International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 9 (2015) pp. 22393-22401

http://www.ripublication.com

A Survey on DNA Cryptography

Rama Devi. K and Dr. S. Prabakaran

Department Of Computer Science & Information Technology
Research Scholar, SRM University, Chennai, India.
Professor, Department Of Computer Science &Information Technology
SRM University, Chennai, India.

Abstract

The intention of this paper is to examine the DNA cryptographic achievements. This paper reviews new approaches that are designed to solve either theoretical or application problems. Initially Adleman solved NP-problem by means of wet DNA experiment in 1994; DNA computing becomes one of the most suitable alternatives to overcome the silicon computer limitations. One of the main applications of DNA computing is DNA Cryptography. Today, many researchers concentrate on this research subject either to improve available methods used in DNA Cryptography itself or to suggest a new DNA cryptographic approach. This paper gives a brief summary of DNA cryptography, it reviews the research achievements in DNA cryptography and touches the strengths and weakness and explores avenues for future research work.

Keywords: DNA, Cryptography, DNA Computing.

I. INTRODUCTION

Currently, we could examine this is a period of information explosion in which information has become a very crucial tactical resource, and so the task of information security has become more significant. One of the key components of cyber security is cryptography. Cryptography is a way of defending the crucial data from unauthorized access. It has emerged as a secure means of transmission of information. It mainly helps in cutting intrusion from third party.

As some of the modern cryptography algorithms (such as DES, and more recently, MD5) are broken, the new directions of information security are being sought to protect the data. The concept of using DNA computing in the fields of cryptography and steganography is a possible technology that may bring forward a new hope for powerful, or even unbreakable, algorithms.

DNA Cryptography is one interdisciplinary research area that is growing fast since DNA codes are implemented in a cryptographic technique. In this project, we do not intend to utilize real DNA to perform the cryptography process; rather, we will introduce a new cryptography method based on central dogma of molecular biology.

This paper gives a brief summary of DNA Cryptography, it reviews the research achievements in DNA cryptography and touches the strengths and weakness and explores avenues for future research work.

II. BIOLOGICAL BACKGROUND OF DNA CRYPTOGRAPHY

A. 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. The main role of DNA molecules is the long-term storage of information. DNA is often compared to a set of blueprints or 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 DNA double helix is stabilized by hydrogen bonds between the bases attached to the two strands. The four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar/phosphate to form the complete nucleotide, as shown for adenosine monophosphate.

B. The genetic code

The genetic code comprise of 64 triplets of nucleotides. These triplets are named as **codons**. With three exceptions, each codon encodes for one of the 20 amino acids used in the synthesis of proteins. That produces some redundancy in the code: most of the amino acids being encoded by more than one codon. The genetic code can be expressed as either RNA codons or DNA codons. RNA codons occur in messenger RNA (**mRNA**) and are the codons that are actually "read" during the synthesis of polypeptides (the process called **translation**). But each mRNA molecule acquires its sequence of nucleotides by **transcription** from the corresponding gene. The DNA Codons is read the same as the RNA codons Except that the nucleotide thymidine (**T**) is found in place of uridine (**U**). So in DNA codons we have (TCAG) and in RNA codons, we have (UCTG).

C. Transcription and translation

A gene is a sequence of DNA that contains genetic information and can influence the phenotype of an organism.

Within a gene, the sequence of bases along a DNA strand defines a messenger RNA sequence, which then defines one or more protein sequences. The relationship between the nucleotide sequences of genes and the amino-acid sequences of proteins is determined by the rules of translation, known collectively as the genetic code. The

genetic code consists of three-letter 'words' called codons formed from a sequence of three nucleotides (e.g. ACT, CAG, TTT).

In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. This RNA copy is then decoded by a ribosome that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA, which carries amino acids. Since there are 4 bases in 3-letter combinations, there are 64 possible codons (43 combinations). These encode the twenty standard amino acids, giving most amino acids more than one possible codon. There are also three 'stop' or 'nonsense' codons signifying the end of the coding region; these are the TAA, TGA and TAG codons.

III. CHRONICLES OF DNA CRYPTOGRAPHY

It is Adleman[1], with his pioneering work [Adleman, 1994]; set the stage for the new field of bio-computing research. His main idea was to use actual chemistry to solve problems that are either unsolvable by conventional computers, or require an enormous amount of computation. By the use of DNA computing, the Data Encryption Standard (DES) cryptographic protocol can be broken [Boneh, et. al, 1995] [2]. The one-time pad cryptography with DNA strands, and the research on DNA steganography (hiding messages in DNA), are shown in [Gehani, et. al, 2000][3]. In 1995, R.Lipton[4] extended his research to solve NP-complete problem. In 2002, a team led by Adleman solved a 3-SAT problem with more than 1 million possibilities on a simple DNA computer after an exhaustive searching.

