DNA Cryptography An New Approach To Information Security

by

Aouragh Karim

Algeria

Aouragh.karim@yahoo.com

Abstract:

DNA cryptography is a new and promising field in information security. It combines classical solutions in cryptography with the strength of the genetic material. By introducing DNA into the common symmetric key cryptography, it is possible to benefit from the advantages of the classical cryptosystems and solve some of its limitations. There are different ways how DNA can be used to secure information content. In this article we purpose a new approach of cryptography that is DNA cryptography. Our aim is to build a secure and confidential data.

Keywords: Approach/cryptography/ DNA/Information/Security.

1. Introduction

Interest in information security existed since ancient times and it is present in our modern life. Techniques to protect information are evolving together with the progress in information technology. Secret information was hidden in books or paintings; it appeared in form of the unintelligible text. Some of the first ciphers based on substitution of letters in written text were Polybius and Cesar ciphers. There are two directions in information protection: cryptography and steganography. These two sciences manipulate information in order to change its meaning or hide its existence. Computer age brought a different interpretation of information and new directions in development of ciphers and cryptographic protocols. Computational power offered the possibility to build new and strong algorithms in cryptography, but it was also a strong tool used by cryptanalysts to break the cryptosystems. This means that the subject of finding new and powerful ciphers is always of interest and new directions in cryptography are explored.

Cryptography provides a range of features for information security. The main aspects treated by cryptography are: confidentiality, data integrity, authentication, and nonrepudiation.

The objectives of this article were to concentrate on the confidentiality part and to find new methods (ciphers) to ensure privacy through the use of DNA.

DNA cryptography consists in the use of genetics and biomolecular computation and it is one of the newest directions in cryptography. Genetic material such as DNA can be used as a vast storage space. This idea is inspired from the fact that DNA is a natural carrier of information which is encoded by a 4 letter alphabet: A, C, G, and T. This alphabet can be easily transposed into the binary alphabet (A – 00, C – 01, G – 10, T - 11). Therefore DNA can be used as a storage media for any kind of information. The methodologies of DNA cryptography are not coded mathematically, thus, it could be too secure to be cracked easily.

2. Cryptography

Cryptography is a method in which we protect data or information and transmit it into an unreadable format. Cryptography play major role to secure ATM transmission,

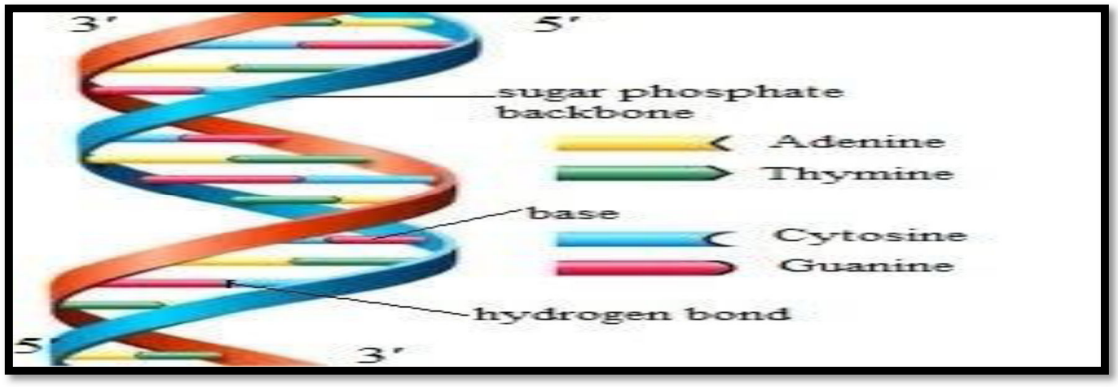
E-commerce, digital media privacy and web data transmission or storage. Modern cryptography work for four major concerns these are non-repudiation, integrity, authentication and confidentiality. In cryptography we use two processes, encryption and decryption.

3. DNA

DNA stands for Deoxyribonucleic acid which store genetic information of the entire living organism ranging from human being to small viruses. It is also called as an information carrier and consists of long polymer of small units called nucleotides.

Further nucleotides consist of three components: Nitrogenous base, five Carbon sugar and Phosphate group. Nitrogenous base consists of four bases: Adenine, Thymine, Cytosine and Guanine (A, T, C, G), all the complex information about organism are stored with the combination of these bases. Adenine and Guanine are called purines, whereas Thymine and Cytosine are called pyrimidines. DNA is a double helix structure as shown in the Figure below.

Figure 1. Double helical structure of DNA.[1]



DNA double helical structure was discovered by the two Nobel laureate Watson and Crick and therefore it is also called a Watson-Crick complementary structure, where A and T form hydrogen bond with each other, whereas C and G forms bond with one another. In this structure of DNA both the strands are antiparallel to each other, means if one strand starts from 3′ to 5′, then another strand is from 5′ to3′.

