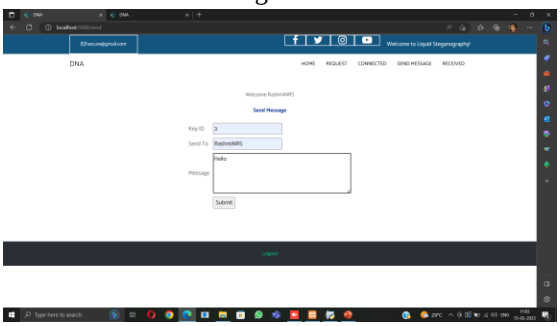


4. Secret Communication Protocol

The secret communication protocol creates and employs DNA Digital Coding in this module.

DNA nucleo-tide	Decimal	Binary
A	0	00
C	1	01
G	3	10
T	3	11

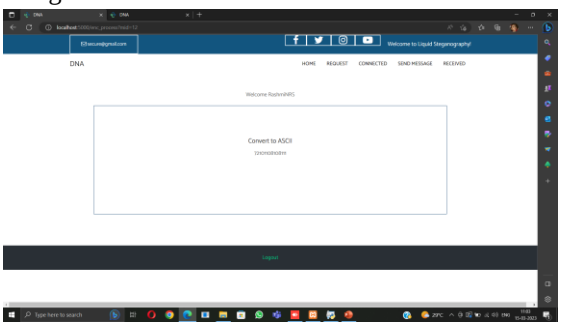
Table 1 DNA Digital Code

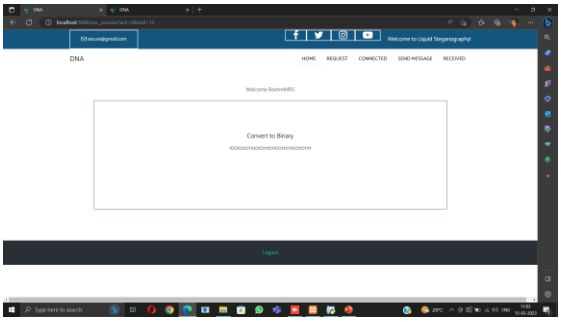


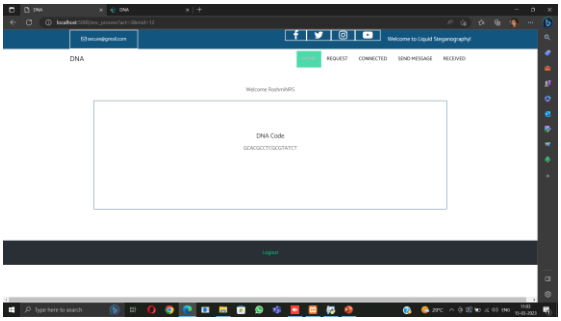
4.1. Encryption of DNA

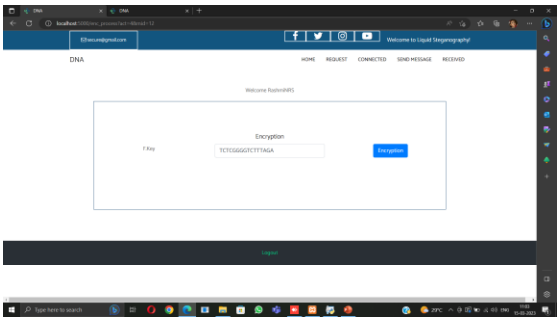
To encrypt and conceal messages inside a DNA sequence, the proposed algorithm takes numerous steps. The following are the steps in the encryption process:

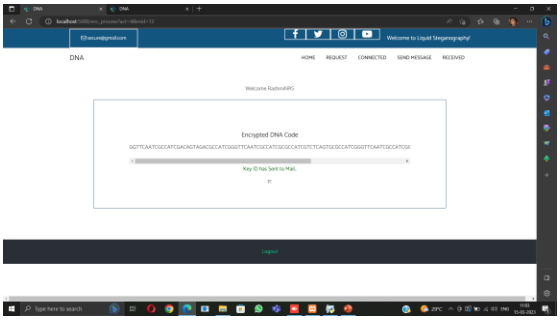
1. Split M into characters, M = m1, m2, m3… mn, and each character is converted into its 8-bit binary equivalent based upon the ASCII standard.
2. Randomly generate a number between 0 and 255 to form K1, and then the key is converted into an 8-bit binary sequence.
3. The last character in M is XORed with K1.
4. The result is XORed with the character preceding the last one in M; the XORing is repeated until all the characters are converted and stored in A.
5. The binary sequence A is converted into a protein sequence.
6. A sample DNA sequence S is selected randomly and converted into a binary bit sequence using Table 1.
7. Generate a random number, which is preferred to be a small number K2, and then divide the DNA sequence S into segments; the segment length should be equal to K2.
8. Add the first binary value of A at the beginning of the first DNA binary segment, and insert the second binary of K1 into the second binary segment, and so on.
9. Combine all of the binary sequences together, and then use them to create a fabricated DNA sequence using Table 1.

