Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc


# Research on decision optimization of network security expert system based on multi-source heterogeneous data

He Li[1,✉]

[1] *School of Cyber Science and Engineering, Southeast University, Nanjing, Jiangsu, 211189, China*

ABSTRACT

With the prosperous development of the Internet, the open network environment has also become a breeding ground for hackers, especially in the huge network system, the multi-stage, large-scale and coordinated network attacks have brought great trouble to the traditional defense means. This study designs a network security expert system based on multi-source heterogeneous data based on the characteristics of huge volume of multi-source heterogeneous network security data, heterogeneous format, and diverse semantics. The system contains five hierarchical structures: perception layer, event layer, alarm layer, attack context and attack pattern layer, and attack scenario layer. Petri nets are used for network security risk analysis and assessment to overcome the shortcomings of traditional defenses that become difficult to handle after modeling the attack scenarios. Incorporating the D-S evidence theory, the outputs of multiple decision engines are applied to the network security posture assessment to analyze the network condition from a global perspective and further enhance the effect of network attack classification. In the simulation experiments of simulated attacks, the monitoring information of the network security expert system has autocorrelation coefficients within two times standard deviation ($\pm 0.1$) after the 0th-order differencing, which indicates that the system is able to accurately assess the potential values of network attacks, such as scanning, brute-force cracking, DoS, and Web.

*Keywords:* cyber-attacks, multi-source heterogeneous data, expert systems, Petri nets, D-S evidence theory

## 1. Introduction

Based on the rapid development of communication technology, social and economic development has gradually entered into the Internet economic system, the network in the social public to bring convenience at the same time, but also produced a certain security risks, the development of social

✉ Corresponding author.
*E-mail address:* lhzjxtei@126.com (H. Li).
Received 25 March 2024; Revised 05 June 2024; Accepted 15 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-477
© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

stability has brought about a negative impact. And multi-source heterogeneous data technology can play a key role in optimizing the expert system used for correlation analysis of network security events.

Multi-source data fusion technology is able to correlate and combine data from multiple sensors of a large system and integrate them for unified evaluation [1-2]. The network attack path analysis method of multi-source heterogeneous data fusion is able to fuse multi-source heterogeneous data to construct the dynamic vulnerability of the network based on the assessment of the importance of the network nodes and effectively detect the attacker's intention to attack [3-5]. It constructs a static attack graph by evaluating the core nodes of the network topology combined with vulnerability information, then establishes the correlation relationship between alert events and calculates the probability that each attack path may occur by adopting the cumulative probability method [6-8]. Network security managers can then observe the path of maximum cumulative probability and take relevant protection strategies in time to safeguard network security. The network security posture assessment system based on multi-source heterogeneous data realizes the comprehensive analysis of network traffic, the reading of network attack characteristics and their core attributes, the mapping from core attributes to attack types, and the fusion of the outputs of multiple decision engines, which effectively improves the effectiveness of decision-making assessment [9-12].

Efforts should be made in today's era to enhance the cyber security situational awareness technology based on heterogeneous data from multiple sources to cope with the large scale and diverse types of data in complex network environments [13].Sahu, A. showed that rule-based and anomaly-based cyber security detection systems have a high rate of false positives, and proposed to improve the cyber detection accuracy by fusing cyber information with physical real-time data using feature transformation and selection techniques and prevent false data or command injection attacks [14]. Anjum, N. et al. proposed data fusion scheme MIND model for fusion of less parameterized datasets using machine learning integrated classifiers, which performs well in terms of resource utilization, complexity, and false alarm rate in the network detection process [15]. Yu, G. F. developed a system for assessing the advanced level of cyber security posture that establishes a multi-objective decision-making model based on multi-source information including expert trust, enterprise preference, etc., to significantly improve the protection capability of network security [16]. Ma, L. et al. explored how to improve the data utility of multi-source data collection, and experiments showed that the data security analysis method that utilizes encrypted symmetric key to extract the data security exchange features has a more perfect performance [17]. Wu, H. et al. studied the security detection method of heterogeneous network with multi-source fusion based on game theory, used Nash equilibrium theory to solve the detection threshold, used the local optimal attack allocation scheme of heterogeneous network as the means of attack resource allocation, and used it as the basis for constructing the non-zero-sum game model between the network attacker and the defender [18]. Zhou, J. et al. improved the traditional federated learning algorithm to deal with multi-source data by adopting the AHP, and introduced the tensor Tucker decomposition theory to capture the high-dimensional features of heterogeneous data, and finally solve the fusion problem of non-interactive multi-source heterogeneous data [19]. Yao, Y. et al. established a framework for fusion of multi-source heterogeneous cybersecurity situational data to abstract multi-source heterogeneous data into analyzable attack events, and use the attack behavior model as the benchmark for rapid response to the massive heterogeneous data and complete security detection [20].

The research constructs a network security expert system based on multi-source heterogeneous data fusion, and organically combines threat detection, multi-source data fusion, and hierarchical network threat assessment methods through an in-depth analysis of the network security posture assessment mechanism. A network security posture assessment system containing risk assessment module, defense decision optimization module, and multi-source fusion module is proposed. A Petri net is utilized for dynamic network risk assessment, and the entire Petri net is cut into two categories based on the characteristics of network attacks, i.e., attack delivery states with a single input and

delivery states with multiple inputs. Finally, in order to verify the feasibility and effectiveness of the network security expert system and the defense decision optimization module, a simulated attack network environment is constructed and simulation experiments are conducted for analysis.

## 2. Multi-source heterogeneous data-based network security expert systems

### *2.1. Modeling of Directed Cyber Attacks*

Network attack modeling refers to modeling and analyzing network attack behaviors, establishing connections between discrete attack events and constructing cognitively compliant attack scenarios, thus assisting in various processes such as attack detection, assessment, and defense. Attack modeling methods are important in the field of cybersecurity research, and the existing network attack modeling methods can be roughly divided into three categories, i.e., methods based on use-case frameworks, methods that represent network attacks from a temporal perspective, and graph-based methods. Commonly used attack modeling methods include Attack Graph model, Pyramid model, etc.

2.1.1. Attack Graph Model. Attack graph modeling, is an attack modeling methodology for assessing and improving network security based on graphical structures. Since the graph model itself is an ideal mathematical structure for expressing and analyzing correlations, it has intuitive and visual characteristics, facilitates formal analysis and is easy for users to understand. Attack graphs represent the process of network state changes caused by an attacker utilizing attack activities, where vulnerability exploits are represented by rectangles and preconditions are represented by rectangles, and preconditions are combined with each other in accordance with and/or relationships, often generated by enumeration of state-change sequences. Where a state-activity sequence is a possible attack path, the attack graph model measures the security of the network through the vulnerability and state transfer of each node, and is widely used in the fields of security risk assessment and the development of network security hardening measures.

An example of an attack graph model is shown in Fig. 1. The main challenge faced by Attack Graph models is the explosion of node states as the number of nodes grows, and the research on Attack Graph models focuses on the efficient generation of attack graphs, and the scalability of the model is the most critical factor affecting its popularization. Mulval is the most widely used attack graph generation tool, which not only provides an attack graph generation and visualization component, but also integrates an inference engine for attack states and potential privilege causality.

