International Journal of Academic Engineering Research (IJAER)
ISSN: 2000-001X
Vol. 3 Issue 1, January – 2019, Pages: 1-5

DNA Computing and Its Application to Information and Data Security Field: A Survey

Omar G. Abood *, Shawkat K. Guirguis
Department of Information Technology, Institute of Graduate Studies and Researches, Alexandria University, Egypt.
** Corresponding author: omar.ghazi88@yahoo.com*

Abstract- *The emergence of the research topic of DNA in the field of Information Storage, Security and Cryptography has promised new heights to the field. DNA computing is a novel method of simulating the bio-molecular structure of DNA and computing. The promise it came with was a brand-new mean of data structure and calculation which would provide a brand-new generation for storage and more importantly, security and cryptography due to the wide array of parallelism in DNA. The field of cryptography would thrive if the DNA is properly used as it can realize several security technologies such as Encryption, Steganography, Signature and Authentication through the DNA molecular as information medium. This work presents the broad idea of using the existing DNA cryptosystems which utilize the use of the four DNA symbols (A, T, C and G) that represent the four binary two-tuples (00, 01, 10, and 11). However, the obstacles facing the DNA cryptography systems stand in the influence of Turing in the corresponding theoretical computing model. Where it stands in the theoretical stage, it seems to have the ability to solve major problems regarding the security of systems through cryptography.*
Keywords*:* DNA, DNA computing, Steganography DNA-Based Cryptography, Information security technology

1. Introduction

DNA computing presents a new method of the simulation of the bio-molecular structure of DNA and computing through the molecular biological technology which remains to be fresh and new with the potential of interdisciplinary growth. It can be used to face the increasing security threats which have turned information security into a main threat nowadays. The encryption of data is needed whilst transmitting to guarantee the security of the information post and during transmission.

Where cryptography is the practice of hiding information and data, cryptography goes hand-in-hand with it in the guarantee of the safety and security of said information. For instance, the use of DNA Computing methods has had the ability to break the widely known Data Encryption Standard (DES) [1][2]. DNA cryptography has thus been employed in the cryptographic field as a mean risen from the research practices of DNA Computing [3][4], where the DNA was used as an information carrier and the modern biological technology was used as a tool for implementation. In this section, we explore DNA, DNA Computing and Security and Cryptography Systems.

A. DNA (Deoxyribonucleic Acid)

All living organisms are presented in the DNA which is a molecule. What DNA does is that it transmits the genetic information required for growth, development and functioning of all living organisms from the Behemoth to the viruses. It is found in the nucleus of the cell (Nuclear DNA); however, only a small fraction of it is located in the mitochondria (Mitochondrial DNA). The genetic information stored in the DNA molecule is in the form of code which consists of four chemical bases. Those chemical bases are Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) which pair up with one another in a way such that each base is having a specific partner. The pairs are shown as follows:

1. Adenine with Thymine (A=T) and vice versa (T=A), and
2. Cytosine with Guanine (C=G) and vice versa (G=C).

Accordingly, they formulate units which are known as "Base Pairs" where each of those bases is also linked to two other molecules namely Sugar molecule and Phosphate molecule. The sugar and phosphate molecule are known as Nucleotide upon unity. In a DNA molecule, deoxyribonucleotides are joined into a polymer by a phosphodiester bond among one ribose and the 3' hydroxyl of the following, in this way left a free 5' end and a free 3' end. Typically, the atoms of the DNA are matched polymers collected from the 5' to the 3' end. The "double-helix" is the natural shape of DNA as both single and twofold DNA string parts can be unified outside the cell. In a twofold helix DNA string, two strands are corresponding as far as grouping, that is A to T and C to G as indicated by Watson-Crick rules, which is one of the best logical revelations of the twentieth century [5]. Figure 1 shows the DNA structure.

