Journal of Information Hiding and Multimedia Signal Processing ©2019 ISSN 2073-4212
Ubiquitous International Volume **10**, Number **2**, March **2019**

A Chaotic Key Expansion Algorithm Based on Genetic Algorithm

Juan Wang [1,2], Bo-Wen Pan[2], Qian-Rui Wang[2], Qun Ding[1,*]

[1]Electronic Engineering Institute
Heilongjiang University
74 Xuefu Road, Harbin, 150001, China

[2]Electronic and Information Engineering Institute
Heilongjiang University of Science and Technology
2468 Puyuan Road, Harbin, 150027, China

* 76115347@qq.com, qunding@aliyun.com

Received July 2018; revised August 2018

Abstract. *The key expansion algorithm is a critical factor that affects the encryption performance of block cipher. In view of the digital chaotic sequence, there are obvious degradation of the dynamic characteristics and the phenomenon of local periodicity, a chaotic key expansion algorithm based on genetic algorithm according to the genetic mechanism that survival of the fittest is proposed in this paper. Taking the digital chaotic sequence as the initial population, using the fitness function to evaluate the merits and demerits of the individual sequence, and performing the operations of the selection, crossover, and mutation, so that the generation and expansion of the key are realized through the generation-by-generation evolution. The simulation and test results show that the key expansion algorithm is safe and efficient, which is easy to implement and can better meet the encryption requirements of lightweight block cipher.*
Keywords: key expansion; chaotic mapping; genetic algorithm; block cipher

1. **Introduction.** The block cipher is generally composed of two parts, namely the encryption round function and the key expansion algorithm. The key expansion algorithm refers to that, the initial master key generates subkeys according to a certain rule, and uses it for each round to perform encryption and decryption together with round function. In recent years, attacks against the key expansion algorithm are attracting increasing attention, the existence of weak key classes will make the block cipher have obvious attack weaknesses under their effect[1, 2]. For example, plaintext pair encrypted with the same key or two different related keys were used to implement a correlation key difference attack[3] for the key recovery, and a sliding attack[4] is performed using the periodicity of each round subkey to attack the cryptographic algorithm. Therefore, it is very important to design a secure and efficient key expansion algorithm for cryptographic security.

Apart from the general characteristics such as strong randomness, high complexity and large key space, the key expansion algorithm should also have independence and sensitivity to effectively resist all kinds of attacks against its defects[5]. At present, the implementation of key expansion is mainly divided into two categories. One is to operate on the master key directly to obtain each round subkey, and the other is to use round-by-round iterative output starting from the master key as each round subkey. The specific implementation method involves the simplest linear operation such as simple replacement and

cyclic shift, XOR, modular addition, and other linear operations that introduce avalanche effect, s-box, parameter control and other non-linear operations to enhance the confusion effect, as well as the improvement and optimization of the above methods [6].

By virtue of the initial value sensitivity, intrinsic randomness, and orbital ergodicity of the chaotic mapping, it would gradually evolve into a more secure and efficient key generation mechanism [7]. Due to the digital chaotic sequence has obvious dynamic characteristics degradation and local periodicity phenomenon [8], using it directly as a key can be easily deciphered by attackers through the probabilistic statistics and other methods. In response to this problem, this paper proposes to use the digital chaotic sequence as the initial population, calculates and orders the fitness function, and performs the operations of the selection, crossover, and mutation, sequentially generates the round subkeys required for block cipher encryption by population iterative optimization. The key expansion algorithm not only has more excellent cryptographic characteristics, but also can improve the efficiency of key generation and expansion, which is an ideal choice for lightweight block ciphers used in wireless sensor network and other fields.

2. Chaotic Mapping.

Chaos is a deterministic and pseudo random process produced by the nonlinear dynamic system [9]. Chaotic attractor has the characteristics of topological transitivity and miscibility and it is sensitive to the system parameter, as a result it fits for the confusion principle in cryptography design. Meanwhile, the divergence of trajectories and the sensitivity to initial value of chaos fit for the diffusion principle in cryptography design [10]. Compared with continuous chaotic system, the discrete chaotic mapping is fast to iterate and easy to control, having incomparable superiority in hardware implementation [11].