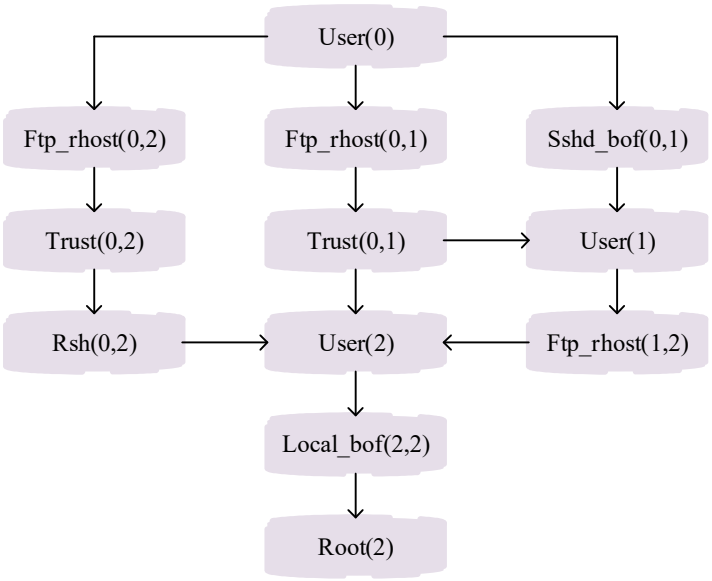
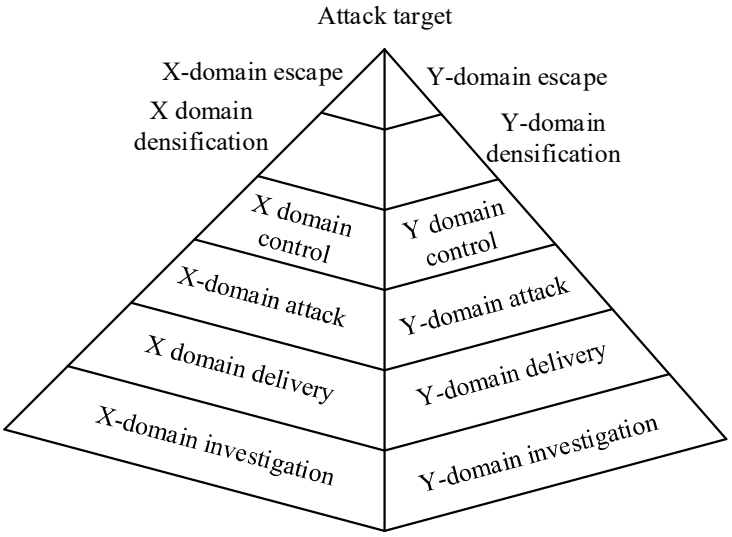
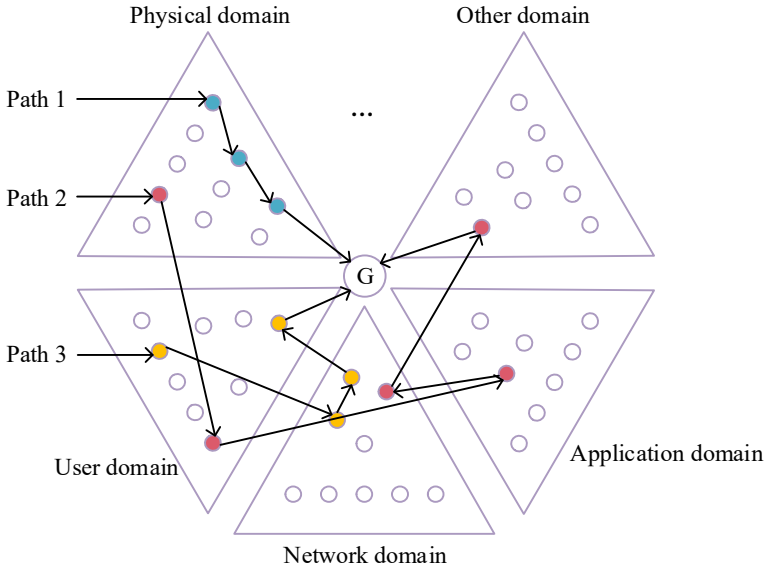



**Fig. 1.** Attack Graph model example

2.1.2. Pyramid model. The attack pyramid model defines the attack domain association method in the attack scenario description, and expresses the attack scenario by building a hierarchical polyhedral structure model, and the pyramid model and its unfolding structure are shown in Fig. 2. The top of the pyramid model is the attack target, and the attack scenario is divided into five parts: physical domain, user domain, network domain, application domain and other domains. The attack implementation process is divided into six stages according to the attack chain model: reconnaissance, placement, attack, manipulation, secret retrieval, and escape, which correspond to the six levels of the pyramid model. The attacker implements the attack activities according to the multi-stage characteristics, and the attack implementation breaks through layer by layer from the bottom to the top. The detection results are dispersed in different domains, and the analysis process is based on mapping the content events of different domains to the model domain, and a detailed description of the attack process can be obtained through model expansion.




(a) Pyramid model




(b) Expansion structure

**Fig. 2.** Pyramid model and its expansion structure

### *2.2. Network Security Expert System Design*

Directed network attack is a major threat to critical infrastructure and national security, the application layer attack through the computer network breaks through the traditional physical protection, the development of new technologies such as cloud computing, Internet of Things, artificial intelligence and other new technologies for network attack detection has brought new challenges and new solution ideas. There are many problems to be solved in the research field, and the detection and scenario reconstruction of targeted network attacks are still open research issues.

Based on the in-depth study of network security detection technology represented by directed network attack (APT), this paper designs a network security expert system (HeteMSD) framework based on multi-source heterogeneous data as shown in Fig. 3. The framework of network security expert system based on multi-source heterogeneous data, from bottom to top are perception layer, event layer, alarm layer, attack context and attack pattern layer, and attack scenario layer, and the characteristics of each layer are summarized as follows:

1) Perception Layer. As the initial data source for directed network attack detection, multi-source heterogeneous security data perception and acquisition is an important basic research topic. Various data sources are comprehensively and systematically summarized to efficiently and reliably collect and transmit security data from a large number of security collectors and devices. The main research contents of the perception layer include data acquisition, data aggregation, data parsing and preprocessing, data fusion, feature extraction and feature selection.

2) Event Layer. The perceptual layer data is used to characterize the original detection data without security semantics, and the data sources of network security are diversified, the format is heterogeneous, and the attributes of different data differ greatly. And the main research content of the event layer includes event aggregation, content security, event correlation, anomaly detection and so on.