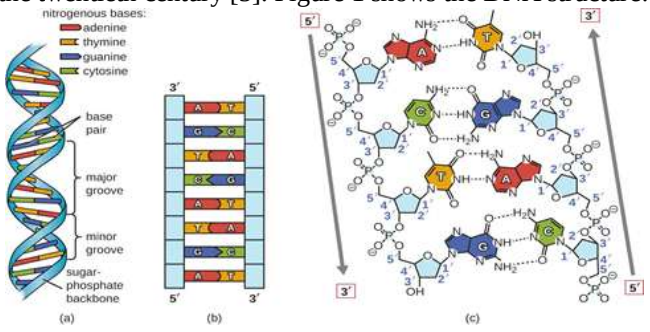

Figure 1: DNA Structure

B. DNA Computing

DNA as a technology for computing was developed by Leonard Adleman of the University of Southen California back in 1994 where he exhibited a proof-of-concept to solve the Seven-Point Hamiltonian path problem in which a DNA molecule was employed as a mean of computation [6]. The method was extended and further developed by Lipton in 1995. A couple of years later, Quyang brought forth a molecular biology-based experimental solution to the maximal clique problem [7]. Following that in the dawn of the new millennium, Liu blueprinted a DNA computing model system which is known as the surface-based DNA computing that had the ability to solve the satisfiability problem [8] which was later re-evaluated by Wu who improved the surface-based method [9].

In 2001, a programmable and autonomous computing machine formed out of biomolecules was designed by Benenson upon which a finite automaton could run [10]. Moving towards the year 2003, researchers of the Weizmann Institute of Science in Rehovot, Israel uncovered a programmable molecular computing device made up of DNA molecules and enzymes instead of the more used then Silicon Microchips [11]. On the 28th of April, 2004, Yaakov Benenson, Ehud Shapiro, Uri Ben-Dor, Binyamin Gil and Rivka Adar at the Weizmann Institute claimed to have developed a DNA computed attached to an input and output module in the journal Nature in which they claimed to have a theoretical ability to recognize cancerous activity in a cell and to have the ability to deliver the anti-cancer drug upon recognition of the activity [12]. A set of Shakespearean sonnets, an audio file of Martin Luther King, Jr.'s speech "I Have a Dream" and a JPEG picture were all stored on a DNA digital data storage the beginning of 2013 [13]. Later the same year, researchers brought into existence a transcriptor which is a biological transistor [14].

Interesting new computing paradigms were enriched by the success of DNA computing [15]. The principles of DNA Computing which were first introduced by Adleman to solve the Hamiltonian path problem follow the process of finding a path that begins at vin, ends at vout and enters every other vertex exactly once on a directed graph. For each vertex i in the graph, a random 20-mer oligonucleotide DNA sequence was generated. The following list presents the process that was directed to solve the Hamiltonian path problem:

1. Generating random paths through the graph,
2. Keeping the paths which begin with vin and end with vout only,
3. Keeping the paths which enter n vertices only if the graph has n vertices,
4. Keeping only the paths which enter all of the vertices of the graph once minimum, and
5. If any paths remain, say "yes", otherwise say "no" [2].

Adleman employed the DNA sequence encoding of all possible answers to the problems, removing the solutions that meet not the necessities through a series of restrictive conditions. The difference between traditional computing and DNA computing is obvious through the Hamiltonian path problem which will further be discussed. The DNA parallelism capabilities are exposed in Adleman's experiment.

2. Cryptography and Security Using DNA Computing

Security is represented through encryption or ciphering text which is the process of converting the plain text into encrypted, non-recordable text [16]. What cryptography is designed for is to hide information. Sensitive information requires the security of cryptography to ensure the security and secrecy of said information. Whereas security has been threatened numerous times in the recent years, DNA cryptography presents a novel approach that hopes to fill-in the gap apparent by the current security and cryptography systems. Table 1 shows a comparison between the Traditional Cryptography and DNA based Cryptography systems. In DNA cryptography, information is carried out through the DNA nucleotides (denoted by the letters A, C, G and T). Various studies in the past have attempted to invent a method of cryptography through DNA computing wherein 2003, Jie Chen employed a DNA cryptographic approach that consisted of a one-time pad, molecular theory and performed encryption and decryption of a 2-dimensional image [17].

Later, in 2004, Ashish Gehani presented another approach that was founded on DNA that also used a one-time pad [18]. This approach laid forth the foundation upon which DNA cryptography was developed on. According to the source of one-time pad, Vernam and Shannon claimed that it had "perfect secrecy". In 2012 [19] presented a lossless image steganography approach to hide a secret image in the cover image. DNA sequencing, Sudoku solution matrix and (t, n) threshold sharing systems were used to accomplish the approach to represent the secret image and cover image respectively. The camouflaging process is used to embed the secret image into cover image and stego image is obtained. The following year witnessed another approach based on DNA based-data embedding by Balado, [20] through the use of substitution mutation modelled. The Kimura model from the molecular evolution studies was used to improve the capacity of the DNA data.

