A Review on Emerging Cryptography Technique: DNA Cryptography

Tushar Mandge
Dept of Information Technology
Shri Vaishnav Institute of
Technology and Science
Indore (M.P.), India

Vijay Choudhary
Dept of Information Technology
Shri Vaishnav Institute of
Technology and Science
Indore (M.P.), India

ABSTRACT

In recent years, information technology has grown much and modern era is of digital information. DNA cryptography is a new born technology based on DNA computing. Biological DNA has extraordinary potential to fulfill the modern computing requirements as well as to secure large scale information. DNA has vast parallelism, high information density and exceptional energy efficiency. DNA cryptography is one of the emerging technologies, which binds biology and information technology together. This paper presents a review on DNA cryptography and its basic encryption techniques.

Keywords

DNA, DNA computing, DNA cryptography, DNA chip, DNA steganography, PCR amplification, DNA digital coding, DNA one-time pad

1. INTRODUCTION

A new area of computing evolved in the year 1994 when Dr. Leonard M. Adleman [1] presented the idea of biological DNA to solve the complex mathematical problem. He found that biological DNA has high computational capability. He solved the directed Hamiltonian path problem (HPP) with seven vertices in graph similar to travelling salesman problem, which shown that DNA molecules has potential to solve large scale combinatorial problems. Richard J. Lipton [2] extended this idea, he used DNA to solve NP-complete problem, which was SAT problem of computer science. In the year 1995, Boneh et al [3] presented an approach to break the DES (Data Encryption Standard) algorithm by methods of DNA computing. In 1999, Tylor Clelland [4] presented a new approach of steganography where secret message is encoded as DNA strands among a multitude of random DNA strands.

DNA computing is a computation technology, where computations are made in the form of DNA Sequences. DNA sequences or genes are the coding sequences which contains information of living things. Any information can be encoded in a DNA sequence and it can be computed using bio-computing tools [5]. DNA has vast parallelism and high storage capacity. Due to DNA's vast parallelism it is capable of doing the combinatorial search among a large number of possible solutions represented by DNA strands. High storage capacity makes it much powerful in comparison to today's computer storage capability. A gram of DNA contains near about 10^{21} bases which is equivalent to 10^{8} tera bytes of data. These features of DNA explored it as much efficient medium to handle the future data storage and computational requirements [6].

The internet and networking technology is growing day by day, a lot of information flows every day on networks, at the same time security threats are also increasing for users. There are various hackers and attackers who always try to break the systems, to steal crucial information or to modify them. So information security has become a very critical aspect of modern computing systems. There are few organizations like government sections, military, banking who can't afford to leak any information which may be confidential and secret. In this case they need a strong information security system which can handle large scale information with high level of privacy and security.

Cryptography is the art and science of achieving security by encoding messages to make them non-readable where data is scrambled in the way that nobody can understand it. Cryptanalysis is the technique of decoding messages from non-readable to readable form without having any idea about how the message was converted from readable to non-readable. Encryption is a process of encoding a plain text message in to cipher text. Decryption is reverse process of encryption where cipher text is transformed back to plain text. There are two types of cryptography mechanism available based on the number of keys used. In Symmetric key cryptography, sender and receiver of information share the same key for encryption and decryption process whereas if different keys are used by sender and receiver then it is called Asymmetric key cryptography. In asymmetric cryptography the key used by sender is known as public key which may be publicly known and the key used by receiver is known as private or secret key which should be kept secret [7].

There are several cryptography algorithms developed in past such as DES, AES, RSA etc. DNA cryptography is a new cryptography technique came in to existence with the development of DNA computing. In which DNA is used as information carrier and biological tools are used as implementation tools. As DNA has vast parallelism and high storage capacity, it explored DNA molecules as a new way of cryptography. Currently DNA cryptography is not in the mature stage that's why very few approaches are introduced yet [8]. In DNA cryptography, message is encrypted in the form of DNA nucleotide sequence which is the combination of four nucleotide bases. There are few key technologies of DNA which are widely accepted such as DNA digital coding and PCR amplification.

