



Information Security Using DNA Sequences

Hiba Safaa¹, Sahar Adill² and Ali Yakoob²

¹College of Science for women, university of Babylon, Babil, hilla

hiba.abid.gsci18@student.uobabylon.edu.iq.gsci

²College of Sciences for Women, University of Babylon dr.sahar.Adill@gmail.com, Hilla, Babel, State.

Received:

30/1/2022

Accepted:

18/10/2022

Published:

31/12/2022

Abstract

Information security is a significant cause for concern, mainly because of the explosive growth in internet usage over the last few years. Due to this growth, there have been occurrences of unauthorized access, which have been reduced thanks to "using a range of secure communication protocols, such as encryption and data concealment". Using DNA's bio-molecular capabilities, the usage of DNA as a carrier for encryption and data concealing has increased in recent years. The realization that DNA may function as a transport medium sparked this movement. In this study, we first examine and briefly outline the evolution of the present DNA coding system. After that, the several ways DNA has been used to enhance encryption techniques are categorized. The benefits and drawbacks of these algorithms and the most recent advancements in DNA-based encryption techniques are discussed. Finally, we provide our thoughts on the potential future of DNA-based encryption algorithms.

Keywords: information Security, DNA, DNA Encryption, Microfluidic, Chaotic map.

الخلاصة

يعد أمن المعلومات من المواضيع المهمة، ويرجع ذلك أساساً إلى النمو الهائل في استخدام الإنترنت على مدى السنوات القليلة الماضية. نتيجة لهذا النمو، كانت هناك حالات وصول غير مصرح به، والتي تم تقليلها بفضل "استخدام مجموعة من بروتوكولات الاتصال الآمن، مثل التشفير وإخفاء البيانات". باستخدام القدرات الجزيئية الحيوية للحمض النووي، ازداد استخدام الحمض النووي كناقل للتشفير وإخفاء البيانات في السنوات الأخيرة. أثار إدراك أن الحمض النووي قد يعمل كوسيط نقل أثار هذه الحركة. في هذه الدراسة، نفحص أولاً ونلخص بإيجاز تطور نظام ترميز الحمض النووي الحالي. بعد ذلك، يتم تصنيف الطرق العديدة التي تم بها استخدام الحمض النووي لتحسين تقنيات التشفير. تمت مناقشة مزايا وعيوب هذه الخوارزميات وأحدث التطورات في تقنيات التشفير القائم على الحمض النووي. أخيراً، نقدم أفكارنا حول المستقبل المحتمل لخوارزميات التشفير القائمة على الحمض النووي.

الكلمات المفتاحية: أمن المعلومات، الحمض النووي، تشفير المعلومات، شريحة مايكروفلودك، الخارطة الفوضوية



1- Introduction

Information security refers to the practice of guarding information being accessed in an unauthorized manner. Most people believe that the purpose of security measures is to keep private data from falling into the wrong hands. At the same time, another popular theory holds that security measures aim to avoid data loss, corruption, or tampering [1]. [2, 3] Cryptography and data hiding, two closely related ideas, are the approaches that see the most widespread use in the disciplines of computer security and communication security, respectively. The development and use of these systems are distinct, although the two sets of protocols aim to ensure data privacy and integrity. Cryptography is the practice of scrambling data so that unauthorized parties cannot read it. Data hiding is a type of hidden writing that conceals the presence of the message that is being covered (Steganography, watermarking). In contrast, cryptography is the method by which the party that possesses the associated secret key modifies the meaning of personal writing. Data hiding is a subset of the broader category of hidden writing. To conceal the existence of the data, the approach for concealing it should result in the original medium maintaining as few of its qualities as is practically possible after the data has been covered. It is common practice to use data concealment rather than encryption when sending information over an unsecured public channel [4, 5, 6]. This is because data concealment is both safer and more sufficient than encryption. "DNA as an informational and computational carrier with the assistance of molecular technologies" has given rise to a new field known as DNA-based cryptography, which has been named after [7]. This new field was given birth as a result of the discovery of the computing potential of DNA.

