IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 July 2022, accepted 28 August 2022, date of publication 5 September 2022, date of current version 16 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3204663

RESEARCH ARTICLE

Network Security Situation Assessment Based on Improved WOA-SVM

RAN ZHANG, MIN LIU, ZHIHAN PAN, AND YIFENG YIN
College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan 450001, China
Corresponding author: Ran Zhang (ranranzh@sina.com)

This work was supported in part by the Key Scientific Research Project for Higher Education Institutions of Henan Province under Grant 21B520021, in part by the Collaborative Innovation Project of Zhengzhou under Grant 2021ZDPY0106, and in part by the Natural Science Foundation of Henan Province under Grant 202300410508.

ABSTRACT Network security situation assessment is an important means of understanding the current network security situation to provide a basis for taking security measures. To address the problem that the accuracy of existing network security situation assessment methods needs to be improved, this paper proposes a network security situation assessment method based on support vector machine (SVM) optimized by whale optimization algorithm (WOA) that is improved by adaptive weight (AW) combined with simulated annealing algorithm (SA). In this method, the SVM is embedded into the fitness function calculation of the improved WOA, and the global optimization characteristics of WOA are used to determine the optimal penalty parameter c and kernel function parameter g of the SVM. To solve the problem of the WOA being prone to falling into local extremum and slow convergence when solving large and complex data problems, an adaptive weight is used to adjust the whale position update coefficient, and a simulated annealing algorithm (SA) is used to increase random search factors to avoid falling into local extremum, so as to improve the global optimization ability. The experimental results show that this method is feasible, can assess the network security situation more accurately, and has better convergence than other assessment algorithms based on an improved SVM.

INDEX TERMS Network security situation assessment, support vector machine, whale optimization algorithm, adaptive weight, simulated annealing algorithm.

I. INTRODUCTION

With the rapid development and application of technologies, such as the Internet, big data, cloud computing, and artificial intelligence, cyberspace security is facing increasing risks and threats. DDoS assaults, APT attacks, the increasing number of high-risk vulnerabilities, the frequent occurrence of data exposure events, the constant appearance of ''gray'' applications, and the security risks brought by high and new technologies are among the most prominent issues. At present, the security problems faced by the network system are mainly as follows: the amount of network security data involved is gradually increasing and becoming increasingly larger; the network security events are constantly fragmented, making it difficult to be perceived, and the obtained security information is scattered and disorderly, administrators need to spend a lot of time and energy to analyze the potential security threats, which is time-consuming and labor-consuming, and half the results with double the effort; and many existing network security systems have limitations in data collection, some of which are limited to one or several aspects of network security data collection, analysis and processing, which is difficult to describe and reflect the network security situation comprehensively. In the face of these new challenges and threats, existing traditional network security defense means, strategies and methods (such as intrusion detection systems, firewalls, anti-virus, access control, etc.) can no longer keep up with the actual security requirements of today's network systems.

Security situational awareness was first applied in the aviation and military fields, and then gradually extended to the network security field. In 1988, based on the concept

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Steven Li.

of situational awareness, Endsley summarized it into three levels, namely ''situation element extraction, situation understanding, and situation prediction,'' among which the core of ''situation understanding'' is ''situation assessment'' [1]. In 1999, Bass applied the mature theory and technology of air traffic supervision ATC situation awareness to cyber security situation awareness, and proposed the concept of cyberspace situation awareness for the first time [2], which laid a foundation for research on network security situation technology. Network security situational awareness refers to ''recognizing and understanding environmental factors within a certain time and space, and predicting future security development trends'' [3]. Based on Endsley's model, the concept of network security situation assessment emerged. The network security situation assessment integrates and analyzes the situation elements and data information extracted from the network, models and assesses the current network security situation, obtains the situation value according to the assessment model, dynamically reflects the current operation status of the network system and the severity of the threat overall, and predicts and forewarns its development trend, so as to provide decision support for network security management. Network security situation assessment has important research significance and application value as a new technology for the next-generation network security and an important component of a new network security defense system.

The main work of this paper is as follows:

1) A network security situation assessment method based on an improved WOA-SVM is proposed, which is optimized by adaptive weight (AW) and simulated annealing algorithm (SAA), which is short for AWSA-WOA-SVM. In this method, the improved SVM model is trained to generate security situation value to assess the current network security situation.

2) The adaptive weight and simulated annealing algorithm are introduced into WOA to overcome its local optimization and slow convergence, so as to improve the global optimization ability of WOA.

3) The improved WOA is used to determine the optimal penalty parameter c and kernel function parameter g of SVM to increase the accuracy of the assessment model based on SVM.

4) The experimental results show that the network security situation assessment method based on the WOA-SVM improved by the AW and SAA has smaller errors, higher accuracy and better convergence than the WOA-SVM algorithm and other improved WOA-SVM algorithms, which can more accurately and effectively assess the current network security situation.

II. NETWORK SECURITY SITUATION ASSESSMENT METHOD BASED ON ARTIFICIAL INTELLIGENCE

Network security situation assessment employs a series of mathematical models and algorithms to analyze the preprocessed original security data and events that are collected from the network and then to obtain quantitative or qualitative network security situation assessment results in the form of a security situation value to reflect the current network security situation, which is based on the established network security situation indicator system and certain prior knowledge. At present, network attacks are becoming increasingly diversified, complicated and random, and the network security situation is a complex and constantly changing nonlinear process. Therefore, the use of artificial intelligence technologies such as machine learning and deep learning is an inevitable development trend for assessing the network security situation. Building a highly accurate, scientific and objective network security situation assessment model is a research focus of network security situation assessment. Many artificial intelligence-based assessment methods for network security have been proposed. They are roughly classified into three types based on their theories: those based on mathematical model, those based on knowledge reasoning, and those based on pattern recognition.

A. BASED ON MATHEMATICAL MODELS

This method assesses the situation primarily by constructing an assessment function, and the key is the construction of the function. The most common method is the analytic hierarchy process (AHP). Reference [4] proposed a network security situation assessment model that uses an alarm verification algorithm in conjunction with a fuzzy inference algorithm to improve the analytic hierarchy process, effectively eliminating the impact of false alarm information and intuitively reflecting the network security situation; however, the data source of this method is relatively single. Reference [5] provided an assessment approach that integrates the AHP with the hierarchical model of context assessment, which simplifies the situation assessment problem and may reflect the entire security condition of the network and better serve high-level decision-making. The disadvantage of the model based on a mathematical model is that there is currently no objective and unified standard for function construction and it is prone to being influenced by subjective human factors, resulting in inaccurate evaluation results.

