Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (2): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

Data Hiding Based DNA Issues: A Review

Sahar Adill Kadhum[1]

1. Computer Science Dept, Babylon University, College of Science for Women
Dr.sahar.adill@gmail.com

Article Information
Submission date: 10 /5 / 2020
Acceptance date: 18 /5/ 2020
Publication date: 30/ 6 / 2020

Abstract
Security of Information are a key concern, particularly with the extension growth of internet usage. This growth comes the incidents of unauthorized access which are countered by the use of varied secure communication techniques, namely; cryptography and data hiding. More recent trends are concerned with DNA used for cryptography and data hiding as a carrier exploiting its bio-molecular properties. This paper provides a review about published DNA based data hiding techniques using the DNA as a safeguard to critical data that transmitted on an insecure channel, to find out the strength and weaknesses points of them. This will help the future research in designing of more efficient and secure data hiding techniques-based DNA.

Keywords: Security, Data hiding, Cryptography, Steganography, DNA

1: Introduction

Security is the protection of information from unauthorized access. The main purpose of security in the modern perspective is to keep sensitive data from being changed, destroyed, and stolen by a third party [1]. The most commonly used methods in communication and computer security fields are cryptography and data hiding which are related concepts [2, 3]. Even though both methods have a similar goal in providing data security and confidentiality, but their implementation and usage are differing.

Cryptography is the science of converting some data to an incomprehensible format. Cryptography changes the meaning of secret writing by the party that owns the associated secret key while data hiding is a hidden form of writing which hides the existence of the embedded message. The data hiding technique should preserve a minimum change in the characteristics of the original media after covering the data to conceal its existence. Thus, data hiding is more secure, sufficient, and often preferred to cryptography in the transmission of data across an insecure, public channel [4, 5,6]. The disclosure of the computational ability of DNA, that utilizes DNA as an informational and computational carrier with the aid of molecular techniques, a new field add to cryptography based on DNA that emerged after [7].

data hiding techniques classified into Steganography and Digital Water Marking. Steganography is the art of hiding sensitive data like text, images etc. in a different media such as image, video and audio in order to prevent attracting attention that data

is there [8]. The main objective behind steganography is to hide data with the minimum difference between the original carrier and the modified one that can be observed by the naked eye, consequently, the attraction decreases and the algorithm security increases. As a result, the hacker cannot reveal the confidential information. On other hand, digital water marking which protects the copyright ownership information from unauthorized users [8,9].

Steganography is more secure and often preferred to cryptography for two reasons: Cryptology is not sufficient when transmitting data within an unsecure, public channel. It is just the science of covered writing, while steganography is hidden writing which concerns hiding the existence of the message [8]. Besides this, Deoxyribonucleic acid (DNA) is being proposed as a use for many computational purposes. A remarkable property of DNA is its vast storage capacity as one gram of DNA is known to store about 108 tera-bytes [7]. However, like every data storage device, DNA requires protection through a secured algorithm. Various biological properties of DNA sequences can be exploited for obtaining successful secured cryptography and steganography processes [6].

For achieving maximum protection and powerful security with high capacity and low modification rate, new data hiding methods have been proposed by researchers based on DNA with the advent of biological aspects of DNA sequences. This leads to a new born research field based on DNA computing [8, 10].

2: Molecular Biology

In molecular biology, genetic information and features are stored in DNA. These genetic materials characterize all the behavioral and physical aspects of an organism as it encodes the genetic instructions used in functioning all known living organisms [8]. DNA is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and viruses [11].

DNA is composed of two long strands known as a double helix; each is made of building blocks called nucleotides. Three components make up a nucleotide: Four bases (A - adenine, G - guanine, C - cytosine, T - thymine), a deoxyribose sugar and a phosphate group as shown in figure (1) [12].

