International Journal of Engineering and Technology Innovation, vol. 14, no. 1, 2024, pp. 44-57

Optimization of SM4 Encryption Algorithm for Power Metering Data Transmission

Yi-Ming Zhang[1], Jia Xu[2,*], Yi-Tao Zhao[1], Qing-Chan Liu[1], Qiu-Hao Gong[2]

[1]Measurement Center of Yunnan Power Grid Corporation, Kunming, China

[2]Faculty of Civil Aviation and Aeronautics, Kunming University of Technology, Kunming, China

Received 03 August 2023; received in revised form 14 September 2023; accepted 15 September 2023

DOI: https://doi.org/10.46604/ijeti.2023.12675

Abstract

This study focuses on enhancing the security of the SM4 encryption algorithm for power metering data transmission by employing hybrid algorithms to optimize its substitution box (S-box). A multi-objective fitness function is constructed to evaluate the S-box structure, aiming to identify design solutions that satisfy differential probability, linear probability, and non-linearity balance. To achieve global optimization and local search for the S-box, a hybrid algorithm model that combines genetic algorithm and simulated annealing is introduced. This approach yields significant improvements in optimization effects and increased non-linearity. Experimental results demonstrate that the optimized S-box significantly reduces differential probability and linear probability while increasing non-linearity to 112. Furthermore, a comparison of the ciphertext entropy demonstrates enhanced encryption security with the optimized S-box. This research provides an effective method for improving the performance of the SM4 encryption algorithm.

Keywords: SM4 cryptographic algorithm, hybrid algorithm optimization, S-box, nonlinearity

1. Introduction

In today's rapidly evolving internet environment, ensuring the security of data transmission for electric power system energy measurement has become particularly important. Wireless sensor networks play a crucial role in the power system by collecting and transmitting environmental data. However, ensuring the stable operation of the power system, prolonging the lifespan of wireless sensor networks, and protecting data transmission security are critical. To address these challenges, Tiwari et al. [1] have proposed the CF-AREOR protocol. An improved version of an energy-efficient opportunistic routing scheme based on adaptive ordering is designed to prolong the lifespan of wireless sensor networks.

In addition, Sharmila et al. [2] have proposed secure key management and authentication protocols that use hybrid approaches to establish critical connections, providing higher security. Through the collaboration of uncrewed aerial vehicles and Internet of Things devices [3], power system and energy measurement data can be transmitted to data centers and other locations, enabling remote real-time monitoring, fault diagnosis, and predictive analysis functionalities. However, potential attack risks such as energy supply interruption, data tampering, or unauthorized information disclosure must be considered during the data transmission process. These security issues can significantly impact the effective operation of the power system. Thus, ensuring the security of data transmission for electric power system energy measurement has become an urgent problem in the academic community. In recent years, extensive research has been conducted to ensure the security and reliability of data transmission through various methods and techniques. For example, secure quantization techniques and sector-based approaches can eliminate errors and improve safety [4].

* Corresponding author. E-mail address: y17787199084@163.com

English language proofreader: Yen-Chun Hsieh

Cryptography [5] has evolved as a widely adopted method to combat viruses and unauthorized attacks, ensuring secure transmission of network data and promoting safe access to remote databases. The SM4 encryption algorithm has gained recognition for its strong security, efficiency, and ease of hardware implementation [6], making it widely used in data encryption transmission processes. The substitution box (S-box) is a critical structure in the nonlinear part of the SM4 cipher. However, although existing S-box designs contain nonlinear components, potential encryption attacks may compromise their nonlinearity, making them susceptible to advanced attacks in the future. To address this problem, scholars have proposed different methods to optimize S-box designs. For example, they enhanced nonlinearity through the tangent-delay elliptical reflection cavity mapping system (TD-ERCS) [7] and utilized knowledge related to the medical field to generate truly random numbers [8-9] to improve the nonlinearity of the initial structure of the S-box.

Additionally, innovative, dynamic S-box generation schemes have been proposed to enhance the encryption strength of cryptographic algorithms [10-12]. However, most existing research has focused on local S-box optimization, posing challenges in finding the global optimal solution for complex problems. Moreover, some methods introduce complex mathematical structures or mapping systems, increasing algorithm complexity and decreasing convergence speed, efficiency, and performance. Therefore, more robust optimization methods are needed to achieve breakthroughs and improve S-box designs.

Increasingly, scholars have discovered that metaheuristic algorithms [13-14] can be applied to S-box optimization due to their simplicity, speed, and reasonable convergence rates [15]. Therefore, researchers have combined chaos systems with metaheuristic algorithms to optimize S-box performance [16-20]. More advanced methods mainly consider improving chaos systems [21-22] to enhance the randomness of S-box chaos mapping and construct highly nonlinear S-box designs. Although these studies have made progress in optimizing S-boxes, the iteration process of chaos systems can affect convergence speed, efficiency, and stability and pose challenges in handling local optima, affecting the performance and security of encryption algorithms.

To address these issues, a multi-objective fitting function is designed in this study to comprehensively evaluate the performance of the S-box and guide the optimization algorithm to select the best S-box structure. In addition, an innovative hybrid algorithm combining genetic algorithm and simulated annealing is proposed to obtain the best S-box model. Both algorithms reduce the number of parameters, making them more practical, and can be further fine-tuned to suit real-world applications. By considering both global and local techniques, this hybrid approach improves the performance and security of the S-box.

The hybrid algorithmic model proposed in this study has a wide range of practical application potential, especially in cryptography and cryptographic algorithm optimization. It can be used to improve cryptographic algorithms, enhance various cryptographic algorithms that can create more robust and secure cryptographic systems, and improve security in IoT devices to prevent unauthorized access and data leakage. It also provides an essential tool for cryptography research. This model is expected to address evolving security challenges and protect data security and sensitive information. The contributions of this paper are as follows:

(1) A novel multi-objective fitting function is proposed for comprehensively evaluating and optimizing the S-box's performance to improve data transmission security.

(2) A hybrid algorithm model combining genetic algorithm and simulated annealing is established to achieve global optimization and avoid local optimum to optimize the S-box structure and enhance its nonlinearity.

(3) The average differential probabilities of the optimized S-box are reduced to 0.0359 and 0.0371, the moderate linear probabilities are reduced to 0.094 and 0.096, and the intermediate nonlinearity is significantly improved to 112. Compared with other algorithms, encryption using the optimized S-box can improve the information entropy of the ciphertext.

2. Introduction of SM4 Password

In the current academic context, power metering data transmission faces problems such as data leakage and network threats, and the security of passwords in digital systems is vital. Therefore, the SM4 cryptographic algorithm proposal is of great academic significance.SM4, as a cryptographic algorithm, will be the core of the academic research in this paper. The following sections will detail the SM4 cryptographic algorithm's technical details.

