A New Molecular Encryption Model Based on Microfluidic Techniques

Pengcheng Ma¹ and Heng Sun¹*

¹ College of Information Science, Jinan University, Guangzhou, Guangdong, 510632, China
*Corresponding author’s e-mail: 71831960@qq.com

Abstract. Non-linearity behaviours of fluids have not been fully understood by human beings. In this paper, we present a new molecular encryption model based on microfluidic techniques. The security of molecular cryptography is based on non-linearity of fluids, rather than conventional mathematical hardness assumptions. As there is no mathematical analysis method to predict non-linear behaviours of fluids, it is unaffected by the attack of mathematical analysis. This method provides a new insight to information security and is hopefully to be a new promising direction.

1. Introduction

Molecular cryptography, which is different from conventional cryptography and quantum cryptography, is an emerging promising direction of information security using molecule to encrypt data. A similar research direction of this field is DNA cryptography, which is to encode messages in DNA sequences to conceal the information. With high density of DNA, ultra large information can be stored in a small amount of DNA material [1]. DNA cryptography has provided some new methods to cryptography. Clelland et. al. [2] put forward a DNA-based, doubly steganography technique for sending messages using microdots. This elegant study first showed an implementation of DNA cryptography and the significance fact of using microdots. Several years later, Gehani et. al. [3] proposed two DNA one-time-pad encryption schemes, a substitution method and an XOR scheme. These approaches present a DNA-based cryptographic method rather than number-based methods. Such studies attracted the attention of researchers to study the potential to create a novel cryptography system based on DNA. Plenty of researches are proposed [4, 5] and makes DNA cryptography a main part of molecular cryptography. Lai et. al. [5] proposed an asymmetric encryption method and put forward a new hardness assumption, which was the biological hardness assumption. Using molecules to encrypt messages provides a security different from conventional cryptography which can resist the attack methods developed in conventional cryptography and may even resist quantum computer attack [5]. However, none of these approaches to date hold an efficient transmission method. DNA can only be transported by physical methods, which are not convenient to use in practice. Microfluidics inspired us and we found fluids can be implemented for cryptography.

Recent studies in microfluidics field provided an important explanation of the interaction among droplets in the channel [6]. Microfluidic technology provides a high efficiency which is very suitable for cryptographic requirements. DNA computation and other applications using microfluidic devices have been proposed by early researchers, which show feasibility of using microfluidics in computer applications. Gehani et. al. provided a model for micro-flow based bio-molecular computation device
Grover et al. developed an integrated microfluidic processor that performs molecular computations using single nucleotide polymorphisms as binary digits [8]. Fuerstman et al. studied the interaction among droplets in the channel and proposed an encoding/decoding microfluidic network [9]. We design a microfluidic selector to perform molecular encryption methods we proposed. A new hardness assumption, which is based on the nonlinearity of fluids, is provided. This method has not yet explicitly been addressed and in addition included in recently published papers. Lastly, this method is the first try to connect molecular encryption method with modern communication systems, which may provide a new insight of information manipulation and technique.

The paper is divided into two parts. The first part will present an encryption algorithm and the second part will present a designed microfluidic selector to implement this algorithm.

2. Methods

2.1. Molecular encryption principle

Several terminologies used in this encryption method, which has never been explained in this field.

1) Molecular code unit. Molecular code is a new encoding scheme which provides a method to encode information in molecular form. The representation of molecular code unit depends on the chosen molecules. For example, we can use DNA triplet to represent each molecular code unit. Multiple molecular code units map a binary number. The coding scheme is the basis of our molecular encryption method. In this paper, DNA molecule is used as an example throughout this paper to explain how to use molecular code unit. Other materials can also be used for molecular code unit representation and the design of each molecular code unit can be different based on the situation.

2) Molecular code pad. The molecular code pad is an integration of all possible molecular code units. The input source can be binary bits, DNA sequences or any other molecular code form. As modern communication systems use binary bits, molecular code can be converted back to its binary form and be transmitted through existed systems. Molecular code units can be implemented using various numbers of DNA bases. For instance, if we use triplets (three DNA bases, ACT, CTG et al), there will be 64 different triplets. These 64 triplets comprise of a molecular code pad. Then, we choose half triplets to represent 0 and others to represent 1 by using random algorithms. For example, we choose the first triplet denoted as 1 and the second as 0, then the choice continues alternatively. In this way, binary bits are encoded into DNA triplets using previously generated substitution table, which is called 0-1 substitution rule. As we can see, each 1 and 0 have 32 potential DNA triplets which means this substitution is onto, and there are \(A_{64\times64} \approx 1.2 \times 10^{99}\) molecular code pads in the chosen scheme. Besides, each triplet has the same probability to be chosen as 0 or 1. Thus, there are \(C_{64\times32}=1.8 \times 10^{18}\) possibilities for 0-1 substitution. The key space is the multiplication of above, which is \(A_{64\times64} \times C_{64\times32} \approx 2.3 \times 10^{107}\). The security of molecular code pad also lies in the usage of nonlinear characteristic of fluids. As the generation of molecular code pad is nonlinear, it cannot be predicted or analyzed out by mathematical analytical methods. The generation algorithm will be present later.

