Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.061525

ARTICLE

A Hierarchical Security Situation Assessment Approach for Train Control System under Cyber Attacks

Qichang Li[1,2,*], **Bing Bu**[1] **and Junyi Zhao**[1]

[1]State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing, 100044, China
[2]Signal and Communication Research Institute, China Academy of Railway Sciences, Beijing, 100081, China

*Corresponding Author: Qichang Li. Email: liqichang@rails.cn

Received: 26 November 2024; Accepted: 27 February 2025; Published: 19 May 2025

ABSTRACT: With the integration of informatization and intelligence into the Communication-Based Train Control (CBTC) systems, the system is facing an increasing number of information security threats. As an important method of characterizing the system security status, the security situation assessment is used to analyze the system security situation. However, existing situation assessment methods fail to integrate the coupling relationship between the physical layer and the information layer of the CBTC systems, and cannot dynamically characterize the real-time security situation changes under cyber attacks. In this paper, a hierarchical security situation assessment approach is proposed to address the security challenges of CBTC systems, which can perceive cyber attacks, quantify the security situation, and characterize the security situation changes under cyber attacks. Specifically, for the physical layer of CBTC systems, the impact of cyber attacks is evaluated with the train punctuality rate and train departure interval indicators. For the information layer of CBTC systems, the system vulnerabilities and system threats are selected as static level indicators, and the critical network characteristics are selected as dynamic level indicators to quantify the real-time security situation. Finally, the comprehensive security situation assessment value of the CBTC systems is obtained by integrating the physical and information layer indicators. Simulation results illustrate that the proposed approach can dynamically characterize the real-time security situation of CBTC systems, enhancing the ability to perceive and assess information security risks.

KEYWORDS: Transportation; train control system; cyber security; hierarchical security situation assessment

1 Introduction

Nowadays, more and more attention has been paid to the research on information security of train control system [1,2]. With the deep integration of urban rail transit informatization and intelligence, the threats faced by the train control system are gradually increasing, heightening the risk of information security. The train control system adopts the universal TCP/IP protocol stack, rendering them vulnerable to cyber attacks such as data spoofing and flood attacks. Moreover, it is difficult to achieve complete enclosed interconnection between the train control system and other information service systems, thus increasing the information security risks associated with train control system [3]. The occurrence of information security incidents could significantly impact critical infrastructure, including urban rail transit [4].

Situation awareness technology is a proactive information security technology that can identify potential information security threats and protect the CBTC systems from serious damage. Different from intrusion detection technology [5,6], situation awareness technology includes not only the detection of cyber

attack, but also the assessment of system security status and the prediction of system security evolution trends [7]. Assessing the security situation of the system is the key to the application of situation awareness technology [8]. However, traditional situation assessment methods suffer from two main drawbacks. First, current situation awareness methods mainly establish static models to assess the system situation (e.g., assess the current security situation), but cannot dynamically characterize real-time situation variations under cyber attacks. Second, the network situation assessment methods do not take advantage of the coupling relationship between the physical layer and the information layer of the CBTC systems, and some inherent characteristics of the CBTC systems are not considered, such as redundant network structures and fail-safe mechanisms [9]. For instance, the system operates with a fixed line cycle, resulting in stable data characteristics. And the system operates 24/7 without interruption, leaving no extra time to update the situation assessment model. Besides, the fail-safe mechanisms will cover up the equipment failures caused by the cyber attacks, making it impossible to assess the real security situation.

Therefore, this paper proposes the hierarchical security situation assessment approach to address the situation assessment challenges within CBTC systems. The proposed assessment approach can dynamically characterize the information security situation from both the information and physical layers of CBTC systems, and utilize the inherent coupling relationship between the two layers to achieve a comprehensive security situation assessment. To the best of our knowledge, this is the first work discussing the application of a hierarchical security situation assessment approach for CBTC systems. The contributions of the paper are as follows:

- In view of the challenges faced by current situation assessment methods, a hierarchical security situation assessment approach is proposed. The hierarchical security situation assessment approach can perceive cyber attacks in real-time and characterize the security situation changes under cyber attacks. Specifically, by analyzing the impact of cyber attacks on the physical layer of the train control system, quantifying the changes in the situation indicators of the information layer under cyber attacks, integrating the situation assessment indicators of the physical layer and information layer of the CBTC systems, the coupling relationship between the assessment indicators is established, and the comprehensive security situation assessment value is obtained.
- In order to dynamically perceive the cyber attacks and reflect the real-time situation variations of CBTC systems, an improved generative adversarial network based artificial immune system (GAN-AIS) is exploited by our approach. Specifically, the changes in antibody concentration of GAN-AIS are used to characterize the dynamic security situation variations of the CBTC systems. The network characteristics are used for immune learning and self-updating, and the GAN neural network is used to balance the antibody population and rapidly clone effective antibodies.

The rest of this paper is organized as follows. Section 2 provides the research background and related studies of the situation assessment approach. Section 3 describes the comprehensive architecture and the detailed mechanism design. The performance of the proposed approach is simulated and verified in Section 4. The results and discussion are introduced in Section 5. And the conclusion is described in Section 6.

2 Background and Literature Review

This section provides background knowledge on artificial immune system and GAN neural network. Moreover, the related works for situation assessment are also discussed.

2.1 Artificial Immune System

Artificial immunity is an advanced technology inspired by the theory of natural immunity and follows the principle of natural immunity. It has made remarkable achievements in the field of information security and cyber attack recognition [10]. Artificial immune system can achieve the system security situation awareness without prior knowledge, and continuously enhance its perception performance by the designed evolutionary mutation mechanisms, to perceive the security situation dynamically and accurately, and ensure the safe operation of the system.

The artificial immune system possesses the characteristics of self-learning, self-organization, self-adaptation, highly distributed and long-term memory [11]. Generally, the immune mechanism can be divided into three main stages. The first is the process of immune cells from immaturity to maturity, known as self-tolerance. The second is the process of immune response, in which immune cells match with antigens and form immune memory. The third is the immune feedback process, the specific antibodies are generated and the immune process can be regulated.

2.2 GAN

GAN is a deep neural network that has been widely used in the field of data learning and data generation [12]. The standard GAN is composed of generator G and discriminator D, the instance samples are generated by the generator and mixed with the real samples, and then randomly sent to the discriminator to distinguish whether the data is real or comes from G. Similar to the adversarial mechanism, optimization objectives, and the dynamic learning process of reinforcement learning [13,14], the core of the GAN network is the adversarial process between the generator G and the discriminator D. For the training process of GAN, the purpose of the generator is to fit the distribution of real samples, so that the discriminator cannot discern real samples or fake data. Through iterative training and optimization, the Nash balance is achieved. During the training process, the parameters of the generator G and the discriminator D are continuously adjusted to achieve better generation effects through dynamic learning. The objective function of GAN can be expressed as

$$\min_{G}\max_{D} V(D,G) = E_{x\sim P(x)}\left[\log D(x)\right] + E_{z\sim P(z)}\left[\log\left(1 - D(G(z))\right)\right] \tag{1}$$

where D and G are discriminator and generator, respectively. x is the sample data, z is the random noise. $p(x)$ is the real data distribution, and $p(z)$ is the latent space distribution.

By learning the distribution of real data, GAN can be used to solve the problem of imbalanced data distribution and data generation. In an artificial immune system, the clone selection of specific antibodies is critical for the evolution and variation of antibody populations. Based on the GAN model, the mature antibody vectors are selected and learned, then the generated antibodies are added to the antibody population, enabling the artificial immune system to benefit from a richer population.

2.3 Literature Review

For the IT network, Yu [15] propose a multi-objective decision method for network security situation grade assessment. The method first defines membership functions of attribute eigenvalue grades, trust transmission methods, multi-path trust integration, and further establishes a multi-objective decision grade assessment optimization model to obtain the network security situation grades. Wang et al. [16] propose a network security information analysis and network security situation (NSS) model based on data mining, the model cannot only detect network security threats but also evaluate the network security situation with system security indicators. Specifically, the threat data are classified and the risk measurement is conducted.

Then the security of each type of threat is measured according to the information characteristics of network attacks. Finally, according to different attack motivations, different evaluation methods are used to obtain comprehensive situation assessment results. Du et al. [17] propose a network security situation assessment model in the SDN environment. The situation elements are mapped to the layers of back propagation (BP) neural network, and the weights and thresholds of the BP neural network are optimized by the cuckoo search algorithm. Specifically, the situation indicators are used as the input data for the BP neural network, and the value of the output layer is the result of the evaluation. Guo et al. [18] propose a defensive random game model, which first analyzes the threat propagation process, establishes the threat propagation-access relationship network, and then designs a random game model of the threat action and protection strategy game. The random game model is used to quantify the network security situation. Yang et al. [19] propose a situation assessment model based on unsupervised learning. The paper applies the variational autoencoder (VAE) and GAN for data feature information extracting and error threshold calculation. Then the situation value is calculated based on the threat probability and threat impact. Specifically, the threat probability is obtained from the test results of each test, and the threat impact is defined by the Common Vulnerability Scoring System (CVSS). Wang et al. [20] propose a network security situation assessment framework based on analytic hierarchy process (AHP). The indicators of basic operation, threat, vulnerability, and asset breakage are considered to establish the assessment model. Three indexes of risk situation, the system operation, and the damage degree are discussed, and the weight factors of the evaluation index are assigned by the AHP process. In the aforementioned research, mathematical models are established to assess the network security situation. However, most of these models adopt subjective evaluation methods or neural network models. Such models are heavily weighted towards subjective factors and lack interpretability. Furthermore, these models analyze the impact of cyber attacks from the network perspective, such as assessing the security risks of the system and obtaining the network security situation value. However, these models do not consider the impact of attacks on the industrial systems operation and cannot be directly applied to train control systems.

For the industrial control systems, Zhang et al. [21] propose a new security situation awareness method for the power grid. First, the power grid security situation evaluation indicators are selected, and then the power grid security situation risks are quantified and divided into different security risk levels. Finally, the power grid security situation prediction model based on deep learning is constructed and the security situation awareness prediction results are verified. Zhang et al. [22] propose a combination active-passive risk source identification method for the terminal area control system, which includes 19 secondary indicators and 3 primary indicators. Then the sparrow search algorithm (ISSA)-extreme learning machine (ELM) framework is used for risk probability prediction. Finally, the overall risk probability obtained by the Bayesian theory and evaluation weight are used to obtain the overall situation assessment value. Zhao et al. [23] propose a wavelet neural network analysis method to obtain the security situation value. The network and control characteristics such as security monitoring, security alarms, security analysis, security audits, and security verification information are selected as indicators. Then the indicators are selected as the sampled dataset to train the wavelet neural network model, and the security situation value of the power control system is set as the output. Li et al. [24] propose a network security situation assessment method for industrial control networks. The information in the sparse data is extracted by the stack autoencoder and the data dimensionality is reduced. The nonlinear mapping relationship of the network status is fitted through the RBF network, and the security situation assessment value of the industrial control network is calculated. However, the above methods do not consider how the security situation of the industrial control system changes when cyber attacks are introduced. Lei et al. [25] propose a smart grid security situation awareness algorithm based on deep reinforcement learning. The situation elements such as network port traffic information, alarm

events in the system logs, node vulnerabilities, and known attacks are acquired for situation assessment. The integrated edge computing and deep reinforcement learning method can minimize processing costs on the premise of minimizing attack detection error rates. However, the proposed algorithm evaluates the system security situation solely based on detection metrics, such as detection rate and false alarm rate. Real-time dynamic security situation assessment methods still require further optimization and research.

For the cyber-physical transportation systems (CTS), Yu et al. [26] conducted a comprehensive analysis of the security threats, attack mechanisms, and defense measures of cyber-physical systems from three dimensions: physical domain, cyber domain, and cyber-physical domain. Cyber-physical systems is an important part of Industry 4.0 and faces many security threats, the article discusses the challenges and future directions of current research. Alsulami et al. [27] propose a transfer learning based intelligent intrusion detection system for autonomous vehicle-cyber physical systems. The actual position and actual speed of the autonomous vehicle, the actual position and actual speed of the preceding vehicle, these one-dimensional data are converted into two-dimensional images for input into the pre-trained convolutional neural network (CNN) model. Then the intelligent intrusion detection system using the CNN model to detect the cyberattacks targeting the physical components of an autonomous vehicle through controller area network (CAN). Abdo et al. [28] propose a comprehensive connected and automated transportation system (CATS) cybersecurity research framework that covers multiple levels from individual vehicles to transportation networks, and emphasizes the importance of threat analysis and risk assessment (TARA) tools in cybersecurity research. Specifically, TARA assesses the likelihood and impact of cyber attacks and combines them to derive the CATS risks. The TARA method includes both qualitative and quantitative methods, the qualitative method relies on expert experience to assess network risk, while the quantitative method is based on probability theory and statistical models to assess the likelihood of threats and risks. The above studies introduced the security issues and security assessment methods faced by different transportation modes. However, some security assessment methods are highly subjective and lack a dynamic security assessment process to deal with cyber threats.