As a kind of chaotic mapping, although the logistic has been widely researched and applied owing to its simple mathematical model, the disadvantages of which are apparent, such as infinite fixed-point attractor, small surjective map interval, narrow parameter ranges and low complexity, etc.[12, 13, 14, 15]. When the system parameter is improperly chosen, the iterative results would tend to be a fixed-value or concentrated distribution, when used as the key directly the security of cryptographic algorithm would be influenced to a certain extent. Therefore, an improved logistic chaotic mapping is proposed in literature [16] and the iterative equation is

$$x_{n+1} = \mu x_n(1-x_n^2)\mathrm{mod}1 \qquad n = 1, 2, 3... \tag{1}$$

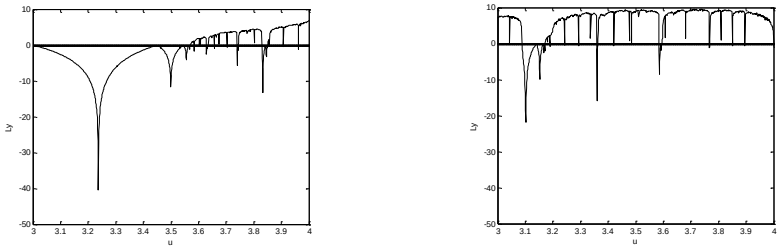
In Eq. (1), the interval of system parameter is $\mu \in (0, 4]$, the interval of initial value is $x_n \in (0, 1]$ and mod1 is the standard modulo 1 operation.

As shown in Figure 1, the Lyapunov exponent of logistic mapping is positive only when $\mu > 3.57$, while the parameter range corresponds to positive Lyapunov exponent of the improved logistic mapping is obviously enlarged. As shown in Figure 2, it is apparent that the attractor structure of the improved logistic mapping possesses a higher level of complexity. Although there is only a trivial improvement, the dynamic characteristics of the improved logistic mapping are significantly enhanced to provide a reliable guarantee for the security of keys and ciphers.

3. Genetic Algorithm.

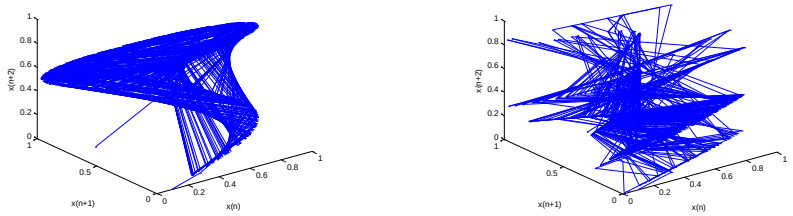
Genetic algorithm [17, 18] is an optimization method that simulates the evolution of biological population, which is widely used due to its many excellent characteristics. Genetic algorithm is essentially a kind of optimization process of population iteration, starting from a random initial population, according to the principle

of survival of the fittest, the next generation population with better performance is produced through the genetic optimization operation of competition, selection, reproduction, mutation and so on.



(a) The logistic mapping (b) The improved logistic mapping

FIGURE 1. Comparison of Lyapunov exponent of chaotic mappings



(a) The logistic mapping (b) The improved logistic mapping

FIGURE 2. Comparison of attractor of chaotic mappings

The genetic algorithm first encodes the parameters, randomly generates a certain number of individuals as chromosomes to form the initial population, and designs fitness function as a criterion for judging the performance of each individual. Individuals with good performance are selected as paternal and maternal individuals with a certain probability, which will participate in the process of crossover and mutation in inheritance, so as to generate a new generation of population [19].

The probability selection operation of genetic algorithm is based on the individual fitness, evaluates the merits and demerits by calculating the fitness value of chromosomes in the population, so as to determine the size of the genetic opportunities and form the intermediate population. When using the genetic algorithm to solve optimization problems, the key lies in whether the selection, crossover, mutation strategy and parameter settings are reasonable. These three operations are the main search operators of the genetic algorithm, which can guarantee the reasonable convergence speed of the algorithm and prevent the possible local convergence under the constraint of the fitness function[20].

The fitness function is generally a mathematical model of the optimization problem and calculates the fitness value of the individual sequence according to certain adaptive conditions. Through the selection operation some candidate individuals are retained and others are discarded, then the excellent individuals will be inherited directly to the next generation. The selection operator directly affects the convergence speed of genetic algorithm and whether the optimal solution can be found. In this paper, the size of the sum of Hamming distances between individual sequences is used as the selection basis for fitness

$$\begin{aligned} d_1 &= d_{12} + d_{13} + d_{14} \\ d_2 &= d_{21} + d_{23} + d_{24} \\ d_3 &= d_{31} + d_{32} + d_{34} \\ d_4 &= d_{41} + d_{42} + d_{43} \end{aligned} \tag{2}$$

In Eq. (2), d_{ij} is the Hamming distance between the individual sequence i and j, and d_i is the sum of the Hamming distances of the corresponding individual sequence and other individual sequences. The greater the sum of the Hamming distances, the lower the correlation of the sequence to other sequences, the greater the probability that it will be selected.