B. BASED ON KNOWLEDGE REASONING

This method primarily constructs models based on specific criteria and empirical knowledge, and applies logical reasoning theory to evaluate it. The evidence theory and graph models are two of the most representative examples. Based on the evidence theory, for example, [6] proposed a network security threat situation assessment method based on unsupervised generation reasoning, which solves the shortcomings of high computational cost, time consuming and low efficiency of the supervised assessment method, and can more intuitively assess the overall situation of network threats; [7] studied a network security situation assessment model based on DS evidence theory, which used principal component analysis (PCA) to preprocess the alarm data, adopted the improved DS evidence theory and combined the credibility of multi-source attack data to improve the alarm recognition rate. Based

on the graph model, for example, [8] proposed a situation assessment method using the Seeker Optimization Algorithm to improve the hidden Markov model, which can more accurately assess the situation of network security, but there were irrelevant and false positive data in situation elements, which need further research on observation sequence; [9] proposed a network security situation assessment method with Markov game model as the core and combined with four-level data fusion, which considered the interaction between attackers and defenders, so it was closer to reality and can assess the network security situation more accurately. The disadvantage of the methods based on knowledge reasoning is that when encountering a large number of data conflicts, the amount of calculation and complexity is relatively large, and it is difficult to guarantee the independence of each assessment indicator, and sometimes it depends on expert experience, which may lead to low accuracy of the assessment.

C. BASED ON PATTERN RECOGNITION

This method is mostly based on machine learning theory for assessment, which is also the method studied in this paper. Its main theories include rough set, neural network, support vector machine and so on. For example, [10] proposed a situation assessment model of CS-BP neural network optimized by DS evidence theory, which introduced conjugate gradient algorithm into CS algorithm, to improve the local search ability of CS algorithm, overcome the local minimum problem and reduce the subjectivity of BPA, and improve the accuracy of situation assessment; [11] proposed a network security situation assessment model based on the optimization of SVM parameters based on the gravitational search algorithm (GSA), which searches the best parameters in SVM through GSA to minimize the error between the generated data and the actual network security situation assessment data. The disadvantage of pattern recognition method is that the learning efficiency becomes low, the optimal parameters are difficult to determine, and the adaptive ability is poor when encountering large-scale and complex data.

At present, the network security situation assessment method has not been standardized, and there are some flaws, such as inaccurate assessment results and poor adaptability. Aiming at the problems of low accuracy and slow convergence in the above assessment methods, this paper proposes a network security situation assessment method that introduces adaptive weight (AW) and simulated annealing algorithm (SA) into whale optimization algorithm (WOA) to optimize support vector machine (SVM), and compares this algorithm with other intelligent algorithms optimized SVM methods for network security situation assessment.

III. NETWORK SECURITY SITUATION ASSESSMENT BASED ON AWSA-WOA-SVM

This paper proposes a network security situation assessment model based on AWSA-WOA-SVM. In this model, WOA is used to determine the optimal parameters c and g of SVM, AW is used to adjust the whale position update coefficient, and SA is used to increase the random search factor to improve its global optimization ability. This assessment model can be divided into three stages, which are shown in FIGURE 1.

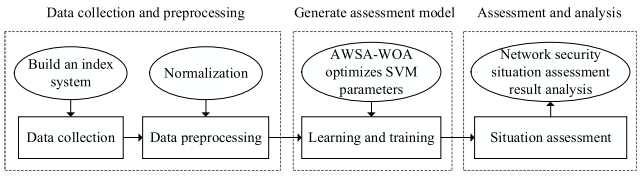

FIGURE 1. Network security situation assessment model based on the AWSA-WOA-SVM.

A. INPUT DATA

The experimental sample data are extracted from the network, the situational indicators are extracted according to the situational indicator extraction principles and the situational indicator system is constructed. Then the extracted sample data are preprocessed. Finally, the obtained data set is divided into two parts: training set and test set.

B. SITUATION ASSESSMENT

Two algorithms of AW and SA are used to improve the WOA and the improved WOA algorithm (AWSA-WOA) is used to perform an optimization search operation on the penalty parameter c and the kernel function parameter g of the SVM to determine the optimal combination of parameters. Then the optimal combination of parameters is assigned to the model.

C. OUTPUT RESULTS

The preprocessed test sample data is input to the final obtained SVM assessment model and the situation assessment results of the test samples are output.

IV. CONSTRUCTION OF THE NETWORK SECURITY SITUATION INDEX SYSTEM

The extraction of situation indicators is a prerequisite for situation assessment, and the construction of the situation index system serves as a foundation for the extraction of situation elements. We must follow specific principles while extracting situation indicators so that we can scientifically and rationally construct a multi-directional and multi-angle reflection of the network's security condition.

A. SELECTION PRINCIPLES OF SITUATION ASSESSMENT INDICATORS

The construction of the index system is a very complicated process, but it is also a key link in situation assessment and prediction. Therefore, it is vital to construct a scientific and logical index system. If the number of indicators selected is large, the complexity and workload of the situation assessment system will grow, resulting in a decline in the system's efficiency and speed. On the contrary, if the number of selected indicators is small, it cannot fully reflect the security status of the whole network. Therefore, before

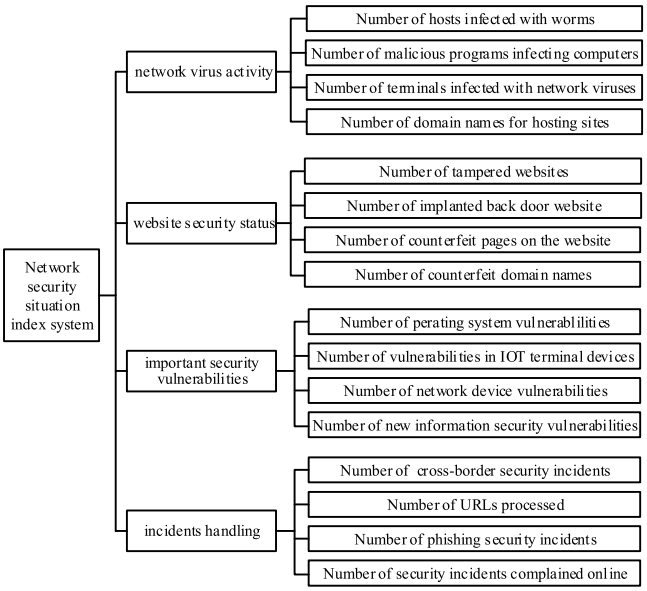

FIGURE 2. Index system of network security situation.

establishing the index system, the corresponding index selection principles should be determined to make the assessment results more accurate, efficient, and comprehensive. The following are some principles for the construction of situation indicators [23].