Some perceptions by different researchers,

- Clelland et al.,
- Gehani et al.,
- Leier et al.,
- Wong et al.,
- Arita et al.,

Clelland et al

Inspired by the micro-dots used during the II world war, Clelland et al. developed an extension of this principle [5]. The scientists produced artificial DNA strands, which contained secret messages. A triplet encodes one character or number. The Clelland algorithm is a simple substitution cipher which encodes characters into DNA sequences using the following encoding function

- $E : X \rightarrow Y$
- $X \in \{A, B, C,..., Z, 0, 1,..., 9, ".",",","," : ","+"\}$
- $Y \in \{xyz : x, y, z \in \{A, C, G, T\}\}$

The decoding function is corresponding $D : Y \rightarrow X$.

Now Clelland et al. ligated two primers with the synthesized DNA sequences, a forward and a reverse primer. These ligated sequences were mixed up with dummy strands. Important preconditions are:

- length of dummy strand = length of message DNA with primers
- #copies of each dummy = #copies of message DNA

The receiver must know the decoding function and the primer to decode the message. The primers are used for the polymerase chain reaction and in the last step the amplified DNA sequence has to be sequenced and decoded. To improve the security one can use dummy strands, which are not random but correspond to words out of a dictionary.

Gehani et al

The original One-Time pad uses the XOR – exclusive **or** ($\oplus$). In the case of DNA, the XOR is very impracticable and therefore it is better to use the properties of DNA. Gehani et al. established a DNA One-Time pad by creating word pairs [3]. The first word is the plain text and the second one is the cipher text. After such a block of plain and cipher text, there is a stop codon (Figure 1). The DNA polymerase completes the plain and cipher text. To encode a message, the plain text is mixed with the DNA sequences. It binds directly to the corresponding complementary sequence. The DNA polymerase creates the cipher text accordingly and the decoding is functionally analogous. The cipher text binds to its complement and the DNA polymerase creates the plain text.

Figure 1

DNA One-Time pad.
Ai: plain text, *Bi*: cipher text (and primer for the DNA polymerase), black box: stop
Modified from Gehani et al. [3].

Leier et al

Leier et al. encoded binary information into DNA sequences. A short DNA sequence represents the binary 1_2, another one represents 0_2 [6]. Further there are another two short DNA sequences, which represent start and end. The fragments have sticky ends and can be ligated (Figure 2). All resulting sequences are like this $s\{0_2|1_2\}e$. The start and end marker have primer sequences on one site for the polymerase chain reaction, which cannot be ligated. Although it seems to be more complicated, it is very similar to the algorithm of Clelland et al. The resulting DNA sequence is mixed with dummy strands and can only be detected and isolated knowing the primer sequences.