2. BIOLOGICAL BACKGROUND

DNA computing and cryptography is based on the biological elements of DNA. DNA is a biological term, Deoxyribo nucleic acid, which is composed of two polynucleotide chains. In 1953, Watson discovered the structure of DNA. Fig.1 shows a DNA structure. It is a bio-molecule which contains genetic information about living thing. It contains the genetic instructions required to form other cells. According to Watson and Crick, DNA is formed by two strands of nucleotides which create a double helix structure. These two strands are bind together with four nitrogen bases A (adenine), C (cytosine), G (guanine), T (thymine). DNA strands have chemical polarity 5' and 3' at each end of a strand. Based on this chemical polarity two single strands combine together in anti-parallel way to form double stranded structure, where base A always makes pair with base T and base C makes pair with G. A DNA molecule formed by two single stranded chains contains hydrogen bonds between bases. There is a double bond between A and T whereas triple bond between C and G as shown in fig.2. The DNA strands that makes bond with each other through A-T and C-G are known as complementary strands. This complementary DNA double helix structure was identified by Watson and crick [5] [9].

2.1 Biological Operations

Biological operations based on DNA molecules which can be helpful to solve computational and mathematical problems through DNA computing are as follows,

2.1.1 Hybridization

Hybridization is the process in which two single stranded DNA chains or sequences are combined together to form a double stranded DNA.

2.1.2 Denaturation

It is the process where double stranded molecule is resolved in to two single stranded DNA.

2.1.3 Ligitation

It is the process where two double stranded DNA sequences are combined together to form a new double stranded DNA sequence [5].

2.1.4 Polymerase chain reaction

Polymerase chain reaction (PCR) is a biological technique used to amplify certain regions of DNA sequences using enzymatic replication. Primer pairs are used for amplification. Primers are the small DNA fragments [9].

2.1.5 Gel electrophoresis

It is a technique of separating or exacting DNA fragments (DNA sequences) of a DNA strand located in the gel where gel acts like filter. Term electrophoresis is how to push the DNA strands through the gel filter. Gel electrophoresis is performed in an electrophoresis box by the application of electric current, where it can make the DNA move. Short strands move through the holes in the gel more quickly than longer strands. Strands of same length move with same speed and ends up group together, in this way, the DNA strands in the sample sort themselves [11].

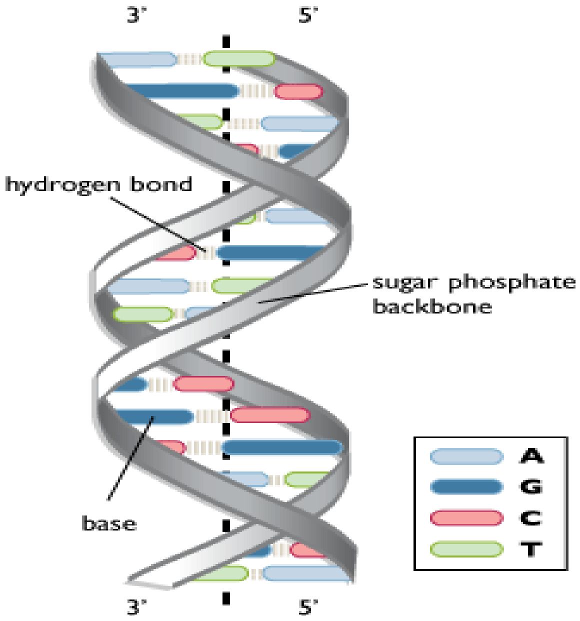

Figure 1: DNA structure [10]

2.2 Basic DNA Terminology

There are few conceptual biological terms which should be known to work with DNA computing and cryptography as described below,

- Codons are the sequence of three nucleotide bases in triplet form.
- Genes are the working subunits of DNA. Each gene contains a particular set of instructions, usually coding for a particular protein or for a particular function. Genes contains coding sequences (exons) and non-coding sequences (introns).
- Chromosome is a large organized structure of DNA coiled around proteins, which contains genes, regulatory elements and other nucleotides sequences. The chromosomes can be thought of as long strings of genes.
- Genome is the unique sequence of each organism which contains DNA content of a cell, including nucleotides, genes and chromosomes [9] [12] [13].

3. DNA CRYPTOGRAPHY

DNA cryptography is a new born technology emerged with the advances of DNA computing. It explored DNA molecules as a strong medium for computation as well as for cryptography purpose to secure data in the form of data encryption, Authentication, digital signatures etc [8]. There are some fundamental techniques which are used for DNA encryption such as DNA digital coding and PCR amplification which has been employed by various researchers, are described below.