Steganography is often favored over cryptography because of the following two reasons: Steganography is more secure than cryptography. When transferring data through an unsecured public channel, encryption on its own is not a good form of data security. The term "ciphering" refers to the science of covered writing, in contrast to the word "steganography," which refers to the practice of concealed writing. Aside from that, deoxyribonucleic acid, also known as DNA, is now considered for several other computers uses. In addition to its enormous capacity for storage [7], DNA also has the potential to store information in the form of data packets that are instructive [8, 9]. DNA, like any other kind of data storage medium, has to have some form of protection in the form of a safe algorithm operating as intended. To successfully carry out secure cryptography and steganography operations, it is necessary to make use of a wide variety of the biological features that DNA sequences possess [6]

The critical distinction between steganography and watermarking is that the former hides the fact that confidential information even exists. Steganography is useless if the presence of hidden data is leaked. Watermarking, on the other hand, makes the existence of hidden data transparent [10].

The novel DNA-based data concealment techniques developed by researchers aim to give the highest possible level of security and protection while preserving high capacity and low modification rates. Numerous innovative data-hiding strategies have been created by scientists in light of the biological properties of DNA sequences. As a result, the field of DNA computing has just emerged [8, 11].

And these are DNA cryptography's end goals [11]: First, DNA computing supports a high level of parallelism, which is crucial for boosting processing performance. DNA molecules are very efficient mediums for transmitting the information. Third, it uses little energy.

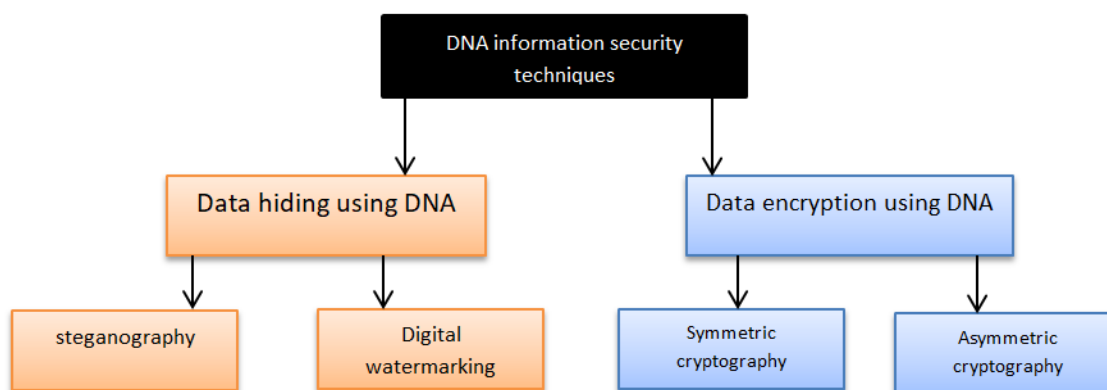


Figure:1 DNA Information security techniques

2- Data Concealing Techniques using DNA

Amany E. et al., 2021 [12] developed and demonstrated the substitution-based method for hiding data in DNA sequences. In the suggested data concealment approach, binary data is encoded and buried inside a DNA sequence. The (a previously suggested DNA replacement technique) is not a substitute for (Least Significant Base method). According to their understanding, the proposed adjustment is based on a straightforward notion that has never been implemented. It was shown that DNA amino acids may be classified and that each DNA codon in one of the groups can contain two hidden bits of information rather than just one, as the Least Significant Base method suggests. It is advised that a procedure similar to the Least Significant Base method be used since both ways are blind and preserve the regular biological structure of DNA in their counterfeit DNA sequences. It is evaluated using a publicly accessible DNA sequence dataset (BALI BASE). According to the evaluation findings, the suggested technique achieved a 50% increase in data concealing capacity compared to the Least Significant Base method. A considerable decrease in the chance of the Least Significant Base method breaking was also seen when the recommended strategy was compared to the old way.

B. Adithya & G. Santhi, 2020 [13] shows that DNA computing and Morse code encoding may be used to keep vital information safe while operating in a hostile environment. An encoding table encodes plaintext into DNA sequences for subsequent examination. RNA configurations now manage the transcription and translation of encoded data. The traditional genetic code is decoded using organic molecules, while the Morse code pattern is used to solve the stego DNA. It can be



shown that when compared to existing techniques like AES and RSA, DNA-based Playfair cipher, and Vigenere encryption, the concealing capacity is excellent, and the execution time is brief.

