Microdots
DNA Steganography

Wissam Makki Alwash
College of Law, University of Babylon

Abstract

As well as new modern encryption algorithms are found or created, the process of breaking them is become more easy and uncomplicated in few years after or even months because of the high evolution in computer and mathematics sciences. The field of information security looks in new guidelines to shield the data it transmits. The idea of using DNA computing in the fields of cryptography and steganography has been recognized as a new possible technology that may bring forward a new hope for unbreakable algorithms. Researches have been performed in both cryptographic and steganographic situations with respect to DNA computing but researchers are still looking at much more theory than practicality. The constraints of its high tech lab requirements and computational limitations combined with the labor intensive extrapolation means, illustrate that the field of DNA computing is far from any kind of efficient use in today's security world. This paper will summarize some of the fundamentals of DNA and DNA computing and its use in the area of steganography and shows some of new ideas to enhance the established method.

الخلاصة

بالرغم من وجود أو اختراع خوارزميات التشفير الحديثة، فإن عملية كسرها أصبحت أكثر سهولة وغير معقدة في سنوات قليلة أو حتى بعد شهور بسبب التطور العالي في مجال علوم الكمبيوتر والرياضيات. إن مجال أمن المعلومات يضع مبادئ توجيهية جديدة لحماية البيانات عن طريق حماية المعلومات التي يقوم بنقلها. ان فكرة استخدام حوسبة الحمض النووي في ميادين التشفير وعلم إخفاء المعلومات قد تم الاعتراف بأنها التكنولوجيا الجديدة المحتملة التي قد تطرح أملا جديدا لخوارزميات غير قابلة للكسر. أجريت أبحاث في حالات كل من التشفير و علم إخفاء المعلومات فيما يتعلق بحوسبة الحمض النووي و لكن الباحثين لا يزالون يبحثون نظريا أكثر مما هو عملي. إن وسائل العمل المكثفة و المجهدة مجتمعة مع متطلبات تكنولوجيا المختبر العالية و القيود الحسابية، بيّنت بأن حقل حوسبة الحمض النووي بعيد كل البعد عن أي نوع من الكفاءة عند استخدامه في مجال الحماية في عالم اليوم. تلخص هذه الورقة بعض أساسيات الحمض النووي ، و حوسبة الحمض النووي واستخدامه في مجال علم إخفاء المعلومات و يبين بعض الأفكار الجديدة التي أنشئت لتعزيز النظريات الموجودة.

Introduction

The theories of using DNA computing in the field of data security methods to ban or thwart the forgery activity appears late in now days. How realistic are these theories and is it feasible to see these technologies changing the security marketplace of today [**Ivars , 2000**].

Before starting to understand the principles of DNA computing we must first illustrate what DNA actually is. All organisms on this planet are made of the same type of genetic blueprint which bind us together. The way in which that blueprint is coded is the deciding factor as to whether you will be bald, have a bulbous nose, male, female or even whether you will be a human or an oak tree [**Donald , 2001**].

Materials and Methodology

Cryptography has been shown recently as a new application of DNA computing [**Clelland *et al.*,1999**]; Clelland et al. have demonstrated a method to steganography by hiding secret messages encoded as DNA strands among a multitude of random DNA. Steganography means hiding of secret messages among other information to conceal their existence [**Kahn, 1967**].

The first material used is the DNA. Inside the cells of any organism there is a material called Deoxyribonucleic Acid (DNA) which is a double-stranded helix of nucleotides which carries the genetic information of a cell. This information is the code used within cells to form proteins and is the building block upon which life is formed. Strands of DNA are long polymers of millions of linked nucleotides. These nucleotides consist of one of four nitrogen bases, a five carbon sugar and a phosphate group. The nucleotides that make up these polymers are named after the nitrogen base that it consists of: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). These nucleotides will only combine in such a way that C always pairs with G and T always pairs with A [**Donald ,2001.**]. For example, a single-stranded DNA segment consisting of the base sequence TAGCCT will stick to a section of another strand made up of the complementary sequence ATCGGA. The links between pairs of bases are responsible for binding together two strands to form the characteristic double helix of a DNA molecule [**Ivars , 2000**]. The two strands of a DNA molecule are antiparallel, where each strand runs in an opposite direction. The combination of these 4 nucleotides in the estimated million long polymer strands can result in billions of combinations within a single DNA double-helix. These massive amount of combinations allows for the multitude of differences between every living thing on the planet from the large scale (mammal vs. plant), to the small (blue eyes vs. green eyes).