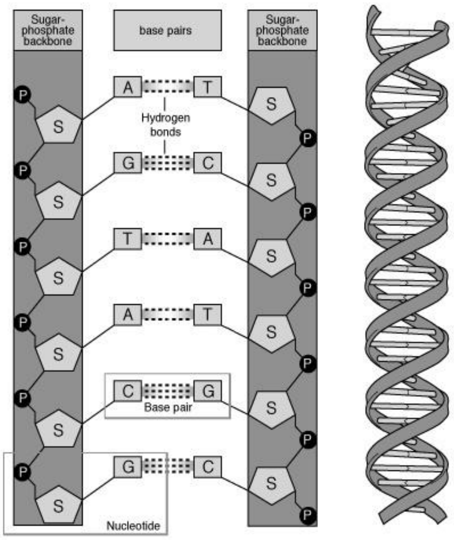

Figure 2: DNA base pairing and hydrogen bonds [14]

3.1 DNA Digital Coding

DNA digital coding provides mapping of DNA bases which plays main role in encrypting and decrypting any information [9]. DNA digital coding is a technique which is based on the concept of binary digital coding, which is encoded by the combination of 0 or 1. DNA coding is based on four nucleotide bases A, C, T, G. To encode the nucleotide bases is to map these units as, A(00),T(01),C(10),G(11) , so according to it the possible encoding patterns are 4!=24. Where A always makes pair with T and C always makes pair with G of two strands [15]. Among these 24 possible patterns only 8 patterns are identified correct according to complementary rule of nucleotide bases, they are 0123/CTAG, 0123/CATG, 0123/GTAC, 0123/GATC, 0123/TCGA, 0123/TGCA, 0123/ACGT, 0123/AGCT and among these 8 patterns, 0123/CTAG, which perfectly reflects the biological characteristics of four nucleotide bases [16]. A DNA coding pattern has shown in Table 1.

A text message can be digitally coded using DNA digital coding technique. Suppose there is a plaintext message "ASTROLOGY", first of all each character of message can be converted to their corresponding ASCII values (A=65, S=83, T=84, R=82, O=79, L=76, O=79, G=71, Y=89). Now, these values can be converted to their binary, base-2,(sequence of bits 0 and 1) or base-4 (sequence of 0,1,2 and 3) representation. Base-4 conversion (65=1001, 83=1103, 84=1110, 82=1102, 79=1033, 76=1030, 79=1033, 71=1013, 89=1121) will be mapped to" TAAT TTAG TTTA TTAC TAGG TAGA TAGG TATG TTCT". Along with DNA Digital coding other encryption steps can be performed to make data much secure such as PCR amplification which uses secret primer pair as key.

3.2 PCR Amplification

Polymerase chain reaction (PCR) is a molecular biology technique of DNA amplification based on Watson–crick complementary model. A DNA sequence is encoded using two primers. Primers are the small DNA fragments. Two primer pairs are used as key for PCR amplification [8]. The message which should be secure is placed between the two primer pairs to encode it to a new sequence. It is much difficult to amplify the message encoded sequence if the PCR primer pairs are not known. This technology requires correctness of the primers of the sequence because different lengths of primers will generate different results, so correct message can't be extracted [15].

Table 1. A DNA digital coding pattern

Digital Codes	**DNA Digital Coding**
00(0)	A
01(1)	T
10(2)	C
11(3)	G

Steps followed in biological PCR operation: [5] [17]

Step1:Denaturation: PCR amplification starts with denaturation process, where a double stranded molecule is resolved in to two single stranded DNA. Sample is heated at 94 to 96 degree Celsius one to several minutes to denature or separate double strand to two single strands.

Step2: Primer Annealing: At this step, temperature is lowered to 50 to 65 degree Celsius for one to several minutes where primers are attached to their complementary sequences. The primers are designed to amplify the DNA regions.

Step3: Primer Extension: The temperature is raised to 72 degree Celsius for one to several minute. Here polymerase enzyme adds nucleotides to the strand of short primer on base of original DNA strand. The DNA strands between primers amplified.

4. DNA ENCRYPTION TECHNIQUES

PCR and DNA digital coding techniques are the most important for DNA cryptography, a brief description about these techniques is already given in this paper. There are few other techniques which are presented by researchers in previous works of DNA cryptography.

4.1 DNA Random One-time Pad Based

Traditional One-Time pad (Vernam Cipher) encryption technique is implemented using set of randomly arranged non-repeating characters as the input cipher text. This set of random characters works as a pad. It is considered highly secure because if an input cipher text used once, it is never used for any other message, due to which it is named as One-Time. The length of the pad should be equal to the length of the plain text in One-time pad. In DNA Cryptography there is a method of DNA one-time pad substitution where One-time pad encryption process uses a random codebook to convert short segments of original plaintext messages to cipher text, which provide a random mapping. Two important points should be noted about codebooks, first a codebook must be truly random, and second thing it must be used only once for any message [6] [7] [18].