3) Message string unit. With further research, using only molecular code pad has some special situations, which are insecure. Suppose a sender successfully generates a molecular pad and finished choosing substitution triplets, but when he uses the substitution triplets, each binary bit (0 or 1) is substituted by the same triplet, like all 1 for ACT and 0 for CTG, then there will be only two possible cases for an eavesdropper. To prevent such situations, we create a message string unit which consists of three segments: P segment (position segment), message segment and confusing message segment as shown in Figure 1 (c). Suppose the string unit is of 256 ng. The three segment each other is 3 ng, 144 ng (6 bytes), 109 ng. The position of each three segments are not fixed, which is in order to hide the position of true secret messages. In this way, the eavesdroppers cannot get the secret messages, so they have no knowledge to analyze the statistical frequency of each character. The P segment is used to help the legal receiver to find the secret message segment and decrypt the message. It is selected by the sender and sent with the secret key. The position of message segment is not fixed, but P segment is always before message segment. P segment and message segment is connected and confusing message
segment cannot be in the middle of these two segments. In order to maximize the utility of this structure, we should generate confusing segments according to the probability of each molecular code units in secret messages. Message string unit provides necessary confusion of secret messages.  

4) Molecular hardness assumption. Different from early works by other researchers, the hardness of this cryptographic method lies in the non-linearity of fluids. Different viscosity number of different fluid materials can lead to different motion velocities. Based on the non-linearity of viscosity, the output will be different from input sequences, which is the same as diffusion of messages. The condition parameters set is the secret key. The message is meaningless to an eavesdropper who has no knowledge of the condition parameters (key), and hence the correct molecular code pad cannot be generated. Besides, mathematical analysis cannot be applied to such encryption method using molecular hardness assumptions. This molecular hardness assumption is a new hardness assumption for cryptography.  

2.2. Cryptographic algorithm

1) Description of algorithm. This algorithm is a one-time-pad stream cipher using molecular code scheme and microfluidic selector as a random pad generator. The input messages can be text, audio or video. 

2) Encryption process. In the first step, the molecular code pad can be chosen by the sender. After the sender generates the molecular code pad, he will choose 32 triplets as 0 and others as 1. After converting input messages to molecular codes, message string units are formed according to the principle described. In the last step, the message string units are transformed back into binary format again by simply map A to 11, T to 00, C to 01 and G to 10. Finally, the plaintexts are encrypted and can be transmitted through traditional networks. 

3) Decryption process. The process of decryption is essentially the same as that of the encryption like modern cryptography. What we should note is that messages are first converted to molecular form and then use the same generated molecular code pad and 0-1 substitution rule to decrypt the messages. The implementation details of generating molecular code pad are described in the key exchange process. 

4) Key exchange process. This algorithm is a secret-key encryption method. The first secret key is the seed and transmitted through a secret channel. After that, the next molecular code pad will use the sequence of code pad generated last time and P segment chosen is transmitted in the secret message. 

![Figure 1](image)

Figure 1. (a) An example of molecular code pad. In this paper, the molecular code units are DNA triplets. The 64 triplets are connected one by one. Each molecular code pad could be different. (b) Flow chart of encryption algorithm. (c) An example of message string unit.
2.3. Microfluidic device design

A key part of our method is the generation of molecular code pad. Fluids at the small length scales that are characteristic of microfluidic systems [6] and viscosity-dominated flows are governed by equation of motion that are non-linear in the velocity of the fluid [9]. With different Reynold number Re and viscosity characteristics, the non-linearity of fluids can be used for random molecular pad generation. The microfluidic platform consists of a droplet generating device as shown in Figure 2(a) and a micro-selector as shown in Figure 2(b). The following part shows our design of microfluidic device for encryption.

1) Droplet generating device. The chip comprises three fluid inlets connected to three micropumps to form droplets of different solutions through T-junction. The droplets are moved by a constant force of micropumps. Before pumping into the microfluidic device, each fluid has been heated in desired temperatures given by sender.

2) Microselector. The microselector incorporates three uniquely modified channels. As the resistance force is different due to environment parameters, the velocity of each droplet is different. And due to the length of the channel each droplet pumped in, the time when they are pumped out is different from the original order. Each channel has one droplet in it and another droplet can be pumped in only when the previous drop was pumped out. The process is parallelized by increasing the number of devices. In this way, the upper bound of time complexity is determined by the droplet which consumes the longest time for each device. When a droplet moves through the outlets of microselector, a sensor will distinguish it and record the sequence on computer. After all droplets move out the micro-selector, a molecular code pad is generated. The used droplets can be collected and be prepared to be used next time.

