Xml security using DNA Technology

Amish S Desai

Department Of Computer Science & Engineering,

Parul institute of Engg & Tech,

Gujarat, India

Abstract

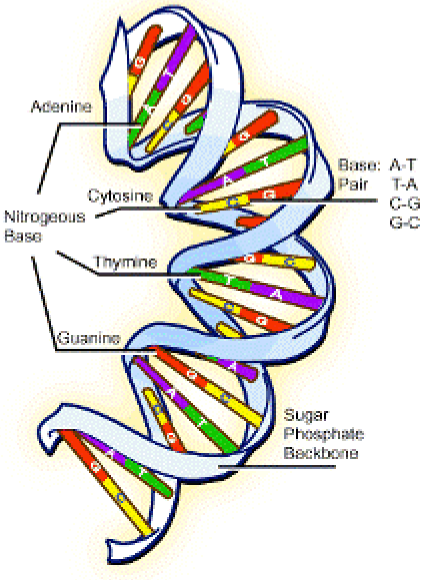
As XML becomes a standard for disseminating data on the internet, applications disseminating XML data on the internet are rapidly increasing ,so flow of XML data on the internet could breach the privacy of data providers unless access to the disseminated XML data is carefully controlled. The methods using encryption have been proposed for such access control .DNA Cryptography is a new born cryptographic field emerged with the research of DNA Computing. The vast parallelism, exceptional energy efficiency and extraordinary information inherent in DNA molecules are being explored for computing, data storage and cryptography. In this paper, we briefly introduce the biological background of DNA cryptography and the principle of DNA computing, summarize the progress of DNA cryptographic research and several key problems, discuss the trend of DNA cryptography, and propose a novel encryption algorithm is devised based on number conversion, DNA digital coding, and use of enzymes , which can effectively prevent attack on xml file and provide excellent security.

1. Introduction

DNA cryptography, a new branch of cryptography utilizes DNA as an informational and computational carrier with the aid of molecular techniques. It is relatively a new field which emerged after the disclosure of computational ability of DNA [1]. DNA cryptography gains attention due to the vast storage capacity of DNA, which is the basic computational tool of this field. One gram of DNA is known to store about 108 tera-bytes. This surpasses the storage capacity of any electrical, optical or magnetic storage medium [2], [3]. DNA is being proposed to use for many computational purposes. For example, Barish et.al. demonstrated a tile system that takes input and produces output using DNA [4]. The method is now also used to solve many NP-complete and other problems. Such as Rothemund et. al. showed that DNA can also be used to compute XOR function which isan essential part of cryptosystems [5]. It is a very potential field of research, as work which has been done in this field suggests that it can put many challenges to the modern cryptosystems [6]. By utilizing DNA cryptography, several methods have been designed to break many modern algorithms like Data Encryption Standard (DES) [7], RSA [8], [9] and Number Theory Research Unit (NTRU) [10], [11]. Traditional cryptographic systems have long legacy and are built on a strong mathematical and theoretical basis. Traditional security systems like RSA, DES or NTRU are also found in real time operations. So, an important perception needs to be developed that the DNA cryptography is not to negate the tradition, but to create a bridge between existing and new technology. The power of DNA computing will strengthen the existing security system by opening up a new possibility of a hybrid cryptographic system. This paper gives a simple comparison between traditional and DNA cryptographic methods. It gives an insight to the benefits which can be achieved with the help of DNA cryptography.

2. DNA

Deoxyribo Nucleic Acid (DNA) is the hereditary material of almost entire living organisms ranging from very small viruses to complex human beings [17]. It is an information carrier of all life forms. DNA is a double helical structure with two strands running anti parallel as shown in Figure 1. DNA is a long polymer of small units called nucleotides. There are four different nucleotides depending upon the type of nitrogenous base they have got.



There are four different bases A, C, T, G called Adenine, Cytosine, Thiamine and Guanine respectively [18], [19].DNA stores all the huge and complex information about an organism with the combination of only these four letters A, C, T and G. These bases form the structure of DNA strands by forming hydrogen bonds with each other to keep the two strands intact. A forms hydrogen bond with T whereas C and G forms bonds with one another [20]. It can be seen in Figure1. Until 1994, DNA was believed to carry only the biological information but it was Adleman who revealed the computational ability of DNA when he solved NP complete Hamiltonian path problem of seven vertices [1]. After that DNA has been used as a computational tool as well [21]. DNA computers deal with the DNA language that consists of four letters A, C, T and G [17], [18]. The computational ability of DNA is now used in cryptography as well. DNA cryptography is a very potential field and if manipulated in appropriate manner can give much harder competition to other fields of cryptography [22].

3.DIFFICULT BIOLOGICAL PROBLEM USED IN THIS SCHEME

DNA is the germ plasm of all life styles. In a double helix DNA string, two strands are complementary in terms of sequences, that is A to T and C to G according to Watsoncrick rules, which is one of the greatest scientific discoveries.Some unresolved difficult biological problem in DNA science might have special value in cryptography and can achieve a new encryption technique. There are more difficult problem are more complex than biological problem. This absolutely different from well studied difficult mathematical problems. Here in our study we selected a typical difficult biological problem to develop an encryption scheme and tried to discuss the security of this scheme.

