DNA Computing and Its Application to Information Security Field

Guangzhao Cui, Cuiling Li, Haobin Li, Xiaoguang Li
Henan Key Lab of Information-based Electrical Appliances, Zhengzhou 450002

Abstract

DNA computing is a new computational paradigm by harnessing the potential massive parallelism, high density information of bio-molecules and low power consumption, which brings potential challenges and opportunities to traditional cryptography. In this paper, on the basis of reviewing the principle of DNA computing and the development situation of DNA computing briefly, we analyze some schemes with secret key searching and introduce the application of DNA computing in encryption, steganography and authentication.

1. Introduction

DNA computing is a new method of simulating biomolecular structure of DNA and computing by means of molecular biological technology which is a novel and potential growth interdiscipline. In a pioneering study, Adleman demonstrated the first DNA computing [1], it marked the beginning of a new stage in the era of information. This approach has been extended by Lipton to solve another NP-complete problem, which is the satisfaction problem [2]. These elegant studies demonstrated how problems corresponding to Boolean formulas can be solved by a massively parallel processing procedure. DNA computing has been proposed to solve difficult combinatorial search problems such as the Hamiltonian path problem (HPP), using the vast parallelism to do the combinatorial search among a large number of possible solutions represented by DNA strands. In 2002, Braich, R. S. etc got the solution of a 20-Variable 3-SAT Problem on a DNA Computer [3] However, DNA computing has many further exciting applications besides the pure combinatorial search. It can simultaneously attack different parts of the computing problem put forward challenges and opportunities to traditional information security technology. For example, in 1995, Boneh et al. demonstrated an approach to break the Data Encryption Standard (DES) [4], [5] by using DNA computing methods. In 1999, Clelland et al. achieved an approach to steganography by hiding secret messages encoded as DNA strands among a multitude of random DNA. DNA and RNA are appealing mediums for data storage due to the very large amounts of data that can be stored in compact volume [6], [7]. They vastly exceed the storage capacities of conventional electronic, magnetic, optical medium. A gram of DNA contains about 10^{21} DNA bases, or about 10^8 tera-bytes. Hence, a few grams of DNA may have the potential of storing all the data stored in the world [8], [9]. Recent research has considered DNA as a medium for ultra-scale computation and for ultra-compact information storage [10], [11]. DNA cryptography is a new born cryptographic field emerged with the research of DNA computing [12], [13], [14], [15] in which DNA is used as information carrier and the modern biological technology is used as implementation tool. The vast parallelism and extraordinary information density inherent in DNA molecules are explored for cryptographic purposes such as encryption, authentication, signature, and so on. The new born DNA cryptography is far from mature both in theory and realization, and this might be the reason why only few examples of DNA cryptography were proposed.

1.1. The Principle of DNA Computing

DNA (Deoxyribonucleic Acid) is the germ plasm of all life styles. It is a kind of biological macromolecule made up of nucleotides. Each nucleotide contains a single base. There are four kinds of bases, which are adenine (A), thymine (T), cytosine (C) and guanine (G). In a double helix DNA string, two strands are complementary in terms of sequence , that is A to T and C to G according to Watson-Crick rules.

The principle of DNA computing makes use of the special double helix of DNA molecular and Watson-Crick rules. The process of DNA computing mostly

978-0-7695-3736-8/09 $25.00 © 2009 IEEE
DOI 10.1109/ICNC.2009.27

includes three steps:1 Encoding all candidate solutions to the computational problem of interest; 2 Controllable reaction under enzymes, which generates all kinds of data pool that includes possible solution to the computational problem; 3 Extracting the problem¡s solution by a polymerase chain reaction (PCR).