1) THE PRINCIPLE OF HIERARCHY

Because the selected indicators have varied effects on the network, the hierarchy principle should be followed when building the indicator system, so that the indicators of each layer have different functions, and the associated quantification and decentralization should be carried out.

2) THE PRINCIPLE OF PROXIMITY

Because the selected indicators affect all aspects of the network, they must have similar influence capability, and there must be a certain connection between the indicators. As a result, these influencing factors should also be comprehensively considered and coordinated.

3) THE PRINCIPLE OF INTEGRITY

The selected indicators should be scientific, complete and independent, and can objectively reflect the relationship between objects and indicators, but also can comprehensively reflect the security status of the whole network.

B. CONSTRUCTION OF ASSESSMENT INDEX SYSTEM

Based on the principles of the above situation indicator construction, combined with the complexity, heterogeneity, diversity, uncertainty, etc. of the existing network, this paper establishes an indicator system of situation assessment, which can be divided into four categories: indicators based on network virus activity, indicators based on website security status, indicators based on important security vulnerabilities, and indicators based on incident handling. Although there are only 4 categories, each category can be further refined to reflect different network environments states [24]. The four categories of indicators can be divided in detail, as shown in FIGURE 2.

V. ASSESSMENT ALGORITHM BASED ON AWSA-WOA-SVM

A. WHALE OPTIMIZATION ALGORITHM

Whale Optimization Algorithm (WOA) is a new type of metaheuristic intelligent optimization algorithm proposed by Mirjalili and Lewis [12]. It searches for the best solution by mimicking the "spiral bubble network" search strategy of humpback whales. The algorithm has the advantages of few adjustment parameters, simple operation and easy understanding [13]. The WOA mainly involves the following optimization steps: surround prey, spiral predation, and search for prey [14].

1) SURROUND PREY

Humpback whales surround their prey when hunting. After the humpback whale has selected the optimal position, other whales will approach this position, and its position update in the iterative optimization process is represented by the following formulas (1) and (2).

$$D = |CM^*(t) - M(t)| \tag{1}$$

$$M(t+1) = M^*(t) - A \cdot D \tag{2}$$

where $M^*(t)$ is the position vector of the optimal solution at the t-th iteration, and $M^*(t)$ will be updated accordingly when a better solution appears in the iteration process; $M(t)$ is the position vector of the solution at the t-th iteration; D is the iterative distance between the optimal solution position and the current solution in the t-th iteration; the coefficient vectors A and C are determined by formulas (3) and (4).

$$A = 2a \cdot r_1 - a \tag{3}$$

$$C = 2r_2 \tag{4}$$

Among them, a is a constant that linearly drops from 2 to 0 in the iterative process, which can be expressed as $a = 2t/maxgen$, $maxgen$ is the maximum number of iterations; r_1 and r_2 are random vectors in [0,1].

2) SPIRAL PREDATION

When the whale (searcher) gradually approaches the prey (the optimal solution), the variable a will decrease accordingly, and the coefficient A will also decrease linearly with the variable a according to formula (3). When A is $[-1, 1]$, the next position of the new whale (searcher) can be any position between the current position and the optimal position (optimum solution). The whale (searcher) will attack the prey (optimal solution) in a spiral way, update the position according to formula (5), and gradually approach the position of the prey.

$$M(t+1) = D' \cdot e^{bl} cos(2\pi l) + M^*(t) \tag{5}$$

where, $D' = |M^*(t) - M(t)|$ represents the position distance between the i-th searcher and the current optimal solution; b is a constant that is used to define the shape of logarithmic spiral, and l is a random number within $[-1,1]$.

Humpback whales not only move in a spiral manner, but also constantly narrow the search range. Therefore, assuming a 50% probability of switching between the contraction surrounding mechanism and the spiral model, the whale position is updated according to formulas (6) and (7).

$$M(t+1) = M^*(t) - A \cdot D \tag{6}$$
$$M(t+1) = D' \cdot e^{bl} cos(2\pi l) + M^*(t) \tag{7}$$

where, $D' = |M^*(t) - M(t)|$ represents the position distance between the i-th searcher and the current optimal solution.

3) SEARCH FOR PREY

Humpback whales will also randomly search for prey, which can improve the algorithm's global search capabilities. The update formula for randomly searching for prey whale position is shown in (8).

$$M(t+1) = M_{rand}(t) - A\,|CM_{rand}(t) - M(t)| \tag{8}$$

where, M_{rand} is a position vector randomly selected from the current population (representing a random whale).

B. IMPROVED WHALE OPTIMIZATION ALGORITHM

From the above optimization search process, it can be seen that the whale optimization algorithm does not involve the steps of jumping out of the local optimum, so it is prone to fall into local extremes and slow convergence when solving problems with relatively large amounts of data or complex problems. The literature [15] uses a chaotic strategy to update the whale position, and the literature [16] uses adaptive weights and optimal neighborhood perturbation to guide the position update. In this paper, the adaptive weights combined with a simulated annealing algorithm will be used to improve the whale optimization algorithm.

1) IMPROVING WHALE OPTIMIZATION ALGORITHM WITH ADAPTIVE WEIGHTS

The adaptive weight (AW) coefficient can not only affect the local search ability of the algorithm, but also affect the global search ability of the algorithm. For example, a relatively large weight coefficient helps the algorithm to jump out of the trap of local optimization, thereby improving the global search ability of the algorithm, while a relatively small weight coefficient is beneficial to the accurate search of local search space, which it can improve the local search ability of the algorithm and the convergence of the algorithm [17]. This paper uses the characteristics of the adaptive weight method to update the whale position, and its calculation formula is shown in (9).

$$\mathrm{k}(t) = 0.2\cos(\frac{\pi}{2} \times (1 - \frac{t}{maxgen})) \tag{9}$$

where, t is the number of iterations, and the maximum number of iterations $maxgen = 100$.