In recent years, there have been few studies on situation assessment for CBTC systems. Lu et al. [29] propose a resilience-based security assessment method based on structural information entropy to measure the security level of CBTC systems. The two-dimensional structure entropy is used to evaluate the performance of the cyber domain, and the impact of cyber attacks on the physical domain of the CBTC systems is calculated according to the timetable and running states. The resilience metrics considering both the cyber domain and physical domain are used to analyze the security level of CBTC systems. Kang et al. [30] propose a multi-dimensional Gaussian hidden Markov model approach to quantify the situation awareness value of CBTC systems. The information features, including CPU usage, RAM usage, disk access rate, and network rate, are considered, and an integrated situation awareness value for CBTC systems is derived using entropy weights. However, the research does not closely integrate with the characteristics of the CBTC systems, and it fails to explore the changes in the information security situation when the system suffered cyber attacks. In this paper, we propose a hierarchical security situation assessment approach that combines physical layer and information layer indicators to quantify the security situation under cyber attacks, and the real-time comprehensive security situation assessment value is calculated through the evaluation indicators from both dynamic and static aspects.

3 Hierarchical Security Situation Assessment Approach

In this section, we first provide the general structure of our hierarchical security situation assessment approach. Then, we describe the situation assessment approach in detail.

3.1 Overall Architecture

As a typical cyber-physical system, CBTC systems integrate physical processes with computing systems, incorporating controllers, actuators, and sensors. Train operations in the physical environment are governed by a computing and communication network, which relies on multi-source sensors for data acquisition. The CBTC architecture is generally divided into two primary layers: the physical layer and the information layer [31]. As depicted in Fig. 1, the physical layer encompasses the state and actions of trains, components, devices, and sensors. The information layer achieves communication and control signal exchange between critical devices such as the zone controller (ZC), computer interlocking (CI), automatic train supervision (ATS), database storage unit (DSU), access point (AP), and vehicle onboard controller (VOBC). This hierarchical structure ensures efficient interaction between the physical and information layers of the system.

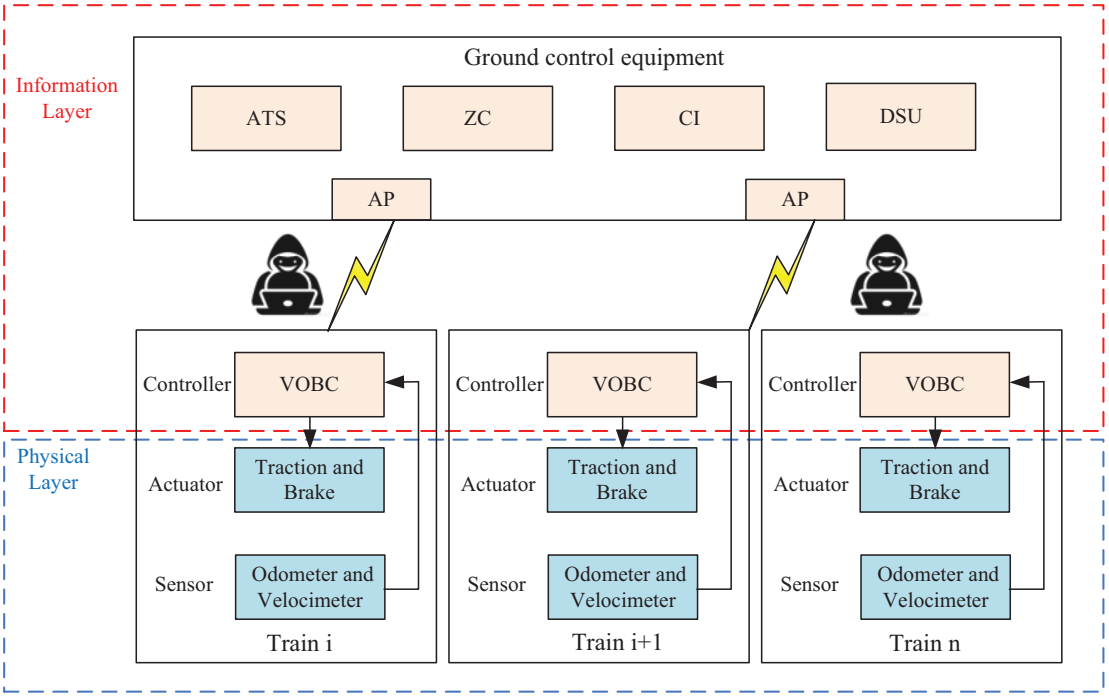


Figure 1: The general structure of CBTC systems

The purpose of CBTC systems is to ensure the safe and efficient operation of trains. The physical layer primarily focuses on the acceleration, speed, and location of the train, as well as the observation of the physical characteristics such as train speed and train location. The information layer is mainly concerned with communication and information exchange between the train and the ground. For example, the ZC calculates the correct movement authorization (MA) and sends it to the VOBC of the train through the train-ground network. Cyber attacks do not directly damage the physical characteristics of trains but can indirectly disrupt normal operations through the communication network. For instance, a DoS attack can obstruct the normal communication between trains and the ground, while a data tampering attack can manipulate MA information transmitted between trains and the ground. The results of cyber attacks may lead to anomalies in the physical layer of the system. Therefore, when conducting situation assessment, it is not only necessary to quantify security situation of the information layer under cyber attacks but also consider its impact on the physical layer of the system.

The general structure of the hierarchical security situation assessment approach is shown in Fig. 2. For the physical layer, the impact of cyber attacks on the system is discussed. Specifically, the train punctuality rate and train departure interval are used to evaluate the impact of cyber attacks on the physical layer. For the information layer, the system vulnerabilities, system threats and network characteristics are the main aspects that need to be assessed and quantified. The specific indicators for each aspect are listed and discussed later.

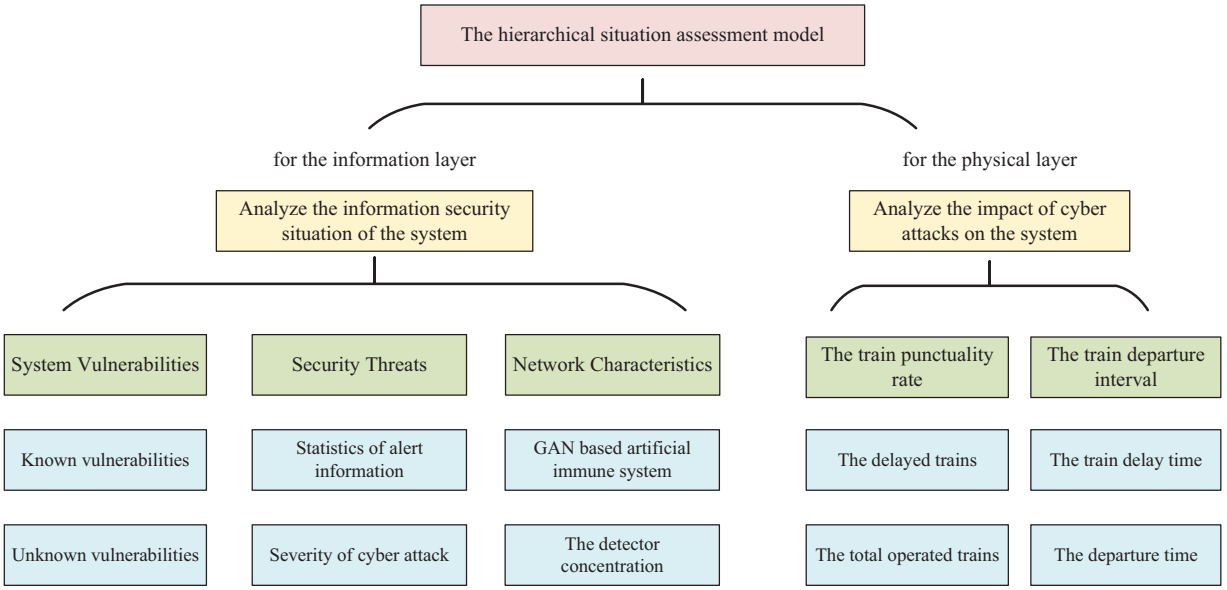


Figure 2: The general structure of hierarchical security situation assessment approach

3.2 Situation Assessment for the Physical Layer of CBTC Systems

In this section, the train control model is established and the Kalman filter method is used to obtain train state observations. The classical Kalman filter method has uncertain observation errors under cyber attacks. Therefore, we first establish the attack model and then introduce the unscented Kalman filter to estimate the train state under cyber attacks. Then, we calculate the changes in the train punctuality rate and departure interval based on the train state observations. Finally, the quantitative impact of cyber attacks on the physical layer of the train control system can be calculated.

3.2.1 Train Control Model

We assume that the train control system is discrete and linear time-invariant, and the control model can be expressed as

$$X_{k+1} = AX_k + BU_k + W_k \tag{2}$$

$$U_k = -CX_k \tag{3}$$

where X_k is the train state vector, U_k is the system input vector, Y_k is the system observation vector. $W_k \sim (0, Q)$ is zero-mean Gaussian random process noise. A and B are the system parameter matrix, C is the feedback gain matrix. The parameter matrix A, B, and C can be calculated according to the train dynamics equation.

The train dynamics equation can be expressed as

$$s_{k+1}^{(i)} = s_k^{(i)} + T \cdot v_k^{(i)} + \frac{1}{2} T^2 \cdot u_k^{(i)} \tag{4}$$

$$v_{k+1}^{(i)} = v_k^{(i)} + T \cdot u_k^{(i)} \tag{5}$$

where s_k^i, v_k^i, and u_k^i represent the location, velocity, and acceleration of the ith train at the time k, respectively, T is the data transmission cycle of VBOC and ZC.

The train state vector X_k and system input vector U_k can be expressed as

$$X_k = \left[\delta x_k^{(1)}, \delta x_k^{(2)}, \ldots, \delta x_k^{(n)}\right]^{\mathrm{T}} \tag{6}$$

$$U_k = \left[\delta u_k^{(1)}, \delta u_k^{(2)}, \ldots, \delta u_k^{(n)}\right]^{\mathrm{T}} \tag{7}$$

where $\delta x_k^{(i)} = \left[\delta s_k^{(i)}, \delta v_k^{(i)}\right]$ is the state vector of train i, $\delta s_k^{(i)} = s_k^{(i-1)} - s_k^{(i)} - S^{(i)}$ and $\delta v_k^{(i)} = v_k^{(i)} - V^{(i)}$ are the deviation of tracking interval and train speed, respectively. $S^{(i)}$ and $V^{(i)}$ represent the optimal tracking interval and train speed. $\delta u_k^{(i)}$ is the acceleration generated by train traction and braking of train i. n is the number of trains.

Then the train dynamics equation can be rewritten as

$$\delta s_{k+1}^{(i)} = \delta s_k^{(i)} + T \cdot \left(\delta v_k^{(i-1)} - \delta v_k^{(i)}\right) + \frac{1}{2} T^2 \cdot \left(u_k^{(i-1)} - u_k^{(i)}\right) \tag{8}$$

$$\delta v_{k+1}^{(i)} = \delta v_k^{(i)} + T \cdot u_k^{(i)} \tag{9}$$

According the Eqs. (4)–(9), the parameter matrix $A = \left(a_{ij}\right)_{n\times n}$, $a_{ij} \in \mathbf{R}^{2\times 2}$, $B = \left(b_{ij}\right)_{n\times n}$, $b_{ij} \in \mathbf{R}^{2\times 1}$, and $C = \left(c_{ij}\right)_{n\times n}$, $c_{ij} \in \mathbf{R}^{1\times 2}$ can be calculated as

$$a_{ij} = \begin{cases} \begin{bmatrix} 1 & -T \\ 0 & 1 \end{bmatrix}, & i = j \\ \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, & i = j-1 \\ \mathbf{0}_{2\times 2}, & others \end{cases} \tag{10a}$$

$$b_{ij} = \begin{cases} \begin{bmatrix} -T^2/2 \\ 1 \end{bmatrix}, & i = j \\ \begin{bmatrix} T^2/2 \\ 0 \end{bmatrix}, & i = j-1 \\ \mathbf{0}_{2\times 1}, & others \end{cases} \tag{10b}$$

$$c_{ij} = \begin{cases} \mathbf{c}^{(\mathbf{i})}, & i = j \\ \mathbf{0}_{1\times 2}, & i \neq j \end{cases} \tag{10c}$$

where $\mathbf{c}^{(i)}$ is the control output generated based on the distance to train $i-1$ and the speed of train i.