2. DNA Computing Challenges Traditional Cryptology

The security of traditional cryptology is usually based complex mathematical problem that we can¡t find a quick algorithm at this stage, such as famous Rivest-Shamir-Adleman (RSA) encryption ,the security of which bases on the difficulty of a large number finding its two prime factors. Once corresponding quick methods to mathematic problems were found, they might be no longer secure. DNA computing provides a parallel processing capability with molecular level, introducing a fire-new data structure and calculating method. It can simultaneously attack different parts of the computing problem, putting forward challenges to traditional information security technology. A number of proposals have been submitted for breaking conventional cryptosystems by DNA computing. It indicated that the cryptosystem using public-key was perhaps insecure.

2.1. Breaking DES

DES is a cipher which based on a Symmetric-key algorithm that uses a 56-bit key. The algorithm was initially controversial with classified design elements, a relatively short key length, and suspicions about a National Security Agency backdoor. DES is now considered to be insecure for many applications. This is chiefly due to the 56-bit key size being too small. Dan Boneh et al. spent nearly 4 months to construct DES^{-1} liquid and then broke DES within a day [4]. They claimed that any symmetric system under 64 bits can be broken with this method. The process to solve this kind of problem is listed as follows: Firstly, encode appropriate binary codes, create initial DNA liquid which contains all possible keys; Secondly, carry out 16 wheels of encryption after pasted known plaintext strands respectively. Lastly, find the solution by searching. Though the idea brought up by Boneh et al. is comparatively simple in theory, what they used in the experiment are mainly extracting, separation, pasting etc. the concrete operation is not easy in real experiments since the method of binary system is comparatively abstract.

2.2. Breaking RSA

In cryptography, RSA is an algorithm for public-key cryptography. The security of the RSA cryptosystem is based on two mathematical problems: the problem of factoring large numbers and the RSA problem. Full decryption of an RSA ciphertext is thought to be infeasible on the assumption that both of these problems are hard, i.e., no efficient algorithm exists for solving them.

Weng-Long Chang et al. have designed integer factorization way of utilizing DNA computing, which can be used to break RSA. On the basis of Adleman¡ thought, Beaver et al. translated large number of decomposition problems into HPP [16], they analysed 1000 bits RSA and concluded that to solve that problem required the acme number to be 106 at least, namely 10^{200000}L liquid to be needed on the grounds of conservative estimation. Obviously it is infeasible . For this, Winfree et al. came up with the idea of computation by self-assembled tiles since DNA tiles can be more easily ¡programmed¡ to incorporate the constraints of a given problem. Brun proposed in theory the systems could compute the sum and product of two numbers using the tile assembly model [17]. He found that the addition and multiplication can be done using $\Theta(1)$ tiles and both computations can be carried out in linear time with the tile assembly model and then he combined those systems to create two new systems with more complex behavior to factor numbers and solve subset sum problems [18], [19]. Xuncai Zhang proposes a scheme which using self-assembly of DNA tiles subtraction [20] which can be used to factor integers.

2.3. Breaking Number Theory Research Unit

Number Theory Research Unit (NTRU) is a novel, efficient public-key cryptosystem proposed by Hoffstein et al, which is based on a new problem, the closest vector problem. Its security comes from the interaction of the polynomial mixing system with the independence of reduction modulo of two relative prime integers p and q. Here, we show now that it is possible to break the NTRU cryptosystem using Self-Assembly of DNA Tilings. Such attack has already been used in Res.[21].The attack involves computation of cyclic convolution product modulo some integer and it is this part that has been shown to be done

efficiently using DNA Wang tiles which effectively breaks the system. However, the main disadvantage of the attack is that it is not fully supported by physical experiments and the Wang tiles that are used for the attack to form 3D self-assembly have not yet been made practically. On their basis Xuncai Zhang uses existing tiles to carry out the attack NTRU cryptosystem. The basic idea is to exploit the massive parallelism possible in DNA operations in order to emulate a non-deterministic device that breaks the NTRU system in polynomial time. Such emulation can be achieved by exponential-order parallelism [22].