2.1. SM4 cryptographic algorithm

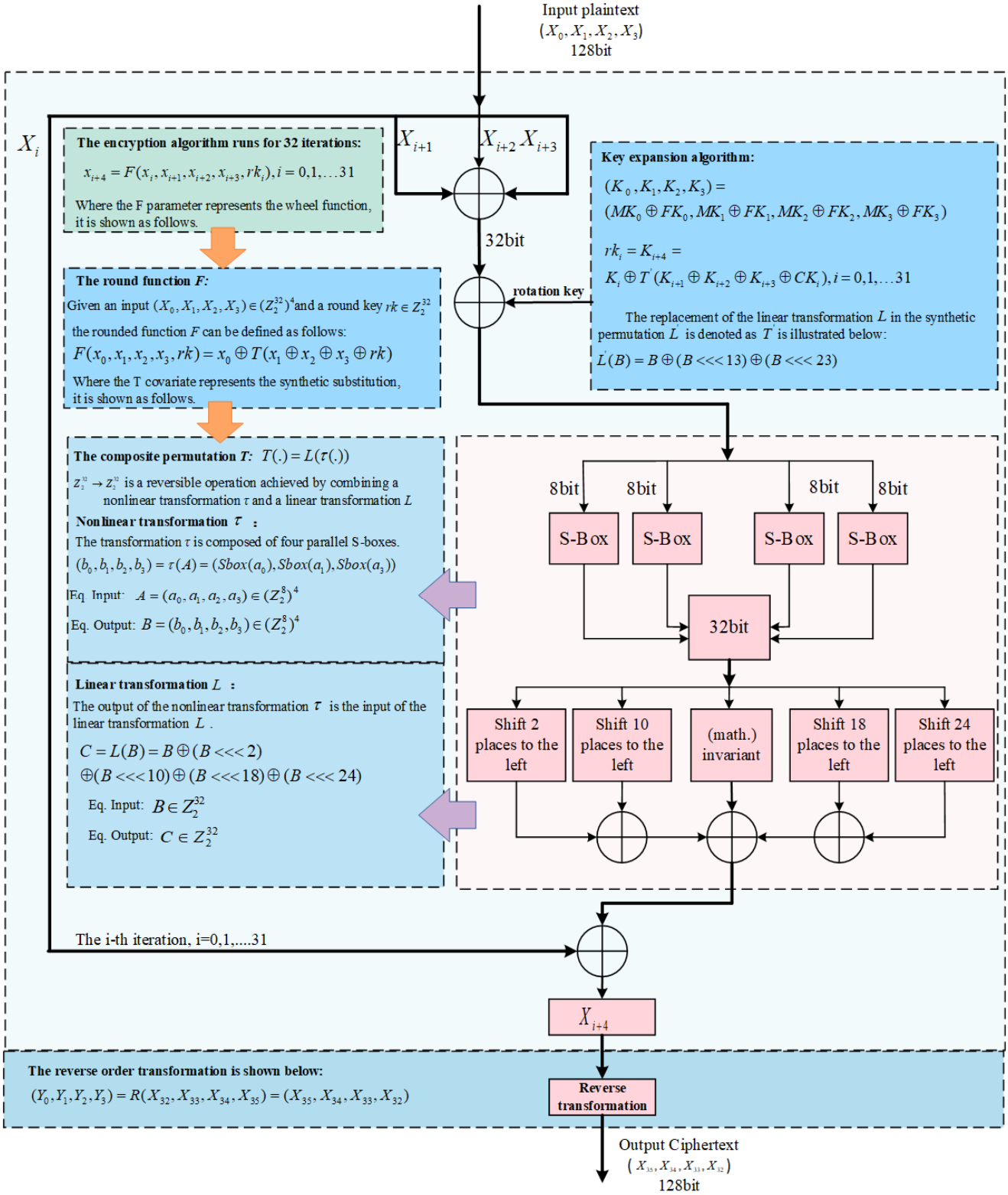

Fig. 1 SM4 cryptographic algorithm structure

The SM4 algorithm [6], also called SMS4, is a symmetric block cipher introduced and developed by China's State Cryptography Administration in 2006. This algorithm is widely used for data encryption and decryption, especially when security is transmitting data. SM4 operates on a Feistel network framework, employing a block size and a key length of 128 bits each. The algorithm comprises 32 rounds of iterations involving substitution, permutation, exclusive or (XOR), and critical addition operations. The key is expanded to guarantee safety and protection, generating 32 subkeys. The encryption and decryption procedures follow a nonlinear iterative structure, although the order of the round keys is reversed during decryption.

In the SM4 cryptographic algorithm, the encryption algorithm comprises 32 iterations and one inverse transformation R. The symbol $\oplus$ represents 32-bit XOR, $\lll$ represents a 32-bit circular left shift, and Z_2^n represents a set of binary sequences of length n bits. It is defined by the variables $(X_0, X_1, X_2, X_3) \in (Z_2^{32})^4$, $(Y_0, Y_1, Y_2, Y_3) \in (Z_2^{32})^4$ representing the ciphertext output and the plaintext input, respectively. During the encryption process, the encryption key has a length of 128 bits and it is denoted as $MK = (MK_0, MK_1, MK_2, MK_3)$, where $MK_i (i = 0,1,2,3)$ represents a word (each word is 32 bits or 4 Bytes). As for the round key, it is denoted as $(rk_0, rk_1, \cdots, rk_{31})$, where $rk_i (i = 0,1, \cdots, 31)$ is a 32-bit word. The round key is generated based on the encryption key. Moreover, $FK = (FK_0, FK_1, FK_2, FK_3)$ represents the system parameters while $CK = (CK_0, CK_1, \cdots, CK_{31})$ representing fixed parameters used in the key expansion algorithm. Note that $FK_i (i = 0, \cdots, 3)$ and $CK_i (i = 0, \cdots, 31)$ are represented as words.

It is worth noting that the decryption key is the decryption process that closely mirrors the encryption process, with the sole distinction being the reversed order of the round key usage. During decryption, the round keys are employed in a specific sequence $(rk_{31}, rk_{30}, \cdots, rk_0)$. Moreover, the structure of the SM4 cryptographic algorithm is depicted in Fig. 1.

2.2. *Introduction of S-Box*

The S-box [6] plays a pivotal role in the block cipher algorithm, substituting or transforming each bit within a block. This transformation brings obfuscation and nonlinearity, thereby greatly enhancing the algorithm's security and capacity to resist cryptanalysis. The design of the S-box aims to amplify the algorithm's nonlinear properties, improving its resilience against linear and differential cryptanalysis. By strengthening the S-box, the algorithm's ability to withstand cryptanalysis attacks is significantly enhanced, leading to an overall improvement in the security level of the encryption algorithm. The initial S-box is displayed in Fig. 2.

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	d6	90	e9	fe	cc	e1	3d	b7	16	b6	14	c2	28	fb	2c	5
1	2b	67	9a	76	2a	be	4	c3	aa	44	13	26	49	86	6	99
2	9c	42	50	f4	91	ef	98	7a	33	54	b	43	ed	cf	ac	62
3	e4	b3	1c	a9	c9	8	e8	95	80	df	94	fa	75	8f	3f	a6
4	47	7	a7	fc	f3	73	17	ba	83	59	3c	19	e6	85	4f	a8
5	68	6b	81	b2	71	64	da	8b	f8	eb	f	4b	70	56	9d	35
6	1e	24	e	5e	63	58	d1	a2	25	22	7c	3b	1	21	78	87
7	d4	0	46	57	9f	d3	27	52	4c	36	2	c7	a0	c4	c8	9c
8	ea	bf	8a	d2	40	c7	38	b5	a3	f7	f2	ce	f9	61	15	a1
9	e0	ae	5d	a4	9b	34	1a	55	ad	93	32	30	f5	8c	b1	e3
a	1d	f6	e2	2e	82	66	ca	60	c0	29	23	ab	d	53	4c	6f
b	d5	db	37	45	de	fb	8e	2f	3	ff	6a	72	6d	6c	5b	51
c	8b	1b	af	92	bb	dd	bc	7f	11	d9	5c	41	1f	10	5a	d8
d	a	c1	31	88	a5	cd	7b	bd	2d	74	d0	12	b8	e5	b4	b0
e	89	69	97	4a	c	96	77	7e	65	b9	f1	9	c5	6e	c6	84
f	18	f0	7d	ec	3a	dc	4d	20	79	ee	5f	3e	d7	cb	39	48

Fig. 2 Initial S-box [6]

The security of the S-box is essential for the overall safety of a cryptographic algorithm. Despite its well-defined construction, attackers still consider it their primary target. Therefore, carefully designing and integrating the S-box structure within the SM4 cryptographic algorithm is vital. Once validated, it protects sensitive data confidentiality and robustly supports secured data transmission. Therefore, this study focuses on analyzing the S-box of the SM4 algorithm from three perspectives: differential property, linear property, and nonlinear property.