The adaptive weight coefficient $k(t)$ is substituted into formulas (6), (7) and (8), the position update formulas of the improved WOA algorithm are shown in (10), (11) and (12).

$$M(t+1) = k(t) * M^*(t) - A \cdot D \tag{10}$$
$$M(t+1) = k\,(t) * D' \cdot e^{bl} cos(2\pi l) + M^*(t) \tag{11}$$
$$M(t+1) = \mathrm{k(t)} * M_{rand}(t) - A\,|CM_{rand}(t) - M(t)| \tag{12}$$

According to the above equation, AW is used to change the update speed of the algorithm. Here, AW adopts a trigonometric function, so it is periodic. When WOA is in the initial stage of iterative optimization, the inertia weight coefficient $k(t)$ is at the minimum value. At this time, the position update speed increases slowly, which improves the local search ability of the algorithm. As the number of iterations increases, the value of $k(t)$ gradually increases, and the update speed of the algorithm gradually increases, so as to enhance the global search ability of the algorithm to a certain extent. Therefore, adjusting the position update speed of the algorithm through the weight coefficient AW can well balance the global and local search ability of the algorithm.

2) USING ADAPTIVE WEIGHTS AND SIMULATED ANNEALING ALGORITHM TO IMPROVE WHALE OPTIMIZATION ALGORITHM

Simulated annealing algorithm (SA) is a global search algorithm extended from the local search algorithm [18], which has the innate advantage of global search. In order to further strengthen the global search capability of WOA, the SA algorithm is introduced in WOA so as to effectively avoid WOA from falling into the trap of local optimum.

In the WOA iterative optimization, WOA is used to determine the individual optimal solution and the global optimal solution, but if the optimal position of the population is at a local extreme, the obtained optimal fitness value will also tend to the local minimum, which will degrade the global search performance of the algorithm. Therefore, in order to avoid falling into the local extremum, the principle of SA is introduced, that is, the sudden jump probability is adopted, and the bad solution is accepted with the probability P_i to help WOA jump out of the local optimum. The probability P_i is determined according to formulas (13) and (14).

$$P_i = \begin{cases} 1 & df < 0 \\ \exp(-\frac{df}{T_t}) & df \geq 0 \end{cases} \tag{13}$$
$$df = fitness(i) - fitnesszbest \tag{14}$$

where, $fitness(i)$ is the current whale fitness value, and $fitnesszbest$ is the global best whale fitness value; df is the difference between them, which is determined according to the positive or negative of the difference. If $df < 0$, it means that the new whale position is better than the optimal position of the original population, so the new position will replace the optimal position of the original population; if $df \geq 0$, it indicates that the new whale position is inferior to the

optimal position of the original population, so a probability $\exp(-df/T_t)$ corresponding to the current temperature T_t will be generated to determine whether to accept the new solution. Among them, the initialization and rate of change of temperature T_t are as formulas (15) and (16).

$$T_t = -\frac{fitnesszbest}{\log(\alpha)} \tag{15}$$

$$T_{t+1} = \mu T_t \ (t \geq 0, 0 \leq \mu \leq 1) \tag{16}$$

where, *fitnesszbest* is the global best whale fitness value, α and μ are the control parameters, t is the number of iterations, $t \geq 0, 0 \leq \mu \leq 1$ and $\alpha \in [0.2, 0.5]$.

C. USING IMPROVED WHALE OPTIMIZATION ALGORITHM TO OPTIMIZE SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a new machine learning method proposed by professors Corinna Cortes and Vapnik in 1995. It is a generalized linear classifier that performs binary classification of data in a supervised learning method. [19]. Compared with traditional neural networks, SVM has advantages in versatility, robust-ness, effectiveness, computational simplicity and theoretical support [20]. However, when SVM is used for pattern recognition or regression prediction, there is no uniform standard for the selection of its parameters c and g.

1) PENALTY PARAMETER C

In the process of using SVM to solve the fitting function, a minimum optimization problem is involved, and the penalty parameter c is introduced. The parameter c is used to adjust the objective function to find the balance between the maximum interval and the minimum relaxation factor, that is, when the sample data is classified incorrectly, the larger the value of the parameter c, the more complex the algorithm will be, thus classification errors will not occur. However, if the parameter c is set too high, the algorithm's generalization ability will be weakened, and the empirical risk may not change. On the contrary, a smaller value of parameter c reduces the complexity of the algorithm, but increases the algorithm's empirical risk. As a result, it is necessary to use intelligent optimization algorithm to find a suitable parameter c, so that the support vector machine can perform better [21].

2) KERNEL FUNCTION PARAMETER G

In this article, the radial basis function (also known as the Gaussian kernel function) is selected.

$$K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}), \quad \sigma > 0 \tag{17}$$

σ is the bandwidth of the Gaussian kernel function. When σ is infinitely close to 0, all sample data will become support vectors. At this time, the classification is also very accurate, but it causes the overfitting phenomenon, and thus the algorithm lacks generalization (that is, it cannot accurately classify new sample data). Conversely, when σ is infinitely close to infinity, all of the sample data will be classified into the same category, which has no experimental significance [22].

D. ASSESSMENT ALGORITHM BASED ON THE SOA-SVM OPTIMIZED BY ADAPTIVE WEIGHTS AND SIMULATED ANNEALING ALGORITHM

At the moment, the most popular SVM optimal parameter selection methods include experience, experimental comparison, grid search and large-scale search. These methods have their own set of benefits and drawbacks. In this paper, the WOA algorithm optimized by adaptive weight and simulated annealing algorithm is used to determine the optimal parameters of SVM for network security situation assessment. The main steps of the assessment algorithm based on the WOA-SVM optimized by adaptive weight and simulated annealing algorithm (AWSA-WOA-SVM for short) are as follows.

Step1. Collect existing network security data as sample data to establish a sample set, then normalize and preprocess the sample data in the sample set and divide it into training set and test set.

Step2. Initialize the population size of the WOA algorithm, upper and lower limit of whale position and its initial position, and initialize annealing temperature, cooling rate and sudden jump probability of the SA algorithm. Initialize the iteration number t=*1*, maximum iteration number maxgen.

Step3. Calculate fitness value. Calculate the individual and group extreme positions and the corresponding best fitness value of the whale. This algorithm takes the mean square error (MSE) between the assessment value obtained by the SVM model and the true value as the fitness function, so the smaller the fitness value, the higher the accuracy.

Step4. When $t \leq maxgen$, update the values of parameters a, r_1, r_2, A, C, b, l, p and k.

Step5. Iterative optimization. When $p < 0.5$, if $|A| \geq 1$, the individual position in the current whale population is updated according to formula (12), and $M_{rand}$is randomly selected from the current whale population; If $|A| < 1$, the spatial position of the individual in current whale population is updated according to formula (10). When $p \geq 0.5$, the spatial position of the current whale individual is updated according to formula (11). Finally, the optimal position and global optimal position of the whale and their fitness values are updated.