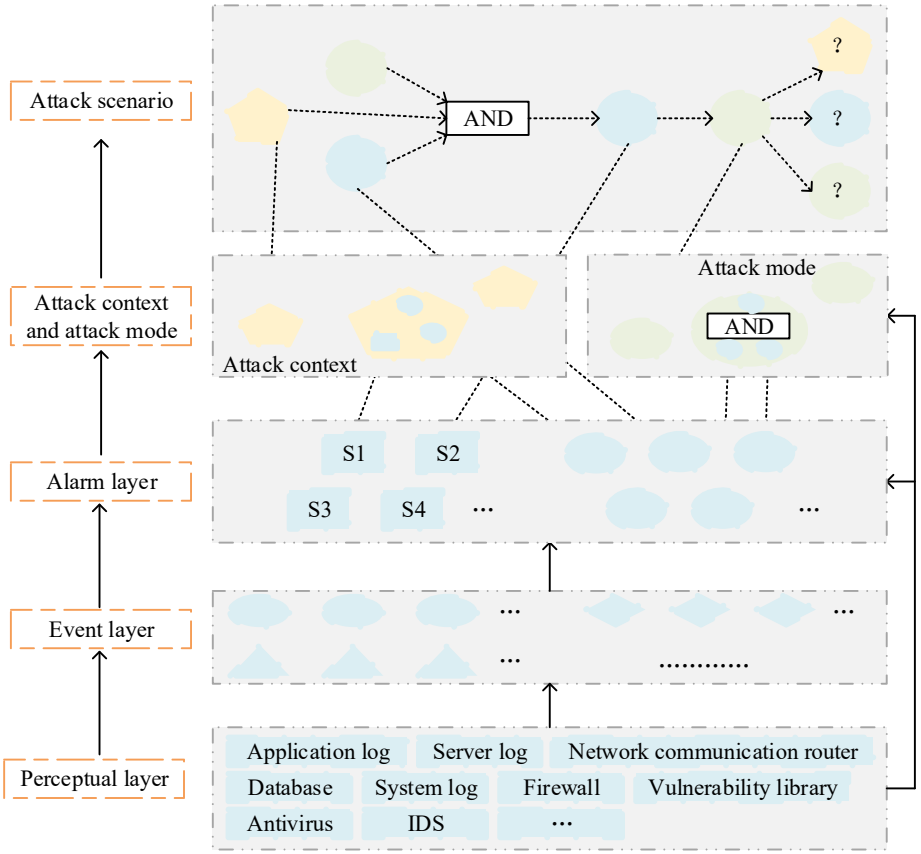


**Fig. 3.** Network security expert system based on multi-source heterogeneous data

3) Alarm Layer. The main research content of the alarm layer includes alarm expression, alarm fusion, alarm correlation, and security event management.

4) Context and Pattern Layer. Various types of alarm data come from a wide range of sources, and the context and attack pattern layer mainly provides the results of aggregating and correlating various types of alarm events, which can be expressed as attack fragments or attack steps.

5) Attack Scenario Layer. The main purpose of the attack scenario layer is to model and reconstruct the scenarios of directed network attacks, and present the reconstructed attack scenarios in a visual way, which is in line with human cognition. The main research content of the attack scenario layer includes threat intelligence expression and reasoning, attack causal analysis, visual analysis and attack planning.

### *2.3. Cyber risk assessment module*

2.3.1. Attacking Petri net generation. In this paper, Petri nets are used for network security risk analysis and assessment, which overcomes the defects of traditional attack trees and attack graphs that become difficult to deal with after modeling the attack scenarios, and has good scalability [21]. According to the algorithm, the Petri net generated by the directed network attack (APT) scenario can be derived as shown in Fig. 4. Firstly, based on the judgment of the attack type of the detected messages, it is found that they belong to 3 different attacks and $s1$, $s2$ and $s3$ belong to Ftp attack ($t1$), $s4$ and $s5$ belong to Rpc attack ($t2$) and $s6$ and $s7$ belong to Lpd attack ($t3$). According to the algorithm, the sub-Petri nets $s1 \rightarrow t1 \rightarrow s8$ are established by adding information number $s1$, followed by $s2 \rightarrow t1 \rightarrow s8$ and $s3 \rightarrow t1 \rightarrow s8$. It can be found that by adding information $s2$ and $s3$, only nodes $s2$ and $s3$ are added, which is due to the fact that these two nodes maintain a logical relationship with node $s1$. Continuing to add $s4$ information just adds attack behavior node $t2$ and resource node $s4$ to the original Petri net. This is because the attack behavior $t2$ contained in $s4$ information maintains the logic or relationship with the attack behavior $t1$ contained in $s1$, $s2$ and $s3$. Similarly, the Petri net generation for the whole attack scenario is completed according to our algorithm.

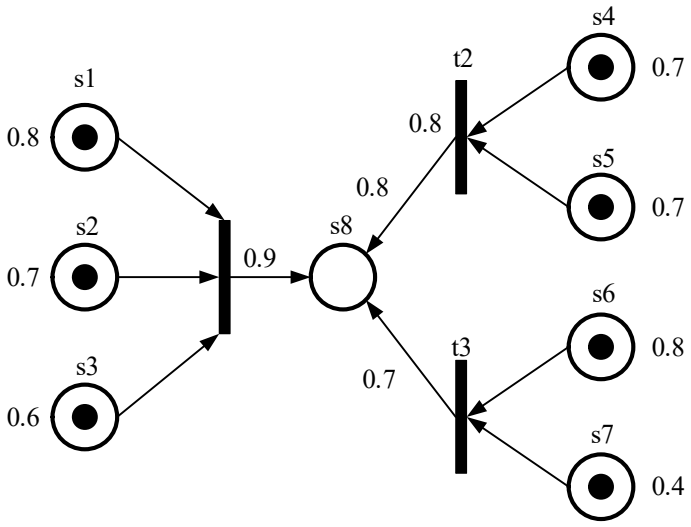



**Fig. 4.** The producted Petri Nets based on the information of attack scenarios

2.3.2. Dynamic risk assessment. In order to utilize Petri nets for dynamic network risk assessment, this paper slices the entire Petri net into two categories based on the characteristics of network attacks,

attack delivery states with a single input and delivery states with multiple inputs. Let the risk of an object being attacked be $F_{risk}$ and $F_{risk} = R$. Let $\mu(\cdot)$ represent the value of the risk function for the existence of a node, e.g., $\mu(s_i)$=10 for a resource node $s$. The risk value is 10 if exploited and $\mu(\cdot)$ is defined by the user based on the importance of the node (confidentiality, integrity, and availability), and the origin of this value is not used as a focus of research in the cybersecurity expert system. Therefore, in the Petri net study, the value is set to 1.

1) Attack delivery state with a single input. In this attack scenario, it is further categorized into two more scenarios based on whether the attack has multiple consecutive different attacks, the single attack scenario and the multiple consecutive attack scenario. The probability of risk and the value of risk for the presence of both attacks at node $s_{i+1}$ in the network can be expressed as $\phi(s_{i+1})$ and $R(s_{i+1})$, respectively:

$$\phi(s_{i+1}) = \phi(s_i) \times \phi(t_i) \tag{1}$$

$$R(s_{i+1}) = \mu(s_{i+1}) \times \phi(s_{i+1}) \tag{2}$$

where, if node $s_{i+1}$ is attacked, both the preconditions and postconditions of the attack behavior node $t_i$ must be satisfied for the attack to work, i.e., both $\phi(s_i)$ and $\phi(t_i)$ must be true.

The probability of risk and the value of risk for the existence of node $s_{i+1}$ in the network can be expressed as $\phi(s_{i+1})$ and $R(s_{i+1})$ respectively:

$$\phi(s_{i+1}) = \phi(s_1) \times \prod_{k=1}^{i} (\phi(t_k)) \tag{3}$$

$$R(s_{i+1}) = \mu(s_{i+1}) \times \phi(s_{i+1}) \tag{4}$$

2) Transfer states with multiple inputs. In this attack scenario, the transfer state is further subdivided into two broad categories logical and relational multiple-input states and logical or relational multiple-input states, based on the dependencies generated by the attack. The probability of risk and the value of risk for the existence of both attacks at node $s_{i+1}$ in the network can be expressed separately:

$$\phi(s_{i+1}) = \min\{\phi(s_1) \times \phi(t_1), \cdots, \phi(s_i) \times \phi(t_1)\} \tag{5}$$

$$R(s_{i+1}) = \mu(s_{i+1}) \times \phi(s_{i+1}) \tag{6}$$

If node $s_{i+1}$ is attacked, all attack preconditions for the attack behavior node $t_i$ must be satisfied. As long as one of the attack conditions is not satisfied, the probability that node $s_{i+1}$ is attacked is zero.