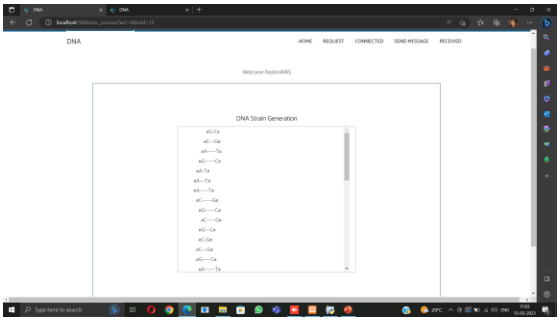


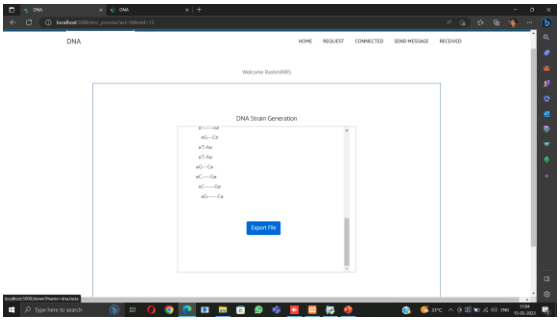








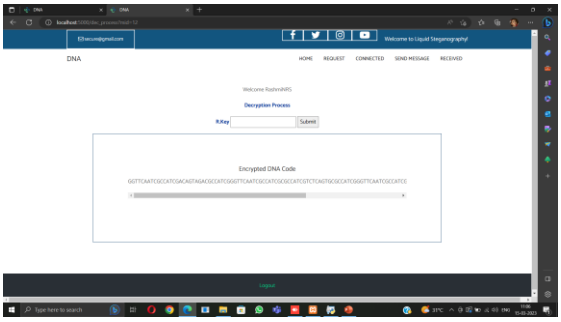


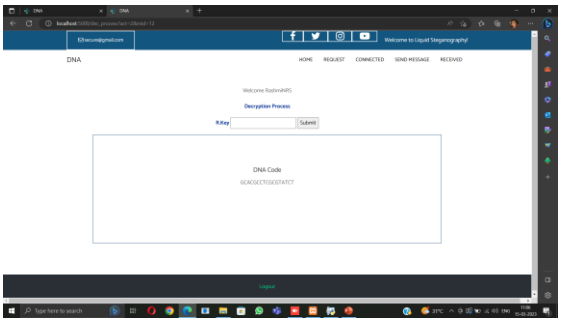


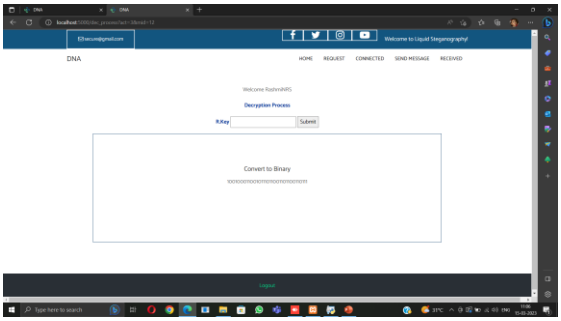
4.2. Decryption of DNA

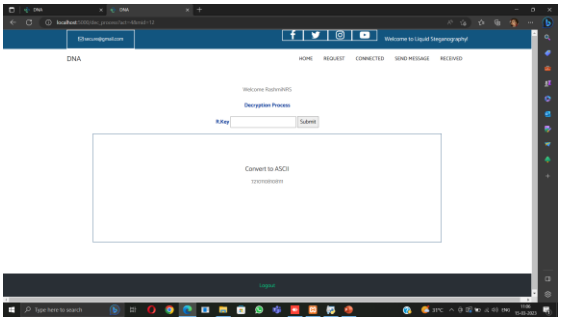
In order to decrypt the message, the receiver must understand K1 and K2. He or she must also receive the original DNA from the sender. The receiver takes the following actions:

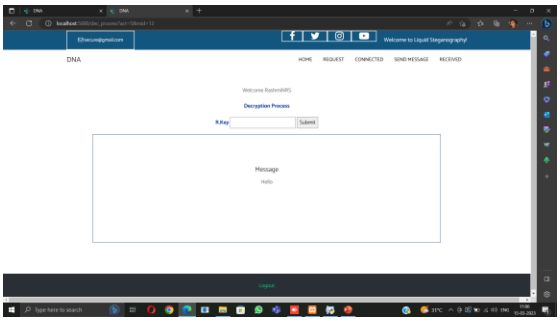
1. Using Table 1, convert the received fake DNA sequence into a binary sequence.
2. Split the binary DNA sequence into segments with lengths equal to K2 + 1.
3. Concatenate the first bit from each segment to produce significant bit B.
4. XOR the first eight binary bits of B with K1, then XOR the next eight bits of B with the previous eight bits of B, and so on.
5. Convert the DNA sequence's binary bits into ASCII text values.











5. Transmission of Data

In this module, the recipient obtains the final key ID, which can be stored and transported in the form of liquid, paper, or plate. This allows it to be camouflaged as a beverage, cosmetic, or other object, making it easy to prevent loss or damage. By leveraging the replication capability of DNA in living cells, the message can be easily duplicated.

6 SYSTEM DEVELOPMENT

In the Secure Hiding and Sharing of Messages process, both the sender and receiver play a role in transmitting a message. The sender and receiver are provided with a key table and key pool in advance. The key must be replaced every time it is used, and when the key table is depleted, a new one is distributed via secret channels. The message is first encoded into a DNA code, then substituted to produce the secret message (SM) using a substitution sequence (Sub) obtained from the key pool via PCR with the key. To introduce randomness into the SM, a substitution algorithm is used with a random substitution sequence. However, since DNA code carries language information, it cannot be entirely randomised, so random SMs are obtained by performing substitutions with a different sequence each time. The SM is then hidden in living cells with random DNA sequences to obtain an encoded message (IC), which is transmitted to the receiver through public channels. The receiver can retrieve the SM and Sub by using the key in PCR and then recover the original message using the DNA substitution square.

7 RESULTS

DNA Steganography has numerous advantages

- reasonable ease of placement and detection by the intended recipient
- its difficulty in detection and erasure by attackers
- its credibility in the event of a dispute, its robustness against filtering, compression, or truncation
- its reasonable overhead, and no significant change in the meaning or function of the original data.

However, to ensure successful hiding of the data, the data being hidden should have error correction/redundancy.

Advantages of DNA Data Storage

The advantages of DNA as a Data Storage medium are

- DNA can store information at an incredibly high density, with a gram of DNA containing billions of nucleotide bases that can be used to store data.
- This allows for an ultra-compact storage medium that can hold vast amounts of information. For example, a single gram of DNA can store up to 108 terabytes of data, and just a few grams could hold all the data in the world.
- This makes DNA an attractive option for long-term storage of digital data, as it has the potential to far outlast current storage media.

The Benefits of DNA Computing

• Performance - Compared to conventional computers with a performance of around 100 MIPS (millions of instructions per second), Adleman showed that combining DNA strands can result in computations equivalent to 109 or better, which is arguably over 100 times faster than the fastest computer.

• Low Storage Needs - DNA has an incredibly high storage density compared to conventional storage media. In other words, a very small volume of DNA can store a large amount of information. This is because DNA molecules are incredibly small and densely packed, with each nucleotide taking up only a tiny amount of space. Conventional storage media, such as hard drives or flash drives, require much more space to store the same amount of information.

• Low Power Requirements - Unlike traditional computers, DNA computation does not require any external energy source to run, as the chemical bonds

that make up DNA occur naturally. This means that DNA computation has significantly lower power requirements than conventional computers.