In 2014, the chaotic logistic map for confusing and diffusing the image pixels, and then a DNA sequence was used as an OTP (one-time-pad) to change pixel values [21]. Later in 2015, Niyat et al. have implemented chaos-based image encryption using a hybrid cellular automaton and a DNA sequence [22].

Another work that was implemented by Das in 2015 [23], focused on the concept of using single stranded DNA as primary cover and analyzed the security loopholes of the traditional algorithms against visual and statistical attacks. The simulated results have proven that the Dual cover steganography provided better security than the previous existing algorithms. In 2016, two studies were implemented using DNA cryptography in the medical field where Akkasaligar et al. have proposed a secure medical image

encryption based on the intensity level through the use of the chaos theory and DNA cryptography [24]. Ochani et al. have also implemented another cryptography with steganography model to impose security through medical images which were made to hide patients related data into other images and transfer them securely [25].

Table 1 Comparison between Current Cryptography and DNA Cryptography methods

Items	Traditional Cryptography	DNA Cryptography
Ideal System	Silicon chip based	DNA chip based
Information Storage	Silicon computer chips	DNA strands
Storage Capacity	1 gm silicon chip contains 16 Mega-bytes	1 gm DNA chip contains 10^8 Tera-bytes
Processing Time	Slow	Fast
Performance Dependency	Implementation and system configuration	Environmental conditions

3. Problems Facing DNA Encryption

Whilst DNA computing presented a brand-new computing mode that has a lot of prospect, it could not deviate from the influence of Turing in the corresponding theoretical computing model. DNA remains to be in the theoretical stage, it has been used on a limited level to resolve certain problems where the varieties of problems have led to the discrepancy of computing schemes. Under the current DNA computing modes, the time complexity of DNA computing compared to the space complexity has not seen much increase with the computational complexity remarkably. That is because the fact that DNA computing has only got the ability to convert the time complexity into space complexity. Then, once the complication of problems breaks the physical limit of DNA segment which operated by the bio-chemical technique, DNA computing is still too far away to reach.

Mathematical cryptography has a tremendous ability of increasing the length of the cipher, thereby it'll prevent the cryptography from powerful attack using DNA computing. It is thus that in terms of the existing DNA computing modes, though DNA computing has massively improved the ability of the cipher break of people, it is disabled to build intimidation to the security of cryptography.

4. The Applications of DNA in Security

There are several applications that rely on DNA in security. Those applications are described in the following sub-sections as the DNA Encryption, DNA Steganography and DNA Certification which have all been employed in the recent years due to the significant ability of DNA computing in both its wide array of parallelism and its incredible storage capacity that remains to be unmatched.

A. DNA Encryption Techniques

In theory, one-time-pads security is entirely secure as decided by the randomness of the cipher-key. There are two steps that the algorithm requires:

1. The data of the cipher-key is random.
2. The cipher-key cannot be recycled for reuse.

However, one-time-pads are known to bring optimum security with two difficulty levels: producing large-scale random cipher-keys that run well with the plain code length and facing the problem of cipher-key saving and distributing. This makes the one-time-pads algorithm impossible. Yet, DNA presents an information carrier that has an enhanced memory density which is one hundred billion to one thousand billion times compared to the universally used disk memory. Gehani [18] used this idea to present the one-time-pads mechanism which was based on DNA in order to blueprint two encryption methods of one-time-pads of DNA sequence. Those methods are presented as:

1. Mapping Substitute: a method which translates the fixed length DNA plain code sequence cell to DNA cryptograph sequence according to the defined mapping graph, we call it mapping substitute.
2. The Exculsiveor: which uses biological molecular techniques to carry through exlusiveor operation of DNA plain code and cipher-key sequence.

Gehani also made use of the super parallelism and incomparable storage capacity of DNA computing in dissymmetric encryption mechanisms. Accordingly, the Gahani way had the ability of increasing the amount of information to obtain greater and more complex data structure of adding precise coding information.

B. DNA Steganography

The transmission of DNA is becoming more and more abundant and brief. It advances the transmission to reduce the costs and increase the security of information [27,28]. DNA steganography has more layers of protection than the simplex code encryption techniques that have provided a new idea for the security of information.

The main idea behind DNA steganography is hiding the information which requires encryption in the large numbers of irrelevant DNA sequence chains. It is thus that the only receiver capable of dining the correct DNA fragment based on the conventional information would be the proper one.