This study focuses on the encryption and decryption methods rather than the decryption mechanism itself, which this study has not adequately explored. Following is a list of the steps in the computer algorithm used to encrypt plaintext into ciphertext. Before beginning the encryption process, all phases of plaintext to DNA conversion must be performed. Take this as an example: Suppose the plaintext is "Tree," and the plaintext is split into two halves at random, as an example. A random element is included to ensure that the plaintext is distributed equally. DNA sequences may be encoded using the binary coding rules A-00, T-01, C-10, and G-11: Intron sequences that match are encoded in DNA as part of the transformation process. The altered DNA sequence is translated into mRNA by replacing Uracil (U) with Thymine (T) on both the left and right flanks of the DNA groups on either side of the DNA groups. It's a method for mimicking the process of transcription that occurs in living things.

When the mRNA sequence is translated into the tRNA sequence, the DNA from the letter set is swapped out for its DNA letter set equivalent. Pairs like "A-U, A-U, and G-C" and "C-G" will change. The tRNA sequence is transferred to the DNA sequence by replacing Uracil (U) with Thymine (T) (T). This is a more natural way to reverse-transcribe the simulation than the previous one. The reverse transcription DNA sequence is transferred twice on the right-hand side of the chromosome, once on each side. There are 256 amino acids in the universal basic amino acids table. However, only 20 of them are included in this research.

According to this research, the concealing capacity has been increased to 2.4 bits per nucleotide (bpm). Currently, 1.5 billion pounds have been hidden, 40% less than the advised way. With the current approach, a 181,432bp reference DNA can hide a message of 35.05kb, but the proposed technique can hide s57.03kb.

By applying the same length reference DNA sequence, the Study approach demonstrates its superiority over the AES cipher and the RSA cipher's concealing capabilities (181,432bp). The AES cipher can cover up to 47.02 kb, while the RSA cipher can cover up to 42.08 kb using the same reference DNA sequence length. This is due to the traditional approaches' usage of extra padding bits in their encryption computation.

Sajib Biswas & Monowar Hossain, 2021 [14] Instead of having to provide a keyword along with the encrypted DNA sequences, offer a solution that is both faster and more secure than the alternatives. The privacy of the communication may be compromised if a secret keyword is communicated before the message. As a result, they used DNA sequences to encode the private keyword. Aside from making it more difficult for the attacker, this also makes it more difficult for them to decipher the secret keyword. There are two primary phases in the recommended procedure. The first step is to encode the message into the DNA using the 10 x 10 Playfair cipher and a modified Caesar cipher. The message and cover DNA are then randomly implanted and covered in the second stage, following which the sequence is altered using a broad, complementary base substitution technique. Afterward, a random number generator inserts the key into the resultant



DNA. Thus, the extraction operation may be carried out without the recipient knowing anything before sending the message. As a result, it gives us more peace of mind than any other kind of secrecy could.

Shah Haris Nabi et al. developed a state-of-the-art approach to data concealment using DNA sequences in 2021 [15]. A DNA sequence database is queried for the reference sequence. 1.63 108 DNA sequences are available to the public at large. Given the abundance of publicly available DNA sequences, it becomes increasingly difficult for an adversary to spot the existence of a covert message. A total of two methods for concealing information have been proposed previously. The first stage in transforming top-secret information into Stego-DNA is to use a substitution table and a reference sequence. The second strategy involves inserting code into DNA. The reference DNA sequence and the DNA's binary encoding are selected using a private key. There is little doubt that secret information may be masked in DNA sequences. In contrast, a DNA sequence may be used to cover up personal files. It is a viable option to study genetic programming and encryption for 126 S. to alter their DNA to keep private information secure. DNA sequences may be used with the proposed method for secure data storage, transmission, and encryption.

3- Encryption Algorithm using DNA

Tarek Hagra et al., 2022 [16] Study entitled " Anti-attacks encryption algorithm based on DNA computing and data encryption standard," An encryption/decryption technique for any form of data based on DES-DNA is presented in this work. There are two variants of the Encrypt algorithm. Both versions of Encrypt use DNA computing instead of binary space to encrypt and decode data. For brute force assaults, the suggested method uses a secret key of 64 nucleotides, making it intractable as 128-bit algorithms. As part of this effort, the S-Boxes used in encrypting were rebuilt so DNA computing could be used to choose S-Boxes. The first selection approach renders Encrypt more vulnerable to differential attacks, whereas the second selection method hides them. A critical sensitivity study shows the suggested algorithm's sensitivity to changes in the secret key. Detailed picture data are employed in statistical analysis because of their features. Data entropy, correlation analysis, and NIST statistical tests show that the suggested method is intractable.