The crossover operation is a process of gene recombination that simulates sexual reproduction in nature, the gene segments of paired chromosomals are interchanged to construct a new individual through a certain crossover probability. The arithmetic crossover refers to the generation of two new individuals through the linear combination of two individuals, the most commonly used of which the binary single-point arithmetic crossover operation is adopted in this paper. A crossover point is first randomly set in the individual sequence, and some genes of the two individuals after the point are interchanged when the crossover is performed

$$\begin{aligned} &\text{individual } A\text{: } 0110\uparrow 000 \rightarrow \text{new individual } A'\text{: } 0110111 \\ &\text{individual } B\text{: } 1100\uparrow 111 \rightarrow \text{new individual } B'\text{: } 1100000 \end{aligned} \tag{3}$$

The mutation operation simulates the genic mutation phenomenon of the organism, which is carried out on individuals of the population through a certain mutation probability. In this paper, the code weight of individual sequence is chosen as the basis of the mutation decision. All individuals in the population are judged with the predetermined mutation probability to determine whether to carry out mutation or not, and to select the mutation position for the individual who mutates, so as to achieve gene improvement. If the length of the binary individual sequence is n, and $[2/5n, 3/5n]$ is preset to the ideal mutation probability interval, when the individual sequence length is $n = 128$, the mutation operation is performed according to the following rule

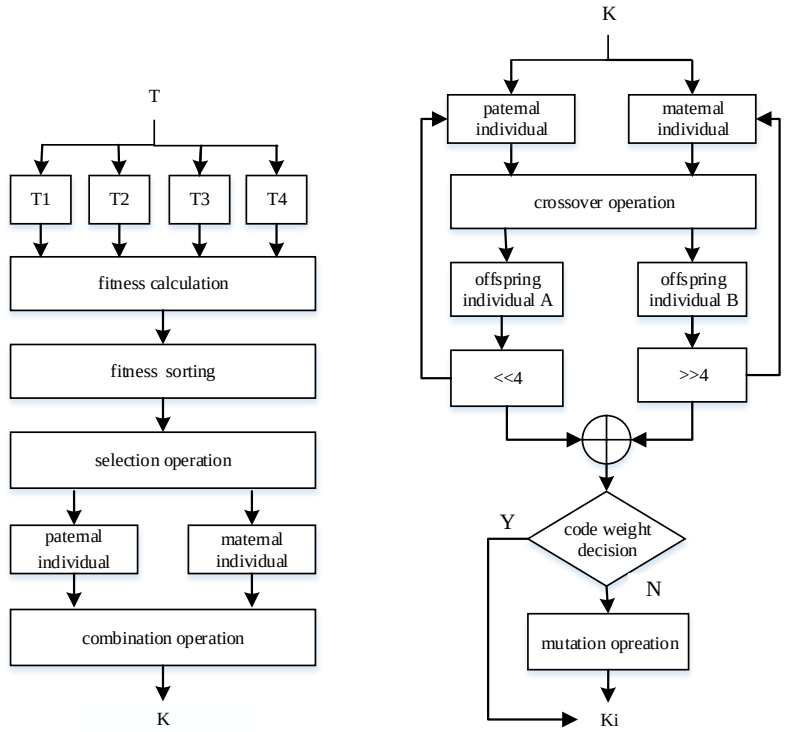
$$D = \begin{cases} C_0(\overline{64-w}) & , \quad w < 51 \\ C & , \quad 51 \le w \le 77 \\ C_1(\overline{w-64}) & , \quad w > 77 \end{cases} \tag{4}$$

In Eq. (4), C is the individual sequence to be improved, D is the improved individual sequence, and w is the code weight of individual sequence. When $51 \le w \le 77$, the individual sequence C does not need to perform mutation operation; when $w < 51$, $C_0(\overline{64-w})$ represents that $(64-w)$ 0s of individual sequence C are reversed from left to right; when $w > 77$, $C_1(\overline{w-64})$ represents that $(w-64)$ 1s of individual sequence C are reversed from left to right.

In the genetic algorithm, the crossover operator is used as the main operator because of its global search ability, and the mutation operator is used as the auxiliary operator because of its local search ability. Through the operation of crossover and mutation which cooperate and compete with each other the genetic algorithm has the equilibrium search ability of balancing global and local.

4. **Key expansion algorithm design.** The main key is generated as shown in Figure 3(a), selecting 512-bit digital chaotic sequence as chromosomes which are divided into four groups averagely. The sum of the Hamming distances of each individual sequence and other individual sequences is used as fitness value and sorted by the size of which. According to the principle that the higher the fitness value and the greater the probability of selection, the two individual sequences with the largest sum of Hamming distances are selected from four individual sequences of chromosomes as the paternal and maternal individuals, and the 256-bit initial main key K is obtained by combining them.