(1) Differential property

It is a commonly used technique in cryptography to investigate the permutation operations in the encrypted algorithms. It aims to assess algorithm security by examining the differences between the input and output pairs. Analyzing the differential properties of the S-box is crucial in cryptographic algorithm design and analysis. Moreover, optimizing the S-box design can improve the algorithm's security and reliability. A lower differential probability signifies more robust differential properties, enhancing the algorithm's resistance against attacks like differential cryptanalysis. Moreover, the differential probability is defined as follows [12]:

$$DP^f(\Delta x \to \Delta y) = \frac{\#\left\{x \in Z_2^{32} \middle| f(x \oplus \Delta x) \oplus f(x) = \Delta y\right\}}{2^n} \tag{1}$$

where f is a function of $Z_2^{32} \to Z_2^{32}$, Δx is the input difference, and Δy is the output difference.

(2) Linearity property

In cryptography, the linearity property refers to the ability to establish relationships through linear operations. Low linearity is crucial for the S-box in the SM4 cryptographic algorithm as it eliminates the linear connection between the input and the output, thereby increasing security and resistance against attacks. This ensures that the result of the S-box cannot be determined by linear operations, thus effectively thwarting the cryptographic attacks. Moreover, the linear probability indicates the linear correlation between inputs and outputs. Ideally, this probability should be close to zero, as this strengthens the algorithm's nonlinearity and overall security. The linear probability is defined as follows [12]:

$$LP^f(w \to v) = \left(\frac{\#\left\{x \in Z_2^{32} \middle| vf(x) = wx\right\}}{2^{n-1}} - 1\right)^2 \tag{2}$$

where w is the input mask, and v is the output bitmask.

(3) Nonlinear property

The nonlinearity of the S-box is characterized by the absence of a simple linear relationship between its inputs and outputs, making it difficult to determine its model through linear operations. The high level of nonlinearity plays a significant role in assessing the S-box's resistance to linear attacks. It enhances the security of the encryption algorithm by diminishing its susceptibility to linear attacks.

Furthermore, improving the nonlinearity of the S-box is crucial to prevent linear attacks, strengthen the encryption algorithm's resilience, and ensure its capacity to withstand other forms of cryptanalysis, such as differential attacks. Finally, the nonlinearity is defined as follows [12]:

$$N_f = \min_{l(x) \in l_n} d_H\left[f(x), l(x)\right] \tag{3}$$

Let l_n be a set consisting of all affine transformations on Z_2. The variable $d_H[f(x), l(x)]$ represents the Hamming distance between $f(x)$ and $l(x)$.

3. Hybrid Algorithm Improvement of the S-Box

In this paper, a combination of genetic algorithm and simulated annealing algorithm is used to improve the S-Box algorithm. To fully evaluate this improvement, a multi-objective fitting function is strategically employed. The multi-objective fitting function is designed to accurately evaluate the performance of the S-Box algorithm to meet multiple objectives. The selected metrics should comprehensively assess the quality of the S-Box, including the primary nonlinear metrics as well as the secondary differential and linear probabilities. A balance of these metrics is essential to help improve the nonlinearity of the S-Box while keeping the differential and linear probabilities low, thus increasing its resistance to various cryptanalysis techniques.

3.1. Design of a multi-objective fitness function

The design of the multi-objective fitting function highlights nonlinearity as the primary metric of resistance to linear and differential cryptanalysis, and the function also includes secondary metrics such as differential probability and linear probability to comprehensively assess the resistance of the S-Box to specific attack methods such as differential and linear cryptanalysis. Moreover, the following equation represents the fitness value, which is used to evaluate the quality of the S-box.

$$F(s) = w_1 \times NL\text{-}w_2 \times DP\text{-}w_3 \times LP \tag{4}$$

where NLrepresents the nonlinearity indicator, DP indicates the differential probability indicator, and LP represents the linear indicator. Moreover, w_1, w_2, w_3 represents the weights of the different performance indicators, which are deployed to balance their importance in the fitness function.

The importance of nonlinearity as a primary performance indicator in S-box design must be carefully balanced, as it directly influences the overall security and strength of the S-box. In opposition, the secondary indicators (e.g., differential probability and linear probability) hold lower significance. Therefore, using a multi-objective fitness function allows a comprehensive evaluation of various S-box structures and helps the optimization algorithm select the most optimal choice.

By collectively considering factors such as nonlinearity, differential attack, and linear attack, the improved S-box structures can be identified, enhancing support for secured data transmission. Furthermore, selecting an appropriate multi-objective fitness function is crucial to enable the optimization algorithm to identify the exceptional S-boxes, optimize targeting and efficiency, and bolster the overall security of data transmission.

3.2. The design of the model

To achieve this aim, a hybrid optimization method that combines a genetic algorithm and a simulated annealing algorithm will be employed. Genetic algorithms are widely used in S-box optimization as they allow for global search and exploration of several individuals in the solution space, thus avoiding local optima. In addition, the parallel computing capability of genetic algorithms is particularly advantageous for tackling large-scale problems, especially those involving multiple parameters in S-box design.

As for the simulated annealing, it complements genetic algorithms by enhancing the efficiency of S-box optimization. This technique is proficient in local search, as it accepts suboptimal solutions to prevent local optimization and increases the possibility of global optimization. Moreover, genetic algorithms conduct a global search to improve initial solutions. At the same time, simulated annealing carries out a local search to further optimize the S-box design and enhance its confidentiality and robustness. This hybrid optimization strategy effectively enhances the performance and security of the S-box, subsequently supporting the safety and robustness of the entire SM4 encryption algorithm. The specific steps involved in this approach are summarized as follows:

(1) Randomly generate a set of initial S-boxes as individuals in the population, where each individual represents a possible S-box. The S-boxes are represented as one-dimensional arrays of 256 elements within the population, where each element corresponds to a combination of input and output for the S-box. These combinations are linearly stored in the one-dimensional array. Assuming the S-box individual is denoted as "p_i" the collection of all S-boxes is referred to as $P = \{p_1, p_2, \cdots, p_{256}\}$;

(2) Calculate the fitness of each S-box using the Eq. (4) fitness function, evaluate and assign a fitness value to measure their quality and superiority;