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (2): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

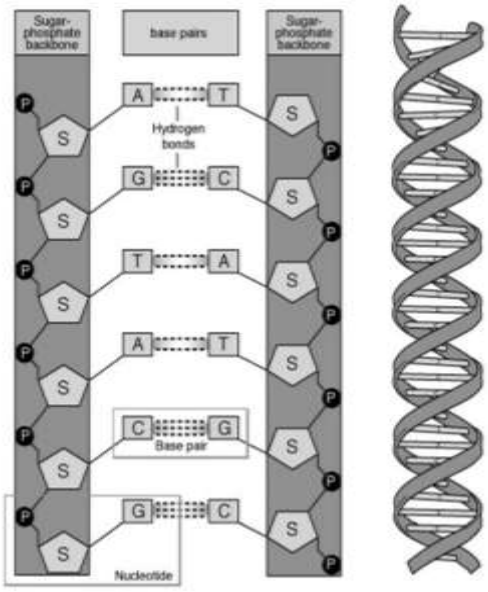

Figure (1): Helical Structure of DNA [12]

The DNA strands have chemical polarity, meaning that on each end of a molecule there are different groups (3‘ - bottom end and 5‘ top end) [13]. As shown in Figure (2) [14], DNA nucleotide is either a purine base or a pyrimidine base. The purine bases are adenine(A) and guanine (G), while the pyrimidines are thymine (T) and cytosine (C) [8]. The standard normal situation allows making hydrogen bonds between adenine and thymine or between cytosine and guanine. This complementary standard rule is known as Watson-Crick base-pairing such that A and T are linked by double hydrogen bonds while C and G are linked by triple hydrogen bonds. This mixture of nucleotides, create long polymer strands of DNA which build massive amounts of combinations of DNA double helix within organisms [8].

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (2): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

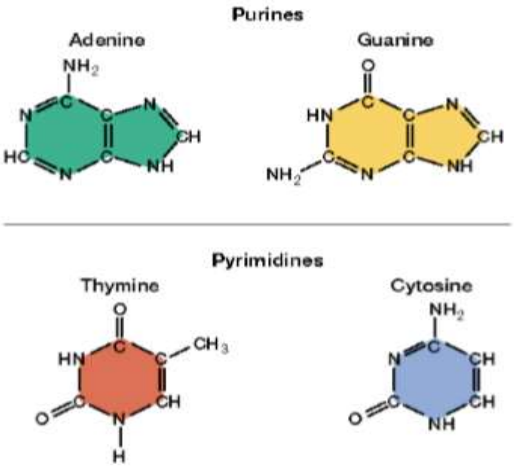

Figure (2). DNA Nucleotide Bases [14]

3: Comparative Study

Some of the recent data hiding based DNA techniques are compared in this section. The main objective is to derive the main goals behind data hiding based DNA to make an effective and reliable contribution towards the data hiding area as shown in table (1). The conclusion behind the derived comparison is to integrate the strengthened points that shall be supported by a proposed DNA based technique. The strengthen points are represented by: first point, encrypt the data by before embedding process in its original format to provide double layered security. The second point applying steganography technique to be more secure. The main feature that shall be supported by a hiding method is the blindness property, in order to avoid sending the original reference sequence to the receiver. The main objective behind the blindness property is to maximize the security degree as possible and to prevent any way can be discovered by an attacker by minimizing the required data that's sent to the receiver as much as possible.

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (2): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

Table (1): Comparative Study of Different Data Hiding Methods Based DNA

No.	Authors	Advantages	Disadvantages	Blind
1	Ref [15] 2010	Insertion technique. High capacity. Easy to implement. Low modification rate.	Length of stego DNA is longer than DNA reference. Payload not equal zero. Need multiple data in extraction process. Does not preserve the functionality of the amino acid. Increase the redundancy. Pure data hiding method.	No
		Complementary Pair rules technique. Easy to implement. Attackers need to know multiple data to crack the secret data.	Payload not equal zero. High modification rate. Does not preserve the functionality of the amino acid. Changing the length of DNA after embedding process. Pure data hiding method.	No
		Substitution technique. High capacity. Payload is equal to zero. Easy to implement. More efficient, more complexity and better performance than previous methods.	. High modification rate. Does not preserve the functionality of the amino acid. Pure data hiding method	No
2	Ref [16] 2011	Carrying out the result of hiding data in cloud to increase the level of confidentiality and complexity. Simple algorithm and high capacity. Payload equal to zero. Preserve the functionality of the biological DNA.	Increase the message size. Security depends on the DNA reference. Pure data hiding algorithm.	No
3	Ref [17] 2012	Mapped between DNA codons and amino to provide security, Encrypt secret message by playfair cipher before hiding. Improve playfair cipher by modifying to 5*5 to avoid its drawbacks where after encryption the diagraphs and the structure of secret text still exists. Providing double layer of security. Hight capacity and better time performance. Provide high probability cracking. Secret key is used.	Expand the length of stego DNA. Not preserve the functionality of the biological DNA. In order to extract the secret message from stego - DNA, it needs to send multiple data to receiver. Payload not equal to zero.	Yes
4	Ref [18] 2012	Propose a cryptographic-data hiding protocol reduce the using of public key as well as for best security. Payload is equal to zero. High capacity. Utilizing the innovative of DNA data hiding to hide the secret key within DNA reference for more security.	Does not preserve the functionality of the biological DNA.	No
5	Ref [19] 2013	Easy to implement. Low modification rate. Does not expand the length of stego DNA. One of the powerful encryption technique. (RSA) is used to encrypt secret data before hiding. Public key is used	Does not preserve the functionality of the biological DNA. Preserve the position of each DNA base which hold the secret data then send to the receiver for extraction. The size of secret data is increased. Low probability cracking.	No