According to the principle of train dynamics, the observation equation can be expressed as

$$Y_k = HX_k + V_k \tag{11}$$

where $Y_k = \left[y_k^{(1)}, y_k^{(2)}, \ldots, y_k^{(n)}\right]^{\mathrm{T}}$ is the system observation vector, and $y_k^{(i)} = \left[s_k^{(i)}, v_k^{(i)}\right]$ is the observed position and speed information of the train. The measurement matrix $H = \left(h_{ij}\right)_{n\times n}, h_{ij} \in \mathbf{R}^{2\times 2}$ is a block diagonal matrix and $h_{ij} = diag\,[1,1]$. $V_k \sim \left(0, \sigma^2\right)$ is zero-mean Gaussian random observation noise.

3.2.2 The Train State Observation

The CBTC systems can be regarded as a discrete-time linear dynamic system and its noise term conforms to the Gaussian distribution. Under the premise that the system is observable, the Kalman filter is the optimal linear estimator that can minimize the mean square state error. The Kalman filter is an online estimator, and each iteration consists of a prediction step and a measurement value update step.

The first stage of the Kalman filter is the prediction stage, which estimates the current state based on the state estimate and input at the previous moment. This stage mainly calculates the state prediction and error covariance prediction, where the state prediction can be expressed as

$$\hat{X}_{k|k-1} = A\hat{X}_{k-1|k-1} + BU_{k-1} \tag{12}$$

where $\hat{X}_{k|k-1}$ is the state at time k predicted based on the estimated state value $\hat{X}_{k-1|k-1}$ at time $k-1$ and the input U_{k-1}.

The error covariance prediction $P_{k|k-1}$ can be expressed as

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \tag{13}$$

where $P_{k|k-1}$ is the predicted state estimation error covariance, represents the uncertainty of the predicted state $\hat{X}_{k|k-1}$, and Q represents the process noise covariance.

The second stage is the update stage. After obtaining the observation value Y_k, the predicted state and the actual observation are combined in the update stage to obtain a more accurate posterior state estimate. The Kalman gain can be expressed as

$$K_k = P_{k|k-1}H^T\left(HP_{k|k-1}H^T + R\right)^{-1} \tag{14}$$

where K_k determines the weight between the predicted state and the observation. The larger K_k is, the greater the importance of the observation. The smaller K_k is, the greater the importance of the prediction. H is the observation matrix, and R is the observation noise covariance.

The state update $\hat{X}_{k|k}$ can be expressed as

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k\left(Y_k - H\hat{X}_{k|k-1}\right) \tag{15}$$

where $Y_k - H\hat{X}_{k|k-1}$ represents the error between the observed value and the predicted value, and $\hat{X}_{k|k}$ is the updated state estimate.

The error covariance update can be expressed as

$$P_{k|k} = \left(I - K_kH\right)P_{k|k-1} \tag{16}$$

where I is the unit matrix. $P_{k|k}$ reflects the uncertainty of the updated state. After the update, the error covariance decreases because the introduction of the observed value reduces the uncertainty of the state.

3.2.3 The Unscented Kalman Filter

When the train control system is suffered by a cyber attack, the system observation equation may deviate. For example, a deviation in the observed position of the train may lead to inaccurate calculation of the movement authorization, resulting in emergency braking or stopping of the train. In that case, the classical Kalman filter will overestimate or underestimate the covariance matrix of the system, causing the state estimate to deviate from the actual value, and making it difficult for the classical Kalman filter to provide accurate state estimation. Therefore, the unscented Kalman filter (UKF) is introduced to estimate the train state under cyber attack.

The unscented Kalman filter is a nonlinear filtering method that nonlinearly maps the distribution of the state (approximately represented by the Sigma point) to the observation space through the observation equation to obtain the approximate statistical characteristics of the observation distribution. The calculation steps can be expressed as

1. *Initialization* We initialize the state vector $\hat{X}_0$ and state error covariance matrix P_0, which can be denoted as

$$\hat{X}_0 = E\left[X_0\right] \tag{17}$$

$$P_0 = E\left[\left(X_0 - \hat{X}_0\right)\left(X_0 - \hat{X}_0\right)^T\right] \tag{18}$$

2. *Sigma point generation* At time t, $2N+1$ Sigma points are generated for the state vector $\hat{X}_{k-1|k-1}$ and covariance $P_{k-1|k-1}$. It can be denoted as

$$\chi_{k-1|k-1}^{(i)} = \begin{cases} \hat{X}_{k-1|k-1} & s = 0 \\ \hat{X}_{k-1|k-1} + \left[\sqrt{(n+\lambda)P_{k-1|k-1}}\right]_s & s = 1, 2, \ldots, N \\ \hat{X}_{k-1|k-1} - \left[\sqrt{(n+\lambda)P_{k-1|k-1}}\right]_s & s = N+1, N+2, \ldots, 2N \end{cases} \tag{19}$$

And the weights of each Sigma point can be expressed as

$$\omega_m^{(i)} = \begin{cases} \dfrac{\lambda}{N+\lambda}, & i = 0 \\ \dfrac{1}{2(N+\lambda)}, & i = 1, 2, \ldots, 2N \end{cases} \tag{20}$$

where n is the state dimension and λ is the scaling parameter.

3. *State prediction* For each Sigma point, the Sigma point at the next moment is predicted by the state equation, that is

$$\chi_{k|k-1}^{(i)} = A\chi_{k-1|k-1}^{(i)} + BU_k \tag{21}$$

Then the state prediction value and covariance are

$$\hat{X}_{k|k-1} = \sum_{i=0}^{2n} \omega_m^{(i)} \chi_{k|k-1}^{(i)} \tag{22}$$

$$P_{k|k-1} = \sum_{i=0}^{2n} \omega_m^{(i)} \left[\chi_{k|k-1}^{(i)} - \hat{X}_{k|k-1}\right]\left[\left[\chi_{k|k-1}^{(i)} - \hat{X}_{k|k-1}\right]\right]^T + Q_k \tag{23}$$

4. *Observation prediction* The observation value of the Sigma point is predicted by the observation equation, that is

$$\hat{Y}_{k|k-1} = \sum_{i=0}^{2n} \omega_m^{(i)} Y_{k|k-1}^{(i)} \tag{24}$$

$$P_{Y_k Y_k} = \sum_{i=0}^{2n} \omega_m^{(i)} \left[Y_{k|k-1}^{(i)} - \hat{Y}_{k|k-1} \right] \left[Y_{k|k-1}^{(i)} - \hat{Y}_{k|k-1} \right]^T + R_k \tag{25}$$

$$P_{X_k Y_k} = \sum_{i=0}^{2n} \omega_m^{(i)} \left[\chi_{k|k-1}^{(i)} - \hat{X}_{k|k-1} \right] \left[Y_{k|k-1}^{(i)} - \hat{Y}_{k|k-1} \right]^T \tag{26}$$

5. *State estimate update* The state estimate can beupdated using Kalman gain, which can be expressed as

$$K_k = P_{X_k Y_k} P_{Y_k Y_k}^{-1} \tag{27}$$

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \left(Y_k - \hat{Y}_{k|k-1} \right) \tag{28}$$

$$P_{k|k} = P_{k|k-1} - K_k P_{Y_k Y_k} K_k^T \tag{29}$$

3.2.4 The Train State Observation under Cyber Attacks

If a data tampering attack is launched at time τ, the system observation equation can be modeled as

$$Y_k = HX_k + V_k + b_k \mathrm{I}_{\{k \geq \tau\}} \tag{30}$$

where b_k represents the bias term caused by the data tampering attack, and $\mathrm{I}_{\{k \geq \tau\}}$ is the indicator function, indicating that the tampering attack is triggered at time τ.

If a denial of service (DoS) attack occurs at time τ, the system observation equation can be modeled as

$$Y_k = D_k \left(HX_k + V_k \right) \tag{31}$$

where $D_k = \mathrm{diag}(d_1, d_2, \ldots)$ represents a diagonal matrix of 0 and 1. If an attack occurs, $d_i = 0$, indicating that the corresponding observation is invalid, and the associated communication channel is interrupted. Otherwise, $d_i = 1$. Note that at time $k < \tau$, $D_k = I_k$.

In the case of data tampering attack, the Sigma points in the state space of the system are projected into the observation space through the observation equation, generating a set of Sigma points corresponding to the observation space. This can be expressed as

$$\gamma_k^{(i)} = H\chi_k^{(i)} + b_k \mathrm{I}_{\{k \geq \tau\}} \tag{32}$$

Thus, the observation equation can be expressed as

$$\hat{Y}_k = \sum_{i=0}^{2n} \omega_m^{(i)} \gamma_k^{(i)} \tag{33}$$

The observation error covariance and state-observation covariance are

$$P_{Y_k Y_k} = \sum_{i=0}^{2n} \omega_m^{(i)} \left(\gamma_k^{(i)} - \hat{Y}_k \right) \left(\gamma_k^{(i)} - \hat{Y}_k \right)^T + R_k \tag{34}$$

$$P_{X_k Y_k} = \sum_{i=0}^{2n} \omega_m^{(i)} \left(\chi_{k|k-1}^{(i)} - \hat{X}_k \right) \left(\gamma_k^{(i)} - \hat{Y}_k \right)^T \tag{35}$$

The state estimation can be updated using the Kalman gain, that is

$$K_k = P_{X_k Y_k} P_{Y_k Y_k}^{-1} \tag{36}$$

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \left(Y_k - \hat{Y}_k \right) \tag{37}$$

$$P_{k|k} = P_{k|k-1} - K_k P_{Y_k Y_k} K_k^T \tag{38}$$

In the case of denial-of-service (DoS) attack, the Sigma points of the system state, that is, the sample points in the state space, can be projected into the observation space through the observation equation, generating a set of Sigma points corresponding to the observation space. That is

$$\gamma_k^{(i)} = D_k \left(H \chi_k^{(i)} \right) \tag{39}$$

Thus, the observation equation can be expressed as

$$\hat{Y}_k = \sum_{i=0}^{2n} \omega_m^{(i)} \chi_k^{(i)} \tag{40}$$

The observation error covariance and state-observation covariance are denoted as

$$P_{Y_k Y_k} = \sum_{i=0}^{2n} \omega_m^{(i)} \left(\gamma_k^{(i)} - \hat{Y}_k \right) \left(\gamma_k^{(i)} - \hat{Y}_k \right)^T + D_k R_k D_k^T \tag{41}$$

$$P_{X_k Y_k} = \sum_{i=0}^{2n} \omega_m^{(i)} \left(\chi_{k|k-1}^{(i)} - \hat{X}_{k|k-1} \right) \left(\gamma_k^{(i)} - \hat{Y}_k \right)^T \tag{42}$$

Similarly, the state estimate $\hat{X}_{k|k}$ can be updated using the Kalman gain K_k.

3.2.5 The Impact of Cyber Attacks on the Physical Layer

The state estimate $\hat{X}_{k|k}$ obtained using the unscented Kalman filter can be used to calculate the actual running time of the train. The train state deviation under cyber attack is defined as the difference between the estimated state and the true state, that is

$$\Delta X_k = \hat{X}_{k|k} - X_k = \begin{bmatrix} \hat{s}_k - s_k \\ \hat{v}_k - v_k \end{bmatrix} = \begin{bmatrix} \Delta s_k \\ \Delta v_k \end{bmatrix} \tag{43}$$

The components of the state deviation ΔX_k are the position deviation Δs_k and the velocity deviation Δv_k, which can be expressed as

$$\Delta s_k = \hat{s}_k - s_k \tag{44}$$

$$\Delta v_k = \hat{v}_k - v_k \tag{45}$$

The train delay time caused by the attack can be expressed as

$$\Delta T_d = \sum_{k=1}^{N_k} \frac{\Delta s_k}{v_k} \tag{46}$$

where k represents the time steps and N_k represents the sum of the time steps of the train operation. That is, the time required for a train to travel from one station to the next is divided into N_k discrete time steps.

Based on the train delay time, the train punctuality rate and the train departure interval can be calculated, which are key indicators for analyzing the system's availability. The train control system relies on a large amount of real-time data (such as train location, speed, signal status, etc.) to dispatch and control train operations. Once the train control system suffers from a cyber attack, train data may be tampered with or lost, causing train delays and reducing train punctuality. The train punctuality rate can be defined as

$$R = \left(1 - \frac{N_d}{N_z}\right) \times 100\% \tag{47}$$

$$N_d = \left\lfloor \frac{\Delta T_d}{T_{threshold}} \right\rfloor \tag{48}$$

where $\lfloor \cdot \rfloor$ represents the floor function (rounding down). $T_{threshold}$ is the threshold for determining delays, usually set to 3 min. N_d represents the number of delays exceeding the threshold, determined as the number of delayed trains. N_z represents the total number of trains operated.