2.4. Breaking International Data Encryption Algorithm

International Data Encryption Algorithm(IDEA) is a symmetric block cipher with a 128-bit key space. The security of IDEA is ensured by confusion and diffusion, which treat with the data in the IDEA. It has been adopted by Pretty Good Privacy. Researchers have finished a great of work in attacking IDEA, such as super-scale integrate circuit calculating and analysis, various parallel computing analysis, as well as other attack schemes against IDEA. But the security of IDEA does not to be threat. DNA computing as research method, in order to illuminate the feasibility of breaking symmetric block cipher with DNA computing, a recursive splicing model DNA algorithm to break IDEA is proposed by Xiutang Geng, and concrete implement process is given in Res.[23].

Although DNA computing is a fire-new computing mode, it can¡t get away from the influence of Turing in the corresponding theoretical computing model. Under the existing DNA computing mode, the time complexity of DNA computing doesn¡t increase with the computational complexity remarkably. But it only converts the time complexity into space complexity. Then, once the complication of problems break the physical limit of DNA segment which operated by the bio-chemical technique, DNA computing is still too far away to reach. Up to now, methods of traditional decryption based on DNA computing can not evade from exponential explode cycle. In spite of that, DNA computing has greatly improved the ability of people to break the cipher.

3. The Application of DNA Computing to Security Field

The advent of a new technique always draw attention of many persons coming form different disciplines, forming a interdisciplines. This is same with the application of DNA computing in security field. It primarily includes: Encryption, Steganography and Authentication.

3.1. DNA Encryption Techniques

Cryptography and data security are critical aspects of conventional computing. They are about communication in the presence of an adversary and encompassed by many problems. An example of a private key encryption method which is secure even in presence of a computationally unbounded adversary is the One-Time-Pads (OTP) cipher. It is known that, OTP is absolutely secure in theory. But practically, key distribution and key generation are critical bottlenecks for the use of OTP cryptosystem. DNA as information carrier has high memory density, which will be better solve the huge cipher key producing and saving problem, providing a even road for the OTP. Prof. Gehani designed two encryption methods utilizing this idea [12]. One method is to translate the fixed length DNA plain code sequence cell to DNA cryptograph sequence according to the defined mapping graph, we call it mapping substitute. The other is called exclusiveor method, which uses biological molecular techniques to carry through exlusiveor operation of DNA plain code and cipher key sequence. It is absolutely secure to use these two methods of OTP encryption mechanism.

Other methods based on DNA computing are presented continuously by many researchers. Andre Leier presented two different cryptographic approaches based on DNA binary strands [24]. The first approach hid information in DNA binary strands and the second approach designed a molecular checksum; Jie Chen proposed carbon nano-tube based message transformation and DNA-based cryptosystem[9]; Kazuo Tanaka used the message-encoded DNA hidden in dummies which can be restored by PCR amplification, followed by sequencing etc operations as a one-way function have constituted a novel method for the key distribution based on the public-key system using DNA[15]; Nini Rao presented a cryptosystem based on recombinant DNA technique [25]; Limin Qin designed an encryption scheme by using the technology of DNA synthesis, PCR amplification and DNA digital coding [26].

At present encryption schemes based on DNA computing mostly utilize its handling step, structure or biochemistry condition. DNA cryptography is still in a

probe stage, to accurately forecast its development in future is hard.

3.2. DNA Steganography

The principle of DNA steganography is to conceal the information which needs encryption in the large numbers of irrelevant DNA sequence chains. This way of decoding like looking for a needle in a heap of hay which make attackers difficult to ascertain the correct DNA fragment. Only the proper receiver can find the correct DNA fragment based on the conventional information in advance between the two parties as well as requires the information which concealed in it. One can argue that steganography is not actually encryption, since plaintext is not encrypted but only disguised within other media.

For DNA steganography, Bancroft et al. successfully hid the famous ¡June 6 invasion: Normandy¡ in DNA microdots, by using an alphabet of exoteric short nucleic acid sequences [27]. Prof. Gehani etc. discussed various modified DNA steganogram methods which appeared to improve the security.

