Annals of the University of Craiova, Mathematics and Computer Science Series
Volume 37(3), 2010, Pages 147–151
ISSN: 1223-6934

Aspects of DNA Cryptography

Calina Popovici

Abstract. Cryptography is both a science and an art, dating back for thousands of years. Science developments have made possible the creation of more sophisticated systems. A combination between biology and cryptography, between DNA and encryption is a quite new and interesting area. Despite its requirements, it looks promising for the future.

2010 Mathematics Subject Classification. Primary 94A60; Secondary 78A70.
Key words and phrases. DNA, cryptography, bioinformatics, Deoxyribonucleic acid .

1. Introduction

Since time immemorial, man has devised schemes to conceal communications, whatever they are. Data security is very important, and often must be maintained at any cost and by any means. From the political level to that of individuals, information is priceless. Businesses trying to defend their trade secrets, unlawful business representatives hide information from the law and they must decipher the message. Phone surveillance has created a veritable psychosis among political figures and businessmen who accused special services of such practices. We come in contact with cryptology every day when we pay taxes, do online shopping or go to work.

Cryptology is very old, but lately it has been granted renewed interest. Cryptology is the study of both cryptography, the use of hidden messages in codes or ciphers, and cryptanalysis, or deciphering coded messages. Cryptology is almost as old as civilization itself, although ciphers and codes prior to the medieval period, or, more precisely, the end of medieval period, tended to be extremely simple, judging by today's standards. By the late 1980s, it was reserved to the military or diplomatic world, but is now publicly accessible. Cryptology contains the art of hiding information and techniques to learn a secret. Cryptography is the art of communicating confidentially through an insecure channel. Cryptanalysis is the art of deciphering these communications when one is not the meant receiver. And cryptology is a combination of these two areas.

Mathematical developments have made possible the evolution of more sophisticated systems. Further improvements in the field of cryptography have accompanied the creation of modern militaries and intelligence services of the nineteenth century. After the world wars and the creation of computers, cryptology entered into a more developed stage, resulting in the creation of codes and ciphers so sophisticated, that just any man, however brilliant, cannot break without the aid of technology. At the same time as computer technology development, transfer of information in digital form has increased rapidly. There are many applications, including email systems, banking systems and data processing, where the information that is transferred must pass through communication channels that may be monitored, listened to by unauthorized

Received July 08, 2010. Revision received September 08, 2010.

persons. While the required degree of security could vary for different applications, it is generally important for all these examples for the substance of communications to pass directly from the desired transmitter to receiver without intermediate parties that can interpret the message which is being transferred.

Bioinformatics is a field of science in which biology, computer science and information technology all merge into a single discipline. The ultimate goal of this science is to enable the discovery of new knowledge in biology and to create a global perspective from which unifying principles in biology can be discerned. There are three major research directions in bioinformatics: the development of new algorithms and statistics with which they could be extracted from a large number of data elements which have common features, the analysis and interpretation of data on different types of nucleotides and amino acid sequences, protein structure, and to develop and implement tools to enable efficient access and manipulation of various types of information.

A rapidly-developing technology is DNA computation or, more generally, biomolecular computation. It has emerged as a viable sub-discipline of science and engineering. [4]

2. DNA

Deoxyribonucleic acid (DNA) is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses. It is composed of the most complex organic molecules. The substance is found in every cell of living beings and is essential for the identity of any organism, from the Euglena viridis, a small unicellular creature bordering between plants and animals, to Homo sapiens sapiens, modern man. The main role of DNA molecules is the long-term storage of information. DNA is often compared to a set of blueprints, like a recipe or a code, since it contains the instructions needed to construct other components of cells, such as proteins and RNA molecules. The DNA segments that carry this genetic information are called genes, but other DNA sequences have structural purposes, or are involved in regulating the use of this genetic information. The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Human DNA consists of about 3 billion bases, and more than 99 percent of those bases are the same in all people. The order, or sequence, of these bases determines the information available for building and maintaining an organism, similar to the way in which letters of the alphabet appear in a certain order to form words and sentences. The German biochemist Frederich Miescher first observed DNA in the late 1800s. But nearly a century passed from that discovery until researchers unraveled the structure of the DNA molecule and realized its central importance to biology.

For many years, scientists debated which molecule carried life's biological instructions. Most thought that DNA was too simple a molecule to play such a critical role. Instead, they argued that proteins were more likely to carry out this vital function because of their greater complexity and wider variety of forms.

The importance of DNA became clear in 1953 thanks to the work of James Watson, Francis Crick, Maurice Wilkins and Rosalind Franklin. By studying X-ray diffraction patterns and building models, the scientists figured out the double helix structure of DNA - a structure that enables it to carry biological information from one generation to the next.

Biological research is constantly growing. A special part in this growth is played by the assimilation of computational methods. The amount of information is enormous and so powerful computational tools are necessary, in subjects such as discovering patterns in large assemblies of sequences of nucleic acids and proteins, biopolymer structure analysis and prediction, genomics, biomolecular sequence analysis.[3]

3. DNA Cryptography

DNA cryptography is a quite new area of cryptography that has emerged with the research of DNA computing, in which DNA is used as information carrier and the modern biological technology is used as implementation tool. The vast parallelism and extraordinary information density that are inherent in DNA molecules are explored for all sorts of cryptographic purposes such as encryption, authentication, signature, and so on.