The sub keys are expanded as shown in Figure 3(b), the single-point crossover operation is performed on the 128-bit paternal and maternal individuals to generate two new offspring individuals A and B. In order to reduce the linearity degree between the individual sequences A and B, cyclic shift permutation are performed on them respectively to improve the diffusion effect, and the XOR operation is performed on the shift outputs of which to obtain the new offspring individual C. In order to improve the balance of individual sequence C, the mutation operation is performed through calculation and decision of code weight to enhance the random characteristics. If the obtained code weight is in the ideal mutation probability interval, the offspring individual C is directly output as the subkey; if the obtained code weight deviates from the ideal mutation probability interval, the genes of offspring individual C will be improved and then output as the subkey. According to the above rules, the subkeys K_i required for each round of encryption can be generated one by one.



(a) Master key generation (b) Sub keys expansion

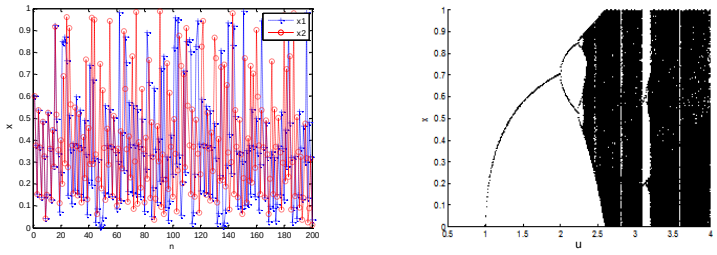
FIGURE 3. Key expansion algorithm

5. **Key expansion algorithm test.** According to the well-known Kerchhoff principle [21], the security of the system does not depend on the confidentiality of the cryptosystem or algorithm, but only on the key. Therefore, testing and analyzing the performance of the key expansion algorithm is necessary.

5.1. **Key space.** The key space, also known as the size of the key, refers to the range of keys to be selected. Theoretically speaking, an exhaustive attack can crack any cryptographic algorithm with enough time and resources. However, in terms of computational complexity, by attacking a cryptographic algorithm exhaustively, the average case needs to try at least half of the keys.

When the key length is r bits, the key space has 2^r elements. The literature [22] has suggested that the key space of the cryptosystem must reach at least 2^{100} to be

considered sufficiently secure. This view has been adopted by the majority of cryptography researchers. As shown in Figure 4, the improved logistic mapping used in this paper is extremely sensitive to the initial value, and the surjective map interval is significantly expanded, combined with the genetic mechanism that survival of the fittest, it can meet the requirements of resisting exhaustive attack.