DNA steganography does not absolutely repulse traditional cryptography and it is possible to construct a hybrid cryptography of them. If we use different traditional encryption methods to preprocess to the plaintext, we can get completely different ciphertext from the same plaintext, it can effectively prevent attack from a possible word as PCR primers. This way has one more layer of protection than the simplex steganography technique, which provides a novel thought for information security and a new orientation for its research. We think the further development needs to do more and deeper research in this area.

3.3. DNA Authentication

Currently, the DNA certification is broadly applied in the field of justice, finance etc. Strictly speaking, DNA certification doesn¡t refer to too many DNA computing techniques, while mainly employ the biological characteristics of DNA. However, if we apply the DNA computing to the DNA authentication , it would improve the complication of algorithm and the level of security.

There are lots of biological genetic engineering under way in recent years, many biological technologies have matured. Researchers can add the DNA certification information to the organ tissue to validate the customer¡s identity and the copyright information. For example, DNA Technology Company of Canada using the DNA sequence to the product certification of the Sydney Olympic game in 2000. All in all, DNA authentication has broad prospect.

4. Conclusion

Although DNA computing creates a molecular computing precedent and broadens the understanding of people to natural computing phenomena, it still stayed in a theoretical stage. There are some problems unresolved successfully about DNA computing:1 Its computing model is mostly just using molecular technique to resolve a certain problem, the varieties of problems result in the discrepancy of computing schemes, there still haven¡t an uniform computing and coding model currently; 2 DNA computing only converts the time complexity into space complexity; 3 There are also error codes in DNA computing, they generate randomly according to probability and can gradually amplified with the increase of the experiment step; 4 DNA liquid is very easy to deteriorate in the process of reaction and even adsorption of the test tube wall may result in fatal error; 5 Most of these proposals implemented computing processes by performing a series of biochemical reactions on a set of DNA molecules, which require human intervention at each step. Thus, the difficulties of such methods for DNA computing are that the large numbers of laboratory procedures and the time consuming, which grow with the size of the problem. Therefore, DNA computing is not very good in resolving real problem according to the available pattern in recent year.

Therefore, in terms of existing DNA computing mode, it is not able to construct real intimidation to the security of cryptography. At the same time, all kinds of encryption scheme pouring out unceasingly based on DNA computing, providing escorting to the DNA molecules¡ bank and DNA molecules¡ information. But, because the security, general, validity and key management of the encrypt mechanism have not been carried out systematic theory analysis. So, DNA cryptography still needs studying exclusively and discussion broadly, its prospect is still uncertain before DNA computing become really mature. But DNA cipher is the beneficial supplement to the existing mathematical cipher, it is a prior choice especially to the lower demand real-time encryption system. Relatively speaking, DNA computing has a brighter development potential in steganography and authentication, which

have a more layer protection than a single encryption. With the rapid development of modern biotechnology, the costly biological experiment has became a normal one. If the molecular word can be controlled at will, it may be possible to achieve vastly better performance for information storage and information security.

Acknowledgments. This paper is supported by the National Natural Science Foundation of China (Nos. 60573190, 60773122), the Natural Science Foundation of Henan Province (Nos. 082300413203) and 2007 Dr. Fund-sponsored project (Nos. 2007BSJJ003).

References

[1] L.M. Adleman, Molecular computation of solutions to combinational problems. *Science*, (1994), 266(4), pp. 1021-1025.

[2] R.J. Lipton, Using DNA to solve NP-complete problems. *Science*, (1995), 268(4), pp. 542-545.

[3] R.S. Braich, N. Chelyapov, and C. Johnson, Solution of a 20-Variable 3-SAT Problem on a DNA Computer, *Science* , (2002), 296, pp. 499-502.

[4] D. Boneh, C. Dunworth, and R. Lipton, Breaking DES using a molecular computer. *In Proceedings of DIMACS workshop on DNA computing*, (1995), pp. 37-65.

[5] L.M. Adleman, P. Rothemund, and S. Roweis, On applying molecular computation to the date encryption strands in DNA based computers, *In Proceedings of 2nd DIMACS Workshop on DNA Based Computers,* (1996), pp. 28-48.