The first use of DNA steganography was done by Bancroft during WWII. Through the use of an alphabet of exoteric short nucleic acid sequence, the decryption of plain code information in DNA chains was possible and adding a special section marked information on the bottom of DNA chains. This kind of DNA has the ability to mix with the same length DNA chains which were split into multiple microdots. However, just a microdot contains DNA molecules which takes count of a hundred million [26].

Accordingly, the attackers, despite their ability to determine the information existing in a microdot among numerous microdots. It is thus that the way to decipher the information lies in looking for a special section of bottom mark which enables the usage of the method of DNA computing to search. Once the DNA chain had been confirmed through the mark, the receivers will adopt PCR to duplicate the same DNA chain along with acquiring the information by deciphering.

C. DNA Certification

DNA certification has little to do with DNA computing techniques except for using the biological characteristics of DNA. It is generally applied in the field of justice and finance among other fields which would certificate the accuracy of biological individuals.

In the beginning of the new millennium, the DNA Technology Company of Canada used the DNA sequence to certificate the products of the Sydney Olympic game. Almost 50 million keepsakes were all marked with a special type of ink from Olympic T-shirts to coffee mugs. The DNA segment used in the ink marks was randomly selected for extraction from an athlete's genome which made it very difficult to fabricate as the athlete was chosen from hundreds of others. This way, a portable scanner was used to identify and verify the information in the ink marking to certify whether or not the keepsakes were authentic.

DNA steganography can be used to appraise the virtues of DNA which enable further and wider certification. Currently, there are numerous biological genetic engineering methods ongoing and upcoming that are being developed. Those methods enable the researchers to add the DNA certification information to the organ tissue through identifying the information of the DNA certification that would be capable of validating and certifying the authenticity of the customer identities and copyright information.

5. The Prospect of DNA Cryptography in the Future of Technology

The future of DNA Cryptography seems to be bright with great potential awaiting had it been utilized right. The development of DNA Steganography and Certification have been rapidly improving through the past few years despite the fact that DNA Computing and DNA Cryptography remain to be in their theoretical cradle. Certification and Steganography have more than a singular layer of protection which has led them to heights beyond those of single encryption that has been employed in most businesses and fields.

Due to the storage capacity and vast parallelism of DNA, it has been shaped to have more advantages than its traditional counterparts in the field of cryptography, security and data encryption. The development of biotechnology and the discovery of a better DNA Encryption design will certainly improve the research on DNA Cryptography in Information Security. However, further research remains to be of essential need to DNA Cryptography in order to reach the limits of the new technology.

At the end, it is recommended that DNA Cryptography would be improved and strengthened through the employment of traditional methods such as AES and DES in order to stand a chance against the ever-growing and significantly powerful Quantum Computing. Through hybridizing DNA Cryptography with other traditional methods, it would be possible to reach a new record in the time it would take to break through the encryption even through the use of Quantum Computing. The complexity of a cipher key hybridized with DNA Cryptography and Traditional Methods would be far more increased.

6. Reference

[1] Boneh, D., Dunworth, C., Lipton, R. Breaking DES Using a Molecular Computer. Technical Report CS-TR,489-95, Department of Computer Science, Princeton University, USA, 1995.

[2] Adleman, L. M., Rothemund, P. W., Roweis, S., Winfree, E. On Applying Molecular Computation to the Date Encryption Strands in DNA Based Computers. Journal of Computational Biology, 1999, 6(1), 53-56.

[3] Gehani A, Labean, T., Reif, J. DNA-Based Cryptography. In: Aspects of Molecular Computing, Springer, Berlin, Heidelberg, 2003, 167-188.

[4] Kazuo, T., Akimitsu, O., Isao. S. Public-key System Using DNA as a One-Way Function for Key distribution. Biosystems, 2005, 81(1), 25-29.

[5] Vijayakumar, P., Zayaraz, G. An Improved Level of Security for DNA Steganography Using Hyperelliptic Curve Cryptography. Wireless Personal Communications, 2016, 89(4), 1221-1242.

[6] Adleman, L. M. Molecular Computation of Solutions to Combinatorial Problems. Science, 1994, 266(5187), 1021-1024.

[7] Ouyang, Q., Kaplan, P. D., Liu, S., Libchaber, A. DNA Solution of the Maximal Clique Problem. Science, 1997, 278(5337), 446-449.