(3) Through the selection operation using the roulette wheel method, the selection probability P_i can be used to determine the probability of each p_i being chosen. In this way, the following equation can select excellent individuals from the current population as the basis for the next generation population;

$$P_i = \frac{F(s)}{\sum_{i=1}^{N} F(s_i)} \tag{5}$$

(4) They perform selection operations to select parent individuals for crossover, generating new offspring. Here, the single-point crossover operation is used. The selected parents are divided into two parts: $p_1 = (p_{11}, p_{12}, \cdots, p_{1n})$ and $p_2 = (p_{21}, p_{22}, \cdots, p_{2n})$, where n is half the number of parents. Let $k(1 \leq k \leq n-1)$ be the crossover point. For each pair of parent individuals (p_{1i}, p_{2i}), two offspring individuals (p_{1i}, p_{2i}) are generated. Specifically, $p_{1i} = (p_{11}, p_{12}, \cdots, p_k, p_{21}, p_{22}, \cdots, p_{2n})$ and $p_{2i} = (p_{21}, p_{22}, \cdots, p_k, p_{11}, p_{12}, \cdots, p_{1n})$. The resulting offspring individuals after the crossover operation are $P_c = (p_{c1}, p_{c2}, \cdots, p_{cn})$;

(5) Performing mutation operations on the generated new individuals introduces random factors to maintain population diversity. Assuming a probability P_m, each offspring undergoes gene mutation, which involves randomly changing some genes. Let's take that the mutation operation changes a specific element s_{ij}. Assuming S represents the set of all possible S-box elements $S = \{0,1,\cdots,255\}$, the mutation operation can be described as $s_{ij} = s_{ij} + delta$, where $delta$ is a randomly selected new value from the set S. The resulting offspring individuals after the mutation operation are $P_m = (p_{m1}, p_{m2}, \cdots, p_{mn})$;

(6) Through iterative optimization, the genetic algorithm continuously iterates and improves the design of the S-box by performing selection, crossover, and mutation operations. This process generates a new generation of S-box individuals. The updated population is $P_m = (p_1, p_2, \cdots, p_N)$, where N represents the population size.

(7) Iterate generation by generation, continuously optimizing the fitness values of the S-box until the maximum specified number of iterations is reached, and then stop. Ultimately, the optimized S-box obtained serves as the result of global optimization.

The genetic algorithm generates a set of S-box individuals after several iterative optimizations, where the highest fitness represents the optimal solution. These optimized S-boxes are used as initial values for the subsequent stage where a simulated annealing algorithm is used for local search and fine-tuning. The process can be summarized as follows:

(1) Represent the S-box as a system and use the simulated annealing algorithm with an initial temperature of $T_0 = 1000$ and a cooling rate of $alpha = 0.99$ for fine-tuning optimization.

(2) A new S-box is randomly generated in each iteration by swapping two elements or applying a random mutation. Let's assume the current S-box is p, and the newly developed S-box is denoted as p'.

(3) Calculate the fitness value $F(p')$ of the new S-box and compare it with the fitness value $F(p)$ of the current S-box.

(4) Based on the acceptance probability function, decide whether to accept the new S-box p' as the current S-box p. The acceptance probability function can be defined using the *Boltzmann* distribution as follows:

$$P = e^{\frac{F(p)-F(p')}{T}} \tag{6}$$

where T represents the current temperature.

(5) At higher temperatures, there is a certain probability of accepting poorer solutions to avoid getting trapped in local optimal solutions. The likelihood of obtaining a new key is exponentially related to the temperature, and as the temperature decreases, the probability of accepting poorer solutions gradually decreases.

(6) As the annealing process progresses, the temperature continuously decreases, causing the probability of accepting poorer solutions to decrease gradually. This allows the algorithm to converge towards the global optimal solution. The temperature T is updated at a rate of alpha, meaning $T = T \times alpha$, where $alpha$ represents the rate at which the temperature decreases;

(7) Eventually, after a certain number of iterations and cooling steps, the simulated annealing algorithm will obtain an optimized S-box representing the global optimal solution.

Fig. 3 presents in detail the process of S-box optimization using genetic algorithms, including the creation of the initial population and the generation-by-generation evolution to obtain an optimized S-box. On the other hand, Fig. 4 demonstrates the process of improving the S-box using the simulated annealing algorithm, which involves the simulated annealing of the S-box in different temperature phases and the gradual tuning of the performance of the S-box in a way that it can be optimized at a higher level.