The train departure interval is an important indicator to measure the efficiency of train services. Once a train suffers from a cyber attack, the train will be delayed and the train departure interval will increase, thus affecting the overall information security of the train control system. The train departure interval can be defined as

$$I = \frac{1}{N_t - 1} \sum_{i=1}^{N_t - 1} \left| F_{i+1} + \Delta T_{d,i} - F_i \right| \tag{49}$$

where F_{k+1} indicates the departure time of train $k+1$ in the train timetable, F_i represents the departure time of train i in the train timetable, $\Delta T_{d,i}$ represents the delay time of the current train i. N_t represents the total number of trains observed within the statistical period.

In summary, from the perspective of the physical layer, the total impact of cyber attack C_p, which consists of R and I, can be expressed as

$$C_p = \omega_{p1} \cdot R + \omega_{p2} \cdot I \tag{50}$$

where ω_{p1} and ω_{p2} represent the weights of train punctuality rate and train departure interval, respectively.

3.3 Situation Assessment for the Information Layer of CBTC Systems

In our hierarchical security situation assessment approach, system vulnerabilities, system threats, and network characteristics are used to quantify the security situation of the information layer. The specific indicators are discussed in detail.

3.3.1 System Vulnerability Analysis

The vulnerability analysis is to assess the changes in the security situation caused by the exposure of system vulnerabilities. Vulnerabilities exist in both the host nodes and system services, and attackers can exploit these vulnerabilities to launch attacks. Specifically, known and unknown vulnerabilities are discussed separately, and an integrated vulnerability evaluation method is proposed.

The probability of successfully implementing an attack is considered to be the probability that the vulnerability is exploited [32]. And the probability of successful exploitation of a vulnerability can be described from the aspects of attack vector (S_AV), attack complexity (S_AC), privileges required (S_PR),

user interaction (S_UI), and $f_v(x)$. It can be denoted as

$$p_{v_i} = S\_AV \times S\_AC \times S\_PR \times S\_UI \times f_{v_i}(x_{v_i}) \tag{51}$$

$$f_{v_i}(x_{v_i}) = \frac{\alpha \cdot k^{\alpha}}{x_{v_i}^{\alpha+1}} \tag{52}$$

where S_AV, S_AC, S_PR, and S_UI are the exploitability subscores obtained from CVSS, respectively. $f_v(x_v)$ is the statistical distribution of vulnerability exploit probability, characterized by the Pareto distribution [33]. x denotes the age of vulnerability v_i, and both $k = 0.00161$ and $\alpha = 0.26$ are constants.

The severity of system host and service vulnerabilities depends on when the vulnerabilities are discovered. Generally, the longer the vulnerability exists, the smaller the impact weights, because these vulnerabilities can be analyzed and patched over time. For the known vulnerabilities, the Common Vulnerability Scoring System (CVSS) is introduced to quantify the vulnerability impact, and the natural logarithm is used to control the magnitude of variables. The vulnerability impact degree of known vulnerabilities is expressed as

$$I_{v1} = 10 \ln\left(1 + \sum_{i=1}^{N_V} p_{v_i} \cdot e^{-\beta_v \cdot Age(v_i)} \cdot C_{v_i}\right) \tag{53}$$

where N_V is the total number of known vulnerabilities, β_v is the parameter that controls how fast the factor decays, $Age(\cdot)$ is the time (days) of vulnerabilities existence, C_{v_i} is the CVSS scores of vulnerability v_i.

The unknown vulnerabilities are those that already exist but have not been exposed. For the unknown vulnerabilities, the probability of vulnerability exposure can be estimated [34]. Generally, the more vulnerabilities there are in a period, the greater the probability of vulnerabilities in the later period. The Bayesian theorem is used to calculate the probability of unknown vulnerability exposure according to the statistical probability of vulnerability exposure over a certain period of time [35]. The calculation is expressed as

$$p(A|B) = p(A) \cdot \frac{p(B|A)}{P(B)} \tag{54}$$

where $p(A)$ is the statistical average probability of industry vulnerability exposure, $p(B)$ is the probability of vulnerability exposure statistics in the past period, $p(B|A)$ is the statistical probability of new vulnerabilities caused by previous vulnerabilities [36].

The threat impact degree of unknown vulnerabilities can be calculated with

$$I_{v2} = p(A|B) \cdot I_{v1} \tag{55}$$

where $p(A|B)$ is the unknown vulnerability exposure probability obtained on the premise that the vulnerability exposure probability is known over a period of time T_e.

Although some unknown vulnerabilities may have no relationship with the previously exposed vulnerabilities. Without losing generality, the probability of new vulnerabilities can still be estimated using vulnerability probability statistics to describe the overall impact of system vulnerabilities on security status. The impact of the system vulnerabilities can be expressed as

$$I_v = \alpha_1 \cdot I_{v1} + \alpha_2 \cdot I_{v2} \tag{56}$$

where α_1 and α_2 are the vulnerability evaluation weights of known and unknown vulnerabilities, respectively.

3.3.2 System Threat Analysis

Threat situation refers to the real-time quantification of the number of alerts and their severity in train control systems. Generally, threats are defined as potential attack behaviors that may harm the train control system. The train control systems face security threats that may compromise the availability, integrity, and confidentiality of devices, networks, and data. These threats, which can originate internally or externally, primarily include malware, resource attacks, content attacks, insider threats, identity and access control threats, and data breaches.

- **Malware** Malicious software such as viruses and worms can spread through train control networks, infecting computers and network devices in the system. This can lead to system paralysis, data loss, or theft, directly endangering train operation safety.
- **Resource Attacks** Resource attacks target the computational or network resources of the train control system. Examples include consuming excessive system computing power or occupying communication bandwidth. Denial-of-Service (DoS) attacks or Distributed Denial-of-Service (DDoS) attacks, for instance, exhaust system resources by sending massive amounts of fake requests, rendering the system unable to operate normally.
- **Content Attacks** The target of the content attack is the content of information exchange in the train control system, and the sensitive data of the train control system is obtained or tampered with by means of monitoring, deletion, tampering and deception. For example, communication data is intercepted and tampered with a man-in-the-middle attack (MITM), impersonating a legitimate communication object, resulting in incorrect train control instructions or data tampering.
- **Insider Threats** Malicious actions from insider personnel exploit their privileges to perform malicious operations, such as modifying control commands or leaking sensitive information. Unauthorized access by insiders can lead to system damage or data theft, causing train operation interruptions or severe accidents such as collisions.
- **Ientity and Access Control Threats** Administrator credentials are stolen, allowing attackers to access the system without authorization and conduct privilege escalation attacks. Additionally, lax internal access control can lead to permission misuse or unauthorized access, allowing illegal operations or data tampering within the system.
- **Data Breaches** There is a large amount of business data and operation data in the train control system, including sensitive information such as train location and running speed. Once this data is leaked, it will lead to privacy exposure or security risks, or even be used for malicious purposes, which may cause serious train safety accidents.

The impact of attacks with varying levels of severity on the system differs; attacks with higher severity indices pose greater threats than those with lower indices. By referencing the SNORT manual, the attack severity is introduced to quantify the cyber attacks. Alert priorities are categorized into High, Medium, Low, and Very Low, corresponding to severity quantification values 4, 3, 2, 1, respectively. High-priority alerts indicate the most severe attacks, while low-priority ones represent the least severe. Additionally, undetected anomalies in system sessions are considered potential risks with very low attack severity and are included in the threat landscape quantification. Typical network attacks faced by train control systems and their corresponding severity indices are listed in Table 1.

Table 1: The severity value of attacks

The typical cyber attacks	Attack severity	g_s
Content attacks, malware, privilege escalation attacks	High	4
DoS attacks, leakage attacks	Medium	3
Probe, Scan	Low	2
Sessions	Very low	1

Note that the high-severity attacks have greater threat impact than low-severity attacks, the exponent calculation is introduced to represent the impact of attacks with different severities. The impact of system threats can be described as

$$T_c = \sum_{j=1}^{N_c} c_j(t) \times 10^{g_j(t)} \times e^{-\alpha(t-t_0)} \tag{57}$$

where N_c represents the number of alert categories, $c_j(t)$ represents the alert numbers of the jth category generated at time t, $g_j(t)$ represents the severity of different types of attacks, $e^{-\alpha(t-t_0)}$ is the time decay factor, α is the adjustment parameter, and t_0 is the initial time of the alert.

The purpose of adding the time decay factor is to make the threat situation quantification meet the real-time requirements of the train control system. The time decay factor can dynamically update the threat situation to ensure that the threat situation value can reflect the current threat situation without accumulating a large amount of outdated alarm information, thus avoiding the excessive impact of past alarms on the current train control status.

3.4 Network Characteristics Analysis Using the GAN-AIS

The characteristic of CBTC systems operation is that it operates on a specific line in a fixed direction according to the train operating diagram, and its information interaction and data flow are stable. The artificial immune system can take advantage of the stable data flow to self-learn the data characteristics of the CBTC systems under normal operation, train the immune system model without prior knowledge or data labels, and realize cyber attack perception and security situation assessment.

In AIS, the antigen is the feature vector of network characteristics data. Let the antigen set be $Ag = \{ag | ag \in S^l\}$, where S^l is the shape space, l represents the dimension of feature strings, including source/destination IP address, source/destination port, protocol type, time, features, traffic, sessions, and so on. The antibody is the detector that recognizes specific antigens. The detector simulates the main functions of immune cells and realizes the main functions such as self-tolerance, cell cloning, and mutation evolution. Let the detector set be

$$D = \{< d, cnt, age, \rho > \mid d \in D, cnt \in \mathbb{N}, age \in \mathbb{N}, \ \rho \in \mathbb{R}\} \tag{58}$$

where d represents the detector, cnt represents the sum of antigens matched by the detector, age represents the age of detectors, ρ represents the concentration of detectors, $\mathbb{N}$ and $\mathbb{R}$ represent the natural number and real number set.

3.4.1 The Mature Process of Detectors

Generally, the purpose of the immune mechanism is to identify unknown foreign antigens and protect the body, and the process of identifying antigens is the process of calculating the affinity between antigens and antibodies. The affinity for the new antigen is the measure of the matching degree between the antibody and the antigen. For the detector with l-dimensional feature vectors, the Minkowski distance is used to calculate the affinity, which can be denoted as

$$D_f(d, ag) = \left(\sum_{i=1}^{l} |d_i - ag_i|^{\lambda} \right)^{\frac{1}{\lambda}} \tag{59}$$

where λ is the parameter of Minkowski distance. When $\lambda = 1$, it is the Manhattan distance and when $\lambda = 2$, it is the Euclidean distance.

The affinity can be calculated as

$$f_a(d, ag) = \frac{1}{D_f(d, ag) + \varepsilon} \tag{60}$$

where ε is a small constant. The closer the Minkowski distance between the antigen and the detector, the greater the affinity. When the affinity is greater than the threshold, the antigen is matched to the detector.

The maturation process of the detector needs to experience self-tolerance. The self-tolerance is an immune response in which the detector does not react to the autoantigens. Let $Self \subset Ag$ represents the set of normal network characteristics, $Non\_self \subset Ag$ represents the set of abnormal network characteristics, and $Self \cup Non\_self = Ag$, $Self \cap Non\_self = \varnothing$. The self-tolerance process can be described as

$$f_t(d, s) = \begin{cases} 0, & \exists s \in Self \wedge f_m(d, s) = 1 \\ 1, & otherwise \end{cases} \tag{61}$$

$$f_m(d, s) = \begin{cases} 1, & f_a(d, s) > \gamma_a \\ 0, & otherwise \end{cases} \tag{62}$$

where $s \in Self$, γ_a is the threshold of affinity.

When the detector d cannot match each self s of the self-set $Self$, that is $f_t(d, s) = 1$, the process can be denoted as the self-tolerance process. When $f_t(d, s) = 0$, it means that the self-tolerance process of the detector has failed, and the detector will be removed. Then the other detectors turn into the mature detectors D_{ma}. It can be denoted as

$$D_{ma} = D_{ma} \cup \{d | d \in D, \forall s \in Self \wedge f_t(d, s) = 1\} \tag{63}$$

3.4.2 The Immune Response of Detectors

In AIS, the immune response is the process of identifying and detecting cyber attacks. Similar to the self-tolerance process, the detection of attack is judged by calculating the affinity of the mature detectors to the antigen. If the affinity between the mature detector and antigen is greater than the preset threshold γ_d, an attack can be detected. If a mature detector can continuously detect cyber attacks, and the cumulative number of matches reaches the threshold N_c, the mature detector turns into the memory detector D_{me}. It can be denoted as

$$D_{me} = D_{me} \cup \{d | d \in D_{ma}, d.cnt \geq N_c \wedge d.age \leq L\} \tag{64}$$

where cnt is the sum of antigens matched by the detector, age is the life generations of detectors, L is the life cycle of memory detectors. The memory detectors have higher detection priority and do not need to go through the maturation process again. However, the memory detector is limited by the life cycle L. When the memory detector reaches the life cycle L, it will be removed from the detector set.