One of the methods used in this paper is Steganography, the branch of information security that attempts to conceal the existence of data through such strategies as invisible inks, secret compartments, and use of subliminal channels [**Alfred , 1997**]. Steganography is one of the oldest methods used for message security. The threat of enemy interception was original motivation to the development of techniques for disguising a message so that only the intended recipient could read it. As early as the 5th century B.C., the Spartans used a *skytale*, the first military cryptographic device (really a form of steganography) for disguising military messages (A skytale consisted of a staff of wood around which a strip of papyrus, leather, or parchment was wrapped. The secret message was written on the strip down the length of the staff. The strip was then unwound and sent on its way. The strip was rewrapped around a staff of the same diameter by the recipient re-forming the message). A somewhat similar device was also in use in China more than 5,000 years ago. Chinese generals sent messages from the battlefront by wrapping a band of silk around a thin pole and then writing on the material. Only someone who had a pole with the same diameter could read the message. To others the message would look like a band of silk with a random pattern of markings. This is one of the earliest documented forms of steganography [**Paul and Christian, 1999.**]. Later examples of steganography are the use of invisible inks and microdots to conceal messages. Steganography has its place in security systems. It is not intended to replace cryptography, but to supplement it. Hiding a message covertly with steganographic methods reduces the chance of the message being detected. However, if that covert message is also encrypted, it requires cryptanalysis to recover the plaintext. Whereas cryptography may not necessarily hide the presence of a secret message, steganography is intended to conceal the very existence of a message. In the digital age, messages can be concealed in graphic images, since most graphics standards specify more gradations of color than the human eye can notice. The steganography algorithm looks for the most complex parts of the image, where neighboring pixels are most different from one another, and adds the data to the least significant bits of the pixels. Without changing the graphical image noticeably, a 64-kilobyte message

can be embedded in a 1024 * 1024 gray-scale picture. Programs are available in the public domain for embedding messages in graphics. There are many other forms of steganography. Messages can be hidden in music audio tracks. The hidden meaning of a unique word or phrase embedded in a routine message, which has been designated to execute some preplanned reaction by the recipient, is also considered a form of steganography. Like the one-time pad, this steganographic protocol is, in itself, unbreakable, assuming perfect insider security on both ends. But, like symmetric cryptography, the key (or meaning in this case) must be prepositioned in advance [**DEPARTMENT OF DEFENSE. 2003.**]. Encoding messages in DNA is a relatively new and still evolving method of steganography.

Another material used as a key to the development of DNA steganography was a technique known as *polymerase chain reaction*, or **PCR**. This technique allows specific regions of a long strand of DNA to be selectively replicated in order to produce several copies of the same message. Certain regions of the DNA strand are sequenced and complementary (primers) are synthesized in the laboratory. Under the right reaction condition, the primers and free DNA bases will interact with the target DNA in **PCR** and replicate the region of interest so that several copies of this region are available for study [**Vinod ,2003**].

The microdot material is a means of concealing messages that was developed by Professor Zapp and used by German spies in the Second World War to transmit secret information. A microdot was a greatly reduced photograph of a typewritten page that was pasted over a full stop in an innocuous letter. The microdot was taken a step further and been used in the development of a DNA-based, doubly steganographic technique for sending secret messages. A DNA encoded message is first camouflaged within the enormous complexity of human genomic DNA and then further concealed by confining this sample to a microdot as illustrated in the following sections [**Clelland *et al.*, 1999**].

The basis strategy or method in DNA steganography is that the encoder and decoder agree on a specific set of primers that will flank the starting and ending sequences of message. Additionally, they agree on a pattern for reading the condons that will be found in the DNA [**Vinod , 2003.**]. First, a 3-base units will be assigned to letters of the alphabet, numerals, and some punctuation marks as shown in figure 1. The encryption key will be used to encode a message (sample message) reading "JUNE6_INVASION:NORMANDY" as a sequence of 69 bases and synthesized the following DNA strand:

{AGTCTGTCTGGCTTAATAATGTCTCCTCGAACGA
TGGGATCTGCTTCTGGATCATCCCGATCTTTGAAA}

Text to DNA Encryption Key			
A = CGA	**K = AAG**	**U = CTG**	**0 = ACT**
B = 7CCA	**L = TGC**	**V = CCT**	**1 = ACC**
C = GTT	**M = TCC**	**W = CCG**	**2 = TAG**
D = TTG	**N = TCT**	**X = CTA**	**3 = GAC**
E = GGC	**O = GGA**	**Y = AAA**	**4 = GAG**
F = GGT	**P = GTG**	**Z = CTT**	**5 = AGA**
G = TTT	**Q = AAC**	**_ = ATA**	**6 = TTA**

H = CGC	**R = TCA**	**, = TCG**	**7 = ACA**
I = ATG	**S = ACG**	**. = GAT**	**8 = AGG**
J = AGT	**T = TTC**	**: = GCT**	**9 = GCG**

Fig 1 - Key used to encode a message in DNA.

