Review on DNA Based Encryption Algorithm for Text and Image Data

Mohit Rusia[1], Reader Hemant Makwana[2]

[1]Dept. of IT, Institute of Engineering & Tech.,
[1]D.A.V.V., Indore, India
[2]Dept. of IT, Institute of Engineering & Tech.,
[2]D.A.V.V., Indore, India

Abstract

Human always learns from the examples previously proposed and implemented objects in real world. Biological examples and their solutions are more helpful in various applications, such as genetic algorithm, swarm computing. The exceptional energy efficiency and extraordinary information inherent in DNA nucleotide sequences are being explored for DNA computing, data storage and cryptography. DNA cryptography is an emerging field of cryptography. In this paper the various encryption algorithm proposed earlier are studied and compared and thus the conclusion for the improvement is proposed.

Keywords- DNA nucleotide sequences, DNA Computing.

1. Introduction

Bioinformatics which includes genetic algorithm, swarm computing is an application of computer science for the management of biological information. This biological and genetic information can be gathered, stored analysed and integrated so as to be applied to gene-based element discovery and development. The need for bioinformatics proficiencies has been explored by applying computer science, mathematics and engineering or the combination of such technologies on the biological data. The field of bioinformatics includes the genetic algorithms, DNA computing, Swarm computing etc. which are discussed as under.

Genetic Algorithms are adaptive search procedures which are listed on Charles Darwin theory of the survival of the fittest. GAs produce a population in such a way that the attribute which is popular, i.e., has higher fitness value is replicated more, which is done by the nature. This is also the fundamental concept behind advancement. So, these algorithms are also referred as the evolutionary algorithms. Genetic Algorithms (GAs) are search procedures to converge to optimal solution based on the theory of survival of the fittest. The main entity of GA is chromosome. Each chromosome suggests a solution to the problem and is composed of a string of cells of finite length. The binary alphabet {0, 1} is often used to represent these cells but integers can be used depending on the application. The fitness value is a rationale or function against which chromosome is tested for its suitability to the problem in hand. The other techniques like the DNA implements the genetic algorithms to solve many other problems like the distribution problem.

DNA, or deoxyribonucleic acid, is the hereditary material in humans and almost all other organisms. Nearly every cell in a human's body has the same DNA. Mostly, DNA is located in the cell nucleus (where it is called nuclear DNA), but a small DNA can be found in the mitochondria (where it is called mitochondrial DNA or mtDNA). The following figure shows the actual DNA structure [2].

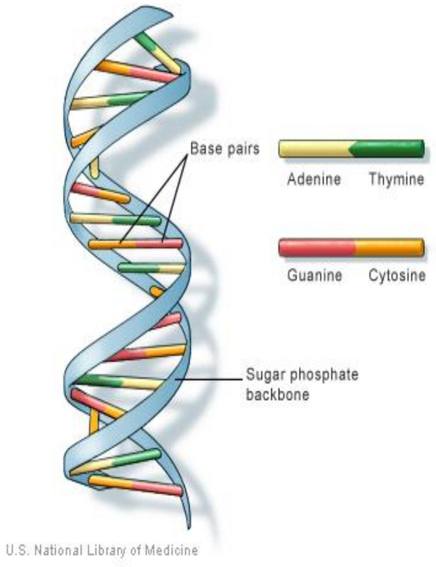

Figure 1. DNA structure [2]

The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T), also known as nucleotide sequences. DNA bases pair up with each other, A with T and G with C, to form units called base pairs. The order, or sequence, of these bases determines the information available for building and maintaining an organism, in a similar way like in letters of the alphabet appear in a certain order to form words and sentences.

All such biometrical technologies helps us in solving any kind of optimisation problem which can be described with the chromosome encoding. It also solves the solution structure and solution parameter problems and pattern recognition problems. These algorithms can be easily transferred to existing simulation and models. But rather it is expensive to use, the genetic algorithm cannot assure constant optimisation response times.