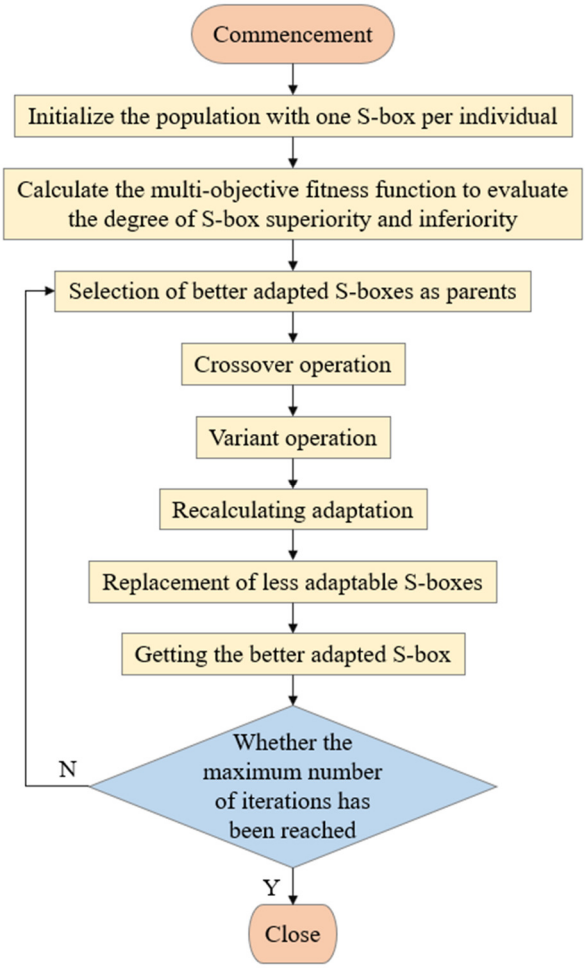

Fig. 3 Genetic algorithm flowchart

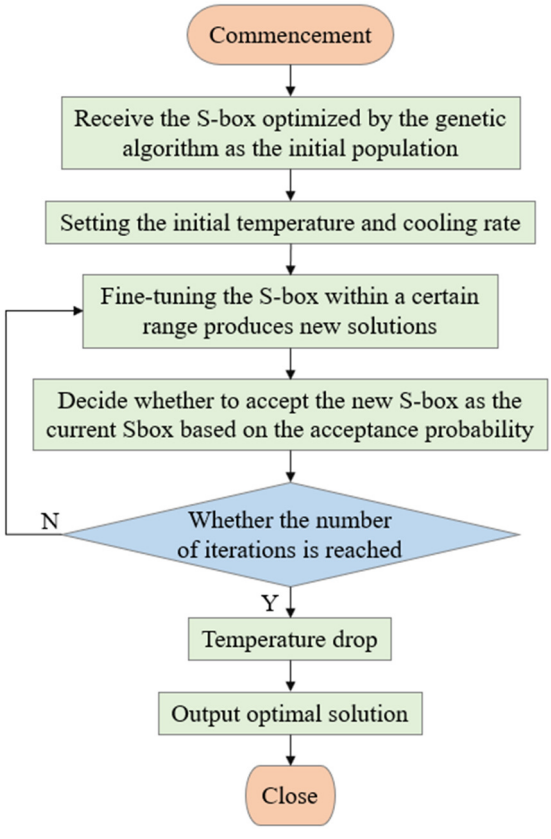

Fig. 4 Simulated annealing algorithm flowchart

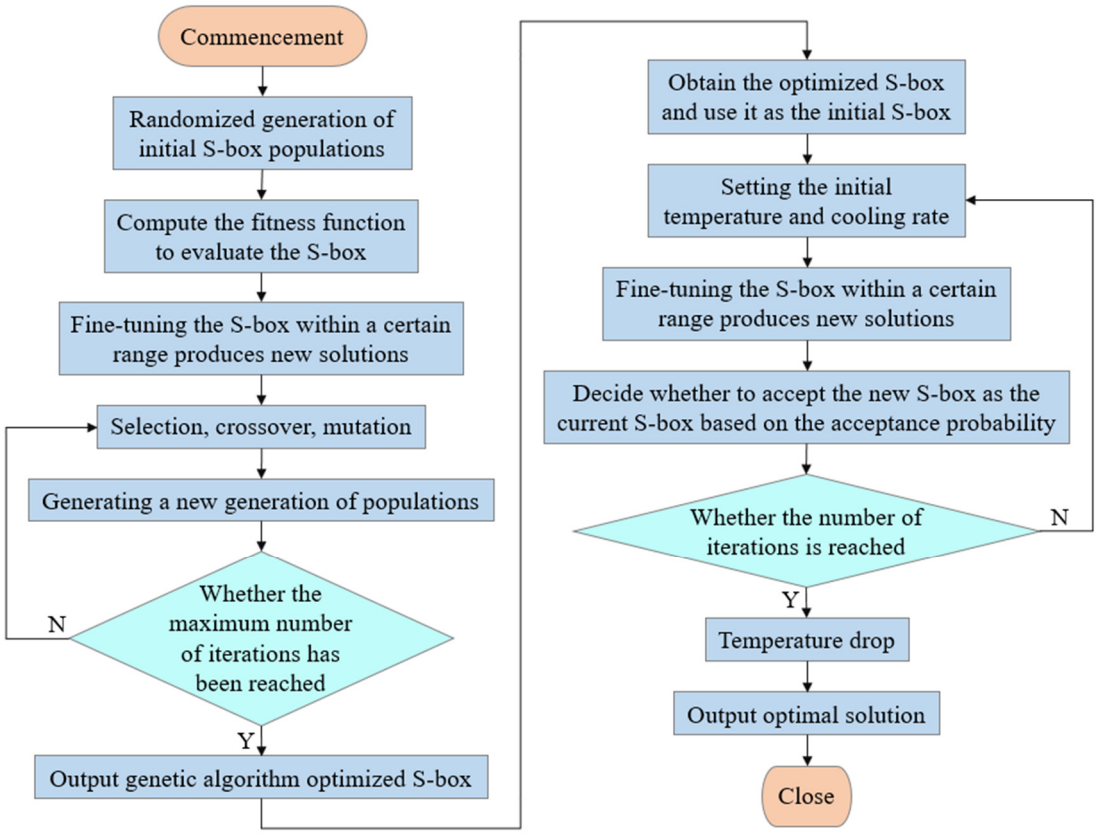

Fig. 5 Flowchart of hybrid algorithm optimization for S-box model

Combining genetic and simulated annealing algorithms is essential for global exploration and local optimization. This integration improves the performance, attack resistance, security, and reliability of the S-box during data transmission of SM4 cryptographic algorithms. Moreover, the integrated process of optimizing the S-box model using this hybrid algorithm is displayed in Fig. 5. The pseudo-code for optimizing the S-box using the hybrid algorithm is shown in Fig. 6.

Algorithm 1 Hybrid Optimization Algorithm

```
// Initializing the S-Box
current_s_box ← Initialize S-Box()
// Genetic algorithms for global search and optimization
Procedure Genetic Algorithm Optimization
// Initializing populations
population ← Initialize Population()
// Iterative optimization process
for generation in range(max_generations) do
    // Calculation of adaptation values
    fitness_values ← Evaluate Fitness(population)
    // Selection of operation
    selected_parents ← Selection(population, fitness_values)
    // Crossover operation
    offspring ← Crossover(selected_parents)
    // Variant operation
    Mutate(offspring)
    // Renewal of stocks
    population ← Update Population(population, offspring)
    // Update current S-box
    current_s_box ← Generate New S-Box(population)
end for
End Procedure
// Simulated annealing algorithm for local search and fine-tuned optimization
Procedure Simulated Annealing Optimization
// Setting the initial temperature and cooling rate
T0 ← 1000
```

Fig. 6 Hybrid algorithm pseudo-code for optimizing S-boxes

```
alpha ← 0.99
// Iterative optimization process
while T0 > 1 do
  for i in range(iterations_per_temperature) do
    // Generate new S-boxes
    new_s_box ← Perturb S-Box(current_s_box)
    // Calculation of adaptation values
    current_fitness ← Evaluate Fitness(current_s_box)
    new_fitness ← Evaluate Fitness(new_s_box)
    // Calculating the probability of acceptance
    probability ← Calculate Acceptance Probability(current_fitness,
    new_fitness, T0)
    if probability >random() then
      // Acceptance of new S-boxes
      current_s_box ← new_s_box
    end if
  end for
  // Drop down the temperature
  T0 ← T0 × alpha
end while
End Procedure
```

Fig. 6 Hybrid algorithm pseudo-code for optimizing S-boxes (continued)

4. Design of an Experiment

To evaluate the improvement effect of the hybrid algorithm optimization on the performance of the S-box, the Python programming language environment was used for simulation experiments. Four types of S-boxes were used in the investigation: the initial S-box, the S-box optimized only by genetic algorithm, the S-box optimized only by simulated annealing algorithm, and the S-box optimized by the hybrid algorithm. The effects of these algorithms on optimizing the S-box can be evaluated by analyzing and testing the performance of the S-box.

To verify the ability of the hybrid algorithm to optimize S-boxes, two sets of hybrid algorithm-optimized S-boxes were set up for testing in this paper, and the experiments were conducted several times and visualized using appropriate graphical tools. These visuals allow us to compare the impact of different optimization methods regarding the S-box performance, facilitating the assessment of performance improvements achieved through hybrid algorithm optimization. The parameter settings are shown in Table 1.

Table 1 Parameter setting situation

Parameter setting situation	Value
Population size	100
Number of genetic algorithm iterations	1000
Variation rate	0.1
Simulated annealing algorithm initial temperature (T_0)	1000
Simulated annealing algorithm cooling rate (*alpha*)	0.99
S-box size	16 × 16

4.1. Analysis of algorithm performance

In the performance analysis section, not only are the three key metrics of differential probability, linear probability, and nonlinearity analyzed in depth computationally but they are also compared in detail with the state-of-the-art. This comprehensive approach allows for an accurate assessment of the performance of the S-box optimization algorithm compared to state-of-the-art algorithms. This comparative analysis highlights the research results in terms of differential probability, linear probability, and degree of nonlinearity. It places these results in a broad cryptographic context to better assess their significance and impact. This further strengthens the academic value and practical application of the paper.

(1) Differential probability

To reduce the differential transmission between input and output, the research on optimized S-box relies on lower differential probabilities to enhance the resistance of the S-box against differential attacks. In this study, the differential probabilities of two groups of optimized S-boxes are 0.359 and 0.371, respectively. The comparison results with the State-of-the-art are shown in Table 2.