Another technique based on one-time pad is DNA XOR one-time pad. Vernam cipher uses a bit sequence S, which used as one-time pad. A copy of S is stored at each side of a sender and a receiver. When a binary message M has to be sent then each bit of message is XORed with the pad bits to produce cipher as Ci=Mi + Ki for i =1, 2… n; where Ki is the bits of bit sequence S which is being XORed with the message bit. The used bits of S are discarded at the source and cipher text C= (C1, C2…, Cn) is sent to destination. At receiver side sequence C is used in the place of M for decryption and bitwise XOR is performed with bits of S. After use the bits of S are discarded [6].

In current scenario due to hardware limitations one-time pad is suitable only for small plaintext messages. In future when DNA chips will replace current silicon chips, obviously due to DNA's high storage capacity small amount of DNA can store large one-time pad. But at same time due to this fact that the length of pad should be same as the plaintext, so obviously mapping through DNA pad may take long time if message length increases.

4.2 DNA Chip Based

Microarray and DNA chip technology changed the way of research in recent years. DNA chip technology is helpful in handling and storing the biological information. It is also capable of doing manipulations of vast amount of genome-sequences [19]. Microarray technology is known by several names like DNA chips, Gene chips or Bio chips. The array can be defined as an ordered sequence of micro spots, where each spot contains a single species of nucleic acid. The microarray technique is based on hybridization process of nucleic acids. A computer chip is a electronic circuit of a small piece of semiconducting material. A semiconductor is a type of material which is neither a good conductor of electricity nor a good insulator. Generally silicon is used to manufacture microchips. These Chips are usually of small size. This small or micro size of chip helps to make modern computers much fast, compact and portable. Production of microarrays starts with choosing probes to be printed on the array. In most cases, probes may be also chosen directly from available databases provided by Gene Bank, UniGene etc.

DNA chips can be used for gene expression analysis [20]. DNA chip has an array of immobilized DNA strands, so that multiple copies of a single sequence are grouped together in a microscopic array [18]. DNA chip consist of several spots embedded on a solid surface such as glass substrate. Each spot consists of cDNA probes or small nucleotide sequences, which binds to the complementary nucleotides, according to Watson-Crick complementary structure. Those nucleotides which bind to these probes are fluorescently labeled, observed and calculated electronically based on the ratio of the binding probe with the DNA sequences on each spot [20]. In encryption process of DNA chip, probes are arranged on a square area of small size of glass or silicon matrix. The receiver uses various strands which contain the fluorescent label to anneal with probes which relate to secret information using fluorescent reaction [21]. DNA chip technology is not only limited to encryption of textual data, it is useful in encrypting images also. Ashish gehani et al. [6] proposed a technique of DNA chip-based encryption and decryption of 2D images using substitution One-time pad. A message can be recovered in decryption process only by using the identical one-time-pad and DNA chip, which used in the encryption process.

As DNA is a biological element, its molecular properties may change due to environmental conditions. So obviously DNA chip's properties can be changed due to changing conditions of environment. In this sense encryption and decryption process will be unstable and may generate different results.

4.3 DNA Steganography

Steganography is a technique that hides the secret message inside other message. A secret message can be hide inside any medium like image, audio or video file. There are various methods found in history, which used for hiding information such as invisible ink, tiny pins pictures on characters, letters placed on specific positions of each word or any graphics image can be used to hide secret message etc. [7]. In basic DNA steganography, input DNA strands which contain message, each one is tagged with random secret key DNA strands and then can be hide inside other random distracter DNA strands. The plaintext is retrieved by hybridization process with the complement of the secret key strands [6]. Viviana I. Risca proposed a DNA steganographic technique in which DNA encrypted message strand is placed between secret primers and hidden in a microdot [22]. Ashish Gehani et al. [6] proposed an improved DNA steganography system by reducing the difference between the plaintext and distracter strands. Due to high storage capacity DNA steganography can be a right option for hiding large scale information. In spite of its simplicity there may be chance that environmental conditions may change the biological property of DNA molecule which can damage DNA sequences. So it may difficult to recover correct plaintext [19].