8 CONCLUSION

DNA is an extremely efficient storage medium. It is lightweight, biodegradable, and uses very little energy. It is now used to spread species, encode protein synthesis, and solve complex computational problems. Who knows what will happen in the future? Recognizing this, techniques for hiding data in this medium to catalogue, annotate, watermark, and/or encrypt information can serve a significant purpose. This project proposes the novel concept of encoding data in DNA. Furthermore, it defines two new and original techniques for hiding data, as well as an evaluation and analysis of their utility for the various functions of hidden data. The first method conceals data in non-coding DNA regions such as non-transcribed and non-translated regions, as well as non-genetic DNA such as DNA computing solutions. In theory, the second can be used to embed information directly into active genetic segments. In addition to a straightforward set of steps for embedding the data, this paper addresses the vulnerability of the codon redundancy technique and provides several additional steps to strengthen a watermark. The practical utility of such a technique is enormous.

REFERENCES

[1] H. R. Yassein, N. M. G. Al-Saidi, and A. K. Farhan, ‘‘A new NTRU cryptosystem outperforms three highly secured NTRU-analog systems through an innovative algebraic structure,’’ J. Discrete Math. Sci. Cryptogr., vol. 25, no. 2, pp. 523–542, 2020.

[2] A. Kumar and S. Tejani, ‘‘S-BOX architecture,’’ in Communications in Computer and Information Science. Singapore: Springer, 2019, pp. 17–27.

[3] A. K. Farhan, R. S. Ali, H. R. Yassein, N. M. G. Al-Saidi, and G. H. Abdul-Majeed, ‘‘A new approach to generate multi S-boxes based on RNA computing,’’ Int. J. Innov. Comput., Inf. Control, vol. 16, no. 1, pp. 331–348, 2020.

[4] A. H. Zahid, M. J. Arshad, and M. Ahmed, ‘‘A novel construction of efficient substitution-boxes using cubic fractional transformation,’’ Inf. Theory, Probab. Statist., vol. 21, no. 3, p. 13, 2019.

[5] M. S. Mahmood Malik, M. A. Ali, M. A. Khan, M. Ehatisham-Ul-Haq, S. N. M. Shah, M. Rehman, and W. Ahmad, ‘‘Generation of highly nonlinear and dynamic AES substitution-boxes (S-boxes) using chaos-based rotational matrices,’’ IEEE Access, vol. 8, pp. 35682–35695, 2020.

[6] M. Mansour, W. Elsobky, A. Hasan, and W. Anis, ‘‘Appraisal of multiple AES modes behavior using traditional and enhanced substitution boxes,’’ Int. J. Recent Technol. Eng. (IJRTE), vol. 8, no. 5, pp. 530–539, Jan. 2020.

[7] J. M. Cheung, ‘‘The designa of S-boxes,’’ Ph.D. dissertation, San Diego State Univ., San Diego, CA, USA, 2010.

[8] F. A. Kadhim, G. H. A. Majeed, and R. S. Ali, ‘‘Proposal new s-box depending on DNA computing and mathematical operations,’’ in Proc. AlSadeq Int. Conf. Multidisciplinary IT Commun. Sci. Appl. (AIC-MITCSA), May 2016, pp. 1–6.

[9] A. H. Al-Wattar, R. Mahmod, Z. A. Zukarnain, and N. I. Udzir, ‘‘A new DNA-based S-box,’’ Int. J. Eng. Technol., vol. 15, no. 4, pp. 1–9, 2015.

[10] A. Majumdar, A. Biswas, A. Majumder, S. K. Sood, and K. L. Baishnab, ‘‘A novel DNA-inspired encryption strategy for concealing cloud storage,’’ Frontiers Comput. Sci., vol. 15, no. 3, Jun. 2021, Art. no. 153807.

[11] L. Jinomeiq, W. Baoduui, and W. Xinmei, ‘‘One AES S-box to increase complexity and its cryptanalysis,’’ J. Syst. Eng. Electron., vol. 18, no. 2, pp. 427–433, Jun. 2007.

[12] A. A. Abdel-Hafez, R. Elbarkouky, and W. Hafez, ‘‘Comparative study of algebraic attacks,’’ Int. Adv. Res. J. Sci., Eng. Technol., vol. 3, no. 5, pp. 85–90, May 2016.

[13] K. Mohamed, M. N. Mohammed Pauzi, F. H. Hj Mohd Ali, S. Ariffin, and N. H. Nik Zulkipli, ‘‘Study of S-box properties in block cipher,’’ in Proc. Int. Conf. Comput., Commun., Control Technol. (I4CT), Sep. 2014, pp. 362–366.

[14] A. A. Abdel-Hafez, R. Elbarkouky, and W. Hafez, ‘‘Algebraic cryptanalysis of AES using Gröbner basis,’’ Int. Adv. Res. J. Sci., Eng. Technol., vol. 3, no. 12, pp. 183–189, 2016.