The probability of risk and the value of risk present at node $s_{i+1}$ in the network can be denoted as $\phi(s_{i+1})$ and $R(s_{i+1})$, respectively:

$$\phi(s_{i+1}) = \max\{\phi(s_1) \times \phi(t_1), \cdots, \phi(s_i) \times \phi(t_i)\} \tag{7}$$

$$R(s_{i+1}) = \mu(s_{i+1}) \times \phi(s_{i+1}) \tag{8}$$

If node $s_{i+1}$ is attacked as long as one of the preconditions is satisfied. Then the most likely probability that node $s_{i+1}$ is attacked is the probability that its preconditions are most likely to be satisfied.

Then according to the above algorithm to calculate the value of the risk that exists in being attacked by the network, the calculation process and results are as follows:

1) Obtain the set of most direct attack behaviors $\kappa(s8) = \{t1, t2, t3\}$ and the set of attack resources $\gamma(s8) = \{(s1, s2, s3), (s4, s5), (s6, s7)\}$ of the attacked node s8.

2) According to the algorithm, it can be judged that there are three logical or relational types of attack behaviors. So it is necessary to calculate the probability of the front resources of these three attack behavior types and the probability of these three attack behaviors, which are denoted as p1, p2 and p3, respectively. $\phi(t_1),\phi(t_2)$ and $\phi(t_3)$.

3) And further judging by the algorithm, it is found that the antecedent resources of the three attack behaviors are logical and relational. So:

$$\begin{cases} p1 = min\{\phi(s_1),\phi(s_2),\phi(s_3)\} = 0.6 \\ p2 = min\{\phi(s_4),\phi(s_5)\} = 0.7 \\ \text{p3=min}\{\phi(s_6),\phi(s_7)\} = 0.4 \\ \phi(t_1) = 0.9 \\ \phi(t_2) = 0.8 \\ \phi(t_3) = 0.7 \end{cases} \tag{9}$$

4) Calculate the attack probability $\phi(s8)$ for the attacked node s8:

$$\begin{aligned} \phi(s8) &= \max\left\{p1\times\phi(t_1),\, p2\times\phi(t_2),\, p3\times\phi(t_3)\right\} \\ &= \max\left\{0.6\times0.9,\quad 0.7\times0.9,\quad 0.4\times0.7\right\} \\ &= 0.63 \end{aligned} \tag{10}$$

5) Calculate the risk value $R(s8)$ of the attacked node s8 and set the value of $\mu(s8)$ to 1. Then $R(s8)=\phi(s8)$.

### *2.4. Defense Decision Optimization Module*

2.4.1. Classification of Cyber Attacks. Type $i$ attacks are defined as positive examples and Type $j$ attacks are defined as negative examples, due to the possible presence of noisy data in the data. Therefore, "soft spacing" is used in the support vector machine to allow a small number of samples to enter the "transition zone" or even split the error, in order to weaken the interference of noisy data as shown in Figure 5. In addition, there is often a nonlinear relationship between network core attributes and network attack types. Therefore, this paper embeds Gaussian kernel function in SVM model:

$$\kappa(x_i,x_j) = \exp(-\frac{\| x_i - x_j\|^2}{2\sigma^2}) \tag{11}$$

The "soft-spaced" support vector machine model based on Gaussian kernel functions is:

$$\begin{aligned} &\min_{\omega,b,\xi} \frac{1}{2}\| \omega \|^2 + C\sum_{i=1}^{m}\xi_i \\ &s.t. y_i(\omega\Phi(x_i)+b) \ge 1-\xi_i \\ &\xi_i \ge 0, i = 1,2,...,m \end{aligned} \tag{12}$$

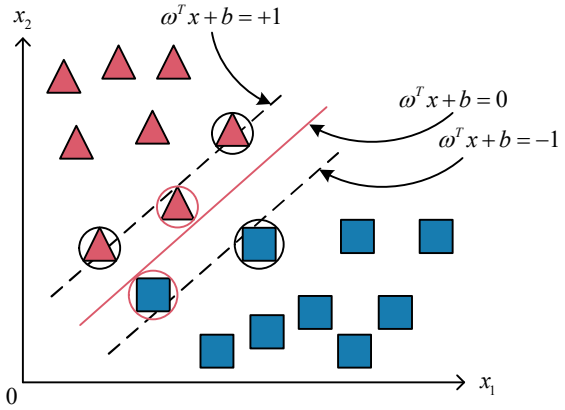



**Fig. 5.** Soft Interval

where $\omega$ is the normal vector, $b$ is the displacement term, $m$ is the amount of data, and $C$ is the penalty factor. For the convenience of solution, the "pairwise problem" is usually obtained by the multiplier method of Lagrange's formula and converted into a minimization problem:

$$\min_a \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j y_i y_j \kappa(x_i, x_j) - \sum_{i=1}^{m} a_i \tag{13}$$

$$s.t.\sum_{i=1}^{m} a_i y_i = 0$$
$$0 \le a_i \le C, i = 1, 2, \ldots, m \tag{14}$$

By solving the above problem, the optimal value $a_i^*$ is obtained, which in turn allows us to find the value of the normal vector:

$$w^* = \sum_{i=1}^{m} a_i^* y_i \Phi(x_i) \tag{15}$$

According to the KKT condition $a_i^{\cdot}(y_i f(\Phi(x_i)) - 1 + \xi_i) = 0$, where the samples corresponding to $a_i^{\cdot} > 0$ are support vectors, and let $S = \{i \mid \alpha_i^* > 0, i = 1, 2, \ldots, m\}$ denote the set of subscripts of all support vectors, the value of the displacement term can be found:

$$b^* = \frac{1}{|S|}\sum_{s \in S}(y_s - \sum_{i \in S} a_i^* y_i \kappa(x_i, x_s)) \tag{16}$$

From this, the model can be derived:

$$f(x) = w^T x + b$$
$$= \sum_{i=1}^{m} a_i^* y_i \kappa(x_i, x) + \frac{1}{|S|}\sum_{i \in S}(y_z - \sum_{i \in S} a_i^* y_i \kappa(x_i, x_z)) \tag{17}$$

Through the above training process, a support vector machine for classifying class $i$ and class $j$ attacks is obtained. Next, the data of two more attacks are selected among the $n$ classes of attacks and the above training process is looped until $c_n^2$ support vector machines are trained.

2.4.2. Optimizing the objective function. Next, using the Platt method, the probability of each support vector machine predicting the type of network attack is calculated. The core idea is to map the distance between a sample point and a support vector machine to a probability using a Sigmoid function. The further the sample is from the support vector machine, the more reliable the output is

and the higher the probability value. Data $(f_k, y_k)$, where $f_k$ represents the distance between the sample and the support vector machine (if the prediction is a positive example, the distance is positive; if the prediction is a negative example, the distance is negative). When the positive example of this support vector machine is a type $i$ attack and the counterexample is a type $j$ attack, the probability value of predicting a type $i$ attack is:

$$q_k = \frac{1}{1+\exp(Af_k + B)} \tag{18}$$

where $A$ and $B$ are parameters that are solved by the following optimization objective:

$$\min -\sum_k t_k \log(q_k) + (1-t_k)\log(1-q_k) \tag{19}$$

$$t_k = \frac{y_k + 1}{2} \tag{20}$$

For positive examples in the dataset, $y_i = 1, \quad t_i = 1$, the optimization objective simplifies to:

$$\min -\sum_k t_k \log(q_k) \tag{21}$$

That is, the optimization objective is to maximize the probability that the prediction is a positive example, and for the counterexamples in the dataset, $y_k = -1, t_k = 0$, the optimization objective simplifies to:

$$\text{min-}\sum_k \log(1-q_k) \tag{22}$$

That is, the optimization objective is to minimize the probability of predicting a positive example. For convenience the probability value of predicting a Type $i$ attack (positive example) is denoted as $r_{ij} = p_k$ and the probability value of predicting a Type $j$ attack (negative example) is denoted as $r_{ji} = 1 - p_k$. Using this method, the probability value of the output result of each support vector machine is calculated.