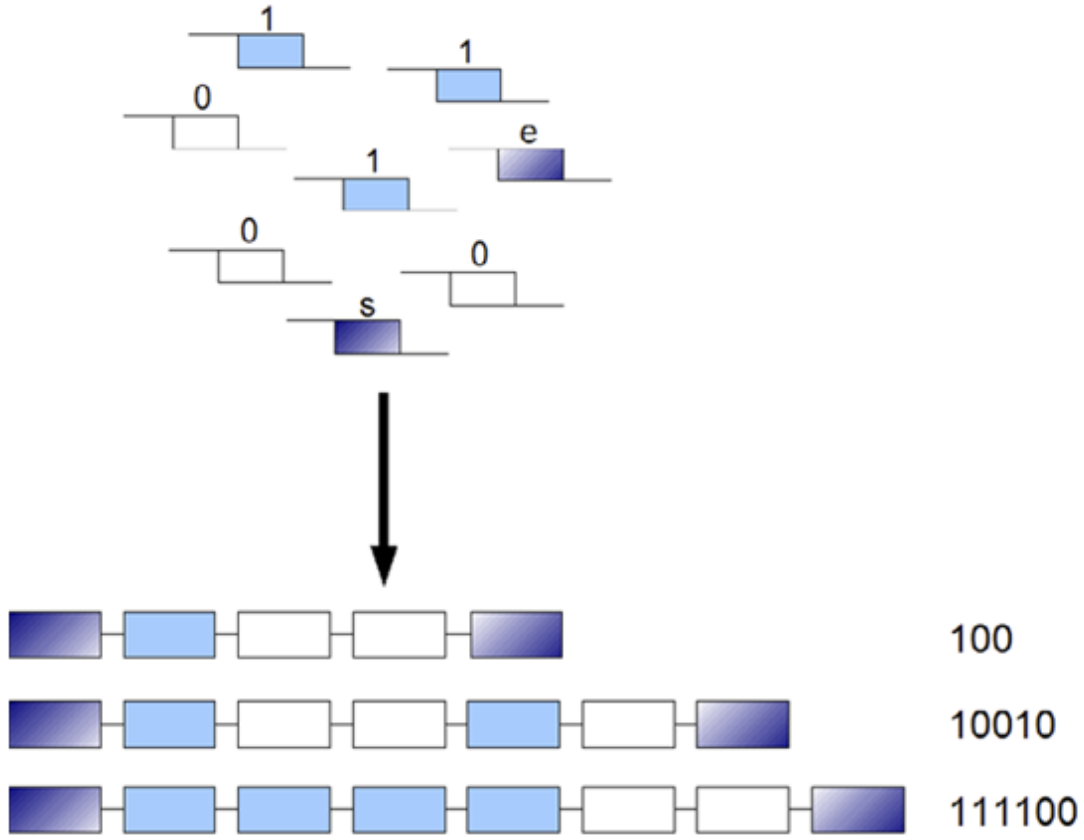

Figure 2

DNA binary strands.
Short DNA strands represent the binary 1_2 (light blue), 0_2 (white), start and end marker (dark blue).These sequences can be ligated to long strands by using the sticky ends. Modified from Leier et al. [6].

Wong et al
Wong et al. developed a steganographic algorithm based on DNA, which is able to store data in living organisms [7]. The data are translated into a DNA sequence which is inserted into a vector. The insert sequence is flanked by two primer sequences which do not exist in the genome yet. This vector is introduced into a cell of a living organism where it coexists and is replicated with the genomic DNA. To extract the data they used a polymerase chain Wong et al. used a substitution cipher similar to Clelland et al. to encode a song text into a DNA sequence and stored it in *Deinococcus radiodurans*. *Deinococcus radiodurans* survive extreme conditions, e.g. ionizing radiation, so the song text can be stored for hundreds of years.

Arita et al
Arita et al. developed a steganographic algorithm based on the degenerative

genetic code. Amino acid codes are redundant so that the translation of mRNA into proteins is a substitution cipher with the following characteristics

- $E : X \rightarrow Y$
- $X \in \{xyz : x, y, z \in \{A, C, G, U\}\}$
- $Y \in \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y, STOP\}$

But the inverse function $D : Y \rightarrow X$ is not injective.

The triplet of threonine is redundant in the third base so mutations in the third base do not exert any influence on the translation of threonine and the translated protein. These mutations are called "synonymous substitutions", in contrast to the "non-synonymous substitutions".

Arita et al. translated each letter of the English alphabet into six codons. A value of 0 means to keep the original base at the third position of a codon, while a value of 1 means to change the third base at that position. Arita et al. added a parity bit to each letter, to keep it odd for possible error detection [8]. They encoded 'KEIO' into the *ftsZ* gene of *Bacillus subtilis* which is essential for cell division and demonstrated as expected that the changed codon sequences did not affect the cell division, colony morphology, growth rate and sporulation frequency of these bacteria. To extract the encoded message one has to know the original sequence so that one can decide whether the codon is the original or the altered sequence.

Comparison to DNA-Encryption Algorithm

Clelland et al., Gehani et al. and Leier et al. produced synthesized DNA sequences which were mixed with dummy strands. These sequences contained a secret message. Knowing the unique primer sequence, the secret message can be read out.

Wong et al. and Arita et al. introduced DNA sequences containing a secret message into living organisms. Wong et al. used a vector which incorporated into the genome of *Deinococcus radiodurans* and Arita et al. used point mutations in redundant codons. Arita et al. used a parity bit for error detection. The disadvantage is that if mutations occur, the hidden information is lost.

A comparative overview of the algorithms and their features is shown in table 1.

Table 1: Comparison of the DNA encryption algorithms

Algorithm	organism	affect	Error detection	Error correction	binary	encryption	utilization
Clelland et al	-	-	-	-	-	-	**20**
Gehani et al	-	-	-	-	-	-	-
Leier et al.	-	-	-	-	+	-	**<=9**
Wong et al	+	+	+	+	-	-	**20**
Arita et al.	+	-	+	-	-	-	**<=5**

- = negative; + = positive;

organism: the use of this algorithm in living organisms; affect: observation that the algorithm exerts an effect on the organism; error detection/correction: the algorithm shows an error detection/correction function; binary: binary information can be encoded; encryption: the use of binary encryption algorithm like AES or RSA; utilization: storage utilization in a 100 bp DNA sequence;

IV. SUPPLEMENTARY DNA CRYPTOGRAPHIC ALGORITHMS A. DNA-PKC

This paper proposes DNA-PKC, an asymmetric encryption and signature cryptosystem [9] by combining the technologies of genetic engineering and cryptology. It is an exploratory research of biological cryptology. Similar to conventional public-key cryptology, DNA-PKC uses two pairs of keys for encryption and signature, respectively. Using the public encryption key, everyone can send encrypted message to a specified user, only the owner of the private decryption key can decrypt the cipher text and recover the message; in the signature scheme, the owner of the private signing key can generate a signature that can be verified by other users with the public verification key, but no else can forge the signature. DNA-PKC differs from the conventional cryptology in that the keys and the cipher texts are all biological molecules. The security of DNA-PKC relies on difficult biological problems instead of computational problems; thus DNA-PKC is immune from known attacks, especially the quantum computing based attacks.