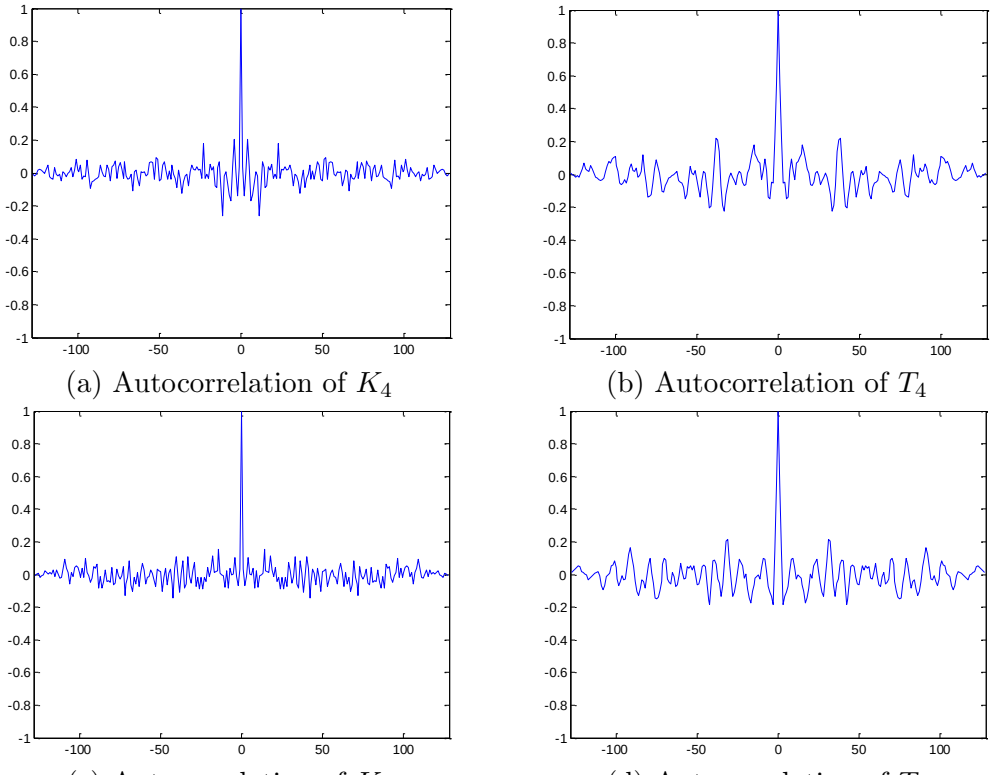


(a) Initial value sensitivity (b) Bifurcation diagram

FIGURE 4. Dynamic characteristics of the improved logistic mapping

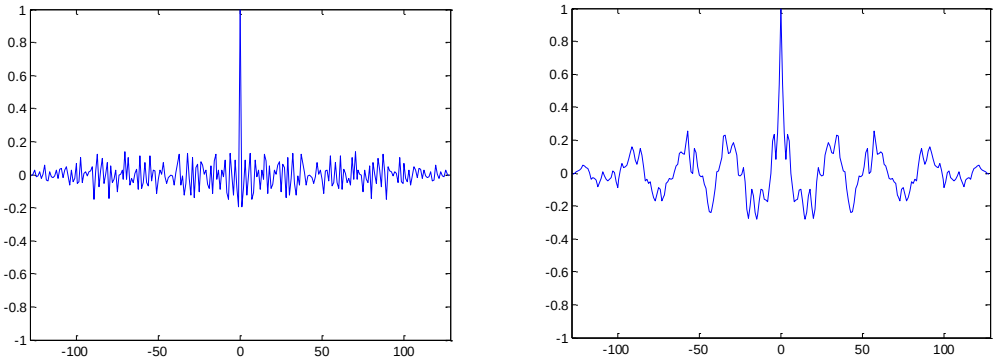
5.2. **Key randomness.** The autocorrelation function is used to describe the degree of correlation between the key values at different times, which is regarded as the important basis for judging the random performance of the key.

Based on the improved logistic mapping, it can be found through the simulations of 4, 8, 12, 16 round subkeys autocorrelation before and after the introduction of genetic algorithm, the combination of the chaotic mapping and the genetic algorithm can significantly enhance the autocorrelation of each round subkey, which effectively alleviates the local periodic phenomenon of digital chaotic sequences.