### *2.5. Multi-source data fusion module*

The multi-source fusion module will fuse the outputs of multiple decision engines to analyze the network conditions from a global perspective and further enhance the effectiveness of identifying the types of network attacks [22]. However, when there is a conflict between the output results of each decision engine (called evidence in multi-source fusion), the fusion effect will be seriously affected. Therefore, the conflicting evidences should be corrected first, and then multi-source data fusion should be performed.

2.5.1. Revision of evidence of conflict. The purpose of conflict evidence correction is to reduce the contradiction between the evidence, so that the fused results are in line with the real situation. Currently, the commonly used correction methods contain the confidence level method and the weighted average method. Here, $M = \{\vec{m_i} | 1 \le i \le n\}$ is used to denote the original data and $M' = \{\vec{m}_i' \mid 1 \le i \le n\}$ to denote the corrected data, where $n$ is the amount of data. The confidence method sets a confidence level $\eta$ and corrects the data based on the confidence level:

$$\vec{m}_i' = \eta \vec{m}_i + (1-\eta) \tag{23}$$

The weighted average method, on the other hand, uses a weighted sum of all the data, corrected for each piece of data:

$$\overline{m_i} = \eta \overline{m_i} + (1-\eta)\sum_{i=1}^{n} \omega_i \overline{m_i} \tag{24}$$

2.5.2. Multi-source data fusion. Multi-source data fusion will fuse the outputs of multiple decision engines to synthesize and analyze the cybersecurity situation. Currently, the main fusion algorithm commonly used in cybersecurity posture assessment is D-S evidence theory.

The D-S evidence theory is a classical fusion algorithm that draws consistent conclusions by correlating and fusing identical events:

$$m(A) = \frac{\sum_{B_i \cap C_j = A} m_1(B_i) m_2(C_j)}{1-k} \tag{25}$$

$$k = \sum_{B_i \cap C_j = \varnothing} m_1(B_i) m_2(C_j) \tag{26}$$

where $k$ is called the conflict coefficient, $B_i$ and $C_j$ are the focal elements, and $m(A)$ denotes the fused result.

## 3. Simulation experiments and analysis of network security expert systems

### *3.1. Experimental environment setup*

In order to verify the feasibility and effectiveness of the network security expert system and defense decision optimization module designed in the previous section, a network environment is constructed and simulation experiments are conducted. Potential attackers enter from the outside and first pass through a firewall, and the traffic passing through the IPFire firewall will then pass through a Snort intrusion detection system. Through the core switch, different functions of the server or end-user hosts are connected respectively. IP different servers run different services, and the services deployed on each server may be HTTP, SSH, MySQL, FTP, etc. The experimental network asset information is shown in Table 1. Among them, a WAF firewall is installed on server number 105. In addition, two devices, 106 and 107, are hosts; the former is a normal host in the network environment that accesses different resources in the network, and the latter is a host that simulates an attack.

**Table 1.** Information on experimental network assets

| Numbering | Open port | Running service | Vulnerability information |
|---|---|---|---|
| 101 | 25,23,58 | FTP, SSH, DNS | CVE-2015-3200, CVE-2019-5520 |
| 102 | 20,3456,24 | FTP, MySQL, SSH | CVE-2015-6520, CVE-2019-6252 |
| 103 | 18,26 | FTP, SSH | CVE-2016-3405, CVE-2019-5735 |
| 104 | 20,23 | FTP, SSH | CVE-2019-18045, CVE-2019-6137 |
| 105 | 60,6050,452 | HTTP, HTTPS | CVE-2017-4523, CVE-2017-15726 |
| 106 | 23 | SSH | MS17-015 |

In addition to the physical topological relationship between network assets, the access relationship structure that each asset satisfies in normal operation is shown in Figure 6. The simulation experiments simulate attacks of network scanning, brute-force cracking, DoS and other attacks, all attacks are launched on host 106 for 120 minutes, and the simulated attacks are divided into four scenarios, with detailed information as follows:

1) Simulated Scanning Attacks. Use Nmap to perform SYN scanning, FIN scanning and UDP scanning on 192.168.1.1/24 network segment for 30 minutes.

2) Simulated brute force attack. Brute force enumeration of SSH, FTP, MySQL, and Web applications in the network using Hydra for 30 minutes.

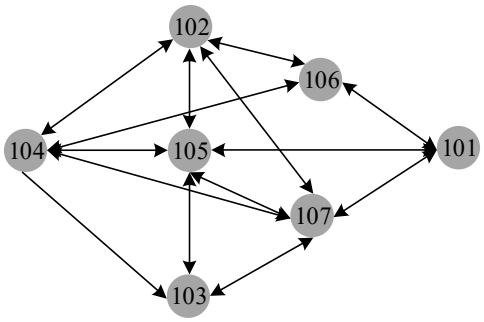



**Fig. 6.** Directed graph of service invocation

3) Simulation of DoS attack. TCP flooding and SYN flooding of the network using LOIC for 30 minutes.

4) Simulated Web Attack. Perform SQL injection, XSS injection, and directory scanning attacks on server 106 respectively for 30 minutes.

### *3.2. Cybersecurity Risk Assessment Model Testing*

3.2.1. System smoothness treatment. Smoothness is an important characteristic of time series data, which describes the statistical characteristics (e.g., mean, variance) that time series data possesses at different moments are stable, and the attack Petri nets generation of the system's cyber risk assessment module is based on the assumption that the data is smooth. In this paper, the first 450 data in the dataset are used as training data, and the network security time series graph is obtained as shown in Figure 7. It can be found that the network risk assessment value (SA) has a tendency to increase and decrease in stages, and the risk assessment time series results in the simulation environment from 0 to 120 min belong to an unsteady time series.