The message sequence will be then sandwiched between two carefully selected oligonucleotide units (primers) consisting of 20 bases each, known only to the sender and the intended recipient:

TCCCTCTTCGTCGAGTAGCA
and the complement of:
TCTCATGTACGGCCGTGAAT

The total length of a single-stranded message molecule was 109 bases. In order to test the efficacy of the technology in decoding hidden messages, a few copies of this molecule were mixed with a huge number of similarly sized fragments of human DNA.

Only a recipient knowing the sequences of both primers would be able to extract the message, using the polymerase chain reaction (PCR) to isolate and make copies of (amplify) the message-containing DNA strand. It would then be a simple matter to determine the sequence of nucleotides in the relevant strand and decode the message. In contrast, an eavesdropper would have to undertake the virtually impossible task of sifting through 420 possible primer sequences to find the correct pair.

Because DNA is a very stable molecule under normal conditions and PCR is a very sensitive analytic technique, a DNA message can be hidden almost anywhere [**Ivars , 2000.**].

The DNA samples were then spread on the filter-paper medium for creating the microdots, which were large enough to cover a normal 16 point font size period at the end of a sentence. Each microdot was determined to contain 10 nanograms (ng) of DNA consisting of both the secret sequence to be decoded and random fragments from the genome. Due to the highly selective nature of PCR technique and the rapidity with which it is able to amplify DNA, the microdot samples were not filtered, but PCR was applied on the whole sample, yielding successful message-specific amplification [**Vinod , 2003.**].

The researchers were able to successfully decipher the message that had been encoded after the PCR amplification was completed. To do so, they took the amplified DNA from the PCR process and subjected it to a treatment known as *gel electrophoresis*. This technique allowed the researchers to specifically isolate the population of secret message DNA, which they then were able to determine the sequence of by using further biochemical assays. By comparing the strand's sequence to the encryption key that had been determined at the start, they were able to successfully decode the message and accomplish the first feat of DNA steganography [**Vinod , 2003.**].

We suggested that this technique could be used in a similar way to the original microdots: to enclose a secret message in an innocuous letter. And it should be possible to scale up the encoded message from the size of the previous discussed simple example, by encoding a longer message in several smaller DNA strands. It

should also use smaller microdots, which could be used for a variety of purposes, including cryptography and specific tagging of items of interest.

Results and Conclusions

All the practical parts of the algorithm are discussed theoretically because there is a great limitations in the lab equipments including the using of the PCR device (in the time of writing this paper, there was only one such device in the country in Baghdad and its use was restricted).

There are three levels of security:

1. Encrypting the message.
2. Hiding it in a human DNA.
3. Using microdot to transfer the message molecule to the receptor.

This will make the process extremely hard to decode if it were intercepted, making it an ideal medium for secret communication.

The field of DNA computing is still in its infancy and the applications for this technology are still not fully understood. The world of information security is always on the lookout for unbreakable encryption to protect the data that we transmit, but it appears that every encryption technology meets its end as the computing technology of our world evolves. Perhaps the DNA computing is viable, but the obstacles that face the field such as the extrapolation and practical computational environments required are daunting.

Clearly, the method of communication via DNA can be made secure, but it requires very sophisticated technology to decode, and this service as one limiting step towards its mass appeal making it widely applicable in the future will therefore require a further advance in technology, which may soon be on the horizon. One of the drawback currently cited by the manufacturers is that there is currently no hand-held reading system or instantaneous method able to decipher the encryption. While this is an added security feature, to some extent, it also limits the feasibility of wide spread applications of the technology.

DNA steganography has emerged as one of the most fascinating advances in cryptography in the last few years. Just a few decades ago, the use of biological molecule in information transfer may have been the stuff that science-fiction novels were made of ; however, with the advent of steganography, such novel ideas that previously resided only in the human imagination could be right the corner.

References

Alfred J. M. 1997.(*Handbook of Applied Cryptography)*, CRC Press, New York.

Clelland C.T., Viviana R., Carter B. 1999. (*Hiding Messages in DNA Microdots)*, Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York.

DEPARTMENT OF DEFENSE. 2003. (*MILITARILY CRITICAL TECHNOLOGIES LIST, SECTION 17: INFORMATION-SECURITY TECHNOLOGY)*, USA, Defense Threat Reduction Agency Ft. Belvoir, VA, October.

Donald N. 2001. (*DNA and DNA Computing in Security Practices – Is the Future in Our Genes?)*, GSEC Assignment Version 1.3.

Ivars P. 2000. (*Hiding in DNA)*, Math Trek Hiding in DNA, Science News Online.

Kahn, D. 1967. (*The Code Breakers)*, Macmillan Publishing Company, New York.

Paul E. P. and Christian B. 1999. (*The Politics of Cryptography)*, Performance Computing.

Vinod N. 2003. (*DNA Steganography)*, National Security, Harvard Science Review, Winter.