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (2): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

6	Ref [20] 2013	The key space is big enough to resist brute force passive intruders. Encrypt the secret data before hiding in host document. The ratio of embedding capacity is 100%. Provide double hiding layers. DNA reference is construct by Chebyshev maps. Substitution method is used in hiding.	Complex calculation	Yes
7	Ref [21] 2014	Increased the capacity and security of the original substitution technique. Improved substitution method.	. Does not preserve the functionality of the biological DNA. Increased the size of secret binary in case of multiplying by 6 if not equal zero will add extra zeros. Expanding the length of stego DNA. sPure data hiding algorithm.	Yes
8	Ref [22] 2014	Sending only the integer value of stego DNA to receiver. High security. It is hard for intruders to know the generated random number seeds. It is hard for intruders to know how many packets are divided, also how many message bits and DNA binary in a packet. Randomly merged secret message bits and DNA reference bits. Secret key is used. High probability cracking.	Increased the redundancy. Increased the message size. Does not preserve the DNA functionality. Expand the length of stego DNA.	Yes
9	Ref [23] 2015	Uses three DNA reference in the proposed algorithm. Encrypt the plain text before hiding. Secret key is used. High probability cracking.	High modification rate. . Does not preserve the functionality of the biological DNA	Yes
10	Ref [24] 2015	. Highly secure and efficient method. Encrypt secret data using RSA algorithm before embedding. Provide double hiding layers. Public key is used.	Does not preserve the functionality of the biological DNA.	Yes
11	Ref [25] 2016	Preserving the information of organism's life. High capacity. Does not expand the length of stego DNA. . XOR and PRBG is used to encrypt the secret data. Errors derived and corrected by Reed-Solomon (RS) code. Secret key is used. High probability cracking.	Not easy to implement. High modification rate.	Yes
12	Ref [26] 2016	. Any type and any size of the secret data and the key can be used. . Applying different encryption techniques and analysing them to select the good one before hiding. Normal key is used to select English letters to generate the playfair cipher grid which can be more secure. High hiding capacity. No redundancy in the process. More simplicity. Good performance with low time execution. sSubstitution method is used in hiding.	Does not preserve the amino acid functionality. High modification rate. Low probability cracking.	No

13	Ref [27] 2017	Encrypt secret data by vigener or playfair cipher. Double amount of data hiding. . High security. After hiding DNA reference will be send in a microdot in a Paper before sending to the receiver. Regenerate another key and DNA sequence then the process of hiding will happen again if the paper is contaminated. reserve the functionality of DNA sequence and avoid any mutations.	Sending multiple data to the receiver for extraction process. High modification rate in non-coding region	No
14	Ref [28] 2017	High security. Double random key generator. Secret key is used. Probability cracking is very high.	. Does not preserve the functionality of DNA. Payload not equal to zero.	No
15	Ref [29] 2017	larger hiding capacity robustness against mutation. secured, undetectable, resistance and preservative to biological functions.	_____	Yes
16	Ref [30] 2018	Combine mathematical operations with DNA computation increase confusion. Robust. secure	obstacles at the infrastructure level as theoretical problems and difficult implementation. Authors focused on these obstacles aiming to make researchers to take advantage of painstaking effort expanded so far in the DNA cryptography	Yes
17	Ref [31] 2019	new steganography method. any format image Hard to detect image.	_____	Yes
18	Ref [32] 2019	Extraordinary Execution in Security, and Parallel Time Performance	_____	Yes
19	Ref [33] 2019	security in concealing the information, not predicting its nature, and preserving the biological structure of the DNA sequence	_____	Yes
20	Ref [34] 2020	More security using Magic Cube for generating crypto keys and hiding locations Enhance Vigener capacity algorithm High capacity. Does not expand the length of stego DNA. preserving the biological structure of the DNA sequence Probability cracking is very high.	_____	Yes