B. Integrating DNA Computing in International Data Encryption Algorithm (IDEA) :

In this paper the author apply DNA computing on well known IDEA algorithm for making it more secure. We are adding a layer of DNA cipher over the basic IDEA algorithm. The cipher now will be in form of DNA sequence which will even hide very existence of the underlying IDEA algorithm. Key space has been extended a bit to make it more immune to cryptanalytic attacks. It can transmit highly confidential data efficiently and securely. The Matlab has been used for implementation [10].

C. A DNA and Amino Acids-Based Implementation of Playfair Cipher :

This paper discusses a significant modification to the old Play fair cipher by introducing DNA-based and amino acids-based structure to the core of the ciphering process. In this study, a binary form of data, such as plaintext messages, or images are transformed into sequences of DNA nucleotides. Subsequently, these nucleotides pass through a Play fair encryption process based on amino-acids structure. The fundamental idea behind this encryption technique is to enforce other conventional cryptographic algorithms which proved to be broken, and also to open the door for applying the DNA and Amino Acids concepts to more conventional cryptographic algorithms to enhance their security features[11].

D. An Encryption Scheme Using DNA Technology:

In this paper [12], an encryption scheme is designed by using the technologies of

DNA synthesis, PCR amplification and DNA digital coding as well as the theory of traditional cryptography. By applying the special function of primers to PCR amplification, the primers and coding mode are used as the key of the scheme. The traditional encryption method and DNA digital coding are used to preprocess to the plaintext, which can effectively prevent attack from a possible word as PCR primers. Biological difficult issues and cryptography computing difficulties provide a double security safeguards for the scheme. And the security analysis shows that the encryption scheme has high confidential strength.

V. CONCLUSION

In this paper, we briefly review the literature of DNA cryptography schemes as special instances of secret sharing methods among participants. The research of DNA cryptography is still at the beginning, and there are many problems to be solved. But the vast parallelism, exceptional energy efficiency and extraordinary information density inherent in DNA molecules endow DNA cryptography special advantages over other kinds of cryptography. Just as Adleman said, "DNA computers" illustrate that biological molecules like nucleic acids, proteins, etc.. This can be used distinctly for non-biological purposes. Hence DNA Cryptography have a great potential in their further exploration.

REFERENCES:

1. M. Adleman "Molecular Computation of solution to combinatorial problems" Science, New Series, Vol. 2Leonard 66, No. 5187. pp. 1021-1024 Nov. 11, 1994
2. Boneh D, Dunworth C, Lipton R J. Breaking DES using a molecular computer. In: DNA Based Computers I. Providence, USA: America n Mathematical Society, 1996. 37–65
3. Gehani A, LaBean TH, Reif JH: DNA-based cryptography. *Dimacs Series In Discrete Mathematics and Theoretical Computer Science* 2000, 54:233-249.
4. R. J. Lipton," Using DNA to Solve NP-Complete problems," Science, vol. 268, pp. 542 545, 1995
5. Clelland C, Risca V, Bancroft C: Hiding messages in DNA microdots. *Nature* 1999, 399:533-534.
6. Leier A, Richter C, Banzhaf W, Rauhe H: Cryptography with DNA binary strands. *BioSystems* 2000, 57:13-22.
7. Wong PC, Wong KK, Foote H: Organic data memory using the DNA approach. *Communications of the ACM* 2003, 46:.
8. Arita M, Ohashi Y: Secret signatures inside genomic DNA. *Biotechnol Prog* 2004, 20:1605-1607.

9. LAI XueJia, LU MingXin, QIN Lei, HAN JunSong & FANG XiWen : Asymmetric encryption and signature method with DNA technology Science China Press and Springer-Verlag Berlin Heidelberg 2010.
10. Pankaj Rakheja :Integrating DNA Computing in International Data Encryption Algorithm (IDEA) *International Journal of Computer Applications (0975 – 8887) Volume 29– No.8, September 2011*.
11. Mona Sabry, Mohamed Hashem, Taymoor Nazmy: A DNA and Amino Acids-Based Implementation of Playfair Cipher *(IJCSIS) International Journal of Computer Science and Information Security, Vol. 8, No. 3, 2010*.
12. Guangzhao Cui,Limin Qin,Yanfeng Wang, Xuncai Zhang: An Encryption Scheme Using DNA Technology 978-1-4244-2724-6/08/$25.00 2008 IEEE.