A Novel S-box Optimization Method Based on Immune Genetic Algorithm

Ding Zhu
School of Computer Science and Technology
Harbin Institute of Technology
Weihai, Shandong province, China
zhudingtop@163.com

Miao Zhang
School of Computer Science and Technology
Harbin Institute of Technology
Weihai, Shandong province, China
zhangmiaozm209@126.com

Xiaojun Tong*
School of Computer Science and Technology
Harbin Institute of Technology
Weihai, Shandong province, China
tong_xiaojun@163.com

Zhu Wang
School of Information
Harbin Institute of Technology
Weihai, Shandong province, China
wangzhu@hit.edu.cn

***Abstract*— S-box is the only nonlinear component in block cipher, and its performance is the key factor to determine the data security of cipher algorithm. Our paper proposes a novel S-box optimization method based on immune genetic algorithm. Firstly, the S-boxes population is generated by chaotic system, then the S-boxes with excellent performance are optimized by a series of operators, including extracting the anti-agent and immune selection. The nonlinear degree criterion, differential uniformity criterion and strict avalanche effect criterion of S-boxes are analyzed. The experimental results show that the optimized S-box has strong characteristics. Compared with the traditional genetic algorithm, this method has faster convergence speed and better resistance to linear attacks and differential attacks. The optimized S-box has a good application in the field of information security prospects.**

Keywords: S-box; chaotic system; data security; immune genetic algorithm

I. INTRODUCTION

Many methods have been proposed to improve the performance of the S-boxes, including the optimization methods based on chaotic mapping and the optimization methods based on intelligent algorithm. The pseudo randomness of chaotic systems and the scalability of intelligent algorithms are conducive to the generation and optimization of S-boxes. For example, Ö zkaynak[1] proposed the use of time-delay chaotic systems to generate chaotic S-boxes; Jakimoski[2] and others proposed an S-box generation method based on exponential and Logistic chaotic maps; Millon[3] and others combined the traditional genetic algorithm with the construction of S-boxes in information security, and the obtained S-boxes have good performance. Szaban[4] and others proposed that cellular automata can be used to simulate the process of constructing S-boxes, and the interaction between S-boxes can be used to optimize S-boxes. The optimized S-boxes have higher security. However, the general intelligent algorithm has some shortcomings, such as in the late stage of S-box optimization, which can't guarantee the convergence of the results. In addition, some methods combine the above two principles. For example, Liu[5] and others proposed an improved immune genetic algorithm based on hybrid chaotic map. Although the convergence effect is good, the optimization effect is general. Therefore, our paper proposes a novel S-box optimization method based on immune genetic algorithm, which solves the shortcomings of traditional genetic algorithm, such as general optimization effect and slow convergence in the later stage of optimization. The optimized S-box can better resist differential attacks and linear attacks.

II. THE PRINCIPLE OF IMMUNE GENETIC ALGORITHM

Immune genetic algorithm[6] was proposed based on biological immune system. The problem to be solved can be compared with antigen in biological immune system, and the solution of problem can be compared with antibody in biological immune system. Forrester applied immune genetic algorithm to the field of information security. The specific steps of immune genetic algorithm including population initialization, vaccine extraction, vaccine replication, calculate the fitness and output optimized individuals.

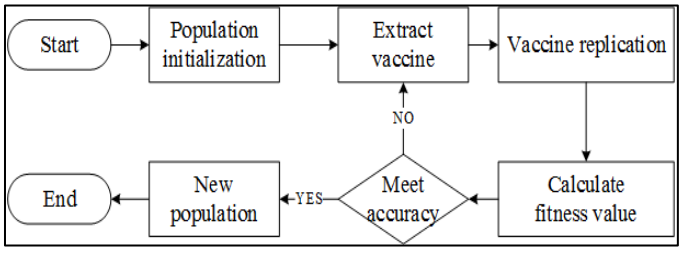

Figure 1. Immune genetic algorithm flow chart.