[6] C.T. Celland, V. Risca, and C. Bancroft, Hiding messages in DNA microdots. *Nature*, (1999), 399, pp. 533-534.

[7] J.P.L. Cox. Long-term data storage in DNA. *Trends Biotechnol*. (2001), 19, pp. 247-250.

[8] G.Z. Cui, Y.L. Liu, and X.C. Zhang, New Direction of Data Storage: DNA Molecular Storage Technology, *Computer Engineering and Applications*, (2006), 42(26), pp. 29-32.

[9] J. Chen. A DNA-based, biomolecular cryptography design. *Circuits and Systems ISCAS apos,* (2003), pp. 822-825.

[10] M. Amosa, G. Paun, and G. Rozenbergd, Topics in the theory of DNA computing, *Theoretical Computer Science*, (2002), 287, pp. 3-38.

[11] G.Z. Xiao, M.Q. Lu, and L. Qin, New field of cryptography: DNA cryptography, *Chinese Science Bulletin*, (2006), 51(10), pp. 1139-1144.

[12] A. Gehani, T.H LaBean, J.H. Reif, DNA-based cryptography, In: *5th DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, MIT, (1999), vol. 54, pp. 233-249..

[13] A. Leier, C. Richter, and W. Banzhaf, Cryptography with DNA binary strands, *Biosystems,* (2000), pp. 13-22.

[14] S.V. Kartalopoulos, DNA-inspired cryptographic method in optical communications, authentication and data mimicking, *in Military Communications Conference*. (2005), pp. 774-779.

[15] T. Kazuo, O. Akimitsu, and S. Isao, Public-key system using DNA as a one-way function for key distribution, *Biosystems*, (2005), pp. 25-29.

[16] D. Beaver, Factoring: The DNA Solution, *Proceedings of the 4th International Conference on the Theory and Applications of Cryptology: Advances in Cryptology*, (1994), pp. 419-423

[17] Y. Brun, Arithmetic computation in the tile assembly model: Addition and multiplication, *Theoretical Computer Science*, (2006), 378, pp. 17-31.

[18] Y. Brun, Nondeterministic polynomial time factoring in the tile assembly model, *Theoretical Computer Science*, (2007), 395, pp. 3-23.

[19] Y. Brun, Solving NP-complete problems in the tile assembly model, *Theoretical Computer Science*, (2007), 395, pp. 31-44.

[20] X.C. Zhang, Y.F. Wang, and Z.H. Chen, Arithmetic Computation Using Self-Assembly of DNA Tiles: Subtraction and Division, *Progress in Natural Science*, (2009), pp. 377-388.

[21] O. Pelletier, A. Weimerskirch, Algorithmic Self-Assembly of DNA Tiles and its Application to Cryptanalysis, *in Genetic and Evolutionary Computation Conference*, N.Y. USA, (2002), pp. 139-146.

[22] X.C. Zhang, Breaking the NTRU Public-key Cryptosystem Using Self-Assembly of DNA Tilings, *Chinese Journal of Computers*,(2008), pp. 2129-2137.

[23] X.T. Geng, Research on Molecular Algorithms in Block Cipher and Graph Theory, *A Dissertation for the Degree of Doctor*, Huazhong University of Science & Technology,(2008).

[24] A. Leier, C. Richter, and W. Banzhaf, Cryptography with DNA binary strands, *Biosystems*, (2000), pp. 13-22.

[25] N.N. RAO, A Cryptosystem Based on Recombinant DNA Technique, *Acta Electronica Sinica* (2004), pp. 1216-1218.

[26] G.Z. Cui, L.M. Qin, and Y.F. Wang, An Encryption Scheme Research with DNA Technology, *Computer engineering and applications*, (2008), pp. 37-42.

[27] C.T. Clelland, V. Risca, and C. Bancroft, Hiding messages in DNA microdots, *Nature*, (1999), 399, pp. 533-534.