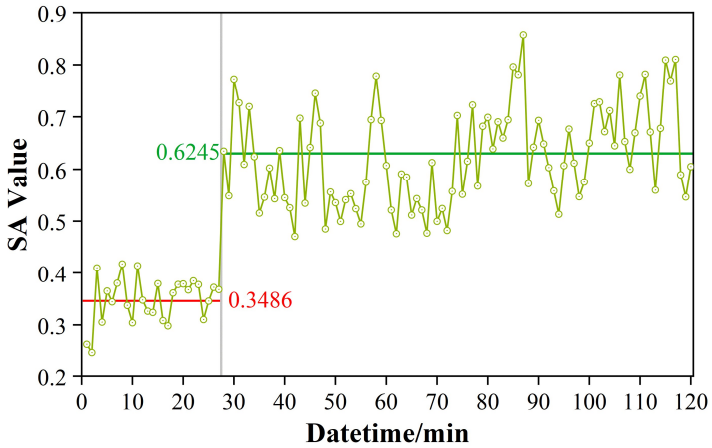
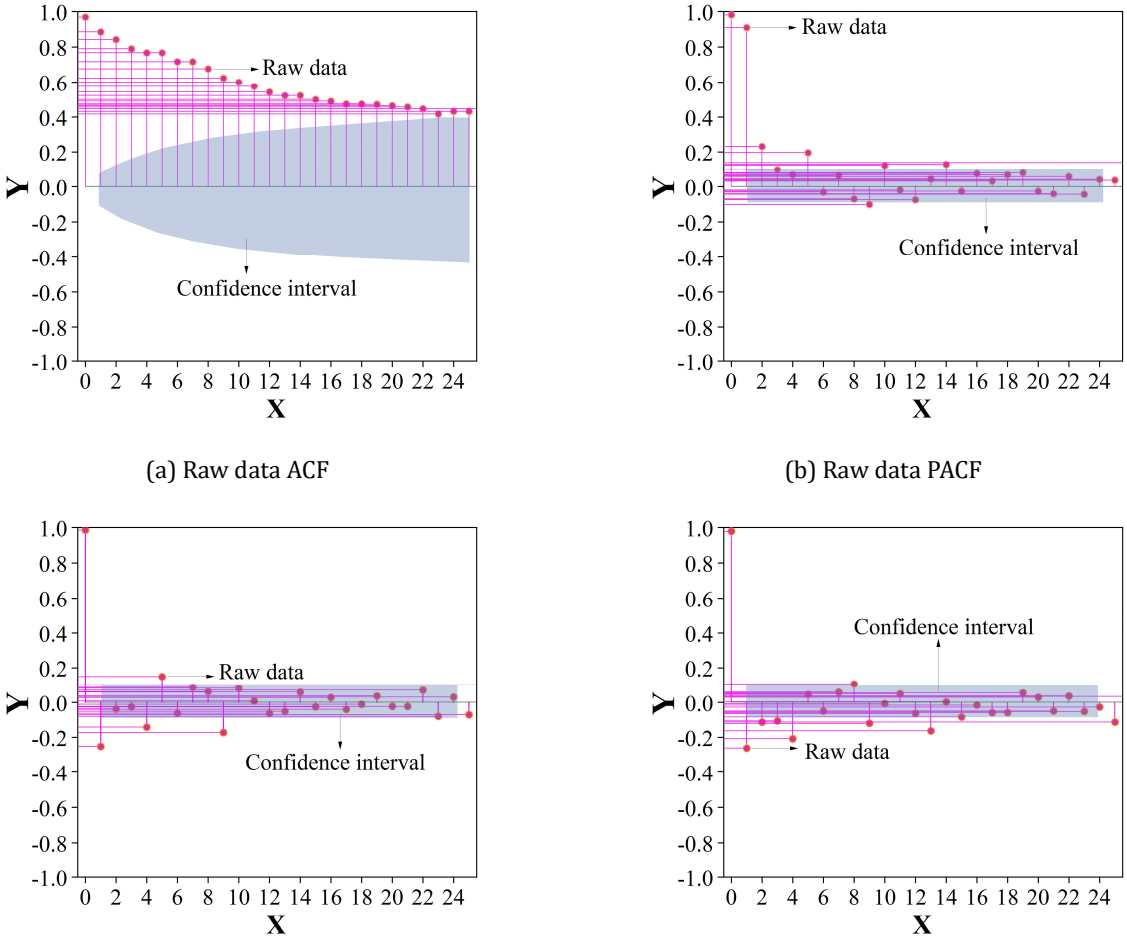



**Fig. 7.** Training data situation curve in dataset

The optimization objective function and partial autocorrelation function of the network defense decision before and after differencing are shown in Fig. 8, where 8(a) and (b) are the autocorrelation function and non-autocorrelation function plots of the original data, and it can be seen that the autocorrelation function gradually tends to be close to 0 but never falls into the confidence interval, which verifies the unstable nature of the cybersecurity posture. The original sequence is first-order differenced, and the autocorrelation function and non-autocorrelation function after differencing are

shown in Fig. 8 (c) and (d), respectively, and it can be seen that the autocorrelation function and the bias correlation function decay to 0 after lagging 1 and 2 periods.




(a) Raw data ACF

(b) Raw data PACF

(c) ACF after differential

(d) PACF after difference

**Fig. 8.** ACF and PACF before and after differential

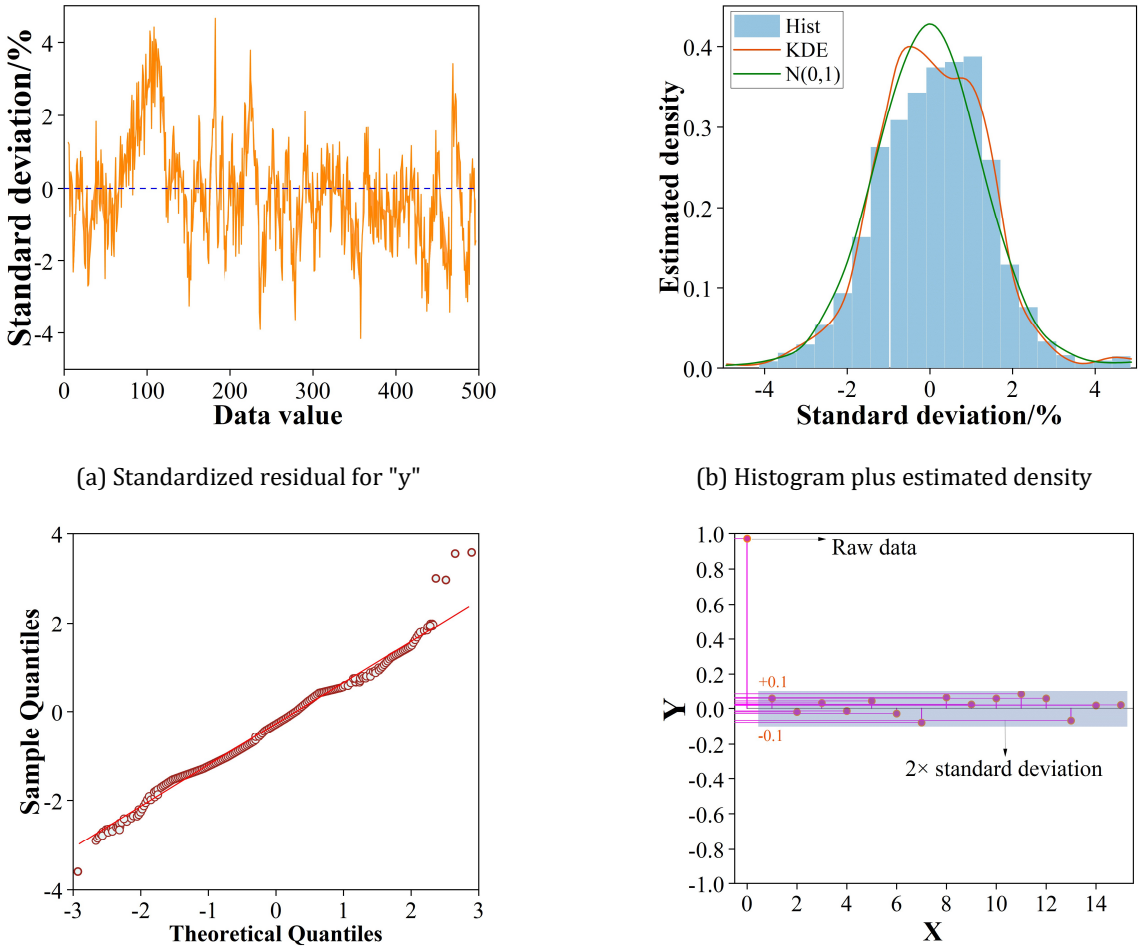
In addition, ADF smoothness test in Python is obtained as shown in Table 2. It can be seen that p is greater than the significant level in the original data and Test Statistic Value cannot reject the original hypothesis at 5% confidence level. These two conditions are achieved after the 1st order differencing, so the time series data satisfy the smoothness requirement only after the 1st order differencing.