I..XOR OPERATION

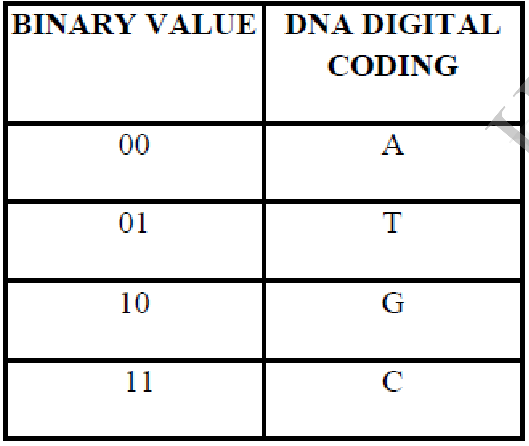
XOR operation is applied over the message and key to increase the repetition of normalized binary bit of 1’s and 0’s,there by gaining high compression factor.

II.Use of Enzymes

In the proposed system databse of enzymes are developed,which take the start and end enzyme randomly and that enzymes are converted into A,C,T,G format.datbase are made with two colum enzyme & Replacement text.

III.DNA DIGITAL CODING TECHNOLOGY

In the information science , the most fundamental coding method is Binary Digital Coding, which is anything can be encoded by two state 0 or 1 and a combination of o and 1,thereare four kind of bases, which are ADENINE(A) and THYMINE(T) or CYTOSINE(C) and GUANINE(G) in DNA sequence. The simplest coding pattern to encode nucleotide bases (A,T,G,C) is by means four digits: 0(00),1(01),2(10),3(11),there are possibly 4!=24 possible pattern by encoding format like (0123/CTAG).



BINARY VALUE	DNA DIGITAL CODING
00	A
01	T
10	G
11	C

IIII. Garbage file

Algorithm generates a random garbage file of thousands of character which is in A,C,T,G format where actual message is stored.

4. PROPOSED SYSTEM

Configuration file or needed file for input.

1.START ENZYME<-Randomely pick by the algorithm from the enzyme database.

2. END ENZYME<-Randomely pick by the algorithm from the enzyme database.

3.ASK FOR KEY INPUT FROM THE USER OR ITS GENERATED AUTOMATICALLY BY THE ALGORITHM.

4.GENERATE A GARBAGE FILE AUTOMATICALLY BY THE ALGORITHM WHICH IS IN ACTG FORMAT.

ENCRYPTION STEPS:

STEP:1 pick the xml file to be encrypted which is Disseminated from the internet.

Step2: left shift the key n times which are taken by the user or randomly generated by the algorithm.

Step3:chop the xml file on the basis of space using tokenizing method.

Step4:convert each token into its binary format from ascii code of each character.

Step5:chop the binary string in the piece of 32 bit.

Step6:convert each 32 bit block into integer format.

Step7:xor each integar with key.

Step8:convert each xored result in to its equivalent binary.

Step9:chop the result by length of 2.

Step10:convert each chopped part into A,C,T,G format where A-00,C-01,T-10,G-11 .

STEP:11 convert <SE>start enzyme & EE> end enzyme into replacement text which are generated randomly by the database which is also in ACTG format.

STEP:12 <SE>+MSG+<EE> the whole message is encrypted and it is in A,C,T,G FORMAT SAY IT IS M.

TO PLACE ENCRYPTED MESSAGE IN GARBAGE FILE

1.PICK A RANDOM USEFUL NUMBER SAY T<SIZE OF GARBAGE FILE.

2.STORE T AT ANY PREDEFINED LOCATION IN GARBAGE FILE.

3.GO TO GARBAGE FILE AT T LOCATION AND PLACE THE MESSAGE M.

4.XOR T WITH KEY K SAY IT IS N.

5.CONVERT N INTO BINARY AND CHOP IT BY LENGTH 2.

6.CONVERT IT INTO A,C,T,G FORMAT.

7.APPEND THIS STRING AFTER M,

DECRYPTION STEPS

1.ENTER START ENZYME,END ENZYME WHICH ARE GENERATED BY THE ALGORITHM.

2.ENTER KEY WHICH ARE USER INPUT OR GENERATED AUTOMATICALLY BY THE ALGORITHM IN ENCRYPTION STEP.

3.INPUT GARBAGE FILE.

4.JUMP ON THE PREDEFINED LOCATION AND RETRIVE THE LOCATION L.

5.JUMP ON L.

6.SKIP THE START ENZYME IF MATCHES.

7.RETRIVE MESSAGE M BY SUBSTRACTING END ENZYME.

8.CONVERT M INTO BINARY FORMAT.

9.CHOP IT INTO 32 BLOCK.

10.XOR EACH BOLCK WITH KEY K.