This survey paper discuses some of the earlier techniques used in the similar domain. The organization of paper will be as follows: Overview Section presents the overview of the security techniques, its advantages and disadvantages on use of it. Literature Survey Section discusses the various approaches made earlier in the same domain by the researchers and the comparative study of these approaches. Finally the conclusion and future work discusses the result and the conclusion of this paper.

2. Overview

Data Security refers to those protective measures that are applied to prevent unauthorized access to data or documents. It is a practice of securing the data from those persons who may misuse it in many ways like disclosure, disruption, modification, perusal, inspection, recording or destruction. Data Security is also known as information security (IS) or computer security. Information flows throughout the network which may be local or of global scope. It is required to secure that information to prevent unauthorized access of it in its path. One need to ensure a right security infrastructure mainly for privacy, integrity and confidentiality in the network for it to be reliable and dependable for information exchange. For that one should encode the data before sending it through various encoding mechanisms available to make it unreadable. This is where the cryptography comes into picture.

Cryptography is the process of converting plain text into cipher text and cipher text into plain text. There are basically two types of cryptography which are namely symmetric and asymmetric cryptography. Here the Plain text refers to a sequence of characters drawn from a finite alphabet, such as that of a natural language and cipher text refers to an encrypted sequence of plain text. In cryptographic processes i.e. Encryption and Decryption is used for security. Here Encryption is the process of scrambling the plaintext using a known algorithm and a secret key, and Decryption is the reverse process, which converts the encrypted message back to the original form using the same or different key. The goal of encryption is to avoid decryption by an adversary who does not know the secret key. A resilient cryptosystem is one for which successful cryptanalysis is not possible by the attacker.

One approach for providing security in a high density information storage area like the DNA can be achieved through the DNA based cryptography technique which uses various ways to secure the data in DNA by converting it into the nucleotide sequences and arranging these in a complex order based on the given data so that it cannot be compromised for many years.

The information security techniques are very advantageous in the manner that by using these one can protect data while the valuable information is in use and while it is being stored. Also the information is kept private which also increases its value and importance for the others also. But there should also be a special concern to be kept while selecting the technique or algorithm for security because as the technology increases the ways to compromise the data security technique also increases.

2.1. Prior Works

There has been much advancement in the use of DNA as a data storage device. One of the most critical steps in the realization of biological data storage is the conversion of digital data to nucleotide sequence. The following few mentioned works which tried to encode the information to be stored in biological sequence in a secure manner.

Battail proposed the idea of using hereditary media as a media for information transmission in communication process [4]. Shuhong Jiao devised a code for DNA based cryptography and stegano-cryptography and implemented in artificial component of DNA [5]. Nozomu Yachie used keyboard scan codes for converting the information to be encoded into hexadecimal value and finally into binary values. The final step was to translate the bit data sequence into four multiple oligonucleotide sequence. This was then mapped with the nucleotide base pairs [6]. Chinese

University of Hong Kong used Quaternary number system to transform the information for mapping it to nucleotides. First they obtained ASCII value of the information and used the mapping table 0=A, 1=T, 2=C and 3=G for the development of nucleotide strand. In this method of encoding nucleotides the number of binary bits used for representing the digital information was same as the nucleotide strand [7]. Simailarly, an approach to store and hide the data with some protective measures was applied by D. Prabhu and M. Adimoola from University of Chennai, India who introduced an encryption algorithm based on number conversion, DNA digital coding and PCR amplification which can prevent attack [1].

3. Literature Survey

1. Encryption through number conversion and DNA digital coding[1]-

This paper used the encryption algorithm by dealing with the bits obtained on number conversion. The intended PCR two primer pairs was used as the key of this scheme. This operation could increase the security of the proposed scheme. Here the complexity of biological difficult problem and cryptography computing difficulties provide a double layer security. And the security analysis shows that encryption scheme has high confidential strength.