*Corresponding author
Xiaojun Tong

978-1-7281-6579-0/20/$31.00©2020 IEEE

III. S-BOX OPTIMIZATION METHOD

A. Optimization of S-box

The S-box population in this paper is generated by iterations of Logistic chaotic system [7], the Logistic chaotic system is denoted by equation 1.

$$x_{n+1} = \lambda x_n (1 - x_n) \tag{1}$$

Where $\lambda \in (0,4)$, $x_n \in [0,1]$, as illustrated in figure 2, the specific steps of the optimization process as follows:

Step 1. Set λ, x_0 as initial values;

Step 2. **S-box construction.** The Logistic chaotic system is iterated 50 times to get the value sequences. All sequence values are mapped in [0,255] By modular operation and rounding operation then S-box population is output;

Step 3. **Encoding**. Each S-box is encoded in binary form;

Step 4. Build the antibody library. 8 S-boxes in AES algorithm are used as antibody library. Then calculate the fitness value of each S-box according to equation 1. According to the fitness value, the S-boxes are sorted in ascending order, that is {S0, S1, S2, S3, S4, S5, S6, S7};

Step 5. According the fitness value, The S-boxes generated in Step 2 are arranged in ascending order according, that is $\{S_k^0\}_{k=0}^{k=n-1}$;

Step 6. Select the S-boxes as excellent individual which meet the performance in $\{S_k^0\}_{k=0}^{k=n-1}$, otherwise go to Step 7;

Step 7. **Selection**. Select the S-box population with a certain size according to the preset selection probability;

Step 8. **Crossover**. According to the preset crossover probability, the S-box population selected in the previous step is converted into a new S-box intermediate population by partial code cross exchange among population;

Step 9. **Mutation**. According to the preset mutation probability, the S-box population generated in the previous step is mutated by the individual coding to generate a new S-box intermediate population;

Step 10. **Build antibody library**. Selected 4 S-boxes randomly individuals in the antibody library as the vaccine in the algorithm process. The antibody library generation reference the design of antibody library;

Step 11. **Vaccination**. The mutation of the S-boxes individuals and the S-boxes in the vaccine binary code bit by bit, then replace the S-boxes individuals with the corresponding position of the four vaccine S-boxes of the same binary code bit. Lastly, generate a new S-box intermediate population;

Step 12. **Immune selection**. Calculate the fitness value of the S-box population generated in the previous step, and compare them with the fitness value of the corresponding S-box individuals in step 4. If the fitness value is larger, the individual will be retained, and the S-box with smaller fitness value in the antibody library will be replaced, then a new antibody library will be generated until all S-boxes are compared. Otherwise, the individual will be deleted and the parent individual will be retained;

Step 13. **Termination conditions**. Go to the Step 3 until the algorithm stops. Lastly, output the optimized S-box.

B. The design of fitness function

The S-box with good performance has certain evaluation criteria, such as nonlinearity criteria, Equiprobable input/output XOR distribution criteria and strict avalanche effect criteria. Combining the above three evaluation criteria, this paper proposes a weighted piecewise fitness function based on immune genetic algorithm as illustrated in equation 2, where N_s is the nonlinearity of S-box, $\{A_1, A_2, A_3, A_4\}$ is the corresponding weight; δ_s is the differential uniformity of S-box, $\{B_1, B_2, B_3, B_4\}$ is the corresponding weight; B_s is the strict avalanche effect, $\{C_1, C_2, C_3, C_4\}$ is the corresponding weight, $F(x)$ is denoted by equation 2.

$$F(s) = \begin{cases} A_1 N_s - B_1 \delta_s - C_1 B_s & N_s < 107 \, and \, \delta_s > 10 \\ A_2 N_s - B_2 \delta_s - C_2 B_s & N_s < 107 \, and \, \delta_s \le 10 \\ A_3 N_s - B_3 \delta_s - C_3 B_s & N_s \ge 107 \, and \, \delta_s > 10 \\ A_4 N_s - B_4 \delta_s - C_4 B_s & N_s \ge 107 \, and \, \delta_s \le 10 \end{cases} \tag{2}$$

The weight of each segment is different according to the performance of S-box in different stages. This paper proposes the scheme as following:

$$A_n = \begin{cases} 1 & n = 1,2 \\ \frac{2}{3} n + A_{n-1} & n = 3,4 \end{cases} \tag{3}$$

$$B_n = \begin{cases} \frac{1}{4} n & n = 1,3 \\ \frac{1}{2} n - \frac{1}{3} B_{n-1} & n = 2,4 \end{cases} \tag{4}$$

$$C_n = \begin{cases} \frac{3}{5} n & n = 1,2,3 \\ 1 & n = 4 \end{cases} \tag{5}$$