Step6. Introduction of SA algorithm. Select a searcher in the neighborhood of the global optimal fitness value and calculate the difference value df according to formula (14). If $df < 0$, the new whale position replaces the original position; if $df \geq 0$, use the probability $\exp(-df/T_t)$ to determine whether to accept the position of the inferior solution, and then update the whale optimal position *gbest* and the global optimal position *zbest* and save.

Step7. Cooling treatment. The temperature is controlled according to formula (16).

Step8. Determine whether the termination condition of the loop is met, that is, whether the maximum number

of iterations is met or the error requirement is met. If it can be met, get the optimal individual *zbest* and assign it to the parameters *bestc* and *bestg* of the SVM; otherwise, skip to Step 3 and repeat the iterative optimization process again.

Step9. The optimal population position $M^*(t)$ obtained by the improved WOA algorithm is assigned to the parameters *bestg* and *bestc* of the SVM respectively. The obtained parameters are substituted into the SVM model to get the final support vector machine model.

Step10. The test data is put into the improved SVM model obtained by training to obtain the assessment value, and the current network situation is analyzed combined with the network security situation assessment level table.

The workflow of parameter optimization algorithm based on AWSA-WOA-SVM for assessment is shown in FIGURE 3.

VI. EXPERIMENTAL SIMULATION AND ANALYSIS

In this experiment, the number of hosts infected with virus, the total number of tampered websites, the total number of websites implanted with backdoors, the number of counterfeit pages of domestic websites and the number of new information security vulnerabilities are selected as the assessment indicators from the four categories of indicators listed above, which comprehensively reflects the threat situation faced by the modern network. The experimental data comes from the network security data published in the third issue of 2016 to the 37th issue of 2020 of the weekly report on network security information and dynamics released by the National Internet Emergency Center, among which 270 data sets are selected as the training set and 10 data sets as the test set. The situation is divided into five levels of excellent, good, medium, poor, and critical, which are respectively expressed by different situation values, as shown in TABLE 1. The experiment is simulated by MATLAB R2019a. The operating system used is Windows 10, the CPU is 1.80GHz, and the memory is 8GB.

To better illustrate the relationship between situation indicators and situation values, TABLE 2 gives the sample data in the weekly reports for periods 1-12 in 2021.

A. DATA PREPROCESSING

Data normalization is helpful to improve the accuracy of situation assessment. This experiment normalizes the sample data according to formula (18).

$$f : x \to y = 2 * \frac{x - x_{min}}{x_{max} - x_{min}} + (-1) \tag{18}$$

where, $x, y \in R^n$, x_{min} is the minimum data in the sample set, and x_{max} is the maximum data in the sample set. The above formula can normalize the sample data to $[-1,1]$, which is important for the network security situation. The result of normalization is shown in FIGURE 4.

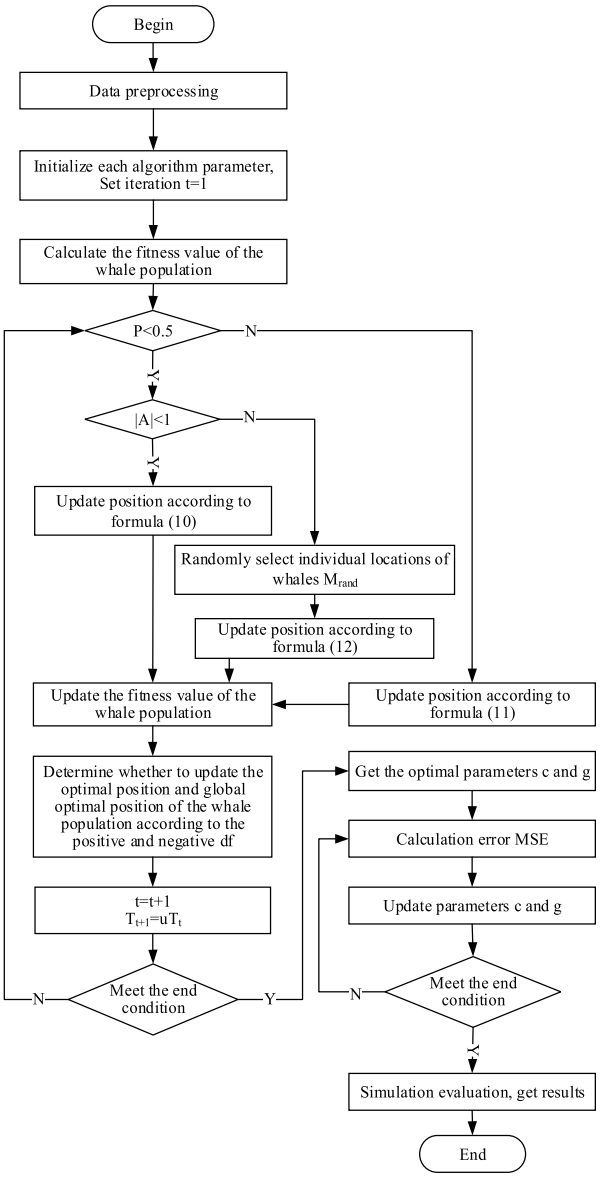

FIGURE 3. Flow chart of SVM parameter optimization algorithm based on the AWSA-WOA for assessment.

TABLE 1. Situation value conversion table.

Excellent	Good	Normal	Bad	Dangerous
5	4	3	2	1

B. ASSESSMENT AND RESULT ANALYSIS

1) ASSESSMENT AND COMPARISON OF RESULTS

FIGURE 5 shows the comparison of the experimental results of the network security situation assessment based on the support vector machine (SVM) optimized by the four algorithms of WOA, AW-WOA, SA-WOA, and AWSA-WOA. It can be seen from FIGURE 5 that the broken line of the situation assessment value obtained by optimizing the SVM based on

TABLE 2. Data samples for periods 1-12 of 2021.

Sample number	actual value	Number of hosts infected with viruses	Number of tampered websites	Number of implanted backdoors	Number of counterfeit websites	Number of new security vulnerabilities
1	4	4335.6	1158	1871	7431	295
2	4	4009.3	3208	1484	4185	268
3	4	3525.1	3657	772	1742	566
4	4	3260.2	3681	855	987	378
5	3	2350.7	4218	1110	303	355
6	4	2950.4	3735	1046	185	457
7	4	1770.2	3407	1294	53	260
8	4	3342.1	3451	1254	78	166
9	4	2406.3	4357	518	226	502
10	4	4073.4	3957	590	243	488
11	4	1721.6	3454	537	2368	617
12	4	1115.6	1840	522	385	679

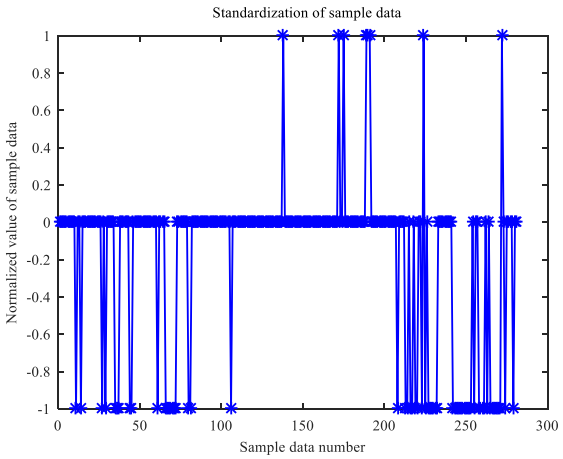

FIGURE 4. Standardization of sample data.