Table 2 Comparison of differential probabilities with State-of-the-art

Substitution box	Differential probability	Substitution box	Differential probability
This article	0.0359	Zhang et al. [7]	0.0391
This article	0.0371	Khan et al. [9]	0.0468
The initial S-box	0.0625	Zahid et al. [10]	0.0390
Only by genetic algorithm	0.0549	Khan et al. [11]	0.0390
Only by simulated annealing algorithm	0.0583	Liu et al. [22]	0.0391

As shown in Table 2, the hybrid algorithm-optimized S-boxes all outperform the S-boxes of other algorithmic techniques in terms of differential probability. This finding highlights the superior performance of hybrid algorithms in improving the performance of S-boxes, especially in terms of differential attacks. This improvement is crucial for improving the security of cryptographic algorithms, as lower differential probabilities mean that it is more difficult to predict the output of the S-boxes, thus increasing the resistance of the cipher.

(2) Linear probability

The S-box must have a low linear probability to achieve a higher level of nonlinearity. In this study, the linear probabilities of two groups of optimized S-boxes are 0.094 and 0.096, respectively. The comparison results with the State-of-the-art are shown in Table 3.

Table 3 Comparison of linear probabilities with State-of-the-art

Substitution box	Linear probability	Substitution box	Linear probability
This article	0.094	Khan et al. [8]	0.117
This article	0.096	Zahid et al. [10]	0.125
The initial S-box	0.125	Khan et al. [11]	0.141
Only by genetic algorithm	0.106	Yang [16]	0.117
Only by simulated annealing algorithm	0.105	Wang et al. [17]	0.085

As shown in Table 3, the hybrid algorithm optimizes the S-box with slightly lower linear probability compared to the chaotic mapping and multi-objective genetic algorithm [17] in this study. However, the hybrid algorithm has a significant advantage compared to other literature.

(3) Nonlinearity degree

The nonlinearity of the S-box plays a vital role in defense against linear approximation attacks. An S-box with high nonlinearity can completely resist attacks such as linear and differential cryptanalysis, thus improving the overall security of the cryptographic algorithm. Table 4 shows the nonlinearity values compared to the state-of-the-art.

Table 4 Nonlinearity value compared with state-of-the-art

Substitution box	Nonlinearity	Substitution box	Nonlinearity
This article	112	Khan et al. [11]	108
The initial S-box	98	Zhao [15]	112
Only by genetic algorithm	103	Yang [16]	107.5
Only by simulated annealing algorithm	102	Wang et al. [17]	111.5
Zhang et al. [7]	107.25	Wang [18]	107.5
Khan et al. [9]	110.5	Si et al. [21]	110.6
Zahid et al. [10]	111.3	Liu et al. [22]	111.25

Comparative results show that the nonlinearity values of the S-boxes optimized using the hybrid algorithm are better or equal to those of the state-of-the-art S-boxes. Consequently, using the hybrid algorithm enhances the complexity of encryption algorithms, significantly increasing the difficulty of attack. To demonstrate more intuitively the advantages of the constructed hybrid algorithm model in improving the performance of the S-box, plotting software was used to unify the computationally derived results in a single figure. The figure shows the performance analysis of three indicators of the hybrid algorithm optimized S-box, the genetic algorithm only optimized S-box, and the simulated annealing algorithm only optimized S-box, as shown in Fig. 7.

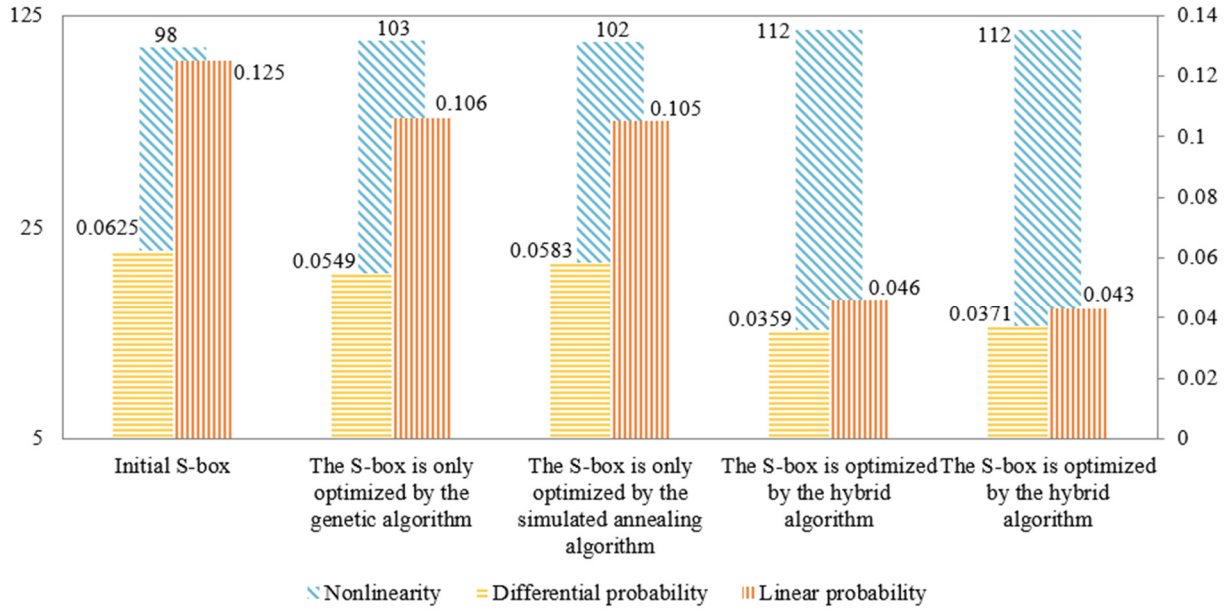

Fig. 7 The summary of the performance analysis

4.2. *Experimental results of data transmission*

Based on the optimized S-boxes obtained from the research above, they were replaced with the original S-boxes in the SM4 encryption algorithm and used in the data transmission process for electric power metering. A particular composite data segment from a specific region was selected as the plaintext, as shown in Fig. 8(a). The data transmission experiment resulted in the ciphertext shown in Fig. 8(b).