C. The design of antibody library

As for S-boxes generated by chaotic system, the function of antibody library is to take immune operate with S-boxes. There is a problem worthy of discussion about how to extract excellent initial antibody in the practical application of immune genetic algorithm. This paper proposes following scheme. Firstly, build an initial antibody library in which stores some strong S-boxes. The antibody library is composed of 8 S-boxes generated by classic AES algorithm, and the S-boxes are arranged according to their fitness values, this is C = {S0, S1, S2, S3, S4, S5, S6, S7}. The performance of S-boxes in the antibody library is excellent, and them have the ability to resist linear attacks and differential attacks. The reason of building antibody library is to avoid the consequences of local convergence after optimization because there are few excellent individuals.

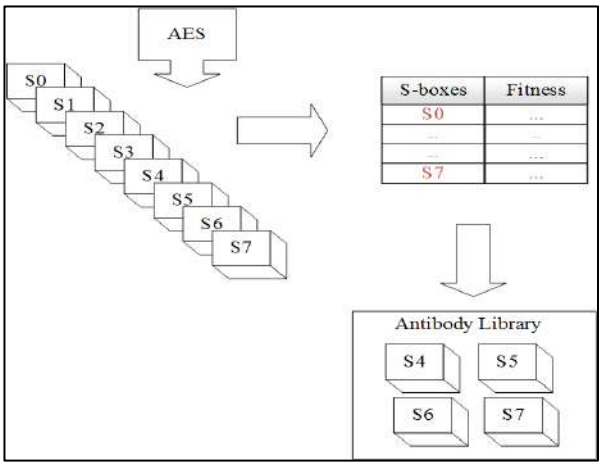

Figure 2. Build antibody library

D. Vaccination and Immune selection

The initial S-box population generated by Logistic chaotic system is performed operations by traditional genetic algorithm, including selection operation, crossover operation and mutation operation. The output of the mutation was recorded as {SM}.Then, 4 S-boxes with the highest fitness are selected in the antibody library, and the binary coding bits corresponding to {SM} are replaced by four binary coding bits with the same antibody to form a new individual, which is recorded as {Sn}. After each iteration, individuals with smaller fitness need to be eliminated in the antibody library and replaced by individuals with larger fitness. Update the antibody library continuously in the whole process of optimization algorithm.

$$M = S_4 \cap S_5 \cap S_6 \cap S_7 \tag{6}$$

$$S_N = S_M \cup M \tag{7}$$

$$C = \begin{cases} \{S_0, S_1, S_2, S_3, S_4, S_5, S_6, S_7\} & if\ F(S_N) \le F(S_M) \\ \{S_0, S_1, S_2, S_3, S_N, S_5, S_6, S_7\} & if\ F(S_N) > F(S_M) \end{cases} \tag{8}$$

Calculate the fitness value of new S-box individual. It can be considered that the individual after vaccination has produced beneficial effect if the fitness value becomes large; Otherwise, it is considered that the individual after vaccination has degenerated, then the new individual will be deleted and the original individual will be retained.

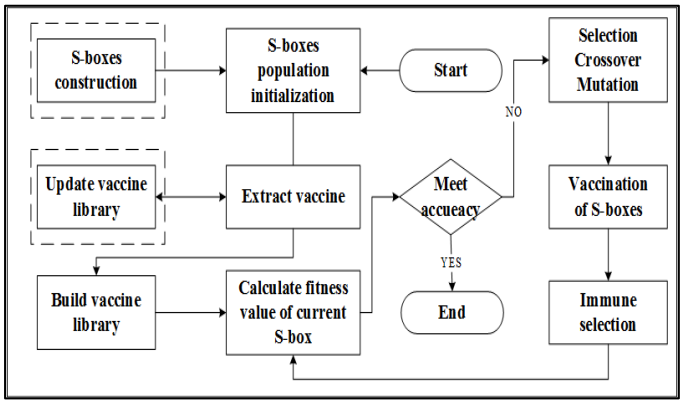

Figure 3. Proposed method folw chart.