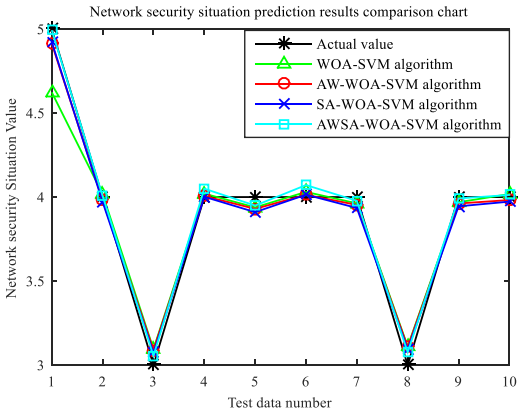

FIGURE 5. Comparison of assessment results based on several different algorithms.

the WOA algorithm is the most volatile and the most unstable among the four algorithms. The situation assessment values obtained by the AW-WOA optimized SVM and SA-WOA optimized SVM are relatively close, their fluctuation and the difference from the assessment value of the emergency center are similar, and they are generally more stable and closer to the actual value polyline than the unimproved WOA optimized SVM algorithm. However, the broken line of the situation assessment value obtained by the AWSA-WOA optimized SVM algorithm is more stable, more consistent with the true value broken line, and less error than the above three algorithms.

TABLE 3 lists the situation value of the WOA-SVM improved by different algorithms for network security situation assessment, and the absolute error between them and the true situation value. TABLE 3 shows that the error of the WOA-SVM algorithm improved by adaptive weight and the simulated annealing algorithm is smaller and its value is closer to the real value than the other algorithms optimized WOA-SVM algorithm, indicating that the AWSA-WOA-SVM algorithm is more accurate for the network security situation assessment.

2) JUDGE ACCURACY AND EFFECTIVENESS

The difference between the true value and situation assessment value are measured using three performance indicators: mean square error (MSE), mean absolute percentage error (MAPE) and mean absolute error (MAE).

MSE indicators:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(x_t - x_t^*)^2 \tag{19}$$

MAPE indicators:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - x_t^*}{x_t}\right| \times 100\% \tag{20}$$

MAE indicators:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|x_t - x_t^*\right| \tag{21}$$

where, x_t and x_t^* represent the true value and situation assessment value respectively, and n represents the number of test data sets.

TABLE 4 shows the mean square error, mean absolute percentage error, and mean absolute error calculated between the situation value obtained from the WOA-SVM network

TABLE 3. Comparison of assessment absolute errors of several different algorithms.

actual value	WOA-SVM evaluated value	WOA-SVM absolute error	AW-WOA-SVM evaluated value	AW-WOA-SVM absolute error	SA-WOA-SVM evaluated value	SA-WOA-SVM Absolute error	AWSA-WOA-SVM evaluated value	AWSA-WOA-SVM absolute error
5	4.626	-0.374	4.914	-0.086	4.923	-0.077	4.993	-0.007
4	4.018	0.018	3.980	-0.020	3.971	-0.029	4.006	0.006
3	3.114	0.114	3.089	0.089	3.074	0.074	3.055	0.055
4	4.014	0.014	4.006	0.006	4.001	0.001	4.050	0.050
4	3.943	-0.057	3.930	-0.070	3.911	-0.089	3.950	-0.050
4	4.029	0.029	4.013	0.013	4.014	0.014	4.071	0.071
4	3.964	-0.036	3.953	-0.047	3.935	-0.065	3.977	-0.023
3	3.128	0.128	3.106	0.106	3.093	0.093	3.077	0.077
4	3.970	0.003	3.962	-0.038	3.945	-0.055	3.992	-0.008
4	4.017	0.017	3.982	-0.018	3.973	-0.027	4.013	0.013

security situation assessment before and after improvement and the real value. As shown in TABLE 4, the MSE, MAPE and MAE values obtained by the net-work security situation assessment algorithm based on AWSA-WOA-SVM are the smallest when compared with those obtained by other improved whale optimization algorithms. This also shows from a macro perspective that the network security situation assessment algorithm based on AWSA-WOA-SVM has higher accuracy and effectiveness.

To show the accuracy of each optimization algorithm more intuitively, the evaluation indicators of different optimization algorithms are shown in the form of histograms in FIGURE 6.

It can be clearly seen from FIGURE 6 that the values of the three assessment indicators of the algorithm proposed in this study are the smallest among these algorithms, which means that the accuracy and superiority of the algorithm proposed in this paper are the highest.

3) CONVERGENCE ANALYSIS

Convergence is a key issue related to whether the algorithm can be implemented. FIGURE 6 shows the changes in the optimal individual fitness value of each optimization algorithm in the iterative optimization process. It can be seen from FIGURE 7 that the individual fitness value searched by the WOA-SVM algorithm at the beginning is relatively large, it begins to fall into local optimization at the fourth iteration, and the fitness value converges to 0.992742, which is the largest in the four algorithms. The optimal fitness value of the AW-WOA-SVM algorithm at the beginning of the iteration is also relatively large, it falls into the local optimization at the third iteration and the fitness value converges to 0.990664. For the SA-WOA-SVM algorithm, its individual fitness value is relatively small compared to the fitness value of the AW-WOA-SVM algorithm at the beginning, but it also falls into the local optimum at the 3rd iteration and fluctuates a little at the 8th iteration, and finally the fitness value of the algorithm converged to 0.990770, which is not much different from the fitness of AW-WOA-SVM. Compared with the above three optimization algorithms, the individual optimal fitness value of the AWSA-WOA-SVM algorithm is the smallest at the beginning, and it changes on the 15th, 19th, 58th, and 65th times. In the end, the optimal fitness value of the algorithm converges to 0.990142. Since this paper uses the mean square error MSE of the security situation assessment value and the true value as the fitness function, the smaller the individual fitness value, the higher the accuracy. Although the WOA-SVM algorithm converges to the local optimal fitness value at the 9th iteration, the AW-WOA-SVM algorithm and the SA-WOA-SVM algorithm converge to the local optimal fitness value at the 3rd iteration, it is obvious that their fitness value is not the smallest. To sum up, the AWSA-WOA-SVM algorithm can achieve smaller individual fitness values faster than other algorithms, which shows that the algorithm has higher accuracy and superiority.