4: Conclusion

Communicating secretly without revealing critical information is very important. The continuous demands for secure techniques lead to create an enormous developing new a security techniques.. Nowadays, DNA has started to be used as a new data carrier as an effective and reliable media. The bio-molecular computational abilities of DNA are exploited by means of cryptography and stenography in order to develop high capacity, secured algorithms, and low cracking probability. In this paper, a comparative study has been done to various recent DNA based data hiding techniques. The main objective behind this comparative study, is to help researchers in the field to do their future work on well and improved secured DNA techniques in more efficient and reliable manners by treating the drawbacks of the already existing algorithms. *

Conflict of Interests.

There are non-conflicts of interest .

References

[1] Singh, G., A study of encryption algorithms (RSA, DES, 3DES and AES) for information security. International Journal of Computer Applications, 2013.

[2] Subhedar, M.S. and V.H. Mankar, Current status and key issues in image steganography: A survey. Computer science review, p. 95-113, 2014.

[3] Hamed, G., et al. Comparative study for various DNA based steganography techniques with the essential conclusions about the future research. in Computer Engineering & Systems (ICCES), 11th International Conference on. IEEE 2016.

[4] Amin, M.M., et al. Information hiding using steganography. in Telecommunication Technology, 2003. NCTT 2003 Proceedings. 4th National Conference on IEEE 2003.

[5] Al-Mohammad, A., Steganography-based secret and reliable communications: Improving steganographic capacity and imperceptibility, Brunel University, School of Information Systems, Computing and Mathematics Thesis, 2010.

[6] G. Hamed et al., "Hybrid Technique for Steganography Based on DNA with N-Bits Binary Coding Rule ," in 7th International Conference on Soft Computing and Pattern Recognition (SoCPaR), IEEE, 2015.

[7] B. Anam et al., "Review on the Advancements of DNA Cryptography, eprint arXiv:1010.0186, 2010.

[8] G. Hamed et al., "DNA Based Steganography: Survey and Analysis for Parameters Optimization," in Applications of Intelligent Optimization in Biology and Medicine, Springer, ISSN:1868–4394, pp. 47–89, 2015.

[9] S. Das et al., "Steganography and Steganalysis: Different Approaches," International Journal of Computers, Information Technology and Engineering (IJCITAE), vol. 2, Issue No. 1, 2008.

[10] I. Peterson, "Hiding in DNA," Proceedings of Muse, 2001.

[11] L. Hood and D. Galas, "The digital code of DNA," Nature, vol. 421, 2003.

[12] The Structure of DNA. Retrieved October 11, from http://ircamera.as.arizona.edu/Astr2016/text/nucleicacid1.htm, 2016.
[13] M. BORDA and O. TORNEA, "DNA Secret Writing Techniques," in 8th International Conference on Communications (COMM), pp. 451– 456, 2010.
[14] Genetics. Retrieved from http://biologydiva.pbworks.com/w/page/47793659, October 11, 2016.
[15] Shiu, H., et al., Data hiding methods based upon DNA sequences. Information Sciences, 180(11): p. 2196-2208, 2010.
[16] Abbasy, M.R., et al., DNA base data hiding algorithm. International Journal of New Computer Architectures and their Applications (IJNCAA), 2(1): p. 183192, 2012.
[17] Atito, A., A. Khalifa, and S. Rida, DNA-based data encryption and hiding using playfair and insertion techniques. Journal of Communications and Computer Engineering, 2(3): p. 44, 2012
[18] Torkaman, M.R.N., N.S. Kazazi, and A. Rouddini, Innovative approach to improve hybrid cryptography by using DNA steganography. International Journal of New Computer Architectures and their Applications (IJNCAA), 2(1): p. 224-235, 2012.
[19] Mitras, B.A. and A. Abo, Proposed steganography approach using DNA properties. International Journal of Information Technology and Business Management, 14(1): p. 96-102, 2013.
[20] Liu, H., D. Lin, and A. Kadir, A novel data hiding method based on deoxyribonucleic acid coding. Computers & Electrical Engineering, 39(4): p. 1164-1173, 2013.
[21] Agrawal, R., M. Srivastava, and A. Sharma. Data hiding using dictionary based substitution method in DNA sequences. in Industrial and Information Systems (ICIIS), 2014 9th International Conference on. IEEE 2014.
[22] Manna, S., et al. Modified technique of insertion methods for data hiding using DNA sequences. in Automation, Control, Energy and Systems (ACES), First International Conference on. IEEE 2014.
[23] El-Latif, E.I.A. and M.I. Moussa, Chaotic Information-hiding Algorithm based on DNA. International Journal of Computer Applications (0975–8887), Vol. 122(10), 2015.
[24] Tank, R.M., H.D. Vasava, and V. Agrawal, DNA Based Audio Steganography. Oriental journal of Computer Science and Technology, 8: p. 43-48, 2015.
[25] Santoso, K., et al., Sector based DNA information hiding method. Security and Communication Networks, 9(17): p. 4210-4226, 2016.
[26] Marwan, S., A. Shawish, and K. Nagaty, DNA-based cryptographic methods for data hiding in DNA media. Biosystems, 150: p. 110-118, 2016.
[27] Marwan, S., A. Shawish, and K. Nagaty, Utilizing DNA Strands for Secured Data-Hiding with High Capacity. International Journal of Interactive Mobile Technologies, 11(2), 2017.