IV. Performance Analysis

In order to verify the cryptography characteristics of S-boxes, according to the performance standard evaluation system of S-boxes, this paper analyzes the performances of the proposed S-boxes included nonlinear criterion, the strict avalanche criterion, differential uniformity, the output bits independence criterion, linear approximation probability.

A. Nonlinear criterion

Nonlinear criterion is an important characteristic of S-Boxes performances evaluation system. The higher the nonlinearity, the stronger the ability of S-box to resist nonlinear attacks. By definition, $f: F_2^n \to F_2$ can be considered as N-Variable Boolean function. The Nonlinear criterion is denoted by

$$N_f = \min_{l \in L_n} d_H(f, l) \tag{9}$$

Where L_n is an affine function set, $d_H(f, l)$ is the Hamming distance between f and l.

B. Equiprobable input/output XOR distribution

Differential probability is a characteristic of S-box, it is a measure of criterion of equiprobable input/output XOR distribution. Differential uniformity also can be expressed by the different approximation probability like as equation:

Different cryptanalysis is introduced by Biham [8]. Differential probability is used to measure the capacity of an encryption function to resist differential analysis. Output variations can be obtained from the input variations. For each XOR input, an equiprobable output can be derived. Set $S(x) = [f_1(x), \ldots, f_m(x)]: F_2^n \to F_2^m$.

$$\eta = \frac{1}{2^n} \max_{\alpha \in F_2^n} \max_{\beta \in F_2^m} \left| \left\{ x \in F_2^n : S(x+\alpha) - S(x) = \beta \right\} \right| \tag{10}$$

is the equiprobable input/output XOR distribution of S(x). In practice, we generally exploit the differential approximation probability to represent the equiprobable input/output XOR distribution. Its expression is:

$$DP_f = \max_{\Delta x \neq 0} \{ x \in X : f(x + \Delta x) \oplus f(x) = \Delta y \} \tag{11}$$

in which X is the input sets, DP_f represents the maximum probability that the output of each given different Δx is equal to Δy.

C. The strict avalanche criterion(SAC)

The strict avalanche effect, which means that each output bit changes with each input bit, and the probability of change is strictly 0.5. The following equation can be used to measure the estimation bias offset of SAC correlation matrix:

$$P_{i,j}(f) = 2^{-n} \sum_{x \in B^n} f_j(x) \oplus f_j(x \oplus e_i) \tag{12}$$

$$S(f) = \frac{1}{n^2} \sum_{1 \le i \le n} sum_{1 \le j \le n} \left| \frac{1}{2} - P_{i,j}(f) \right| \tag{13}$$

D. Performance Comparation

Comparing the performance of the optimized S-box with other papers, as illustrated in Table I, it is calculated that the mean value of SAC is 0.5005 which is so close to 0.5, and the offset is 0.005. It proves that the proposed S-box has better SAC property near the ideal values. The proposed S-box has an average of the nonlinearity 112.00 which is superior to any other algorithm. The differential approximation probability of proposed S-box is 0.03600. The above data show that the optimized S-box has good performance in nonlinearity, differential uniformity and strict avalanche effect, which represents the S-box can resist nonlinear attacks and differential attacks.

E. Convergence analysis

The following Figure 1 shows that a comparison of the proposed method with the traditional genetic algorithm about convergence rate in equal conditions. The fitness value of the proposed S-box is the best when the number of generations close to 1005 times. However, The S-box with traditional genetic algorithm needs 4300 times. As illustrated in figure 1, the convergence rate with the proposed method is faster than that with the traditional genetic algorithm.

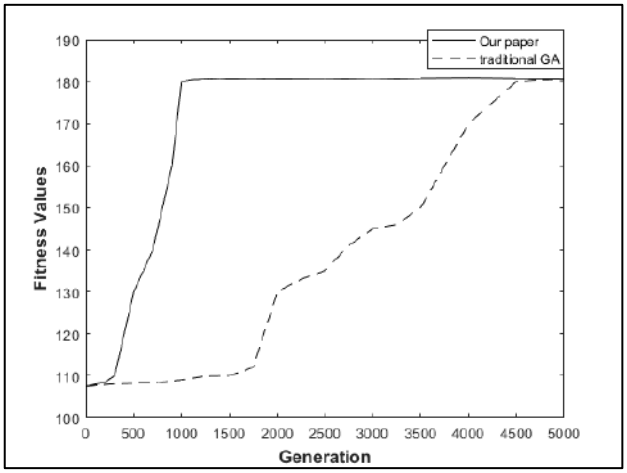

Figure 4. Convergence comparsion.