[15] W. Al Sobky, H. Saeed, and A. N. Elwakil, ‘‘Different types of attacks on block ciphers,’’ Int. J. Recent Technol. Eng. (IJRTE), vol. 9, no. 3, pp. 28–31, Sep. 2020.

[16] E. W. Afify, R. Abo Alez, A. T. Khalil, and W. I. Alsobky, ‘‘Performance analysis of advanced encryption standard (AES) S-boxes,’’ Int. J. Recent Technol. Eng., vol. 9, no. 1, pp. 2214–2218, 2020.
[17] J. H. Cheon and D. H. Lee, ‘‘Resistance of S-boxes against algebraic attacks,’’ in Proc. Int. Workshop Fast Software Encryption (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2004, pp. 83–93.
[18] J. Cui, L. Huang, H. Zhong, C. Chang, and W. Yang, ‘‘An improved AES S-box and its performance analysis,’’ Int. J. Innov. Comput., Inf. Control, vol. 7, no. 5, pp. 2291–2302, 2011.
[19] E. W. Afify, R. Abo Alez, A. T. Khalil, and W. I. Alsobky, ‘‘Algebraic construction of a powerful substitution box,’’ Int. J. Recent Technol. Eng. (IJRTE), vol. 8, no. 6, pp. 405–409, Mar. 2020.
[20] W. I. E. Sobky, A. R. Mahmoud, A. S. Mohra, and T. El-Garf, ‘‘Enhancing Hierocrypt-3 performance by modifying its S-box and modes of operations,’’ J. Commun., vol. 15, no. 12, pp. 905–912, 2020.
[21] M. Chakraborty, S. RoyChatterjee, and K. Sur, ‘‘Study on S-box properties of convolution coder,’’ in Proc. Int. Ethical Hacking Conf., vol. 1065. Singapore: Springer, 2019, pp. 119–128.
[22] N. A. Azam, U. Hayat, and M. Ayub, ‘‘A substitution box generator, its analysis, and applications in image encryption,’’ Signal Process., vol. 187, Oct. 2021, Art. no. 108144.
[23] I. Hussain, T. Shah, H. Mahmood, and M. A. Gondal, ‘‘A projective general linear group based algorithm for the construction of substitution box for block ciphers,’’ Neural Comput. Appl., vol. 22, no. 6, pp. 1085–1093, 2013. 66428 VOLUME 10, 2022 H. A. M. A. Basha et al.: Efficient Image Encryption Based on New Substitution Box Using DNA Coding and Bent Function
[24] F. Özkaynak and A. B. Özer, ‘‘A method for designing strong S-boxes based on the chaotic Lorenz system,’’ Phys. Lett. A, vol. 374, no. 36, pp. 3373–3738, 2010.
[25] R. Guesmi, M. A. Ben Farah, A. Kachouri, and M. Samet, ‘‘A novel design of chaos based S-boxes using genetic algorithm techniques,’’ in Proc. IEEE/ACS 11th Int. Conf. Comput. Syst. Appl. (AICCSA), Nov. 2014, pp. 678–684.
[26] G. Ivanov, N. Nikolov, and S. Nikova, ‘‘Cryptographically strong S-boxes generated by modified immune algorithms,’’ in Proc. Int. Conf. Cryptogr. Inf. Secur. Balkans. Cham, Switzerland: Springer, 2016, pp. 31–42.
[27] M. Ahmad, E. Al-Solami, A. M. Alghamdi, and M. A. Yousaf, ‘‘Bijective S-boxes method using improved chaotic map-based heuristic search and algebraic group structures,’’ IEEE Access, vol. 8, pp. 110397–110411, 2020.
[28] D. Lambić, ‘‘A new discrete-space chaotic map based on the multiplication of integer numbers and its application in S-box design,’’ Nonlinear Dyn., vol. 100, no. 1, pp. 699–711, 2020.
[29] F. Özkaynak, ‘‘On the effect of chaotic systems in performance characteristics of chaos based s-box designs,’’ Phys. A, Stat. Mech. Appl., vol. 550, Jul. 2020, Art. no. 124072.
[30] Q. Lu, C. Zhu, and X. Deng, ‘‘An efficient image encryption scheme based on the LSS chaotic map and single S-box,’’ IEEE Access, vol. 8, pp. 25664–25678, 2020.