ChengyeZou, et al., 2022 [17] Study entitled "A novel image encryption algorithm based on DNA strand exchange and diffusion" According to this study, a new kind of ergodic tent-dynamic cross-coupled map lattice with larger information entropies, a more comprehensive chaotic range, and better ergodicity than other types of cross-coupled map lattices may be used for picture encryption algorithms based on chaos. Encryption schemes traditionally include image confusion and diffusion. Encryption using DNA coding and sequences is employed to create short and long DNA strands, which are then used to perform the functions of DNA strand swapping and diffusion, respectively. Unlike ordinary picture confusion, DNA strand exchange randomly swaps out segments of two different DNA strands. While the Watson-Crick base pairing characteristic classifies image diffusion, DNA strand diffusion is separated into two categories, each using a



distinct DNA computation operation. Their encryption technique can withstand various assaults thanks to statistical and security research. The suggested algorithm's efficiency and safety have been shown in experiments.

Weiyu Ran et al., 2022 [18] Study entitled "A double scrambling-DNA row and column closed-loop image encryption algorithm based on the chaotic system." Using image coding as a basis, they provide an approach for the dynamic updating the double-stranded loop of DNA, rows, and columns. They propose an improved method in the encryption process's scrambling and diffusion phases. They used the Hilbert curve to scramble and overcome the problem of storing neighboring pixels in space during the scrambling step. Using the Hilbert-Knuth-Durstenfeld shuffle technique, the image was shuffled doubly, improving the efficiency of data storage in memory. They used binary diffusion of coding rows and columns of DNA in the diffusion stage to enhance the current diffusion strategy of the closed block. To achieve a closed-loop dynamic update of the cipher system, the initial line of the ciphertext is updated as soon as the last line of the ciphertext is created. Because of the inherent non-linearity of the DNA process, the input and output of the cipher are not a simple linear connection, which increases the sensitivity and security of the cipher system when integrating DNA cipher, plaintext, ciphertext, and critical flow. This study uses SHA-256 to create a master flow because the chaotic system is susceptible to the starting value. Although there is noise in the source image, this method can provide a strong encoding result. Since the proposed encryption technology has excellent reliability and security, as shown in the paper simulation results and security analysis in Part IV, the algorithm offers promising application potential.

Chengye Zou et al., 2021 [19] Study entitled "Encryption Algorithm Based on DNA Strand Displacement and DNA Sequence Operation" In this study, they proposed a DSD-implemented Rössler chaotic system based on an ideal formal chemical reaction network. The ability to compare DNA strand concentration and derive the binary values of chaotic sequences is an improvement over the DSD analog circuit that was previously used. This technique transforms the DSD analog circuit into a digital course, which is simpler to implement in the future. It has been shown via security research that ciphertext and decrypted text are sensitive to critical changes. That ciphertext is random-like, making encryption schemes resistant to statistical and brute-force attacks. Through a thorough investigation, the suggested encryption method can somewhat withstand noise, concentration fluctuation, and inaccuracy in response rate control.

Authors Yangming Hu and Xiaoqiang Zhang's 2021 [20] Method for encrypting multiple images using a combination of a three-dimensional scrambling model and research on dynamic DNA coding The MIE strategy used in this research to improve security, encryption capacity, and efficiency was inspired by the random and ever-changing DNA coding found in living organisms. This study develops a 3D scrambling model and an extended Zigzag transformation through dimensionality reduction. The proposed technique is superior through testing and algorithm



analysis. Medical records, architectural blueprints, and identification documents are some types of sensitive information that may benefit from the proposed method's added layer of security. It's also secure against the most common attacks, including guessing games, differential attacks, chosen-plaintext attacks, and statistical analysis. The following are only four advantages: To ensure the method's safety, the parameters for scrambling and diffusion are paired with the user key and hash value. Due to the ample critical space, the approach is resistant to brute-force attacks. To increase the system's encryption capability, we develop a 3D scrambling model with an excellent scrambling effect and support for an endless number of inputs. The investigation of the encryption quality shows that the suggested approach provides a high level of security for the encrypted data and can encrypt an infinite quantity of data. DNA's strong computational parallelism and excellent storage density are two more features that might significantly improve encryption efficiency.