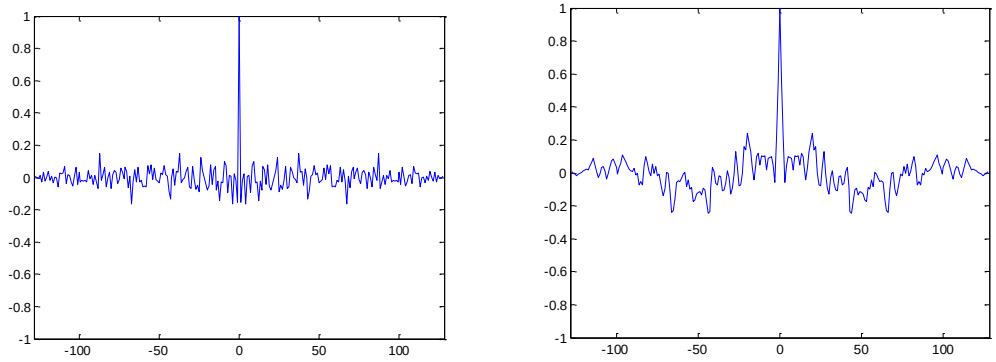


(a) Autocorrelation of K_4 (b) Autocorrelation of T_4

(c) Autocorrelation of K_8 (d) Autocorrelation of T_8



(e) Autocorrelation of K_{12} (f) Autocorrelation of T_{12}

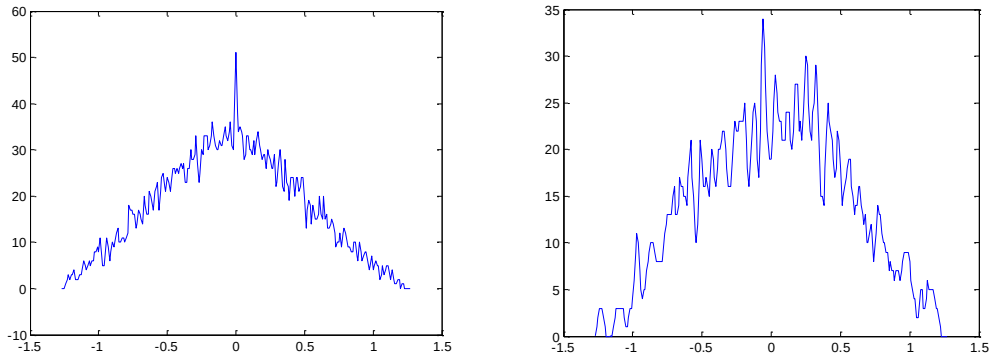


(g) Autocorrelation of K_{16} (h) Autocorrelation of T_{16}

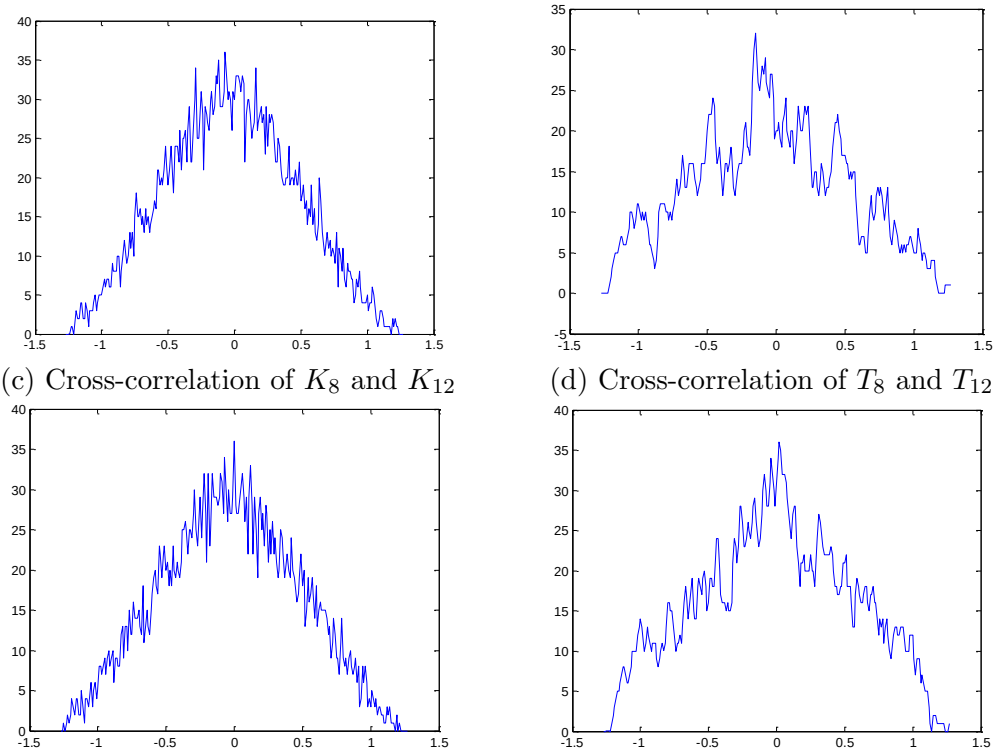
FIGURE 5. Comparison of autocorrelation of key expansion algorithm

5.3. **Key correlation.** In the key expansion algorithm, statistical independence between round subkeys is difficult to achieve, and only the round subkeys can be made as uncorrelated as possible. In mathematical statistics, the cross-correlation function is used to represent the correlation of two random sequences, and it is used here to analyze the degree of correlation between round subkeys.

The cross-correlations between round subkeys before and after the introduction of genetic algorithm are shown in Figure 6. The introduction of genetic algorithm can effectively reduce the cross-correlation between the round subkeys. Meanwhile, because the key expansion has a single direction of operation, even if the first round subkey is intercepted, the initial master key is derived by 2^{128} attempts, which can achieve the intensity of brute force cracking, and the key security is significantly improved.



(a) Cross-correlation of K_4 and K_8 (b) Cross-correlation of T_4 and T_8



(c) Cross-correlation of K_8 and K_{12} (d) Cross-correlation of T_8 and T_{12}

(e) Cross-correlation of K_{12} and K_{16} (f) Cross-correlation of T_{12} and T_{16}

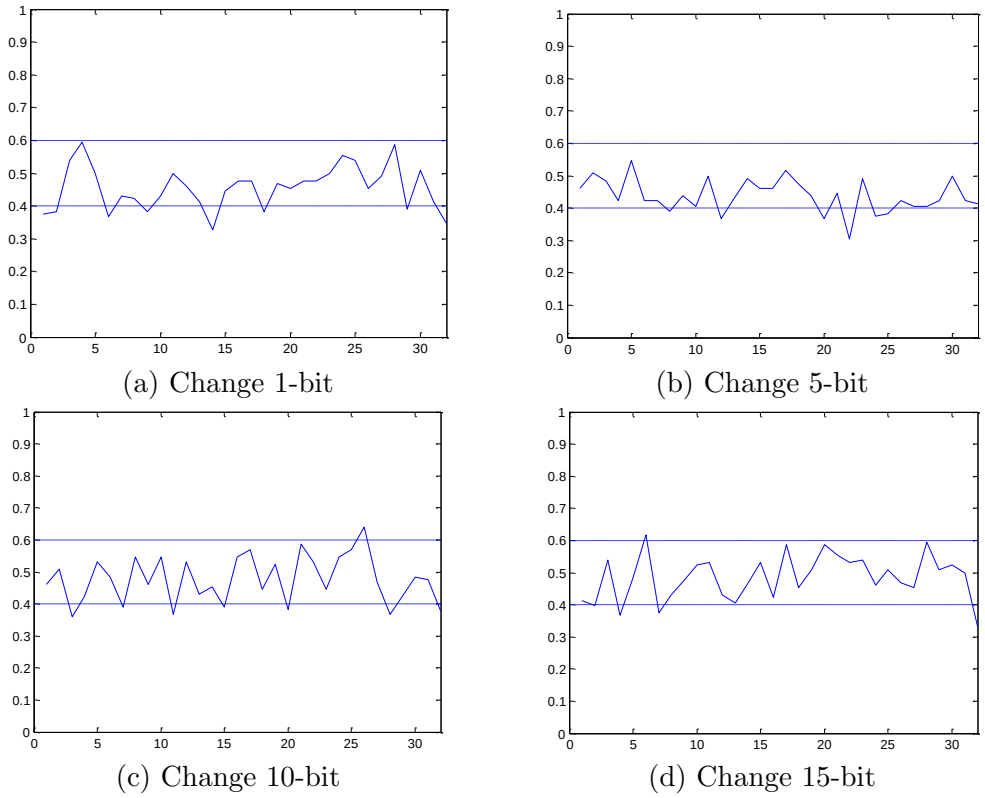
FIGURE 6. Comparison of cross-correlation of key expansion algorithm