**Table 2.** ADF test results of time series data stabilization process

| Parameter | Raw data | First-order difference |
|---|---|---|
| Test Statistic Value | -2.8342 | -8.0556 |
| p-value | 0.0812 | 1.3514e-10 |
| Critical Value (1%) | -3.5085 | -3.5085 |
| Critical Value (5%) | -2.9433 | -2.9433 |
| Critical Value (10%) | -2.4581 | -2.4581 |

3.2.2. Residual test results. The residual test determines whether the risk assessment module in a cybersecurity expert system is able to capture all the information in the cyberattack data. If the model has uncaptured structures or patterns in the data, then the residuals will show some patterns or regularities. The optimal model ARIMA (1,1,2) for the objective optimization function confirmed by

the above steps and the test results are shown in Figure 9. Figure 9 (a) shows the distribution of the residuals after standardization, and it can be observed that the data distribution fluctuates above and below 0, and the trend is consistent with a mean of 0 and a constant variance. Figure 9(b) shows the histogram estimation of the residuals, where the orange curve is the kernel density estimate (KDE) obtained from the histogram, the green curve represents the normal distribution of the standardized (N(0,1)), and the orange curve is next to the yellow curve, indicating that the residuals of the model roughly conform to the normal distribution. Figure 9(c) shows the quantile-quantile plot (QQ), and it can be seen that it is almost a straight line except in the head and tail, indicating that the sample basically conforms to the normal distribution. Figure 9 (d) shows the autocorrelation plot of the standardized residuals, where the horizontal axis indicates the order of the lag, the vertical axis indicates the magnitude of the correlation coefficient, and the shaded portion represents the range in which the standard deviation is a factor of two. It can be seen that none of the autocorrelation coefficients exceeds the range of 2 times the standard deviation (±0.1) after order 0, indicating that the correlation between the residuals and the lagged values is low.




(a) Standardized residual for "y"

(b) Histogram plus estimated density

(c) Normal Q-Q

(d) Correlogram

**Fig. 9.** Results of model diagnosis

In summary, it can be concluded that the risk assessment module in the network security expert system designed in this paper has passed the test and can be used to assess the potential values of network attacks such as scanning attacks, brute force attacks, DoS attacks and Web attacks.

### *3.3. Optimization Analysis of System Attack and Defense Strategies*

Firstly, in order to analyze the evolution law of the cyber security expert system in different initial states, set $\varepsilon = 1/2$, $\omega = 1$. The initial states of the cyber security expert system are set as, state 1 ($p_{11} = 0$, $p_{21} = 0.6$, $q = 0.1$), state 2 ($p_{11} = 0.2$, $p_{21} = 0.2$, $q = 0.1$) and state 3 ($p_{11} = 0$, $p_{21} = 0$, $q = 0.1$). Simulation is carried out using Matlab 2021 to obtain the phase diagram distribution of the system. Where, $p_{11}$, $p_{21}$ and $q$ are the strategy selection probabilities of the risky attacker, conservative attacker, and defender respectively.

1) The evolutionary phases of the attack and defense decisions of the network security expert system are shown in Figure 10, and it can be seen that the system will eventually evolve to the equilibrium point. Take the initial state 1 ($p_{11} = 0$, $p_{21} = 0.6$, $q = 0.1$) as an example, at the initial moment of the adventurous attackers to choose DOS attack strategy of the group proportion of 0.8, conservative attackers to choose "DOS attack" strategy of the group proportion of 0.8, the defender to choose the "upgrade patch" strategy of the group proportion of 0.8. After the continuous learning and improvement of the strategy of "upgrading" strategy of 0.8. After continuous strategy learning and improvement, the system eventually evolves to a stable equilibrium (1,1), which corresponds to a stable strategy, i.e., the group of adventurous attackers adopts the "DoS attack" strategy, and the group of conservative attackers adopts the "Sniffer attack" strategy, and the group of conservative attackers adopts the "Sniffer attack" strategy. "The optimal security defense strategy for the defender group is "upgrade patch".

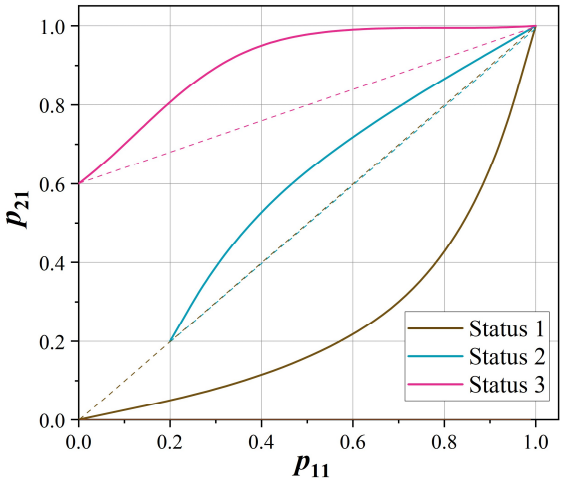



**Fig. 10.** Phase diagram of system attack and defense strategy

This result is consistent with the fact that risky attackers prefer active DoS attacks, which are more profitable. Conservative attackers, on the other hand, prefer passive "Sniffer Attacks" with relatively low risk. Defenders use the "upgrade patch" strategy, can not affect the user's normal service under the premise of increasing the difficulty of the attacker to invade, improve their own defense revenue.

2) Fig. 11 shows the evolution path of the conservative attacker's strategy, which sets the initial moments of the system $p_{11} = 0.2$, $q = 0.1$, i.e., the adventurous attacker chooses strategy $\{A_1, A_2\}$ randomly with mixed probability {0.2,0.9}. The defender randomly chooses strategy $\{D_1, D_2\}$ to implement with mixed probability {0.2,0.9}, and the conservative attacker at the initial moment with probabilities $p_{21} = 0.1$, 0.3, 0.5, 0.7, and 0.9, respectively. The evolution results of the choice of strategy "DoS attack" show that the finite rational conservative attackers eventually stabilize at choosing the pure strategy "Sniffer attack" with probability 1. In the initial stage of evolution, i.e., at the moment t<1, some groups try other strategies, and the proportion of groups choosing the DoS strategy is in an increasing state. However, through learning and strategy improvement, after t>1, the proportion of the group choosing the Dos strategy begins to decline, and after five games, the probability of choosing the

DoS strategy is 0. The result shows that even if a few individuals mutate their strategies to Dos, the gain of the mutated individuals is smaller than the average gain of the group. Eventually, they will give up the irrational DoS strategy and switch to the pure strategy Sniffer, and the results verify that the proposed attack and defense evolution can dynamically portray the trajectory of strategy evolution.

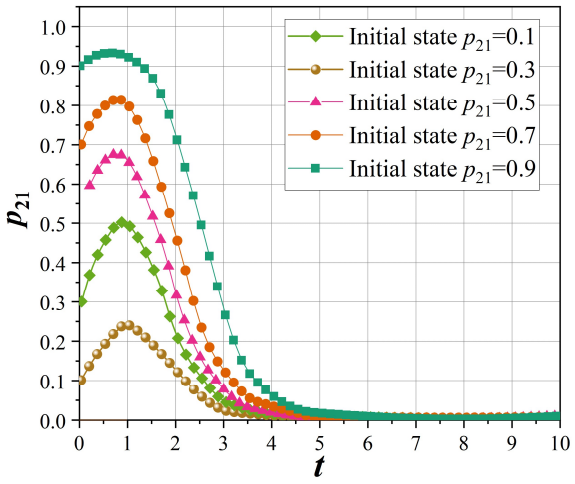



**Fig. 11.** Evolutionary path of conservative attacker strategy

In summary, from the results of the numerical experiments, it can be concluded that:

1) The designed network security expert system is applicable to realistic incomplete posture information situations, where the defender can correctly predict the relationship between the attacker's type and its selected strategy by speculating the probability distribution of the attacker's type. And the defense strategy is selected based on the attributes of its own type, as well as the relationship between the type of the attacker and the adopted strategy, which maximizes the expected gain of the defense. And the a priori probability distribution of different attacker types only affects the convergence speed of the system, and does not affect the final stable evolutionary equilibrium strategy, which enhances the effectiveness and guiding value of network security attack and defense decision-making.