[28] Malathi, P., et al., Highly Improved DNA Based Steganography. Procedia Computer Science, 115: p. 651-659, 2017.
[29] El-Moursy. A.E., Elmogy. Mohammed, Atwan. Ahmad; " DNA-based cryptography: motivation, progress, challenges, and future", JOURNAL OF SOFTWARE ENGINEERING & INTELLIGENT SYSTEMS, ISSN 2518-8739, Volume 3, Issue 1, April 2018.
[30] Hamad.Safwat, Elhadad. Ahmed, and Khalifa. Amal; "DNA watermarking using Codon Postfix technique", IEEE TRANSACTIONS ON JOURNAL NAME, MANUSCRIPT ID, 2017.
[31] F. ALAA KADHIM, ALI. RASHA SUBHI; " HIDDEN ENCRYPTED TEXT BASED ON SECRETE MAP EQUATION AND BIOINFORMATICS TECHNIQUES", Journal of Theoretical and Applied Information Technology, ISSN: 1992-8645, E-ISSN: 1817-3195, Vol. 96, No 1, January 2019.
[32] Ahmed. Zena, Adil. Saher, and Al-Alak. Saif M.KH; "ECC Based Blind Steganography-DNA for Hidden Information", Journal of Engineering and Applied Sciences 14 (Special Issue 8), ISSN: 1816-949X, PP:10240-10244, 2019
[33] Kahdum. Ghaith Dehiaa, Al-Bawee. SaharAdill, Jasim. Mahdi Nsaif; "A Proposed DNA Postfix Hiding Method", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-4, November 2019
[34] Kahdum. Ghaith Dehiaa, Al-Bawee. SaharAdill, Jasim. Mahdi Nsaif; " Multilevel Security System Based on DNA and Magic Cube", MSC thesis, Babylon university- Science Collage for Women, 2020.

الخلاصة

يعد أمن المعلومات مصدر قلق رئيسي ، لا سيما مع نمو استخدام الإنترنت. بسبب هذا النمو ظهرت حالات اختراق للبيانات المرسلة منها الوصول غير المصرح به التي يتم التصدي له باستخدام تقنيات اتصال آمنة متنوعة وهي ؛ التشفير وإخفاء البيانات. تتعلق الاتجاهات الحديثة بالحمض النووي المستخدم في التشفير وإخفاء البيانات كحامل للبيانات من خلال استغلال خصائصه الجزيئية الحيوية. تقدم هذه الورقة استبيانًا حول البحوث المنشورة المستندة إلى الحمض النووي لاخفاء البيانات المهمة كحامي لها والمنقولة عبر قناة غير آمنة لمعرفة نقاط القوة والضعف فيها. لمساعدة البحث المستقبلي في تصميم تقنيات أكثر كفاءة وأمانًا للاخفاء في الحمض نووي.

الكلمات الدالة: الامنية, اخفاء البيانات, التشفير , اخفاء المعلومات