3.4.3 The Mutation Evolution of Detectors

The evolutionary mutation process of detectors is an important way to maintain the diversity of detectors and improve the quality of detectors. The clone selection algorithm (CSA) [37] is used to realize the cloning and mutation process of the detector. In the CSA, the memory detectors with high affinity are selected for the cloning process. The selected detectors are stimulated to mutate according to the mutation mechanism, and the mutated detectors are classified into the immature detector set to start a new life cycle. When applying the traditional mutation algorithm to attack detection, certain challenges may arise. For instance, in the event of a singular type of attack, the diversity within the detector population might be insufficient, leading to limited effectiveness in the cloning and mutation processes. Additionally, the random mutation pattern may fail to maintain the dominant gene generation, resulting in an extended maturation process for mutated detectors.

The GAN simulates the cloning process of mature detectors by learning the characteristics of mature detectors and generating detectors with a small number of detector samples. The GAN can also simulate the mutation process of the detector by generating a large amount of synthetic data [38]. In the mutation process, a certain number of dominant features can be preserved by training the GAN model, and new feature communities can be generated to enrich the diversity of detector features. Specifically, the generator G and discriminator D are used for detector generation and detector discrimination, respectively. The mutated mature detectors D_{mu} can be denoted as

$$D_{mu} = D_{mu} \cup \{d | d \in G(z), \forall s \in Self \wedge f_t(s, d) = 1\} \tag{65}$$

where z is the random noise, $G(z)$ is the mature detector generator of the GAN that achieves Nash equilibrium.

The goal of the generator G is to generate detector profiles that closely resemble the real detector profiles, while the purpose of the discriminator D is to accurately detect the real detector profiles from the generated ones. With a period of training, the generator has a stronger ability to generate realistic detector profiles. The discriminator is more sensitive to the difference between the real and the generated detector profiles. The competitive relationship between the generator G and discriminator D promotes mutual evolution. With the GAN-based artificial immune system, the diversity of the detector population is enriched, and the convergence process after population mutation is accelerated. The cloned and mutated detectors re-enter the collection of immature detectors and start a new life cycle.

3.4.4 The Concentration Representation of Detectors

In the human immune system, the human body will trigger an immune response to the invasion of viruses. The body immune system cannot only detect the virus but also reflect the severity of the virus infection, for example, by the number of white blood cells. Due to the different severity of virus infection, the human body will trigger different degrees of immune response, and the intensity of immune response will change with the degree of infection. This process can be characterized by the concentration of antibodies.

Similar to body immunity, our proposed GAN-AIS method cannot only perceive cyber attacks, but also characterize the current security situation of the system. Because of the redundant structure of the train control system, isolated attacks may not affect the operation of the system, but the detector concentration

can still be used to reflect the security situation of the system, thus characterizing potential cyber-attacks. The ratio of the number of detectors satisfying the threshold of affinity between detectors to the number of all detectors is defined as the detector concentration. The calculation process can be described as

$$\rho_i = \frac{\sum_{j \in S, j \neq i} ds_{ji}}{N_d}, i \in S \tag{66}$$

$$ds_{ji} = \begin{cases} 1, & |f(j,i)| \leq \gamma_\varepsilon \\ 0, & others \end{cases} \tag{67}$$

$$S = D_{ma} + D_{me} + D_{mu} - D_d \tag{68}$$

where γ_ε is the affinity threshold between detectors, N_d is the total number of the detectors, S is the set of useful detectors, $D_d = \{d | d \in D_{me} \wedge d.age > L\}$ is the detectors that reach the end of life.

3.4.5 The Analysis of Network Characteristics

The data interaction and information communication are the basis for supporting the business functions of CBTC systems. The analysis of network characteristics can quickly and accurately perceive various attacks and characterize the current network security situation of the system. The mathematical description of the network security situation can be expressed as

$$F_c = 1 - \frac{1}{1 + \ln\left(1 + c \cdot \sum_{j=1}^{N_j} \sum_{i=1}^{N_i} \rho_{i,j}\right)} \tag{69}$$

where $c \in \mathbb{Z}^+$ is the adjustment factor, N_i is the number of detectors in the set S that are judged by the affinity between detectors, and the detector concentration is not equal to zero (the concentration of some isolated detectors is 0), N_j is the classification of cyber attacks.

3.4.6 Computational Complexity Analysis of GAN-AIS

In this section, we will analyze the computational complexity of the GAN-AIS approach. For the mature process of detectors, assuming that there are N_d detectors and N_{ag} antigens, and the number of detectors undergoing self-tolerance is M, then the computational complexity of the detector maturation stage can be expressed as $O\left(N_d \cdot \left(N_{ag} + M\right) \cdot l\right)$. For the immune process of detectors, assuming the number of mature detectors is N_{ma}, then the computational complexity of the immune response stage can be expressed as $O\left(N_{ma} \cdot N_{ag} \cdot l\right)$. For the mutation evaluation process of detectors, the mutation detectors are generated by a generative adversarial network (GAN). In the generator training phase, assuming the number of iterations is k, the computational complexity of each iteration is $O\left(N_{ma} \cdot l\right)$. In the discriminator training phase, assuming the number of iterations is k, the computational complexity of each iteration is $O\left(N_{ma} \cdot l\right)$. Then the complexity of generating the mutation detector is $O\left(k \cdot N_{ma} \cdot l\right)$. The computational complexity of the mutation stage is $O\left(k \cdot N_{ma} \cdot l\right)$. For the concentration calculation process of detectors, assuming that the number of useful detectors is N_u, then the computational complexity of the detector concentration calculation stage is $O\left(N_d \cdot N_u\right)$. The computational complexity of the network characteristics analysis stage is $O\left(N_j \cdot N_i\right)$. Therefore, the total computational complexity of the GAN-AIS approach is $O\left(\max\left(N_d \cdot \left(N_{ag} + M\right) \cdot l, N_{ma} \cdot N_{ag} \cdot l, k \cdot N_{ma} \cdot l, N_d \cdot N_u, N_j \cdot N_i\right)\right)$.

3.5 The Information Security Situation of the Information Layer

The system vulnerabilities, security threats, and network characteristics indicators are used to quantify the information security situation of the information layer. Specifically, the information security situation is composed of I_v, T_c and F_c. It can be expressed as

$$H_c = \omega_{c1} \cdot I_v + \omega_{c2} \cdot T_c + \omega_{c3} \cdot F_c \tag{70}$$

where ω_{c1}, ω_{c2} and ω_{c3} are the weights of three indicators, respectively. The detailed algorithm is formulated in Algorithm 1.

Algorithm 1: The algorithm of the information security situation

Input: The system vulnerability, system alerts, network characteristic data of CBTC systems
Output: The information security situation value H_c

Initialize: γ_a, γ_d, γ_ε, T_e.
for *each vulnerability* V_i *of the system* **do**
    Calculate the probability of successful exploitation of a vulnerability p_v.
    Calculate the impact degree of known vulnerabilities I_{v1}.
end
for *each period of time* T_e **do**
    Estimate the impact degree of unknown vulnerabilities I_{v2}.
    Calculate the impact degree of system vulnerabilities I_v.
end
for *each period of time* T_e **do**
    Count and classify system alert information $c(t)$.
    Retrieve the severity of the cyber attack $g(t)$.
    Calculate the impact of system threats T_c.
end
while *the iterative stop criterion not reached* **do**
    for *each antigen Ag at preset time intervals* **do**
        Steps 1: the mature process of detectors.
        Calculate the affinity of antigen *Ag* and randomly initial detector *D*.
        Complete the self-tolerance process of the detector *D*.
        Generate the mature detectors set D_{ma}.
        Steps 2: the immune response of detectors.
        Perceive and detect the cyber attacks.
        Generate the memory detectors set D_{me}.
    end
    Steps 3: the mutation evolution process of detectors.
    Generate the mutated detectors set D_{mu} and update the memory detectors set D_{me}.
    Steps 4: the concentration representation of detectors.
    Calculate the concentration of detector ρ.
    Steps 5: the analysis of system network characteristics.
    Obtain the network security situation value F_c.
end
Calculate the information security situation H_c.
return H_c.

3.6 The Situation Assessment for CBTC Systems

Both physical layer and information layer indicators are used for the hierarchical security situation assessment approach. For the physical layer, the main focus is on the impact of cyber attacks on the operation of the CBTC systems, and the total cost is measured from the perspective of time delay and economic loss. For the information layer, changes in system situation are analyzed and characterized from aspects of static and dynamic indicators. The comprehensive situation assessment value SA can be expressed as

$$SA = \omega_1 \cdot C_p + \omega_2 \cdot H_c \tag{71}$$

where ω_1 and ω_2 are the weights of two indicators, respectively.

4 Experiments and Evaluation

In this section, we present our simulation experiments from the perspective of experimental design, attack description, as well as data preprocessing and performance metrics.

4.1 Experimental Design and Attack Description

The simulation experiments are conducted on the security simulation platform of the National Key Laboratory. As shown in Fig. 3, the simulation platform is built with simulation software and real signal equipment, and supports both wired and wireless train-ground communication modes. The real line data of Beijing Metro Line 7 are used to simulate the tracking operation of multiple trains. In addition, the platform can simulate cyber attacks such as DoS attacks and data spoofing attacks, and restore the operating scenarios and operating status of the system under cyber attacks.

Figure 3: The security simulation platform

The experimental parameters of the hierarchical security situation assessment approach are listed in Table 2. According to the parameters, the overall impact of cyber attacks on the physical layer C_p, the

security situation of information layer H_c, and the comprehensive situation assessment value SA can be calculated by Eqs. (50), (70) and (71), respectively.

Table 2: The experimental parameters of hierarchical security situation assessment approach

Symbol	Values	Description
N_t	5	The number of trains
N_s	15	The number of stations
N_v	10	Number of known vulnerabilities
β_v	0.1	The decay factor
T_e	90	Observation period (days) for unknown vulnerabilities
ω_{p1}	0.5	Weight indicators of train punctuality rate
ω_{p2}	0.5	Weight indicators of train departure interval
ω_{v1}	0.8	Weight indicators of known vulnerabilities
ω_{v2}	0.2	Weight indicators of unknown vulnerabilities
ω_{c1}	0.3	Weight indicators of V_c
ω_{c2}	0.3	Weight indicators of T_c
ω_{c3}	0.4	Weight indicators of F_c
ω_1	0.5	Weight indicators of C_p
ω_2	0.5	Weight indicators of H_c

Cyber attacks are introduced to verify the effectiveness of our approach. The CBTC systems mainly suffer two types of attacks: DoS attack and data spoofing attack [39]. The DoS attack may infiltrate the CBTC network, and occupy excessive service resources with seemingly legitimate service requests, thus obstructing the system normal operation. The primary objective of the data spoofing attack is to manipulate the payload content within data packets, aiming to disrupt the functionality of CBTC physical equipment.

In the case of a DoS attack, the primary targets are the communication link between the ground wireless AP and VOBC. The authentication attack on the wireless network, known as an authentication DoS, is launched by using the Mdk3 wireless attack tool. This tool depletes AP authentication request resources by simulating randomly generated MAC addresses and sending a large number of authentication requests to the target AP. Consequently, the AP becomes unable to respond to legitimate communication requests. Meanwhile, the TCP DoS attack on wired backbone networks, specifically a TCP SYN Flood, is launched using the LOIC DoS attack tool. The attack floods the service connection queue with TCP data packets, causing network congestion. As a result, devices are unable to respond properly, disrupting the normal operation of trains.

In the case of a data spoofing attack, the payload contents of data packets are modified with the intent of disrupting train operations. The communication link between the ground wireless access point (AP) and VOBC in CBTC systems is used to transmit control commands and train status information. If this communication data is tampered with, it can compromise the safe operation of the train. A typical example is a Man-in-the-Middle (MitM) attack, where an attacker intercepts and modifies data packets between the targeted parties by setting up forwarding on a middleman device. In this paper, the Bettercap attack tool is employed to intercept and tamper with the communication data. By deploying lightweight Bettercap modules, the attacker manipulates MAC address tables and redirects communication, allowing them to act as middlemen to monitor or modify data between the target host and the gateway.