I. Moussa & E. I. Abd El-Latif, 2019 [21] In a paper titled "Information Hiding Using Artificial DNA Sequences Based on the Gaussian Kernel Function," the researchers developed a two-round encryption system. The most recent encryption method, known as the Data Encryption Standard algorithm, is identical to this system (DES). Here, we use two keys to encrypt the plaintext. Elliptic Curve Cryptography (ECC) and the Gaussian Kernel Function form these two essential keys (GKF). Using a random injective mapping, a new key is created for each character that appears twice in a given key. The second DNA sequence encrypts the message randomly using GKF numbers when it comes down to it.

4. DNA computing challenges and future work

DNA cryptography still needs in-depth research and open discussion before adequately developing. Also unknowable is its future.

Additionally, error codes exist in DNA computing; they are probabilistically generated at random and might gradually rise as the experiment step increases.[22]

The advancement of biotechnology and the appearance of a more effective DNA cipher design alternative will give information security research a new direction. However, further research is required to determine DNA cryptography's security, viability, and stability. [23] .

| Data Concealing using DNA | Data encryption using DNA |
|---|---|
| Increase the concealing capacity | Using DNA ciphers instead of binary ciphers increases the security of cryptographic algorithms. |
| DNA sequences are capable of concealing sensitive information | DNA encryption is one of the rapidly evolving technologies in the world |



5- Conclusions

DNA Information security techniques It is a novel technique for mimicking the biomolecular structure of DNA and computing using molecular biology technologies. A novel strategy for dealing with the intractable NP-complete issue is offered for the first time. Traditional encryption faces both obstacles and possibilities as it adopts this new computational technique, which uses the immense parallel processing power and high memory density of biomolecules. "DNA cryptography is a new area of encryption that has emerged in recent years due to advances in DNA computing." Using DNA molecules as an information medium, it may implement various security mechanisms, including encryption, steganography, signature, and authentication.

In this article, I covered the topic of steganography and outlined some critical distinctions between steganography and encryption. In addition, we reviewed the many different methods for concealing data in steganography and presented a comparative analysis of these methods. I also went through various mechanisms, such as cryptography, steganography, hashing, and authentication, that may be used to facilitate the safe transport of data. In addition, I discussed a variety of strategies for steganographically concealing data. There are still issues, even though some progress was made with the article's coding systems. The quest for coding based on chaotic systems and DNA sequences has a long way to go, despite the increasing in-depth examination of both. The following are some restrictions on the scope of this study: To accomplish speedy processing of the algorithm, chaotic coders must consider the algorithm's time and space complexity. This is especially true for the mass processing data on a large-scale data platform. Future research will find a happy medium between cryptographic security and algorithm complexity to overcome these barriers. To extend the application and applicability of the algorithms, it is necessary to broaden the coding area to include additional forms of multimedia data, such as color images, medical images, remote sensing photos, and other ideas. To boost encoding speed and accomplish multiple coding simultaneously, it is essential to think about the massive parallelism of DNA coding or to employ other coding methods.

Conflict of interests.

There are non-conflicts of interest.

References

- [1] Singh, G. (2013). A study of encryption algorithms (RSA, DES, 3DES, and AES) for information security. *International Journal of Computer Applications*, 67(19).
- [2] Subhedar, M. S., & Mankar, V. H. (2014). Current status and key issues in image steganography: A survey. *Computer science review*, 13, 95-113.
- [3] Hamed, G., Marey, M., El-Sayed, S. A., & Tolba, M. F. (2016, December). Comparative study for various DNA-based steganography techniques with essential conclusions about future research. In *2016 11th International Conference on Computer Engineering & Systems (ICCES)* (pp. 220-225). IEEE.