4. DNA COMPUTING (RELATED WORK)

In 1994, Adleman [2] laid the foundation of DNA computing by giving solutions to the combinatorial problems using molecular computation , one of which is “Hamiltonian path” problem. He solved the instance of graph containing seven vertices by encoding it into the molecular form by using an algorithm and then computational operations were performed with the help of some standard enzymes. This was solved by brute force method. In 1995, Lipton [3] extended the work of Adleman by solving another NP-complete problem called “satisfaction” by using DNA molecules in a test tube to encode the graph for 2 bit numbers . In 1996,Dan Boneh et al. [4] applied the approaches of DNA computing used by

Adleman and Lipton, in order to break one of the symmetric key algorithm used for cryptographic purposes known as DES (Data Encryption Standard). They performed biological operations on the DNA strands in a test tube, such as extraction, polymerization via DNA polymerase, amplification via PCR and perform operations on the DNA strands which have the encoding of binary strings. Then DES attack is planned by generating the DES-1 solution, due to which key can be easily guessed from the ciphertext and further evaluate the DES circuit, lookup table and XOR gates. By using their molecular approach they broke DES in merely 4 months. In 1997, Qi Ouyang et al. [5] applied the approaches of DNA molecular theory in order to generate the solution for maximal clique problem,

which is another NP-complete problem. Thus shows the efficiency of DNA: to solve Hard-problems and vast parallelism inherent in it which makes the operations fast.

5. DNA encoding scheme

In the field of information science, the most basic encoding method is binary encoding. This is because everything can be encoded by the two states of 0 and 1. However, for DNA there are four basic units:

1. Adenine (A);

2. Thymine (T);

3. Cytosine (C);

4. Guanine (G).

The easiest way to encode is to represent these four units as four figures:

1. A(0) –00;

2. T(1) –01;

3. C(2)–10;

4. G(3)–11.

Obviously, by these encoding rules, there are 4! = 24 possible encoding methods. For DNA encoding, it is necessary to reflect the biological characteristics and pairing principles of the four nucleotides. Based on this principle, we know that:

A(0) – 00 and G(3) – 11 make pairs,

T(1) – 01 and C(2) – 10 make pairs.

In these 24 programs, there are only 8 programs

0123/CTAG.

0123/CATG.

0123/GTAC.

0123/GATC.

0123/TCGA.

0123/TGCA,

0123/ACGT,

0123/AGCT match the DNA pair of a complementary principle. The coding scheme should be consistent with the weight of a molecular chain, so we get that 0123/CTAG is the best encoding scheme.

Table 1. DNA Digital Coding.

DNA nucleotide	Decimal	Binary
A	0	00
C	1	01
G	2	10
T	3	11

6. Results

The DNA sequences consists only four alphabets: Adenine (A), Cytosine (C), Guanine (G), Thymine(T).

Mathematically, this means we can utilize this 4 letter alphabet $\sum=\{A,G,C,T\}$ to encode information, which is more than enough considering that an electronic computer needs only two digits, 1 and 0,for the same purpose.

7. Implications

- **Speed:** Conventional computers can perform approximately 100 MIPS (millions of instruction per second).

Combining DNA strands as demonstrated by Adleman made computations equivalent to 10^9

Or better, arguably over 100 times faster than the fastest computer.

- **Minimal power requirements:** There is no power required for DNA computing while the computation is taking place. The chemical bonds that are the building blocks of DNA happen without any outside power source.

There is no comparison to the power requirements of conventional computers.

8. Conclusion

DNA cryptography is a promising and rapid emerging field in data security. The conventional binary data uses two digits '0' and '1' to code information. But for DNA molecules, which is the natural transporter of information, data is encoded by four bases viz. '*A*', '*T*', '*G*' and '*C*'. A few grams of DNA molecules have the capacity to restrain all the data stored in the world. Adleman has explored [Adleman, 1994] how the huge parallelism of DNA strands can concurrently attack the different aspects of the toughest combinatorial problem and solve it in polynomial time. DNA cryptography merges the massive parallelism and storage capacity of DNA molecules with the traditional methodologies of cryptography. At present, big tech giants, such as Microsoft, are taking initiative to commercialize DNA computers in near future. Hopefully, in coming ten to thirty years the virtually unhackable DNA cryptography techniques will be an effective alternative to classical cryptosystem.

9. References

[1] Ashish Kumar Kaundal and A.K Verma, 2014, "DNA Based Cryptography: A Review", International Journal of Information & Computation Technology, Volume 4, Number 7, pp. 693-698.

[2] Adleman. M. L (1994), Molecular Computation of Solutions to Combinatorial Problems, Science, vol**.** 266, pp. 1021- 1024.

[3] Lipton. R. J (1995), Using DNA to Solve NP Complete Problems, Science, Vol.268, pp.542 -545.

[4] Ashish Gehani, LaBean Thomas and John Reif (2004), DNA-based cryptography, In Aspects of Molecular Computing, Springer Berlin Heidelberg, pp. 167-188.

[5] Boneh. D (1996), Breaking DES using Molecular computer, American Mathematical Society, pp 37-65.