11.WE GOT ORIGNAL MESSAGE.

5. Advantages of using DNA Cryptography

1. The biggest advantage of public key cryptography is the secure nature of the private key. In fact, it never needs to be transmitted or revealed to anyone.

2. Moreover, encrypting it along the DNA sequence makes it more secure. A gram of DNA contains 10^21 DNA bases = 10^8 tera-bytes. A few grams of DNA may hold all data stored in world.

3. Since DNA is used for encryption, Signature authorization is not needed. DNA replaces the cause of Digital signatures and digital timestamps. 4. Can work in a massively parallel fashion: DNA is modified biochemically by a variety of enzymes, which are tiny protein machines that read and process DNA according to nature's design. There is a wide variety and number of these "operational" proteins, which manipulate DNA on the molecular level. For example, there are enzymes that cut DNA and enzymes that paste it back together. Just like a CPU has a basic suite of operations like addition, bit-shifting, logical operators (AND, OR, NOT NOR),

etc. that allow it to perform even the most complex calculations, DNA has cutting, copying, pasting, repairing, and many others. And note that in the test tube, enzymes do not function sequentially, working on one DNA at a time. Rather, many copies of the enzyme can work on many DNA molecules simultaneously.

5. Large storage: A gram of DNA contains about 1021 DNA bases, or about 108 tera-bytes. Hence, a few grams of DNA may have the potential of storing all the data stored in the world.

6. The main goal of the research of DNA cryptography is exploring characteristics of DNA molecule and reaction, establishing corresponding theories, discovering possible development directions, searching for simple methods of realizing DNA cryptography, and laying the basis for future development.

7. Input and output of the DNA data can be moved to conventional binary storage media by DNA chip arrays.

6. Conclusion

Due to the development and advance of science, the possibilities are rapidly growing. DNA in connection to cryptography is a fast developing interdisciplinary area. The Proposed method adds some artificial features to make the resulting cipher texts difficult to break. The theoretical analysis shows that this method is powerful against certain attacks, especially against flodding and men-in-middle attacks. The future of this area looks very promising, seeing as DNA is a medium for ultra-compact information storage

References

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

[2] G. Cui, Y. Liu, and X. Zhang, “New direction of data storage: DNA molecular storage technology,” Computer Engineering and Application, vol. 42, no. 26, pp. 29–32, 2006.

[3] J. Chen, “A DNA-based, biomolecular cryptography design,” in IEEE International Symposium on Circuits and Systems (ISCAS), 2003, pp. 822–825.

[4] R. Barish, P. Rothemund, and E. Winfree, “Two computational primitives for algorithmic self-assembly: copying and counting,” Nano Letters, vol. 5, no. 12, pp. 2586–2592, 2005.

[5] P. Rothemund, N. Papadakis, and E. Winfree, “Algorithmic self-assembly of DNA sierpinski triangles,” PLoS Biology, vol. 2, no. 12, pp. 2041–2053, 2004.

[6] G. Cui, L. Qin, Y. Wang, and X. Zhang, “An encryption scheme using DNA technology,” in IEEE 3rd International conference on Bio-Inspired Computing: Theories and Applications (BICTA08), Adelaid, SA, Australia, 2008, pp. 37–42.

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

[8] D. Beaver, “Factoring: The DNA solution,” in 4th International Conferences on the Theory and Applications of Cryptology. Wollongong, Australia: Springer-Verlag, Nov. 1994, pp. 419–423.

[9] Y. Brun, “Arithmetic computation in the tile assembly model: Addition and multiplication,” Theoritical Computer Science, Science Direct, Elsevier, vol. 378, no. 1, pp. 17–31, 2007.

[10] O. Pelletier and A. Weimerskirch, “Algorithmic self-assembly of DNA

tiles and its application to cryptanalysis," in Proceedings of the Genetic
and Evo- lutionary Computation Conference 2002 (GECCO02), New
York, USA, 2002, pp. 139–146.
[11] X. C. Zhang, "Breaking the NTRU public key cryptosystem using selfassembly
of DNA tilings," Chinese Journal of Computers, vol. 12, pp.
2129–2137, 2008.
[12] T. Kazuo, O. Akimitsu, and S. Isao, "Public-key system using DNA as
a one-way function for key distribution," BioSystems, Elsevier Science,
vol. 81, no. 1, pp. 25–29, 2005.
[13] N. Galbreath, Cryptography for Internet and Database Applications:
Developing Secret and Public Key Techniques with Java. New York,
USA: John Wiley and Sons, Inc., 2002.
[14] A. Menezes, P. Oorschot, and S. Vanstone, Handbook of applied
cryptography. CRC Press, 1996.
codes," in 8th IMA International Conference on Cryptography and
Coding, Cirencester, UK, Dec. 2001, pp. 1–8.
computation: Solving the elliptic curve discrete logarithm problem
over gf(2n)," Journal of Biomedicine and Biotechnology, Hindawi., vol.
2008, pp. 1–10, Apr. 2008.