4.2 Data Preprocessing and Performance Metrics

The method in [40] is used to extract the features from the security feature data set. Since each situation indicator had a different value range, the Min-Max Normalization method is used for normalization, which can be expressed as

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{72}$$

where x^* is the normalized value of the situation indicator, x is the original value, $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the situation indicator, respectively.

The detection rate (DR), false alarm rate (FR) and F1_score are introduced to evaluate the effectiveness of perceiving cyber attacks, which can be described as

$$DR = Recall = \frac{TP}{TP + FN}, Precision = \frac{TP}{TP + FP} \tag{73}$$

$$FR = \frac{FP}{FP + TN} \tag{74}$$

$$F1\_score = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{75}$$

where TP represents the number of cyber attacks that are sensed and actually launched, FP represents the number of cyber attacks that are sensed but no actually launched, TN represents the number of cyber attacks that are neither sensed nor actually launched, FN represents the number of cyber attacks that actually launched but not sensed.

5 Results and Discussion

In this section, comparative experiments are conducted to evaluate the proposed approach. When the system suffers from different types of attacks, both the physical layer and information layer indicators are used to reflect the changes in the system security situation.

5.1 Experiments for Perceiving DoS Attack

First, we present the performance evaluation of the approach for perceiving DoS attacks. The batch size, epochs and noise dimension of the GAN model are 256, 100 and 100, respectively. The loss function and optimizer of the GAN model are binary_crossentropy and Adam, respectively. The activation functions of the GAN model generator G and discriminator D are tanh and sigmoid, respectively. The experimental parameters are listed in Table 3.

Table 3: The experimental parameters for perceiving DoS attacks

Symbol	Values	Description
λ	2	Parameter of Minkowski distance
ε	0.001	Constant of affinity calculation
γ_a	0.75	Affinity threshold of self-tolerance process
γ_d	0.25	Affinity threshold of immune response process

(Continued)

Table 3 (continued)

Symbol	Values	Description
γ_ε	0.5	Affinity threshold between detectors
N_c	5	Matching number of detectors
c	2	Adjustment factor

The detection rate and false alarm rate for perceiving the authentication DoS attacks are shown in Fig. 4. With the increase of the detector life cycle, the detection rate increases gradually, the false positive rate increases at first and then decreases. When the life cycle of the detector is set to 5 generations, the detection rate is 64.84%, and the false alarm rate is 0.08%. When the life cycle of the detector is set to 25 generations, the detection rate is increased to 96.83%, and the false alarm rate is 0.16%. When the life cycle of the detector is set to 50 generations, the detection rate is increased to 97.09%, and the false alarm rate is 0.06%. The detection rate of the detector with a life cycle of 25 generations is similar to that of 50 generations, and the false alarm rate is maintained at a low level. It can be noted that the time required for the maturity and evolution of the detectors with a life cycle of 25 generations is significantly lower than that of 50 generations. Therefore, it is considered that the detector with a life cycle of 25 generations is the optimal detector, and the performance of the detector is the best.

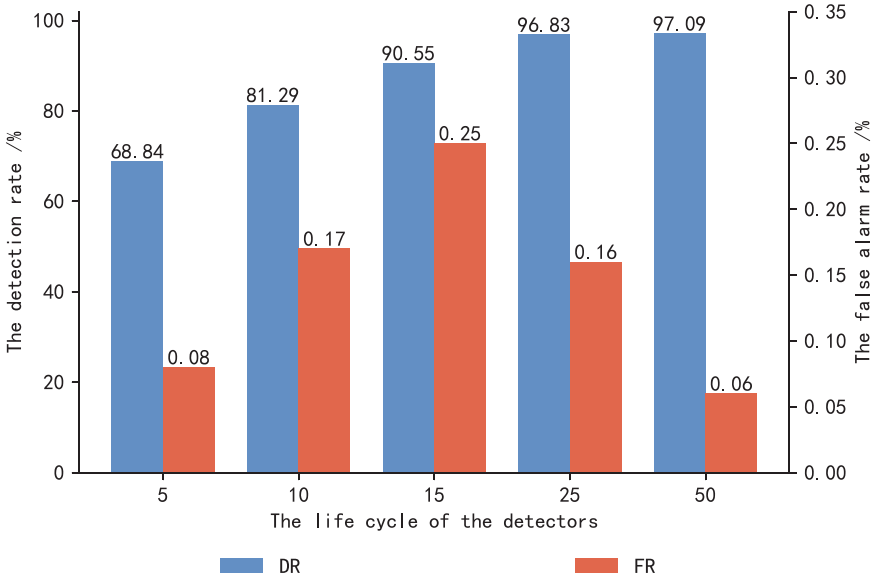


Figure 4: The detection rate and false alarm rate for perceiving the authentication DoS attacks

The detection rate and false alarm rate for perceiving the TCP SYN Flood attacks are shown in Fig. 5. It can be seen that when the life cycle of the detector is short, the false alarm rate is low. With the increase of the detector life cycle, the detection rate increases gradually, and the false positive rate increases at first and then decreases. The overall change trend is similar to the result of perceiving authentication DoS attack. When the life cycle of the detector is set to 25 generations, the detection rate is 98.59%, and the false alarm rate is 1.08%. It has a high detection rate and a low false alarm rate. It can be considered that the detector with a life cycle of 25 generations is the optimal detector.

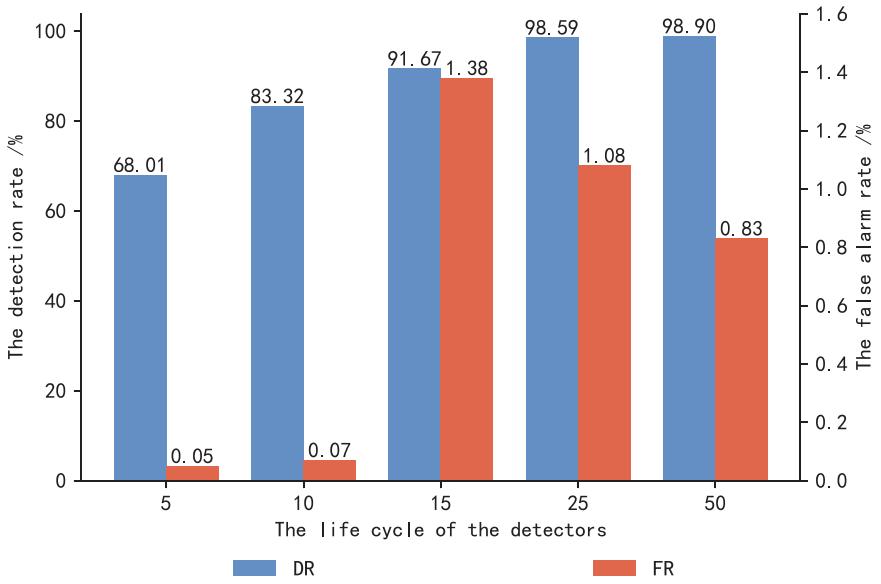


Figure 5: The detection rate and false alarm rate for perceiving the TCP SYN Flood attacks

As shown in Table 4, the different methods are introduced to compare the ability to perceive the TCP SYN Flood attack. The GAN-AIS method is compared with unsupervised learning methods such as AIS [41], convolutional autoencoder and one-class SVM (CAE-OCSVM) [42], synthetic minority oversampling technique and Random Forest (SMOTE-RF) [43], principal component analysis and K-means clustering (PCA-K-Means) [44] and variational autoencoder (VAE) [45] in terms of the detection rate, false alarm rate, F1_score and computation time indicators.

Table 4: Performance comparison of DR, FR, F1_score and computation time

Method	Detection rate (%)	False alarm rate (%)	F1_score	Computation time (s)
GAN-AIS	98.59	1.08	0.9921	490
AIS	97.09	1.90	0.9841	428
CAE-OCSVM	84.89	4.36	0.8969	821
SMOTE-RF	90.90	0.48	0.9499	796
PCA-K-means	92.45	1.94	0.9585	413
VAE	95.09	1.10	0.9575	923

It can be seen from Table 4, the detection rate of GAN-AIS is 98.59%, the false alarm rate is 1.08%, and the F1_score is 0.9921. It has the highest detection rate, a relatively low false alarm rate, and a high F1_score. The reason is that the GAN-AIS approach cannot only learn the network characteristics of the train control system but also further optimize and enrich the detector population through the GAN model. A higher detection rate usually increases the probability of false alarms. However, the proposed GAN-AIS approach can keep the false alarm rate within a relatively low range and maintain a high F1_score value. In addition, compared with other methods, the GAN-AIS approach has a relatively short computation time. Especially when compared with the variational autoencoder (VAE) method, which has a relatively close detection rate, the computation time of GAN-AIS can be reduced by nearly half. Moreover, compared with

methods that have a similar computation time, GAN-AIS has a higher detection accuracy. To sum up, the GAN-AIS approach has outstanding comprehensive perception performance.

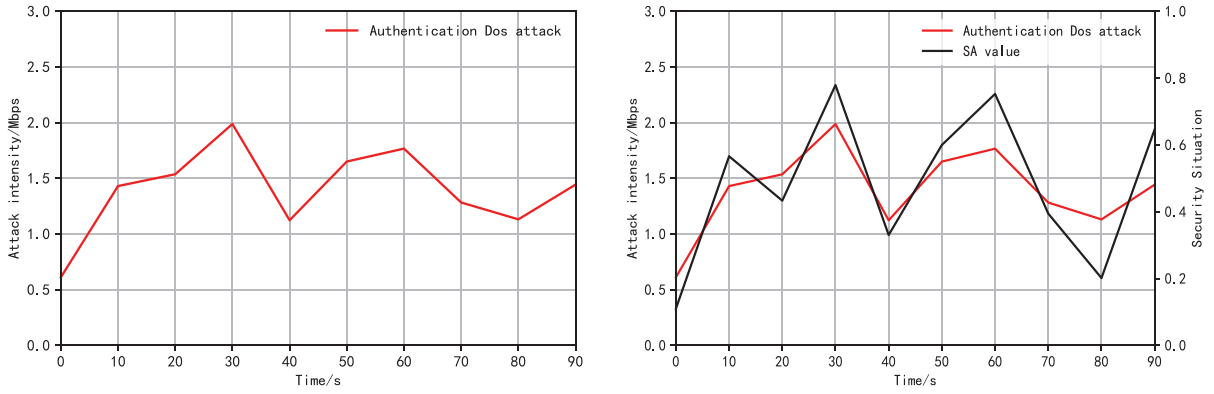
5.2 The Experiments of Situation Assessment under DoS Attack

We assume that due to the redundancy and security mechanism of the system, the DoS attacks will not impact the normal operation of the train, the comprehensive situation assessment value *SA* can be calculated without considering the impact of cyber attacks on the physical layer. Limited by the relatively closed environment, the CBTC systems are rarely updated and upgraded. Therefore, most of the existing vulnerabilities are outdated, but these vulnerabilities can still reflect the security situation of the system. In our experiments, the system vulnerabilities are scanned through Nessus software, and the reference date for vulnerability age calculation is 31 December 2019. The CVSS information of 10 typical vulnerabilities and the corresponding attack severity are listed in Table 5.

Table 5: The CVSS information of vulnerabilities and the corresponding attack severity

No.	Vulnerabilities	CVSS score	*S_AV*	*S_AC*	*S_PR*	*S_UI*	g_s	Release date
1	CVE-2019-2518	7.5	0.85	0.44	0.62	0.85	4	23/04/2019
2	CVE-2019-0708	9.8	0.85	0.77	0.85	0.85	4	06/05/2019
3	CVE-2018-17190	7.5	0.85	0.44	0.62	0.85	4	19/12/2018
4	CVE-2016-8735	9.8	0.85	0.77	0.85	0.85	4	06/04/2017
5	CVE-2016-0714	8.8	0.85	0.77	0.62	0.85	4	25/02/2016
6	CVE-2019-2753	4.6	0.85	0.77	0.62	0.62	3	23/07/2019
7	CVE-2019-0199	7.5	0.85	0.77	0.85	0.85	3	10/04/2019
8	CVE-2018-3299	8.2	0.85	0.77	0.85	0.62	3	16/10/2018
9	CVE-2019-17052	3.3	0.55	0.77	0.62	0.85	2	01/10/2019
10	CVE-2018-8207	4.7	0.55	0.44	0.62	0.85	2	14/06/2018

The change in authentication DoS attack during the 90-s observation period is simulated, and the intensity of the authentication DoS attack and the corresponding security situation are shown in Fig. 6. With the intensity of the attack changes, the security situation of the CBTC systems information layer changes, affecting the comprehensive security situation of the CBTC systems. The attack intensity is set from 0.6 M/s to 2.0 M/s, and the security situation value varies from 0.1 to 0.78. The trend of the security situation value curve is similar to that of the attack intensity curve. The greater the intensity of the DoS attack, the higher the security situation value, and vice versa. Moreover, due to the existence of memory detectors, when the same attack persists, the change in situation assessment values is more drastic than that of attack intensity.