5. CONCLUSION

A new chapter of technology has been opened with the development of DNA computing. Due to DNA's vast parallelism and storage capacity, it is capable to provide a strong support for computing and cryptography. DNA cryptography is a new born technology which is in the development phase and it is in initial stage of research. In this paper, a review on DNA computing and DNA based cryptography is presented. DNA encryption techniques are also described. This technology is far away from the actual realization because biological problems related to DNA cryptography can be performed only in lab with biological tools and methods. There are few techniques available like PCR amplification and digital coding, which can be applied on digital information to perform DNA cryptography operations based on nucleotide sequence. It is possible to provide hybrid security by integrating DNA cryptography with traditional cryptosystems. DNA technology can overcome the problem of current traditional cryptosystems which are limited because of hardware limitations e.g. one-time pad is suitable only for small length of data. In future, DNA chips or Gene chips can replace currently available silicon chips which will exceptionally increase the computational and storage capacity. So there will be no issue about hardware limitations for cryptosystems. The disadvantage for DNA cryptography is that it can be affected by environmental conditions. As biological properties of a molecule may change due to environment or atmosphere changes, so it is possible that the recovery of correct information may be difficult. There is need of more work and research on DNA computing and cryptography to enhance the technical issues and to provide actual realization of this technology.

6. REFERENCES

[1] L. Adleman, “Molecular computation of solutions to combinatorial problems”, Science, JSTOR, vol. 266, 1994, pp.1021–1025.

[2] R. J. Lipton, “Using DNA to solve NP-complete problems”, Science, vol. 268, 1995, pp.542-545.

[3] D. Boneh, C. Dunworth, and R. Lipton, “Breaking DES using a molecular computer”, Proceedings of DIMACS workshop on DNA computing, pp. 37–65, 1995.

[4] Taylor Clelland, “Hiding messages in DNA microdots”. Nature Magazine vol.399, June 1999.

[5] S. Jeevidha, Dr. M. S. Saleem Basha and Dr. P. Dhavachelvan, “Analysis on DNA based cryptography to secure data transmission”, IJCA, vol. 29–no.8, September 2011.

[6] A. Gehani, T. LaBean, and J. Reif, “DNA-based cryptography”, Lecture Notes in Computer Science, Springer, 2004.

[7] Atul Kahate, Cryptography and network security, New Delhi: Tata McGraw Hill, 2012.

[8] D.Prabhu, M.Adimoolam, “Bi-serial DNA encryption algorithm”, Cornell university library, http://arxiv.org/abs/1101.2577, 2011.

[9] Monica Borda and Olga Tornea, “DNA secret writing techniques”, IEEE conference, 2010.

[10] Cyber bridge, http://cyberbridge.mcb.harvard.edu, 2012.

[11] Genetic Science Learning Center, University of Utah, http://learn.genetics.utah.edu, 2012.

[12] News-medical, http://www.news-medical.net/health/ Genes -What-are-Genes.aspx, 2012.

[13] Center for genetics education, http://www.genetics.edu.au/Information/Genetics-Fact-Sheets/Genes-and-Chromosomes-FS1, 2012.

[14] Access excellence, http://accessexcellence.com, 2012.

[15] Yunpeng Zhang and Liu He Bochen Fu**,** Research on DNA Cryptography**,** http://www.intechopen.com, 2012.

[16] G. Cui, L. Qin, Y. Wang, X. Zhang, “An Encryption Scheme Using DNA Technology”, IEEE, 2008.

[17] DNA learning center, CSH, http://www.dnalc.org/ resources/animations/pcr.html, 2012.

[18] J. Chen, “A DNA-based, Biomolecular Cryptography Design”, ISCAS’03, Proceedings, 2003.

[19] B. Anam, K. Sakib, Md. A. Hossain, K. Dahal, “Review on the Advancements of DNA cryptography”, http://arxiv.org/abs/1010.0186, 2010.

[20] Magdalena Gabig, Wegrzyn, “An introduction to DNA chips: principles, technology, applications and analysis”, acta biochimica polonica, vol. 48, no. 3, 2001, pp. 615-622.

[21] Shihua Zhou, Qiang Zhang, Xiaopeng Wei, “An Image Encryption Algorithm Based on DNA Self-Assembly Technology”, IEEE, 2010.

[22] M. Borda, O. Tornea and T. Hodorogea, “Secret writing by DNA hybridization”, acta technica napocensis, electronics and telecommunications, vol. 50, no. 2, pp. 21-24, 2009.