TABLE 4. Comparison table of accuracy.

Evaluation index	WOA-SVM	SA-WOA-SVW	AW-WOA-SVM	AWSA-WOA-SVM
MSE	0.018	0.0037	0.0036	0.0019
MAPE	2.050%	1.414%	1.355%	1.008%
MAE	0.0818	0.0526	0.0494	0.036

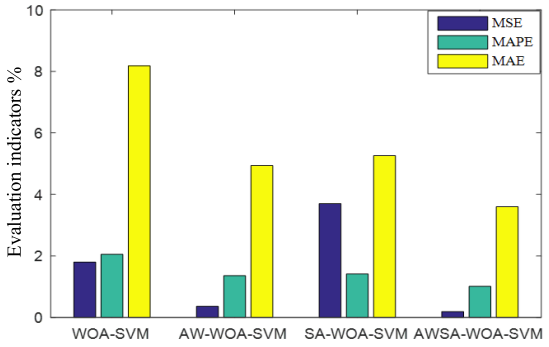

FIGURE 6. Comparison of assessment index values of the four different algorithms.

4) COMPLEXITY ANALYSIS

The complexity of an algorithm is an important indicator for evaluating its performance, which includes time complexity and space complexity. The time complexity of an algorithm represents the total time required to complete it.

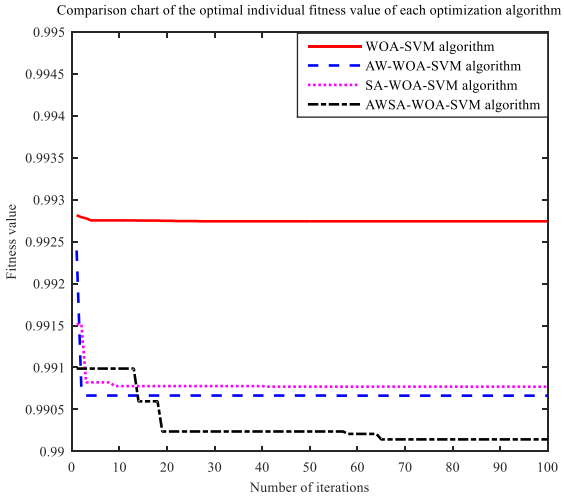

FIGURE 7. Curve of optimal fitness value of the four optimization algorithms.

Generally, the number of basic operations of the algorithm is considered as a measure of time complexity. The population size of the whale optimization algorithm designed in this study is *sizepop*, the maximum number of iterations is *maxgen*, and the dimension of the problem is *dim*, so the time complexity of the WOA is $O(sizepop * maxgen * dim)$. The AW-WOA-SVM, SA-WOA-SVM and AWSA-WOA-SVM algorithms are all completed in the WOA iteration cycle without additional cycles, thus their time complexity is $O(sizepop * maxgen * dim)$. Therefore, the time complexity of the WOA-SVM algorithm before and after the improvement is the same and belongs to the same level. The space complexity of an algorithm is a measure of the amount of storage space temporarily occupied by the algorithm during its execution. The space complexity of an algorithm is defined as the storage space consumed by the algorithm, which is also a function of problem scale n. In this study, the *sizepop* of the whale optimization algorithm and the problem dimension *dim* determine its space complexity. Because the scale and dimensions of all optimization algorithms have not changed, they have the same space complexity $O(sizepop * dim)$.

VII. CONCLUSION

This paper proposes a network security situation assessment method based on an improved WOA-SVM optimized by an adaptive weight and simulated annealing algorithm, which solves the problems that the penalty parameter c and kernel function parameter g of SVM are difficult to select and WOA tends to fall into local extremum and converges slowly. This method introduces the adaptive weight and simulated annealing algorithm into the WOA, uses an adaptive weight algorithm to enhance its local optimization ability, and uses a simulated annealing algorithm to improve its global optimization and convergence ability. The improved WOA makes the selection of SVM parameters more accurate, and makes the assessment result more in line with the actual situation and more effectively reflect the current network security situation. The comparative experimental results show that the assessment result of the AWSA-WOA-SVM algorithm can more accurately reflect the current network security situation, and has better stability and convergence. In the future, other intelligent assessment algorithms will be studied to determine more accurate and efficient network security situation assessment methods.