![Figure 2. (a) Conceptual design of droplet generating device. The intended triplets are pumped in based on the sequence. (b) Conceptual design of microselector. The droplets move through the channels in different velocities and move out in different sequences from the origin.](image)

3. Results and Discussion

An encryption process has been presented. In the example of this paper, each character contains 8 bits, which is 24 ng, then a secret message segment can consist of 6 characters. As for the security of this scheme, the first question is whether the eavesdropper can get the secret key. If the eavesdropper cannot get the secret key, then the system is secure. Although an eavesdropper by incidence has the correct molecular code pads, he should also know the position of secret messages, otherwise he can only use brute force attack to find the secret message. This proves although the eavesdropper obtains the secret key, he still cannot decrypt the message, which means the system is secure. The total number of key space depends on the number of molecular code pads and the number of 0-1
substitution rules. In our example, 32 triplets are mapped to 0 and others are mapped to 1. Each triplet has the same probability to be chosen and for each encrypted word of binary length N, so there are $32^N$ possibilities. What is the most important is that each bit has the same probability 1/32, which satisfy perfect secrecy and One-time-pad condition in Shannon’s theory [10], and if the eavesdropper does not have the correct molecular code pad and 0-1 substitution rule, the secret message cannot be found and then be analyzed by statistical characteristic. To obtain the correct molecular pad, the eavesdropper has to obtain the secret key, which is back to the first question. Another advantage of this method is that there is no worry about that statistical pattern will be revealed as the number of messages increase, because each bit has the same probability using 0-1 substitution. Molecular cryptography has two security folds, mathematical level and physical level. Different from modern cryptography hardness assumptions, molecular cryptography has its own hardness assumptions, which is non-linearity of fluids. Existed DNA encryption methods are like the first case described in Shannon’s work [10]. Our method shows how to transmit molecular data by inventing an encoding scheme using currently existed communication systems, like the Internet.

4. Conclusions and Perspectives
In this study, a microfluidic molecular encryption method has been present by using molecular code scheme. Using molecule to encrypt data is different from conventional encryption methods, and hence there is promising future. By converting encrypted messages into binary form, messages can be transmitted through today’s communication systems. This encryption scheme is not based on the advantage of the vast storage capacity of DNA, but with non-linearity of fluids, more data can be processed with one molecular code pad generation. Using microfluidic chip to implement this encryption scheme can also provide high-efficiency and be low-cost. With further research of this method and with new discovery in microfluidic field, more molecular characteristics can be used to design encryption methods.

Acknowledgments
This research was supported by the Fundamental Research Funds for the Central Universities (grant No. 21617407), "Climbing Program" Special Funds (grant No. pdjhbb0057), and Jinan University Funds for “Challenge Cup” National Undergraduate Curricular Academic Science and Technology Works (grant No. 17112059 &.18113037).

References
[1] Church, G.M.; Gao, Y.; Kosuri, S. (2012) Next-Generation Digital Information Storage in DNA. Science, 337: 1628.
[2] Clelland, C.T.; Risca, V.; Bancroft, C. (1999) Hiding Messages in DNA Microdots. Nature, 399: 533-534.
[3] Gehani, A.; LaBean, T.; Reif, J. (2004) DNA-Based Cryptography. LNCS. 2950: 34-50.
[4] Chang, W.L.; Guo, M.Y.; Ho, S.H. (2005) Fast Parallel Molecular Algorithms for DNA-Based Computation: Factoring Integers. IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, 4: 149-163.
[5] Lai, X.J.; Lu, M.X.; Qin, L.; Han, J.S.; Fang, X.W. (2010) Asymmetric Encryption and Signature Method with DNA Technology. Science China Information Science. 53: 506-514.
[6] Stone, H.A.; Stroock, A.D.; Ajdari, A. (2004) Engineering Flows In Small Devices: Microfluidics Toward a Lab-on-a-chip. Annu. Rev. Fluid. Mech. 36: 381-411.
[7] Gehani, A.; Reif, J. (1999) Micro flow bio-molecular computation. Bio Systems. 52: 197-216.
[8] Grover, W.H.; Mathies, R.A. (2005) An integrated microfluidic processor for single nucleotide polymorphism-based DNA computing. Lab Chip, 5: 1033-1040.
[9] Fuerstman, M.J.; Garstecki, P.; Whitesides, G.M. (2007) Coding/Decoding and Reversibility of Droplet Trains in Microfluidic Networks. Science. 9: 828-832.
[10] Shannon C.E. (2014) Communication Theory of Secrecy Systems. Bell System Technical Journal. 28: 656-751.