[8] Liu, Q., Wang, L., Frutos, A. G., Condon, A. E., Corn, R. M., Smith, L. M. DNA Computing on Surfaces. Nature, 2000, 403(6766), 175.

[9] Wu, H. An Improved Surface-Based Method for DNA Computation. Biosystems, 2001, 59(1), 1-5.

[10] Benenson, Y., Paz-Elizur, T., Adar, R., Keinan, E., Livneh, Z., Shapiro, E. Programmable and Autonomous Computing Machine Made of Biomolecules. Nature, 2001, 414(6862), 430.

[11] Lovgren, S. Computer Made from DNA and Enzymes. National Geographic News, February 24, 2003. (Last accessed: 4/7/2018).
https://news.nationalgeographic.com/news/2003/02/0224_030224_DNAcomputer.html

[12] Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., Shapiro, E. An Autonomous Molecular Computer for Logical

Control of Gene Expression. Nature, 2004, 429(6990), 423.

[13] DNA stores poems, a photo and a speech|Science News, by Rachel Ehrenberg, 6:41PM, JANUARY 23, 2013. (Last accessed: 4/7/2018).https://www.sciencenews.org/article/dna-stores-poems-photo-and-speech

[14] Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., Endy, D. Amplifying Genetic Logic Gates. Science, 2013, 340(6132), 599-603.

[15] Ouyang, Q., Kaplan, P. D., Liu, S., Libchaber, A. DNA Solution of the Maximal Clique Problem. Science, 1997, 278(5337), 446-449.

[16] Omar, G. A., Mahmoud, A. E., Shawkat, K. G. Investigation of Cryptography Algorithms Used for Security and Privacy Protection in Smart Grid. Proceedings of IEEE 9th International Middle East Power Systems Conference, (MEPCON), Cairo, Egypt, Dec. 19-21, 2017, 644-649.

[17] Chen, J. A DNA-Based Biomolecular Cryptography Design. Proceedings of IEEE Proceedings of the 2003 International Symposium, (ISCAS'03) In Circuits and Systems, May, 2003, III-822-III-825.

[18] Gehani, A., LaBean, T.H., Reif, J.H. DNA-Based Cryptography. In Aspects of Molecular Computing, Springer, Berlin, Heidelberg, DNA Based Computers V. Providence, 2004, 167-188.

[19] Chakraborty, S., Roy, S., Bandyopadhyay, S. K. Image Steganography Using DNA Sequence and Sudoku Solution Matrix. International Journal of Advanced Research in Computer Science and Software Engineering, 2012, 2(2).

[20] Balado, F. Capacity of DNA Data Embedding Under Substitution Mutations. IEEE Transactions on Information Theory, 2013, 59(2), 928-941.

[21] Mokhtar, M. A., Gobran, S. N., & El-Badawy, E. S. A. Colored Image Encryption Algorithm Using DNA Code and Chaos Theory. Proceedings of IEEE International Conference on Computer and Communication Engineering, Kuala Lumpur, 2014, 12-15.

[22] Niyat, A. Y., Hei, R. M. H., Jahan, M. V. Chaos-Based Image Encryption Using a Hybrid Cellular Automata and a DNA Sequence. International Congress on Technology, Communication and Knowledge (ICTCK), Mashhad, 2015, 247-252.

[23] Das, P., Deb, S., Kar, N., Bhattacharya, B. An Improved DNA Based Dual Cover Steganography. Procedia Computer Science, 2015, 46, 604-611.

[24] Akkasaligar, P. T., & Biradar, S. Secure Medical Image Encryption Based on Intensity Level Using Chao's Theory and DNA Cryptography. Proceedings of IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, 1-6.

[25] Ochani, A., Jadhav, D., Gulwani, R. DNA Image Encryption Using Modified Symmetric Key (MSK). Proceedings of IEEE International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 2016, 1-4.

[26] Vinodhini, R. E., Malathi, P. DNA Based Image Steganography. In Computational Vision and Bio Inspired Computing Springer, Cham, 2018, 819-829.

[27] Karl, L. DNA Computing: Arrival of Biological Mathematics. The mathematical intelligencer, 1997, 19(2), 9-22.

[28] Kamei, T., Kishii, N., Kurihara, K., Kobayashi, T., Iwamoto, H., Tsuboi, H. DNA-Containing Inks and Personal Identification System Using Them Without Forgery. Jpn. Kokai TokkyoKoho,2002, 8.