- [4] Amin, M. M., Salleh, M., Ibrahim, S., Katmin, M. R., & Shamsuddin, M. Z. I. (2003, January). Information hiding using steganography. In *4th National Conference of Telecommunication Technology, 2003. NCTT 2003 Proceedings*. (pp. 21-25). IEEE.
- [5] Al-Mohammad, A. (2010). *Steganography-based secret and reliable communications: Improving steganographic capacity and imperceptibility* (Doctoral dissertation, Brunel University, School of Information Systems, Computing and Mathematics Theses).
- [6] Hamed, G., Marey, M., El-Sayed, S. A., & Tolba, M. F. (2015, November). Hybrid technique for steganography based on DNA with n-bits binary coding rule. *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)* (pp. 95-102). IEEE.
- [7] Anam, B., Sakib, K., Hossain, M., & Dahal, K. (2010). Review on the Advancements of DNA Cryptography. *arXiv preprint arXiv:1010.0186*.
- [8] Hamed, G., Marey, M., El-Sayed, S., & Tolba, F. (2016). DNA-based steganography: survey and analysis for parameters optimization. In *Applications of intelligent optimization in biology and medicine* (pp. 47-89). Springer, Cham.
- [9] Das, S., Das, S., Bandyopadhyay, B., & Sanyal, S. (2011). Steganography and Steganalysis: different approaches. *arXiv preprint arXiv:1111.3758*.
- [10] [3] Khan, A., Siddiqua, A., Munib, S., & Malik, S. A. (2014). A recent survey of reversible watermarking techniques. *Information sciences*, 279, 251-272.
- [11] Raj, B. B., Vijay, J. F., & Mahalakshmi, T. (2016). Secure data transfer through DNA cryptography using a symmetric algorithm. *International Journal of Computer Applications*, 133(2), 19-23.
- [12] El-deeb, A., Elsis, A., & Youssef, A. (2021). A Substitution-Based Method for Data Hiding in DNA Sequences. *IJCI. International Journal of Computers and Information*, 8(1), 87-105.
- [13] Adithya, B., & Santhi, G. (2021). Data Hiding Using Deoxyribonucleic Acid (DNA) Computing With Morse Code Cryptosystem. *European Journal of Molecular & Clinical Medicine*, 7(10), 1719-1725.
- [14] Biswas, S., & Hossain, M. (2021). Fully Blind Data Hiding by Embedding Within DNA Sequences Using Various Ciphering and Generic Complimentary Base Substitutions. In *Proceedings of International Joint Conference on Advances in Computational Intelligence* (pp. 1-13). Springer, Singapore.
- [15] Nabi, S. H., Sarosh, P., Parah, S. A., & Mohiuddin Bhat, G. (2021). Information Embedding Using DNA Sequences for Covert Communication. In *Multimedia Security* (pp. 111-129). Springer, Singapore.
- [16] Nabi, S. H., Sarosh, P., Parah, S. A., & Mohiuddin Bhat, G. (2021). Information Embedding Using DNA Sequences for Covert Communication. In *Multimedia Security* (pp. 111-129). Springer, Singapore.
- [17] Zou, C., Wang, X., Zhou, C., Xu, S., & Huang, C. (2022). A novel image encryption algorithm based on DNA strand exchange and diffusion. *Applied Mathematics and Computation*, 430, 127291.
- [18] Ran, W., Wang, E., & Tong, Z. (2022). The chaotic system is a double scrambling-DNA row and column closed-loop image encryption algorithm. *Plos one*, 17(7), e0267094.
- [19] Zou, C., Wei, X., Zhang, Q., Zhou, C., & Zhou, S. (2021). Encryption algorithm based on DNA strand displacement and DNA sequence operation. *IEEE Transactions on NanoBioscience*, 20(2), 223-234.
- [20] Zhang, X., & Hu, Y. (2021). Multiple-image encryption algorithm based on the 3D scrambling model and dynamic DNA coding. *Optics & Laser Technology*, 141, 107073.
- [21] Abd El-Latif, E. I., & Moussa, M. I. (2019). Information hiding using artificial DNA sequences based on Gaussian kernel function. *Journal of Information and Optimization Sciences*, 40(6), 1181-1194.



- [22] Cui, G., Li, C., Li, H., & Li, X. (2009, August). DNA computing and its application to the information security field. In 2009 fifth international conference on natural computation (Vol. 6, pp. 148-152). IEEE.
- [23] Cui, G., Qin, L., Wang, Y., & Zhang, X. (2007, April). Information security technology based on DNA computing. In 2007 *International Workshop on Anti-Counterfeiting, Security, and Identification (ASID)* (pp. 288-291). IEEE.