2) In order to analyze the strategy learning and improvement process, by improving the replication dynamic mechanism that describes biological evolution, adding the selection strength factor to objectively reflect the randomness of selection, and portraying the strategy evolution mechanism of attacking and defending parties, it is believed that the player is in the process of continuous trial and error, as well as under the interactions with other participants, and gradually reaches the evolutionary stable state. When a small number of individuals due to the limitations of learning ability, reference to the low-yield strategy, will not affect the overall trend of strategy evolution, will only affect the convergence speed of the system, that is, it requires more repetitions of the game in order to find a stable optimal strategy, so the method in this paper can more scientifically and accurately reflect the process of the optimal defense strategy formation, the selected optimal defense strategy has a strong anti-jamming ability, and improve the The optimal defense strategy selected has strong anti-interference ability, which improves the scientific and practicality of the network security expert system.

## 4. Conclusion

As an important research direction in the field of network security, the network security posture assessment of multi-source heterogeneous data fusion is to a certain extent more adaptable to be used for realistic needs. This paper focuses on the research of multi-source heterogeneous data fusion and posture assessment technology, and designs a network security expert system based on multi-source

heterogeneous data (HeteMSD) based on the formalized representation of directed network attack and its detection process, in order to improve the relevance of multi-source heterogeneous network security data. By describing the process of targeted network attack detection based on multi-source heterogeneous data, a generalized data analysis process for targeted network attack detection is given, and a correlation analysis method based on multi-source heterogeneous data is summarized and proposed to provide model and general framework support for subsequent targeted network attack detection and traceability.

The main research goal of this paper is to establish an effective network security posture assessment and prediction method to comprehensively, objectively and accurately reflect the network security status, and to be able to predict the future network security trends for the purpose of advance prevention. Through the network security expert system designed in this research, the relevant tasks have been basically accomplished and stage-by-stage results have been achieved, but there are still studies worth further exploration:

1) In the real network environment, for servers storing sensitive files their asset weights are correspondingly higher, and accordingly they should be considered to have a greater degree of influence on the overall network security posture.

2) As network data continues to increase, the amount of data to be processed by the security posture assessment system also grows. In order to improve real-time performance, distributed computing technology can be applied to the posture assessment process to improve computational speed and efficiency. Therefore, situational assessment in the context of distributed computing is a direction worthy of future research.

## References

[1] Liu, W., & Zhu, J. (2021). A multistage decision-making method for multi-source information with Shapley optimization based on normal cloud models. Applied Soft Computing, 111, 107716.

[2] Liu, H., Tang, X., Xu, T., & He, J. (2023, November). A Survey of Homogeneous and Heterogeneous Multi-source Information Fusion Based on Rough Set Theory. In International Artificial Intelligence Conference (pp. 235-246). Singapore: Springer Nature Singapore.

[3] Zhang, S., Su, X., Shi, P., Du, T., & Han, Y. (2023). Threat Modeling and Application Research Based on Multi-Source Attack and Defense Knowledge. Computers, Materials & Continua, 77(1).

[4] Du, J., Guo, R., Suo, G., & Zhang, X. (2018, December). A Multi-source Alarm Information Fusion Processing Method for Network Attack Situation. In IOP Conference Series: Materials Science and Engineering (Vol. 466, No. 1, p. 012050). IOP Publishing.

[5] Xu, H. Y., Yue, Z. H., Wang, C., Dong, K., Pang, H. S., & Han, Z. (2017). Multi-source data fusion study in scientometrics. Scientometrics, 111, 773-792.

[6] Li, Y., & Yu, X. (2023, October). The multi-mode FBN multi-source heterogeneous data fusion technology based on AES algorithm. In 2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT) (pp. 1198-1202). IEEE.

[7] Zhao, B., Zuo, S., Liu, L., Li, C., & Xu, M. (2024). Violation Intelligent Recognition Algorithm Based on Multi-source Data Fusion. Journal of Electrical Systems, 20(10s), 1820-1834.

[8] Wu, D., Song, J., Bian, Y., Zheng, X., & Zhang, Z. (2021). Risk perception and intelligent decision in complex social information network. Industrial Management & Data Systems, 121(1), 99-110.

[9] Shi, L. L., Liu, L., Wu, Y., Jiang, L., & Ayorinde, A. (2020). Event detection and multi-source propagation for online social network management. Journal of Network and Systems Management, 28, 1-20.

[10] Zhou, M., Han, L., Lu, H., & Fu, C. (2021). Intrusion detection system for IoT heterogeneous perceptual network. Mobile networks and applications, 26, 1461-1474.

[11] Chen, Y., Lai, Y., Zhang, Z., Li, H., & Wang, Y. (2023). MDFD: A multi-source data fusion detection framework for Sybil attack detection in VANETs. Computer Networks, 224, 109608.

[12] Yin, K., Yang, Y., Yang, J., & Yao, C. (2022, April). A network security situation assessment model based on BP neural network optimized by DS evidence theory. In Journal of Physics: Conference Series (Vol. 2258, No. 1, p. 012039). IOP Publishing.

[13] Zhang, D., Qian, K., Wang, W., Fang, F., Wang, C., & Luo, X. (2020, December). Network security situation awareness technology based on multi-source heterogeneous data. In Proceedings of the 2020 International Conference on Cyberspace Innovation of Advanced Technologies (pp. 420-424).

[14] Sahu, A., Mao, Z., Wlazlo, P., Huang, H., Davis, K., Goulart, A., & Zonouz, S. (2021). Multi-source multi-domain data fusion for cyberattack detection in power systems. IEEE Access, 9, 119118-119138.

[15] Anjum, N., Latif, Z., Lee, C., Shoukat, I. A., & Iqbal, U. (2021). Mind: A multi-source data fusion scheme for intrusion detection in networks. Sensors, 21(14), 4941.

[16] Yu, G. F. (2024). A multi-objective decision method for the network security situation grade assessment under multi-source information. Information Fusion, 102, 102066.

[17] Ma, L., & Li, Y. (2022, September). Multi-source Data Collection Data Security Analysis. In International Conference on Advanced Hybrid Information Processing (pp. 458-472). Cham: Springer Nature Switzerland.

[18] Wu, H., & Wang, Z. (2018). Multi-source fusion-based security detection method for heterogeneous networks. Computers & Security, 74, 55-70.

[19] Zhou, J., & Lei, Y. (2023, January). Multi-source Heterogeneous Data Fusion Algorithm Based on Federated Learning. In International Conference on Soft Computing in Data Science (pp. 46-60). Singapore: Springer Nature Singapore.

[20] Yao, Y., Sun, Y., Liu, Z., & Meng, X. (2020, May). A Data Fusion Framework of Multi-Source Heterogeneous Network Security Situational Awareness Based on Attack Pattern. In Journal of Physics: Conference Series (Vol. 1550, No. 6, p. 062025). IOP Publishing.

[21] Xinlei Li & Di Li. (2014). A Network Attack Model based on Colored Petri Net. Journal of Networks(7),1883-1891.

[22] Yang Weiman,Gu Jianfeng,Wang Xinggui & Wang Weinian. (2024). Attention-parallel multisource data fusion residual network-based open-circuit fault diagnosis of cascaded H-bridge inverters. Journal of Power Electronics(6),875-886.