REFERENCES

[1] M. R. Endsley, "Situation awareness global assessment technique (SAGAT)," in *Proc. IEEE Nat. Aerosp. Electron. Conf.*, vol. 3, May 1988, pp. 789–795, doi: 10.1109/NAECON.1988.195097.
[2] T. Bass and D. Gruber, "A glimpse into the future of ID," *Mag. USENIX SAGE*, vol. 3, no. 24, pp. 40–45, 1999.
[3] J. Gong, X. Zang, Q. Su, X. Hu, and J. Xu, "Network security situation awareness review," *J. Softw.*, vol. 4, no. 28, pp. 1010–1026, 2017.
[4] M. Cui, H. Feng, B. Liu, and L. Lin, "Simulation of an improved hierarchical network security situation assessment model," *Comput. Simul.*, vol. 11, no. 36, pp. 284–289, 2019.
[5] H. Wang, Z. Chen, X. Feng, X. Di, D. Liu, J. Zhao, and X. Sui, "Research on network security situation assessment and quantification method based on analytic hierarchy process," *Wireless Pers. Commun.*, vol. 102, no. 2, pp. 1401–1420, 2018.
[6] H. Yang, F. Wang, and W. Lv, "Network security threat assessment method based on unsupervised generation reasoning," *J. Tsinghua Univ., Sci. Technol.*, vol. 6, no. 60, pp. 474–484, 2020.
[7] Z. Zhao, T. Zhou, and H. Wang, "Quantitative evaluation model of network security situation based on D-S evidence theory," in *Proc. 6th Int. Conf. Dependable Syst. Their Appl. (DSA)*, Harbin, China, Jan. 2020, pp. 371–376, doi: 10.1109/DSA.2019.00057.
[8] X. Li and Y. Duan, "Network security situation assessment method based on improved hidden Markov model," *Comput. Sci.*, vol. 47, no. 7, pp. 287–291, 2020.
[9] X. Li, Y. Lu, S. Liu, and W. Nie, "Network security situation assessment method based on Markov game model," *KSII Trans. Internet Inf. Syst.*, vol. 12, no. 5, pp. 2414–2428, May 2018, doi: 10.3837/tiis.2018.05.027.
[10] J. Qiang, F. Wang, and X.-L. Dang, "Network security based on D-S evidence theory optimizing CS-BP neural network situation assessment," in *Proc. 5th IEEE Int. Conf. Cyber Secur. Cloud Comput. (CSCloud), 4th IEEE Int. Conf. Edge Comput. Scalable Cloud (EdgeCom)*, Shanghai, China, Jun. 2018, pp. 153–159, doi: 10.1109/CSCloud/EdgeCom.2018.00035.
[11] Y. Chen, X. Yin, and A. Sun, "Network security situation assessment model based on GSA-SVM," in *Proc. Int. Conf. Comput., Commun. Netw. Technol. (CCNT)*, vol. 2, Jun. 2018, pp. 161–167.
[12] S. Minrjlalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.
[13] Z. Yan, J. Zhang, J. Zeng, and J. Tang, "Nature-inspired approach: An enhanced whale optimization algorithm for global optimization," *Math. Comput. Simul.*, vol. 185, pp. 17–46, Jul. 2021.
[14] M. Huang, X. Zhan, and X. Liang, "Improvement of whale algorithm and application," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Dalian, China, Oct. 2019, pp. 6–8, doi: 10.1109/ICCSNT47585.2019.8962426.
[15] L. Liu, K. Bai, Z. Dan, S. Zhang, and Z. Liu, "A whale optimization algorithm based on global search strategy," *Small Microcomput. Syst.*, vol. 9, no. 41, pp. 1820–1825, 2020.
[16] D. Zhu, H. Chen, and X. Wang, "Whale optimization algorithm based on adaptive weighting and simulated annealing," *Acta Electronica Sinica*, vol. 5, no. 47, pp. 992–999, 2019.
[17] S. S. Reddy and M. S. G. Prasad, "Improved whale optimization algorithm and convolutional neural network based cooperative spectrum sensing in cognitive radio networks," *Inf. Secur. J., A Global Perspective*, vol. 30, no. 3, pp. 160–172, May 2021, doi: 10.1080/19393555.2020.1825882.
[18] A. Roshani and D. Giglio, "Simulated annealing algorithms for the multi-manned assembly line balancing problem: Minimising cycle time," *Int. J. Prod. Res.*, vol. 55, no. 10, pp. 2731–2751, May 2017, doi: 10.1080/00207543.2016.1181286.
[19] C. Cortes and V. Vapnik, "Support vector network," *Mach. Learn.*, vol. 3, pp. 273–297, Sep. 1995.

[20] S. Ding, Z. Zhu, and X. Zhang, "An overview on semi-supervised support vector machine," *Neural Comput. Appl.*, vol. 28, no. 5, pp. 969–978, May 2017, doi: 10.1007/s00521-015-2113-7.
[21] E. Tuba and Z. Stanimirovic, "Elephant herding optimization algorithm for support vector machine parameters tuning," in *Proc. 9th Int. Conf. Electron., Comput. Artif. Intell. (ECAI)*, Jun. 2017, pp. 1–4, doi: 10.1109/ECAI.2017.8166464.
[22] W. M. P. D. Haryoko, "Optimization of parameter support vector machine (SVM) using genetic algorithm to review Go-Jek's services," in *Proc. 4th Int. Conf. Inf. Technol., Inf. Syst. Elect. Eng. (ICITISEE)*, Nov. 2019, pp. 301–304, doi: 10.1109/ICITISEE48480.2019.9003894.
[23] G. Wu, L. Chen, Z. Si, and L. Bai, "Research on optimization model of network security situation assessment index system," *Comput. Eng. Sci.*, vol. 5, no. 39, pp. 861–869, 2017.
[24] Y. Sun, L. Chen, L. Yin, Y. Guo, X. Meng, and P. Zhang, "Construction of situation assessment indicator system based on latitude and longitude lines of information security," in *Proc. IEEE 4th Int. Conf. Data Sci. Cyberspace (DSC)*, Hangzhou, China, Jun. 2019, pp. 100–105, doi: 10.1109/DSC.2019.00023.

RAN ZHANG was born in Henan, China. She received the Ph.D. degree in computer system architecture from Xi'an Jiaotong University, China, in 2003, and the Ph.D. degree in engineering.

She taught at the Beijing University of Technology, from 2004 to 2010, and she has been an Associate Professor with the Zhengzhou University of Light Industry, since 2011. She has published more than 50 papers in these fields, obtained two invention patents, and compiled an information security planning textbook. Her research interests include network information security, intrusion detection, cloud computing, network security situational awareness, and artificial intelligence security.

Dr. Zhang is a member of China Computer Federation and Chinese Association for Artificial Intelligence. She won the Science and Technology Progress Award of Henan Province in 2014.

MIN LIU was born in Henan, China, in 1995. She received the master's degree in computer science and technology from the Zhengzhou University of Light Industry, in June 2021.

Her research interests include network information security and situation prediction.

Ms. Liu is a member of China Computer Federation. She won the second-class scholarship for two consecutive years and won the title of excellent master's graduate.

ZHIHAN PAN was born in Henan, China, in 1996. She received the B.S. degree in computer science and technology (3G software) and the M.S. degree in computer technology from the Zhengzhou University of Light Industry, China, in 2019 and June 2022, respectively.

From 2019 to 2021, she has participated in the school volunteer work many times during the summer vacation. Her research interests include network information security and machine learning.

Ms. Pan is a member of China Computer Federation. She won a scholarship for outstanding students (Zhengzhou University of Light Industry).

YIFENG YIN was born in 1971. He received the M.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2001 and 2009, respectively. His Ph.D. work has focused on security virtual S-box technologies for authentication service.

He is currently a Professor at the School of Computer and Communication Engineering, Zhengzhou University of Light Industry. His research interests include information security and cryptography.

Prof. Yin was honored as an Expert Member of Henan Smart City Planning and Construction Committee, in 2014. He was invited to serve as a Reviewer for journals of *Information Sciences*, *Journal of Systems and Software*, *IET Communications*, *Cryptography*, and *Communications-Discrete Structures*. He currently holds the expert of China Torch Program Alternative Technology Expert's Database, CCF, and Chinese Association for Cryptologic Research.