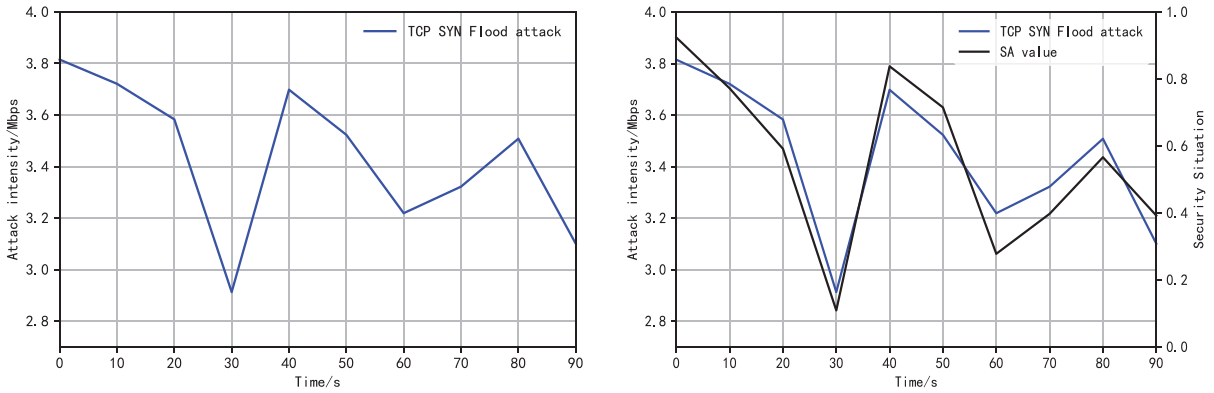


(a) The intensity of authentication DoS attack

(b) The change of security situation with different attack intensity

Figure 6: The changes in the intensity of authentication DoS attack and the corresponding security situation

As shown in Fig. 7, the TCP DoS attack intensity ranges from 2.9 to 3.8 M/s, and the security situation value varies from 0.1 to 0.92. The trend of the security situation value curve is similar to that of the attack intensity curve. It can be concluded that the security situation curve can characterize the dynamic changes in the security situation in real-time.



(a) The intensity of TCP SYN Flood attack

(b) The change of security situation with different attack intensity

Figure 7: The changes in the intensity of TCP DoS attack and the corresponding security situation

5.3 The Experiments of Situation Assessment under Data Tampering Attack

In the simulation experiment, the MitM data tampering attack disrupts the train operation by tampering with the data of the normal operation of the train. Due to the fail-safe mechanism of the CBTC systems, if the data received by the train fails to pass verification, the train will initiate an emergency braking procedure. A total of 15 stations and 5 trains are used to simulate the train operation process. The train tracking interval is set to 3 min to simulate train operation during the morning rush period, and the station stop time is set to 30 s. To simplify the experiment, only the upward trains are considered. The data tampering attack is executed during the train operation, and the train operation diagram is affected. The train operation diagram under the cyber attack and the change of situation assessment value are shown in Figs. 8 and 9, respectively.

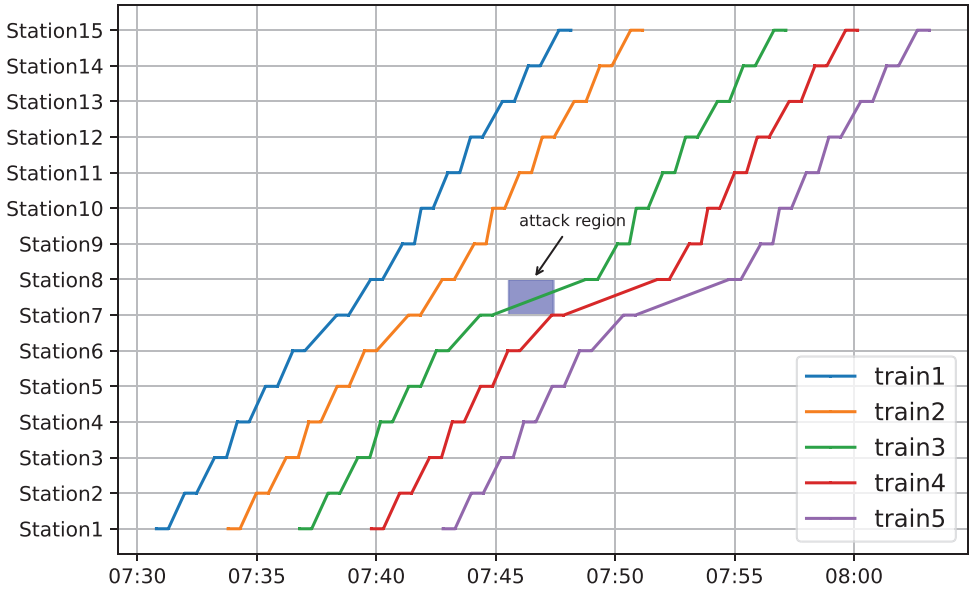


Figure 8: The train operation diagram during morning rush period under a cyber attack

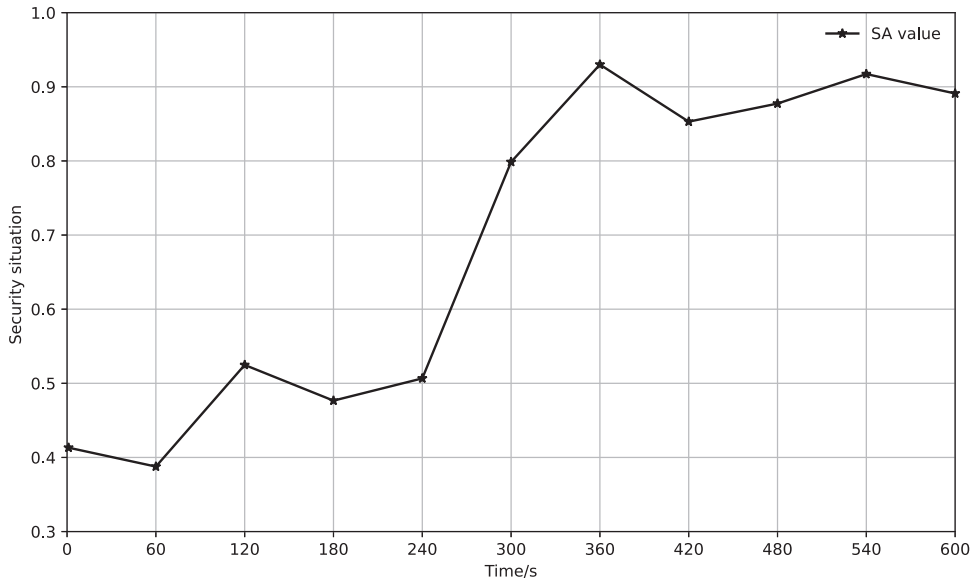


Figure 9: The change of situation assessment value with multiple trains affected by a cyber attack

It can be seen from Fig. 8, the first two trains ran normally according to the train schedule, while the third train suffered a cyber attack when it was running on the track between Station 7 and Station 8. Due to the necessary remedial measures (e.g., equipment restarts and patch upgrades), the third train resumed normal operation 3 min later. But it still affects the normal operation of subsequent trains. The impact of cyber attacks on the physical layer of the train control system can be characterized through the quantitative value of situation assessment. As shown in Fig. 9, the situation assessment value is kept at a low value at first. At the 40th second of the observation period, the train control system suffered a cyber attack, and the system situation value increased significantly. The reason is that the attack had an impact on the physical layer of the train control system, resulting in the time delay and economic losses of several successive trains, which

led to the increases of situation assessment value. With the subsequent mitigation or elimination measures taken, the situation becomes stable. However, as the departure interval of the trains is very short, subsequent trains continue to be affected, and the security situation remains at a high level.

For comparison, another experiment is conducted. The train tracking interval is set to 8 min to simulate the train operation in the off-peak period, and the other experiment settings are the same. The train operation diagram under cyber attack and the change of situation assessment value is shown in Figs. 10 and 11, respectively.

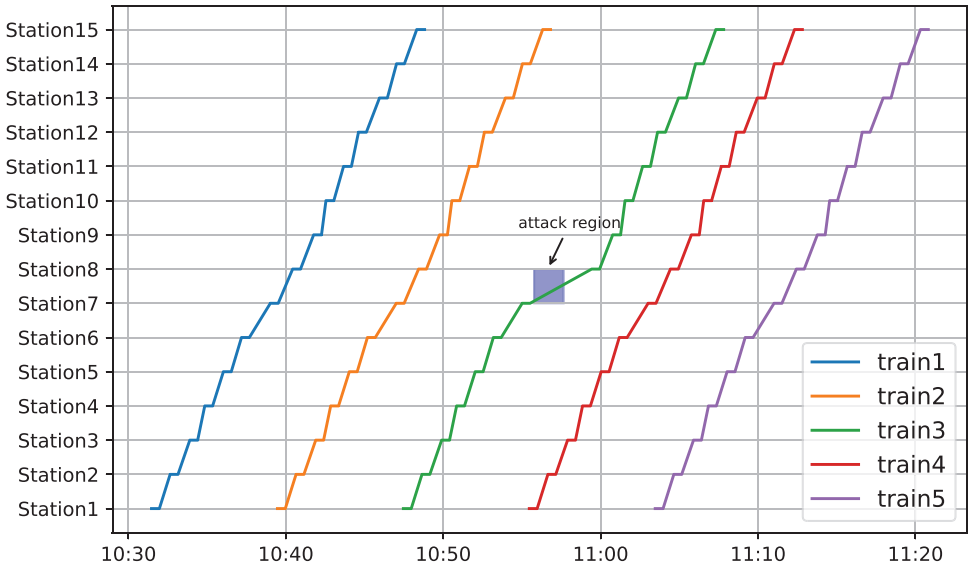


Figure 10: The train operation diagram during flat peak period under a cyber attack

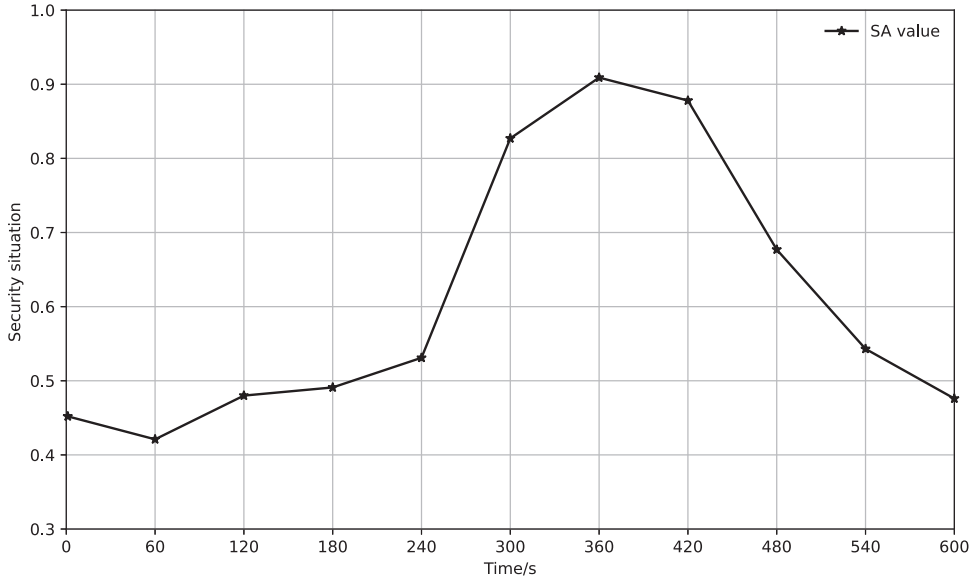


Figure 11: The change of situation assessment value with a single train affected by a cyber attack

It can be seen from Fig. 10, the first two trains run according to the train schedule, while the third train suffered a cyber attack when running on the track between Station 7 and Station 8. Due to the large train operation interval, the subsequent trains are not affected by the third train, and the third train also resumes normal operation 3 min later. As shown in Fig. 11, the situation assessment value initially remains at a low value. At the 40th second of the observation period, the physical layer of the train control system suffered a cyber attack, and the system situation value increased significantly. With the mitigation of the cyber attack, the train operation returned to normal. The security situation value of the train control system decreased to the normal level.