Table 1. The comparison between various approaches in DNA Cryptography domain

S. No.	Paper Title and Authors	Strengths	Security Approach adopted	Weaknesses
1.	Bi-serial DNA Encryption Algorithm (BDEA)[1] by D.Prabhu, M.Adimoolam	The intended PCR two primer pairs was used as the key of this scheme that not independently designed by the sender or receiver. This operation could increase the security of encryption method.	Number conversion, DNA Digital coding, PCR amplification	Because of the use of two primers using the PCR amplification, the computational complexity is higher.
2.	DNA secret writing Techniques[10] by Monica BORDA, Olga TORNEA	The cryptanalysis is hard because of chromosome indexing and use of OTP.	One-Time-Pad (OTP), DNA XOR OTP and DNA chromosomes indexing	Use of steganography technique to encrypt data can be compromised if attacker knows that data exists.
3.	A Pseudo DNA cryptography Method[8] by Kang Ning	The theoretical analysis results that this method is powerful against certain attacks, especially against brute force attacks and also through the translation and transcription cipher text is more complex.	Transcription, Splicing, Translation -mRNA form of data into protein according to genetic code table and key send to the receiver in a secure channel[3]	Length of the cipher text is much higher than plaintext and also the partial information exists after encryption and can be compromised easily.
4.	DNA-Based Cryptography[9] by Ashish Gehani, Thomas LaBean, and John Reif	DNA substitution and XOR methods are based on one-time-pads, which are in principle unbreakable.	DNA substitution and one-time pads and the DNA Steganography techniques	Stenographic techniques rests on the assumption that the adversary is unaware of the existence of data and DNA substitution is not that much secure method.
5.	Comparison and Performance Evaluation of Modern Cryptography and DNA Cryptography [11] by Angeline, Priyadarshini, Thiruthuvadoss	Highly secure as it uses the Triple DES and also the time is reduced. Also the algorithm is simple to perform the operation in a less time.	Triple DES security algorithm with One-Time Pad (OTP)	Consumes more memory. The random number generator may generate the same number after a certain period.

2. DNA Secret Writing through XOR, OTP and BMC[10]-

In this paper the two principles, in combination are adopted-first is bimolecular computation (BMC) and second is algorithms for DNA cryptography like XOR operations and the use One-Time Passwords or Pads (OTPs).

This scheme discussed the application of such hybrid technique of cryptography in security and also the high efficiency and randomness in DNA.

3. A Pseudo DNA Cryptography[8]-

The pseudo cryptography was developed by Ning Kang. In this method, the original information is converted into a DNA nucleotide sequence. This in-turn is converted into two forms of DNA namely Spliced form and Protein form. For this, introns are cut into specific patterns. Actual DNA sequence is not used by this method. Instead it uses the mechanisms of DNA functions, hence its name is Pseudo DNA cryptography.

4. DNA based cryptography[9]-

This paper use OTP as the basic key for encryption. The researchers details the procedures for two DNA one-time-pad encryption schemes as: (i) a substitution method using libraries of distinct pads, each of which defines a certain, randomly generated, pair-wise mapping; and (ii) the XOR scheme utilizing molecular computation and indexed random key strings. Such methods can be applied either for the encryption of natural DNA or for artificial DNA for encoding binary data.

5. Comparison and Performance Evaluation of Modern Cryptography and DNA Cryptography [11]-

Here the Triple DES (TDES) algorithm is used with the DNA cryptography scheme and then the comparison is done among process running time, key size, complexity and cryptographic strength. They also used OTP as the key for encryption process enabling high confidentiality of the plaintext. The paper concludes that along with the practice of Triple DES methods, the DNA methods of cryptography can also be included in practice so that, with the practical implementations of the DNA cryptosystem, the enhanced ways of attaining the security for a huge message with less time can be possibly be attained.

3.1. Proposed Solution

The various approaches used in these papers are beneficial as well as having some weaknesses. So on the basis of such schemes a new approach can be used which encrypt the original data by using various transformations like the bit conversion of the data and then applying the hexadecimal conversion which maps with the MD5 (message digest) applied over the original data. Thus this scheme secures the data from attackers and then afterwards maps the data with the nucleotide sequences, thus making data more secure and finally a compression technique can be introduced which deals with the decrease in the cipher text resulted out. This algorithm reduces the overhead of the large size of encrypted file and also secures the data through the mapping table which changes with the plaintext applied over the algorithm.