5.4. **key sensitivity.** Key sensitivity refers to that as the key iteration progresses, a small change in the initial master key will cause an avalanche effect on each round subkey. The avalanche effect means that only one bit of input change will cause about half of the output bits to change. It is an ideal property of the key and cryptographic algorithm to detect the degree of key diffusion.

Key sensitivity is as shown in Figure 7, when the initial master key changes by 1 bit, the change rate of Hamming distance between the initial master key and the round subkeys change only up and down around 0.5, and it will be closer to 0.5 as the change bits of the initial master key increases, which shows that the key sensitivity can fall into a ideal confidence interval, and even small changes can spread quickly through the iteration of key expansion, making the round subkeys and the encrypted ciphertext completely different.

5.5. **NIST test.** In order to further verify the performance of the key expansion algorithm, the SP800-22 test package developed by the National Institute of Standards and Technology (NIST) is used for the random performance detection of round subkeys. SP800-22 contains a total of 16 standard tests to detect the different properties of random sequences. Each index is obtained through a certain test algorithm, and the p-value of which is reflected as the test result. the greater the p-value, the better the randomness of the test sequences.

As shown in Table 1, NIST tests are carried out on round subkeys of SM4 key expansion algorithm in literature[23], SM4 key expansion algorithm based on chaotic mapping in literature[24] and key expansion algorithm based on chaotic mapping and genetic algorithm in this paper. By comparison, we can see that the key expansion algorithm in this paper makes the randomness of round subkeys significantly enhanced.



(a) Change 1-bit (b) Change 5-bit

(c) Change 10-bit (d) Change 15-bit

FIGURE 7. Sensitivity of key expansion algorithm

TABLE 1. NIST test of key expansion algorithm

test item	P-value in Literature[23]	P-value in Literature[24]	P-value in thist paper
Approimate Entropy	0.994038	0.999242	1.000000
Block Frequency	0.550177	0.283013	0.859684
Cumulative Sums	0.629223	0.546062	0.623533
	0.301120	0.519702	0.549204
FFT	0.121488	0.121488	0.123812
Frequency	0.317311	0.381574	0.576150
Linear Complexity	0.919689	0.985608	0.919689
Longest Run	0.222186	0.228193	0.296950
Rank	0.693720	0.693720	0.693720
Runs	0.826678	0.881524	0.891044
	0.006450	0.400350	0.498961
Serial	0.056433	0.855688	0.498531

6. **Conclusions.** It is very important to design a secure and efficient key expansion algorithm for cryptographic security. In this paper, a chaotic key expansion algorithm based on genetic algorithm according to the genetic mechanism that survival of the fittest is proposed, which takes the digital chaotic sequence as the initial population, calculates and orders the fitness function, and performs the operations of the selection, crossover, and mutation, so that the generation and expansion of the key are realized through the generation-by-generation evolution. The simulation analysis and performance tests show that the algorithm does not require complicated operation to solve the dynamic

characteristics degradation and local periodicity phenomena of digital chaotic sequences, which can not only meet the application requirements of lightweight block ciphers, but also have certain reference significance for exploring new key expansion algorithms.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (no.61471158), the Modern Sensing Technology Innovation Team Project of Heilongjiang Province (no.2012TD007), Science and Technology Innovation Foundation of Harbin (no.2017RAQXJ082) and the Basic Scientific Research-related Subsidy Project of Heilongjiang Provincial Colleges and Universities.

REFERENCES

[1] E. Razali, and R. C. W. Phan, On the existence of related-key oracles in cryptosystems based on block ciphers, *Lecture Notes in Computer Science*, vol. 4277, pp. 425-438, 2006.