239. 200	239. 900	239. 300	0. 822	0. 489	0. 512	0. 4275	0. 0110	0. 9780
239. 900	239. 600	239. 800	0. 550	0. 384	0. 449	0. 3188	0. 0286	0. 9600
239. 600	240. 500	239. 700	0. 556	0. 353	0. 455	0. 3106	0. 0423	0. 9480
239. 900	240. 100	239. 600	0. 446	0. 604	0. 449	0. 3452	0. 0329	0. 9590
239. 900	240. 700	240. 100	0. 503	0. 387	0. 458	0. 3089	0. 0295	0. 9530
239. 700	240. 700	240. 000	0. 504	0. 349	0. 371	0. 2817	0. 0335	0. 9570
239. 400	240. 400	239. 700	0. 475	0. 319	0. 352	0. 2639	0. 0241	0. 9590
239. 900	240. 600	239. 900	0. 510	0. 498	0. 399	0. 3280	0. 0421	0. 9690
240. 400	240. 900	240. 700	0. 425	0. 374	0. 324	0. 2599	0. 0219	0. 9610
240. 400	240. 900	240. 100	0. 481	0. 412	0. 273	0. 2698	0. 0402	0. 9600
240. 400	241. 100	240. 400	0. 511	0. 360	0. 407	0. 2966	0. 0522	0. 9630
240. 800	241. 400	241. 200	0. 499	0. 426	0. 278	0. 2801	0. 0360	0. 9640
241. 300	241. 500	241. 000	0. 384	0. 356	0. 275	0. 2353	0. 0212	0. 9590
241. 400	242. 100	241. 700	0. 442	0. 366	0. 264	0. 2502	0. 0306	0. 9640
240. 900	242. 100	241. 600	0. 442	0. 363	0. 271	0. 2529	0. 0279	0. 9650
241. 400	241. 700	241. 100	0. 420	0. 291	0. 274	0. 2284	0. 0253	0. 9590
241. 400	241. 700	241. 100	0. 357	0. 378	0. 292	0. 2380	0. 0352	0. 9580
241. 200	242. 300	241. 400	0. 387	0. 344	0. 287	0. 2364	0. 0318	0. 9590
241. 400	242. 100	241. 700	0. 392	0. 337	0. 253	0. 2284	0. 0353	0. 9610
241. 000	241. 900	241. 400	0. 337	0. 359	0. 208	0. 2090	0. 0316	0. 9560
241. 200	242. 200	241. 600	0. 358	0. 428	0. 424	0. 2862	0. 0275	0. 9770
241. 300	242. 400	241. 900	0. 368	0. 288	0. 466	0. 2643	0. 0281	0. 9720
241. 400	241. 600	241. 900	0. 319	0. 596	0. 393	0. 3091	0. 0288	0. 9770
241. 400	241. 800	241. 600	0. 334	0. 539	0. 422	0. 3052	0. 0366	0. 9740
240. 700	241. 000	240. 800	0. 333	0. 634	0. 467	0. 3392	0. 0287	0. 9810
240. 400	240. 800	239. 900	0. 390	0. 582	0. 902	0. 4440	0. 0336	0. 9850
240. 000	240. 500	239. 200	0. 443	0. 490	0. 976	0. 4514	0. 0343	0. 9850
239. 900	240. 100	239. 100	0. 809	1. 037	1. 093	0. 7014	0. 0124	0. 9950
239. 700	239. 900	239. 300	0. 503	0. 545	0. 644	0. 3991	0. 0252	0. 9840
239. 400	239. 700	239. 200	0. 586	0. 533	0. 545	0. 3742	0. 0233	0. 9810
238. 600	238. 800	238. 600	0. 521	0. 580	0. 706	0. 4222	0. 0519	0. 9780
238. 300	238. 700	237. 900	0. 745	0. 790	0. 837	0. 5601	0. 0227	0. 9900
237. 600	238. 200	237. 600	0. 601	0. 645	0. 619	0. 4372	0. 0146	0. 9850
236. 800	237. 200	237. 100	0. 877	0. 892	0. 892	0. 6259	0. 0153	0. 9920
237. 400	237. 900	237. 500	0. 679	0. 615	0. 736	0. 4737	0. 0174	0. 9810

(a) Plaintext data fragment

```
QF9xf3RBTLiXU+wW5qcxTk+olYYKQCvWPjb4yxA0FPU78jQMKZ6lQq7WqS4rjAWvtJvte0X
!YqtJkYxaNN+Mb6ZyjAd/5V7CbZCQ1NcaJ2wnSwlK1CA+CoqV7NzgsBcXzZiyu9hqwyblZl
Q7Y1T4DQUt+jKz9yjoTo+Kl5r4LSM9XD0AjDMkpSVm0yKD5z9PxkBGl7P8FF5MdhQF7uUE
C7ob5eUe1SLUQu4+sQu0s4lqGhMO9H7h99itd+lugoMWxfGdf2moq+qGJURhR6sq8/I28
/gLtg2bF9u8imbGd3CARWv888zYqcKlN85CBpdzBgO0lwfZHrtfV4km5sWN7sUR0+KQ2M
CZgCVd+utUrlxOhF3gINpR1j886aSZ69Sy+qTB6NjTKmljmuqLZjJxu/lrJ2ssTeHwE0GlFTfeA
DuBre9eTPE6J5urAwnYgf4EcMjUf7FBDOLs9RntnrKHYv9MeGavaoNYewdbBLFElH++FMi34
:7BuoQ9/SdZxGh3uxtkANrBUnV8MDssFYHHn/ywaBlMZG3rYYQtB8sQA9ZLENNLTQhEqTl
Y1wsR+57Qbng6GkPk8EVjg5ZD+3jA6ekC+6NXs3huqDkcakszIhfkouqrvsGkgAEvF1fxoyAc
'GJ4f3zi8AFqDkLVgXBsDz4EVrQKCvE/iwZZKJTjS0KL0oir0NrJdsipB5M32rzDtfyOUIM0kAGI
)j2Y7w0EKBmpIhc7KHEccLBV3dbclfTF5dYVl00UPrrVr+eyvlvhsoWapamuUwLXGj+GCtY/D:
:WZYWzskFMXLDvTMijxCnOfL97KbHxYbv+dOuIFd26jaXLC3WQvb8smF3MKJBlwslQHm/'
AmWB2BKZD0JQZ+Sm2M2pkmXfkLT1w9eM40wMtX3ltWumDAJCnC6gDB/ExLqg7VR4dI
fj2HkBmvjUdUhKvr6ll/+c0RFkpg9DXT3eHojPAFyMpDcfXy0aizPaIsl66KifdBhNQqbIB2w8E
6eN5MUTURAH4DHcFmVHEzKyfZmo5u83di7/LUK6EusYXFHokqpR45MsexMLWoXZB9My
4KfMbm0/4ohtDrzWtSFt9QsaXcxOSG4BlQ5+oPNNaIhEEEnd0iB5N8ttPeMCjbe+wvGWAz
Ww0ExGz6+nlqk95o7la2Q8o84swlAzKdbe0GopgwiwE4TF5wbE3fcVm4XPawgnnqz0Cg8P
:4bNgqNNgQwNYY0ol26/a7O4JVGLi+P23wWkAQha+iaqHhhPSdQQRU2c03WSWcNtHu
IZuSi7Xgoyo4hPG5T9ZpQvpFzdMDD4t1s46F5XdHHgd0x5gYzXJfyOJMtlxpJNrYSLC2UoiY
goTR/Hmn+AGtniV4EDzv+0t4K7wENv2JSlYSRszC9pkHLKLx9CSAbtizFi3C2qT+85ZwhrjX/
CcyUbEkOJfMmJBHmlub7VIouD8CmjprK1qDK2+Lr8LP9g9lwUQy+btxup6WPWfXlzHfFOL
lblhESGtjcGd8xdrw47lMOo+/uup6XYTPgZ6Bs/tD3bwUMzEMzSh26gex/xOytdAFgq7Ro52
isM1VSVrH36f1hWmhKF7mRwOFkn1V0G6Ykwqne0FlviPkAl HYhlNIR9/Fpl2Rrn2+1OdIep
```

(b) Ciphertext data fragment

Fig. 8 Data encryption before and after comparison

To further verify whether the optimized S-boxes in this study improve the security of encryption, information entropy will be calculated and analyzed on the ciphertext data. Information entropy is a metric used to measure the amount of information or uncertainty. It is usually used in cryptography to evaluate the strength of cryptographic algorithms or keys as well as the randomness and unpredictability of the encryption process. Information entropy is calculated as follows.

$$H(X) = \sum_{x_i \in X} P(x_i) \log P(x_i) \tag{6}$$

In the formula, X represents the sample space and $P(x_i)$ denotes the probability of occurrence of $X = x_i$. The ciphertext is divided into units of bytes. Hence, when the ciphertext follows a uniform distribution, the ideal value of the average information entropy should be close to 8. The information entropy of the ciphertext calculated based on the above transmission experimental data is 7.998, close to the ideal value, indicating that the ciphertext has a high-security performance.