By comparison, it can be concluded that attacks on the train system during the morning peak and off-peak periods have different impacts on the system security situation. During the morning peak, an attack affects a larger number of subsequent trains, resulting in a sustained high security situation value. In contrast, during off-peak periods, the attack causes delays to a single train without affecting the operation of subsequent trains. This is reflected in the security situation, where the situation value initially rises but then returns to normal.

6 Conclusion

This paper proposed a hierarchical security situation assessment approach for CBTC systems, enabling the detection of cyber attacks and evaluating of security situations from the information and physical layers. Specifically, the security situation of the physical layer was assessed from the perspective of the impact of cyber attacks, with train punctuality rate and train departure interval serving as key indicators to quantify the security situation under cyber attacks. Meanwhile, the security situation of the information layer was evaluated at both static and dynamic levels, with system vulnerabilities and system threats acting as static indicators, and network characteristics of CBTC systems serving as dynamic indicators. By integrating the evaluation results from both the information and physical layers, a comprehensive security situation assessment value was obtained.

Experimental results demonstrate that for DoS attacks, the proposed approach achieved the highest detection rate, a low false alarm rate, and a high F1_score, with a detection rate of 98.59%, a false alarm rate of 1.08%, and an F1_score of 0.9921. Additionally, the antibody concentration of the GAN-AIS model was used to dynamically characterize the security situation of CBTC systems in real-time. The higher the attack intensity, the greater the antibody concentration and the higher the security situation value. For the data tampering attack, when the trains were running during the morning rush period, the data tampering attack would affect the state of multiple trains, causing the system security situation value to change from low to high and remain at a high level. In contrast, when the trains were running during the off-peak period, the data tampering attack primarily affected the operation state of a single train, the security situation value initially increased and then gradually decreased with the mitigation of the cyber attack. There are still some limitations in our study. The evaluation was conducted using a limited number of cyber attack scenarios, with relatively simple train operation modes and simulated attack conditions. Future research could explore more sophisticated adversarial attacks or combined cyber-physical attack strategies to further enhance the robustness of the proposed approach.

Acknowledgement: The authors thank anonymous reviewers for their comments and requirements that have helped to improve the paper.

Funding Statement: This work was supported in part by the project of the State Key Laboratory of Advanced Rail Autonomous Operation (RAO2023ZZ004), in part by the Beijing Natural Science Foundation-Fengtai Rail Transit Frontier Research Joint Fund (L211002), in part by the Foundation of China State Railway Group Corporation Limited

under Grant L2021G003, in part by the Scientific and Technical Research Fund of China Academy of Railway Sciences Corporation Limited under Grant 2021YJ094, and in part by the Project I23L00200 and Project I24F00010.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Qichang Li, Bing Bu; data collection: Qichang Li, Junyi Zhao; analysis and interpretation of results: Qichang Li, Bing Bu, Junyi Zhao; draft manuscript preparation: Qichang Li, Bing Bu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Yu Z, Wang H, Chen F. Security of railway control systems: a survey, research issues and challenges. High-speed Railway. 2023;1(1):6–17. doi:10.1016/j.hspr.2022.12.001.
2. Soderi S, Masti D, Hämäläinen M, Iinatti J. Cybersecurity considerations for communication based train control. IEEE Access. 2023;11:92312–21. doi:10.1109/ACCESS.2023.3309005.
3. Kour R, Patwardhan A, Thaduri A, Karim R. A review on cybersecurity in railways. Proc Institut Mech Eng Part F: J Rail Rapid Transit. 2023;237(1):3–20. doi:10.1177/09544097221089389.
4. Wu W, Bu B. Security analysis for CBTC systems under attack-defense confrontation. Electronics. 2019;8(9):991. doi:10.3390/electronics8090991.
5. Muneer S, Farooq U, Athar A, Ahsan Raza M, Ghazal TM, Sakib S. A critical review of artificial intelligence based approaches in intrusion detection: a comprehensive analysis. J Eng. 2024;2024(1):3909173. doi:10.1155/2024/3909173.
6. Roshan K, Zafar A, Haque SBU. Untargeted white-box adversarial attack with heuristic defence methods in real-time deep learning based network intrusion detection system. Comput Commun. 2024;218:97–113. doi:10.1016/j.comcom.2023.09.030.
7. Zhang J, Feng H, Liu B, Zhao D. Survey of technology in network security situation awareness. Sensors. 2023;23(5):2608. doi:10.3390/s23052608.
8. Yang H, Zhang Z, Xie L, Zhang L. Network security situation assessment with network attack behavior classification. Int J Intell Syst. 2022;37(10):6909–27. doi:10.1002/int.22867.
9. Guo E, Bu B. CBTC systems resilience evaluation based on resource state model under DoS attacks. In: 2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC); 2021; Guiyang, China: IEEE. p. 451–6.
10. Zhang Z, Ning H, Shi F, Farha F, Xu Y, Xu J, et al. Artificial intelligence in cyber security: research advances, challenges, and opportunities. Artif Intell Rev. 2022:55;1029–53.
11. Bejoy B, Raju G, Swain D, Acharya B, Hu YC. A generic cyber immune framework for anomaly detection using artificial immune systems. Appl Soft Comput. 2022;130:109680. doi:10.1016/j.asoc.2022.109680.
12. Kumar V, Sinha D. Synthetic attack data generation model applying generative adversarial network for intrusion detection. Comput Secur. 2023;125:103054. doi:10.1016/j.cose.2022.103054.
13. Wu Q, Wang W, Fan P, Fan Q, Zhu H, Letaief KB. Cooperative edge caching based on elastic federated and multi-agent deep reinforcement learning in next-generation networks. IEEE Trans Netw Serv Manag. 2024;21(4):4179–96. doi:10.1109/TNSM.2024.3403842.
14. Shao Z, Wu Q, Fan P, Cheng N, Chen W, Wang J, et al. Semantic-aware spectrum sharing in internet of vehicles based on deep reinforcement learning. IEEE Internet Things J. 2024;11(23):38521–36. doi:10.1109/JIOT.2024.3448538.
15. Yu GF. A multi-objective decision method for the network security situation grade assessment under multi-source information. Inf Fusion. 2024;102:102066. doi:10.1016/j.inffus.2023.102066.

16. Wang J, Zhang K, Li J. Network awareness of security situation information security measurement method based on data mining. J Intell Fuzzy Syst. 2024;46:209–19. doi:10.3233/JIFS-233390.
17. Du Z, Yao H, Fu Y, Cao Z, Liang H, Ren J. Network situation assessment method based on improved BP neural network. Electronics. 2023;12(3):483. doi:10.3390/electronics12030483.
18. Guo X, Yang J, Gang Z, Yang A. Research on network security situation awareness and dynamic game based on deep Q learning network. J Internet Technol. 2023;24(2):549–63.
19. Yang H, Zeng R, Wang F, Xu G, Zhang J. An unsupervised learning-based network threat situation assessment model for internet of things. In: Security and communication networks. New York, NY, USA: John Wiley & Sons, Inc. 2020. doi:10.1155/2020/6656066.
20. Wang H, Chen Z, Feng X, Di X, Liu D, Zhao J, et al. Research on network security situation assessment and quantification method based on analytic hierarchy process. Wirel Pers Commun. 2018;102:1401–20. doi:10.1007/s11277-017-5202-3.
21. Zhang H, Jie SLBD, Yang Y, Zhang R, Lang Q, Zhu LBD. Power Grid Security Situation Awareness Method based on Deep Learning. In: 2023 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics); 2023; Danzhou, China: IEEE. p. 776–80.
22. Zhang Z, Zhang Y. Security situation assessment for terminal area control system operation based on BN-ISSA-ELM. Appl Sci. 2024;14(23):11384. doi:10.3390/app142311384.
23. Zhao J, Li X, Cao Y, Liu J, Yan J, Li C. Analysis and application of intelligent power control system cyber security situation awareness based on wavelet neural network. In: Journal of Physics: Conference Series; 2021; Wuxi, China: IOP Publishing. vol. 2078.
24. Li X, Wang H. Industrial control network security situation assessment based on SAE-RBF. In: 2022 11th International Conference of Information and Communication Technology (ICTech); 2022; Wuhan, China: IEEE. p. 43–7.
25. Lei W, Wen H, Wu J, Hou W. MADDPG-based security situational awareness for smart grid with intelligent edge. Appl Sci. 2021;11(7):3101. doi:10.3390/app11073101.
26. Yu Z, Gao H, Cong X, Wu N, Song HH. A survey on cyber-physical systems security. IEEE Internet Things J. 2023;10(24):21670–86. doi:10.1109/JIOT.2023.3289625.
27. Alsulami AA, Al-Haija QA, Alturki B, Alqahtani A, Alsini R. Security strategy for autonomous vehicle cyber-physical systems using transfer learning. J Cloud Comput. 2023;12(1):181. doi:10.1186/s13677-023-00564-x.
28. Abdo A, Chen H, Zhao X, Wu G, Feng Y. Cybersecurity on connected and automated transportation systems: a survey. IEEE Trans Intell Vehicles. 2024;9(1):1382–401. doi:10.1109/TIV.2023.3326736.
29. Lu R, Dong H, Wang H, Cui D, Zhu L, Wang X. A resilience-based security assessment approach for CBTC systems. Complexity. 2021;2021(1):2175780. doi:10.1155/2021/2175780.
30. Kanghao Z, Hongwei W, Dongliang C. A quantitative situation awareness approach for cbtc systems based on multi-dimensional Gaussian hidden Markov model. In: 2020 Chinese Automation Congress (CAC); 2020; Shanghai, China: IEEE. p. 3488–92.
31. Li Y, Zhu L, Wang H, Yu FR, Liu S. A cross-layer defense scheme for edge intelligence-enabled CBTC systems against MitM attacks. IEEE Trans Intell Transp Syst. 2020;22(4):2286–98. doi:10.1109/TITS.2020.3030496.
32. Huang K, Zhou C, Tian YC, Yang S, Qin Y. Assessing the physical impact of cyberattacks on industrial cyber-physical systems. IEEE Trans Ind Electron. 2018;65(10):8153–62. doi:10.1109/TIE.2018.2798605.
33. Yan K, Liu X, Lu Y, Qin F. A cyber-physical power system risk assessment model against cyberattacks. IEEE Syst J. 2022;17(2):2018–28. doi:10.1109/JSYST.2022.3215591.
34. Ahmed MS, Al-Shaer E, Khan L. A novel quantitative approach for measuring network security. In: IEEE INFOCOM 2008—The 27th Conference on Computer Communications; 2008; Phoenix, AZ, USA: IEEE. p. 1957–65.
35. Feng N, Wang HJ, Li M. A security risk analysis model for information systems: causal relationships of risk factors and vulnerability propagation analysis. Inf Sci. 2014;256:57–73. doi:10.1016/j.ins.2013.02.036.

36. Zhang XQ, Xu JY, Gu CH. Information security vulnerability association analysis based on ontology technology. J East China Univ Sci Technol. 2015;40(1):125–31.
37. Batur Şahin C, Abualigah L. A novel deep learning-based feature selection model for improving the static analysis of vulnerability detection. Neural Comput Appl. 2021;33(20):14049–67. doi:10.1007/s00521-021-06047-x.
38. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: an overview. IEEE Signal Process Mag. 2018;35(1):53–65.
39. Yin B, Bu B, Gao B, Li Q. A hybrid intrusion detection method using improved stacking ensemble algorithm and false positive elimination strategy for CBTC. In: 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC); 2022; Macau, China: IEEE. p. 4253–8.
40. Li Q, Bu B, Zhao J. A novel hierarchical situation awareness model for CBTC using SVD entropy and GRU with PRD algorithms. IEEE Access. 2021;9:132290–300. doi:10.1109/ACCESS.2021.3112166.
41. Shi Y, Li T, Li R, Peng X, Tang P. An immunity-based IOT environment security situation awareness model. J Comput Commun. 2017;5(7):182. doi:10.4236/jcc.2017.57016.
42. Binbusayyis A, Vaiyapuri T. Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM. Appl Intell. 2021;51(10):7094–108. doi:10.1007/s10489-021-02205-9.
43. Wu T, Fan H, Zhu H, You C, Zhou H, Huang X. Intrusion detection system combined enhanced random forest with SMOTE algorithm. EURASIP J Adv Signal Process. 2022;2022(1):39. doi:10.1186/s13634-022-00871-6.
44. Chapagain P, Timalsina A, Bhandari M, Chitrakar R. Intrusion detection based on PCA with improved K-means. In: International Conference on Electrical and Electronics Engineering; 2022; Singapore: Springer. p. 13–27.
45. Talaei Khoei T, Kaabouch N. A comparative analysis of supervised and unsupervised models for detecting attacks on the intrusion detection systems. Information. 2023;14(2):103. doi:10.3390/info14020103.