Table I. Performance Comparation with others S-boxes

S-boxes	*Performance Comparation*		
	Average Nonlinearity	***DP***	***SAC***
Proposed	112.00	0.03600	0.5005
Çavusoglu[9]	106.25	0.03910	0.5039
Lambic[10]	106.75	0.03910	0.5010
AES[11]	107.25	0.01560	0.5049

V. COCLUSION

In this paper, S-boxes population are generated by Logistic chaotic system, and the optimized S-box with excellent performance is output by the antibodies extraction and immune selection. The proposed method overcomes the shortcomings of slow convergence of traditional genetic algorithm. Security analysis experiments show that the optimized S-box has better cryptographic characteristics and can resist linear attacks and differential attacks. Optimized S-box has a good application prospect in information security.

ACKNOWLEDGMENT

This work was supported by the following projects and foundations: project ZR2019MF054 supported by Shandong Provincial Natural Science Foundation, the National Natural Science Foundation of China (No.61902091) and Innovation Research Foundation of Harbin Institute of Technology (HIT.NSRIF.2020099)， the Foundation of Science and Technology on Information Assurance Laboratory (No.KJ-17-004), Equip Pre-research Projects of 2018 supported by Foundation of China Academy of Space Technology (No. WT-TXYY/ WLZDFHJY003)， 2017 Weihai University Co-construction Project.

REFERENCES

[1] Özkaynak, F.; Yavuz, S. Designing chaotic S-boxes based on time-delay chaotic system. Nonlinear Dyn, vol 74, pp.551–557, 2013.

[2] Jakimoski,G., Kocarev, L.: Chaos and cryptography: block encryption ciphers based on chaotic maps. IEEE Trans. Circuits Syst, vol.1, pp.163–169, 2001.

[3] Ilvanov, G., Nikolov, N. & Nikova, S. Reversed genetic algorithms for generation of bijectives-boxes with good cryptographic properties. Cryptogr Commun, vol.8, pp.247-276, 2016.

[4] Wang Y, Wong K W, Li C, Li Y. A novel method to design S-box based on chaotic map and genetic algorithm. Physics Letters A, vol.376, 2012, pp.827-833.

[5] Liu, Aijun,Yang, Yu,Liang, Xuedong,Yao, Hao. Improved Immune Genetic Algorithm Based on Hybrid Chaotic Maps and Its Application. Journal of Convergence Information Technology, vol.7, 2012..

[6] Fatima Benbouzid-SiTayeb,Malika Bessedik,Mohamed Reda Keddar,Abd Errahmane Kiouche. An effective multi-objective hybrid immune algorithm for the frequency assignment problem. Applied Soft Computing Journal, pp.85, 2019.

[7] Junchao Wang, Kaining Han, Shengwen Fan, Ying Zhang, Honghao Tan, Gwanggil Jeon. A logistic mapping-based encryption scheme for Wireless Body Area Networks. Future Generation Computer System, vol.110, pp.57-67, 2020.

[8] Biham E, Shamir A. Differential cryptanalysis of des-like cryptosystems. J Cryptol, vol.4, pp.3–7, 1991.

[9] U. Çavusoglu, A. Zengin, I. Pehlivan, and S. Kaçar, A novel approach for strong S-box generation algorithm design based on chaotic scaled Zhongtang system, Nonlinear Dynamics, vol.87, pp.1081–1094,2017.

[10] D. Lambic, Dragan. A new discrete-space chaotic map based on the multiplication of integer numbers and its application in S-box design, Nonlinear Dynamics, pp.1–13, 2020..

[11] J. Daemen, V. Rijmen, Design of Rijndael: AES-.e Advanced Encryption Standard, Springer Science and Business Media, Berlin, Germany, 2002.