[2] T. Isobe, A single-key attack on the full gost block cipher, *Lecture Notes in Computer Science*, vol. 6733, no. 01, pp. 290-305, 2011.

[3] Y. Dai, and S. Chen, Cryptanalysis of full pride block cipher, *Science China(Information Sciences)*, vol. 60, no. 05, pp. 169-180, 2017.

[4] A. Biryukov, and D. Wagner, Slide attack, *Lecture Notes in Computer Science*, vol. 1636, pp. 245-259, 1999.

[5] C. M. Chen, L. L. Xu, T. Y. Wu, and C. R. Li, On the security of a chaotic maps-based three-party authenticated key agreement protocol, *Journal of Network Intelligence*, vol. 1, no. 02, pp. 61-66, 2016.

[6] J. L. Huang, Studies on Key Schedules and Single-key Attacks of Block Ciphers, Ph.D. Thesis, *Shanghai Jiao Tong University*, ShangHai, China, 2014.

[7] C. M. Chen, K. H. Wang, T. Y. Wu, and E. K. Wang, On the security of a three-party authenticated key agreement protocol based on chaotic maps, *Data Science and Pattern Recognition*, vol. 1, no. 2, pp. 1-10, 2017.

[8] Y. B. Zheng, Y. Song, B. X. Du, and J. Pan, A novel detection of periodic phenomena of binary chaotic sequences, *Journal of Physics*, vol. 64, no. 23, pp. 230501-230501, 2012.

[9] C. L. Fan, and Q. Ding, Arm-embedded implementation of h.264 selective encryption based on chaotic stream cipher, *Journal of Network Intelligence*, vol. 3, no. 01, pp. 9-15, 2018.

[10] X. F. Liao, D. Xiao, and Y. Chen (eds.), Theory and Applications of Chaotic Cryptography, *Beijing Science and Technology Press*, Beijing, China, 2009.

[11] R. M. Yin, J. Yuan, X. M. Shan, and X. Q. Wang, Weak key analysis for chaotic cipher based on randomness properties, Science China(Information Sciences), vol. 55, no. 05, pp. 1162-1171, 2012.

[12] X. Y. Yang, G. Wang, and X. T. Gu, Analysis of logistic chaotic sequences and application simulation, *Designing Techniques of Posts and Telecommunications*, vol. 46, no. 12, pp. 19-22, 2003.

[13] D. L. Zheng, G. Zhao, and G. B. Xu, Logistic mapping digital-flow chaotic strange attractor and its parameter analysis, *Journal of University of Science and Technology Beijing*, vol. 24, no. 03, pp. 350-352, 2002.

[14] Q. Lu, X. H. Lin, J. Li, and X. Y. Bao, Improved method for chaotic frequency hopping sequence, *Journal of Data Acquisition and Processing*, vol. 25, no. 01, pp. 122-125, 2010.

[15] W. Zhang, H. M. Xie, and B. P. Wang. Novel piecewise logistic chaotic spread spectrum communication algorithm, *Computer Science*, vol. 40, no. 01, pp. 59-62, 2013.

[16] J. Wang, and Q. Ding, Excellent performances of the third-level disturbed chaos in the cryptography algorithm and the spread spectrum communication, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 07, no. 16, pp. 826-835, 2016.

[17] G. N. Xuan, D. W. Wang, and R. W. Cheng (eds.), Genetic Algorithm and Engineering Design, *Beijing Science and Technology Press*, 2000.

[18] G. L. Chen, X. F. Wang, D. S. Wang, and Z. Q. Zhuang, Genetic algorithm and its application, *Post and Telecom Press*, vol. 11, no. 07, pp. 59-62, 1999.

[19] X. P. Wang, and L. M. Cao (eds.), Genetic Algorithm:Theory, Application And Software Implementation, *Xi'an Jiao Tong University Press*, 1998.

[20] H. Zhou, and T. L. Huang, Image encryption technology based on genetic algorithm and its implementation, *Journal of Guilin University of Electronic Technology*, vol. 35, no. 03, pp. 228-231, 2015.

[21] D. R. Stinson, Cryptography:theory and practice, *Crc Press*, vol. 20, no. 01, pp. 65-65, 1995.
[22] G. Alvarez, and J. S. Li, Some basic cryptographic requirements for chaos-based cryptosystem, *International Journal of Bifurcation and Chaos*, vol. 16, no. 08, pp. 2129-2151, 2006.
[23] C. X. Shen, H. G. Zhang, D. G. Feng, and Z. F. Gao, A review of information security, *Chinese Science*, vol. 37, no. 02, pp. 129-150, 2007.
[24] C. F. Wang, and Q. Ding, Sm4 key scheme algorithm based on chaotic system, *Journal of Physics*, vol. 66, no. 02, pp. 76-84, 2017.