4. Conclusion

Now, the hybrid cryptography has achieved a new goal in the domain of bioinformatics. The bioinformatics when worked especially upon DNA sequences, then the possibility of storing the huge amount of data in a secure manner increases. There are many variants in the method of applying hybrid cryptography technique which can be applied over the DNA sequences as discussed in the comparison table [Table 1]. These variants benefits in one or the other way to protect data from being compromised in any way.

The solution proposed helps in benefiting over the memory constraint as it reduces the size of the cipher text. Also, since its complexity is not very high so it can be a very efficient method to be applied for encrypting the data. The future enhancement can be that the method can be implemented for diverse kind of data.

10. References

[1] D. Prabhu, M. Adimoolam, "Bi-serial DNA Encryption Algorithm (BDEA)", Submitted on 13 Jan 2011, arXiv:1101.2577

[2] http://www.indy.gov/eGov/County/FSA/Pages/DNA.aspx

[3] Grasha Jacob, A. Murugan, "DNA based Cryptography: An Overview and Analysis", *Int. J. Emerg. Sci.*, 3(1), 36-42, March 2013 ISSN:2222-4254 © IJES

[4] Battail, G.: Heredity as an Encoded Communication Process. *IEEE Transactions on Information Theory* 56(2), 678–687 (2010)

[5]Jiao, S., Goutte, R.: Code for encryption hiding data into genomic DNA of living organisms. In: Signal Processing ICSP 2008. pp. 2166–2169(2008)

[6]Yachie, N., Sekiyma, K., Sugahara, J., Ohashi, Y., Tomita, M.: Alignement- Based Approach for Durable Data Storage into Living Organisms. Biotechnol. Prog. 23, 501–505 (2007)

[7]Chinese University of Hong Kong, http://www.cuhk.edu.hk/cpr/pressrelease/101124e.htm

[8]Ning Kang, A pseudo DNA cryptography Method, http://arxiv.org/abs/0903.2693, 2009

[9]A. Gehani, T. LaBean, and J. Reif, “DNA-Based Cryptography”, Lecture Notes in Computer Science, *Springer*, 2004.

[10]Borda M.E, Tornea O, “DNA secret writing Techniques” *IEEE* conferences 2010

[11]Angeline Priyadharshini, Thiruthuvadoss, “Comparison and Performance Evaluation of Modern Cryptography and DNA Cryptography”, Department of System on Chip Design Masters of Science 2012

[12]Asha Cherian, Surya R. Raj, Abey Abraham, “A Survey on different DNA Cryptographic Methods”, IGNOU, India, *International Journal of Science and Research (IJSR)*, India Online ISSN: 2319-7064

[13]J. Ziv, A. Lempel, “A universal algorithm for sequential data compression”, *IEEE Trans. Inf. Theory*, IT-23:337-343, May 1977.

[14]Pankaj Rakheja, Amanpreet kaur “A Unique Cryptographic Mechanism for Encoding Data Using DNA Structure”, in International Conference on Network Communication and Computers (ICNCC 2011) organized and sponsored by IACSIT, *The Institute of Electrical and Electronics Engineers (IEEE)*, Singapore Institute of Electronics and other organizations.

[15]Atul Kahate, “*Cryptography and Network Security*”, Tata Macgraw Hill,2009

[16]Jonathan P.L. Cox, “Long-term data storage in DNA”, Dept. of Chemistry, University of Bath, Bath, UK BA2 7AY, *TRENDS in Biotechnology* Vol.19 No.7 July 2001

[17]Komal Kumbharkar, Nazma Shaikh , Shraddha Yemale, Prof. Kanchan Doke, “An improved Symmetric key cryptography with DNA based strong cipher”, Computer Engineering, Bharati Vidyapeeth College of Engineering, Mumbai University, Navi Mumbai, India, *IJAIR* Vol. 2 Issue:3-ISSN:2278-7844.

[18]*Gorti VNKV Subba Rao, Md.Sameeruddhin Khan, Mr.A.Yashwanth Reddy, Mr.K.Narayana,* “ Data Security in Bioinformatics”, *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 3, Issue 11, November 2013 ISSN: 2277 128X.