5. Conclusions

The structure of the S-box in the traditional SM4 cipher algorithm has been extensively researched, but being nonlinear can enhance its safety robustness. Moreover, the existing optimization methods often do not consider global optimization. Therefore, this paper proposes a hybrid algorithm to effectively enhance the nonlinearity of the S-box while achieving global optimization. The contributions of this paper are as follows:

(1) This paper presents a new multi-objective fitting function design for evaluating the performance of S-boxes and guiding the selection of optimal structures. The incorporation of these metrics improves the efficiency, relevance, and security of data transmission for S-box optimization, taking into account nonlinearity and resistance to attacks.

(2) This study presents a hybrid algorithm that optimizes the S-box model to achieve global optimization and avoid local optima. The algorithm combines the strengths of genetic algorithms and simulated annealing algorithms. The former enables extensive searching in the solution space, overcoming local optima, while the latter fine-tunes the solutions to enhance their convergence and performance. The parameter settings for these algorithms are easily adjustable and implemented, leading to an overall improvement in the optimization effectiveness.

(3) This study compares the differential probability, linear probability, and nonlinearity of different S-box optimization methods. The results show that the hybrid algorithm significantly reduces the average differential probability of the S-box to 0.0359 and 0.0371, the average linear probability to 0.094 and 0.096, and dramatically improves the average nonlinear probability to 112. The optimized S-box is put into a cryptographic algorithm, and a section of load data is selected for encryption. Its ciphertext information entropy is calculated to be 7.998, very close to the ideal value of 8.

In the future, the hybrid algorithm model will be further studied and applied. Through research and development in practice, using the model in accurate encryption algorithms can improve security and performance. Meanwhile, the model can enhance device security and deal with growing cybersecurity threats in areas such as the Internet of Things. Applying the hybrid algorithm model in real-world scenarios contributes positively to cryptography research and cryptosystem development and meets the needs of information security and data protection. Looking ahead, in-depth research on hybrid algorithmic models will be conducted to promote their widespread application and enhance security protection.

Acknowledgments

This work was supported by the science and technology project of Yunnan Power Grid Co., Ltd. providing funding under the project number YNKJXM20220010.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] R. Tiwari, P. Chithaluru, K. Kumar, M. Kumar, and T. Stephan, "A Cooperation Federated Forwarder Selection Technique for Maximizing Network Lifetime in Wireless Sensor Network," Sādhanā, vol. 48, no. 3, article no. 135, 2023.
[2] Sharmila, P. Kumar, S. Bhushan, M. Kumar, and M. Alazab, "Secure Key Management and Mutual Authentication Protocol for Wireless Sensor Network by Linking Edge Devices Using Hybrid Approach," Wireless Personal Communications, vol. 130, no. 4, pp. 2935-2957, June 2023.
[3] A. Banerjee, S. K. Gupta, P. Gupta, A. Sufian, A. Srivastava, and M. Kumar, "UAV-IoT Collaboration: Energy and Time-Saving Task Scheduling Scheme," International Journal of Communication Systems, vol. 36, no. 14, article no. e5555, September 2023.
[4] H. L. Liu, "Research on the Security Quantification Method of Data Transmission in Heterogeneous Super Super Density Networks," Computer Simulation, vol. 38, no. 1, pp. 150-153, 2021.
[5] Y. Liu, "Application of Data Encryption Technology in Network Security Transmission," Cyberspace Security, vol. 14, no. 03, pp. 41-44, June 2023.
[6] S. He, H. Li, and F. Li, "Optimization and Implementation of SM4 on FPGA," Journal of Xi'an University of Electronic Science and Technology, vol. 48, no. 03, pp. 155-162, April 2021.
[7] X. Zhang, K. Wei, and W. Jiang, "The Nonlinearity Optimization Algorithm of S-box Based on TD-ERCS Sequence," Information Network Security, vol. 21, no. 01, pp. 10-18, January 2021.
[8] M. F. Khan, K. Saleem, M. A. Alshara, and S. Bashir, "Multilevel Information Fusion for Cryptographic Substitution Box Construction Based on Inevitable Random Noise in Medical Imaging," Scientific Reports, vol. 11, no. 1, article no. 14282, 2021.
[9] M. F. Khan, K. Saleem, M. M. Hazzazi, M. Alotaibi, P. K. Shukla, M. Aqueel, et al., "Human Psychological Disorder towards Cryptography: True Random Number Generator from EEG of Schizophrenics and Its Application in Block Encryption's Substitution Box," Computational Intelligence and Neuroscience, vol. 2022, article no. 2532497, 2022.
[10] A. H. Zahid, M. Ahmad, A. Alkhayyat, M. J. Arshad, M. M. U. Shaban, N. F. Soliman, et al., "Construction of Optimized Dynamic S-Boxes Based on a Cubic Modular Transform and the Sine Function," IEEEAccess, vol. 9, pp. 131273-131285, 2021.
[11] M. F. Khan, K. Saleem, M. Alotaibi, M. M. Hazzazi, E. Rehman, A. A. Abbasi, et al., "Construction and Optimization of TRNG Based Substitution Boxes for Block Encryption Algorithms," https://doi.org/10.48550/arXiv.2206.09424, June 19, 2022.
[12] L. Zhang, L. He, and B. Yu, "Design and Analysis of Large-Scale S-Boxes with SPS Structure," Journal on Communications, vol. 44, no. 2, pp. 27-40, February 2023.
[13] S. A. Susan T and N. Balasubramanian, "A Hybrid Metaheuristic Algorithm for Stop Point Selection in Wireless Rechargeable Sensor Network," International Journal of Engineering and Technology Innovation, vol. 13, no. 4, pp. 296-312, October 2023.
[14] A. J. P. Delima, "An Enhanced K-Nearest Neighbor Predictive Model through Metaheuristic Optimization," International Journal of Engineering and Technology Innovation, vol. 10, no. 4, pp. 280-292, September 2020.
[15] J. Zhao, "Research on Security Protection Strategy for Data Transmission Based on Internet of Things," M.S. dissertation, Department Communications Engineering, Jilin University, Changchun, May 2020.
[16] S. Yang, "Research on Cryptographic Algorithms Based on Chaotic Systems and Optimized S-Boxes," M.S. dissertation, Harbin Institute of Technology, Harbin, June 2022.
[17] Y. Wang, M. Wang, and J. Gong, "Optimal Design Method of 8x8 S-box Based on Multi-objective Genetic Algorithm," Journal of Southwest Jiaotong University, pp. 1-10, June 2022.
[18] L. Wang, "Optimized Design of Chaotic S-Box and Its Application in Image Encryption," M.S. dissertation, Changsha University of Science and Technology, Changsha, June 2021.
[19] Zhengquan Li, Yawen Lu, Ruiqing Qin, Lirong Tan, and Bin Gu, An S-Box Design Method Based on Hyperchaotic Systems with Genetic Particle Swarm Algorithm, Chinese Patent, CN202210563082.5, August 2022. (In Chinese)
[20] X. Tong, D. Zhu, and M. Zhang, An S-Box Optimization Method Based on Improved Genetic Algorithm, Chinese Patent, CN202010571146.7, October 2020.
[21] Y. Si, H. Liu, and M. Zhao, "Constructing Keyed Strong S-Box with Higher Nonlinearity Based on 2D Hyper Chaotic Map and Algebraic Operation," Integration, vol. 88, pp. 269-277, January 2023.
[22] X. Liu, X. Tong, Z. Wang, and M. Zhang, "Efficient High Nonlinearity S-Box Generating Algorithm Based on Third-Order Nonlinear Digital Filter," Chaos, Solitons & Fractals, vol. 150, article no. 111109, September 2021.