Nowadays, the field of biology and that of cryptography have come to combine. DNA has come to be viewed as a means for storing information and computing on a giant scale. The study of DNA can be applied in molecular cryptography systems that are based on DNA and one-time-pads, a type of encryption that, if used correctly, is virtually impossible to crack. The size of the one-time-pad can limit the practical utilization of such a cryptographic system. There are various procedures for DNA one-time-pad encryption schemes.[2]

From a cryptographic point of view, DNA is very powerful. The binding capabilities of nucleotide bases (A-T, C-G) offer the opportunity of creating self-assembly structures that are an excellent means of executing computations. Another advantage is that DNA has a huge storing capacity, but on the other hand practically using the implementations requires a lot of time and resources. Simple and effective algorithms are necessary.[5]

DNA has a random character and so the cryptography which is based upon it is in principle unbreakable. When discussing a DNA-based cryptosystem, carbon nanotube-based message transformation is to be discussed.[1]

There are several DNA-based algorithms that have been practically applied. An example is DNA-based watermarking using the DNA-Crypt algorithm. Watermarks that are based on DNA sequences can also be used in order to identify the unauthorized use of genetically modified organisms that are protected by patents. Existing DNA cryptographic and steganographic algorithms use synthetic DNA sequences to store binary information. They may alter the DNA sequence when they are used on living organisms. The DNA-Crypt algorithm and image steganography are based on the same watermark-hiding principle, namely using the least significant base in case of DNA-Crypt and the least significant bit in case of the image steganography. It can be combined with binary encryption algorithms like AES, RSA or Blowfish. Mutations, which cause changes in the reading frame, are problematic and are not appropriate for DNA steganography. Mutations, which change a non-synonymous codon to a synonymous codon or vice versa are more important as these mutations cause errors in the encrypted information. The relevance of these errors depends on the encrypted information. If the encrypted information is an image, e.g. a logo, there would be only a linear colour shift in the image, which is not very relevant and can be corrected very easily.

Another example is a multi-level image encryption algorithm based on chaos and DNA coding. Chaos has been given a lot of attention when discussing image encryption. In this case the location of pixels and pixel values are changed by using a combination of chaos and DNA coding. The algorithm begins by changing the digital image into DNA coding matrix based on DNA code rules. Afterwards, the matrix is divided into four sub-matrixes, which are then modified by a logistic chaotic sequence. The sub-matrixes are then scrambled by adding new logistic sequences and getting the new sub-matrixes. The last step consists of joining all the scrambled DNA sub-matrixes in a DNA sequence and decoding the scrambled DNA matrix to an image matrix.[6]

An approach presents the way in which DNA binary strands can be used for steganography, which encrypts by hiding information, in order to provide rapid encryption and decryption. It is shown that DNA steganography based on DNA binary strands is secure under the assumption that an interceptor has the same technological capabilities as sender and receiver of encrypted messages. Another approach is based on steganography and a method of graphical subtraction of binary gel-images. It can be used to constitute a molecular checksum and can be combined with the first approach to support encryption.

Using DNA in relation with cryptography is a new and exciting study direction. Unfortunately, it needs a lot of resources, it has high tech laboratory requirements and several computational limitations. Thus, the efficient use of DNA cryptography is still difficult from a practical point of view.

4. Algorithms

4.1. RSA. Generate two large prime numbers, p and q

```
Let n = pq
Let m = (p-1)(q-1)
Choose a small number e, coprime to m
Find d, such that de
Publish e and n as the public key.
Keep d and n as the secret key.
Encryption
C = Pe
Decryption
P = Cd
```

4.2. DNA criptography algorithm. Step 1: The binary data, text or image, is used under the form of ASCII code.

Step 2: The numerical values are arranged into a string and they are grouped, being taken several digits at once.

Step 3: The numbers that have resulted are encrypted using a public key, with the result of another set of numbers.

Step 4: The encrypted sequence is transformed in binary.

Step 5: The binary form is transformed in DNA by using substitution. A is considered to be 00, C is 01, G is 10 and T is 11.

Step 6: All sequences are bound together, into a single strand, the ciphertext.[5]

5. Conclusions

Man has always needed some form of cryptography in order to conceal and protect information. Due to the development and advance of science, the possibilities are rapidly growing. DNA in connection to cryptography is a fast developing interdisciplinary area. More and more ideas are put into practice. The future of this area looks very promising, seeing as DNA is a medium for ultra-compact information storage.

References

[1] J. Chen, A DNA-based, biomolecular cryptography design, *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, 25-28 May 2003.

[2] A. Gehani, T. LaBean and J. Reif, *DNA-Based Cryptography*, Department of Computer Science, Duke University.

[3] A.K. Konopka and M.J.C. Crabbe, *Compact handbook of computational biology*, 2004.

[4] J.H. Reif, The emerging discipline of biomolecular computation in the US, *NEW GENERATION COMPUTING* **20**, no. 3, 217–236.

[5] O. Tornea and M. E. Borda, DNA Cryptographic Algorithms, *International Conference on Advancements of Medicine and Health Care Through Technology, IFMBE Proceedings* **26** (2009), 223–226.

[6] Q. Wang, Q. Zhang and C. Zhou, Key Laboratory of Advanced Design and Intelligent Computing, Dalian University, Ministry of Education, Dalian, 116622, China A multilevel image encryption algorithm based on chaos and DNA coding, *Bio-Inspired Computing, 2009. BIC-TA '09. Fourth International Conference on*, 16-19 Oct. 2009, Beijing, China.

E-mail address: calina.popovici@gmail.com