Research on a Network Security Situation Intelligent Awareness and Prediction Model Based on Big Data Technology and Its Supporting Applications

Changyao Yang and Yuan Yan*

Department of Information Science, Zhanjiang Preschool Education College, Zhanjiang, 524084, China
E-mail: yany1229@outlook.com
**Corresponding Author*

Received 29 December 2025; Accepted 27 February 2026

Abstract

The traditional network security situational awareness method is difficult to deal with high-speed multi-source data flow because it relies on a centralized data processing architecture, resulting in poor real-time performance and weak data association. Therefore, building a perception prediction model that can fuse multi-source data in real time, understand the internal structure of the network, and have cognitive reasoning capabilities similar to experts is of great theoretical significance for realizing active intelligent network security defense. The study hypothesizes that the integration of distributed flow processing, network structure mode decomposition, and a cognitive decision-making framework can build a security perception model with high-precision perception and forward-looking prediction ability. The core of this method is to integrate the above components into a unified model of "data structure cognition" three-tier collaboration: Kafka+spark stream processing layer is responsible for real-time data fusion and feature supply. The Structural Modal Modeling and Decomposition (SMMD) layer analyzes the functional topology of the network to achieve fine-grained

Journal of Cyber Security and Mobility, Vol. 15_2, 365–390.
doi: 10.13052/jcsm2245-1439.1524

situation decoupling; The Estimation Memory Control (EMC) cognitive layer simulates the closed-loop process of expert evaluation, experience reuse, and predictive decision-making. This architecture realizes the deep coupling of data-driven, structural understanding, and cognitive intelligence, which is different from the existing situational awareness framework that mainly relies on a single data dimension or lacks explicit cognitive reasoning. To verify this hypothesis, several experiments are designed and implemented. Firstly, a distributed stream processing framework is built based on Kafka+spark to realize the real-time fusion and feature extraction of multi-source security data. Secondly, a safety perception prediction model combining the EMC framework and SMMD is proposed. The network functional topology is analyzed by structural mode decomposition, and the EMC framework is introduced to simulate the expert cognitive decision-making process. The core findings are as follows: Experiments on the Canadian Institute for Network Security network intrusion detection dataset (cic-ids-2017) show that the accuracy of the model in predicting the macro situation level is 93.7%, and the F1-Score for identifying five types of attacks is up to 97.2%. This performance is superior to the mainstream baseline models of LSTM, TCN, and GBDT. In the verification of the real network range, the model can shorten the average response time of high-risk threats to 3.5 minutes, improve the attack containment rate to 95.2%, reduce the false positive intervention rate to 8%, and improve the analysis efficiency by about 83%. The conclusion is that the proposed model is superior to the mainstream methods in perception accuracy, response speed, and operation and maintenance efficiency. This study provides effective technical support for the construction of an active and intelligent network security protection system.

This study also recognizes that the model has some limitations: Its performance depends on high-quality labeled data for initial training and pattern library construction; When dealing with large-scale networks (such as more than 10,000 devices), the computational cost of mode division and state estimation needs to be further optimized; In addition, the prediction ability of the model to the new attack mode (zero day) that has not appeared in the training data remains to be explored. Although the model has been validated on the cic-ids-2017 standard dataset and enterprise-wide, its ability to generalize to other network architectures (such as cloud native, IoT) and more complex real-world operational environments is a direction that needs to be evaluated before actual deployment in the future.

Keywords: Network security, situational awareness, structural modal decomposition, cognitive decision-making, security perception prediction.

1 Overview

With the popularization of technologies such as cloud computing and the Internet of Things (IoT), the structure of cyberspace has become increasingly complex, and cyberattacks have shown new characteristics of scale, concealment, and intelligence. Traditional static and passive defense systems are unable to cope with increasingly severe security challenges. Network Security Situation (NSS) awareness technology has therefore become the core of building an active and elastic defense system [1–3]. Its goal is to understand and predict the overall security status of the network in real time from massive and dynamic security data to provide support for accurate and efficient response decisions. Currently, academia and industry have conducted extensive research in NSS sensing. Chen Z proposed a Radial Basis Function (RBF) neural network prediction model based on a simulated annealing algorithm and a hybrid hierarchical genetic algorithm optimization to address limited prediction accuracy caused by the use of a single algorithm in traditional NSS sensing methods. The predicted situation value of the optimized RBF network in 15 samples was very close to the actual value, indicating that the model has a good prediction effect and can provide effective support for network security maintenance [4]. Oladosu S A et al. proposed a unified conceptual security framework driven by artificial intelligence and incorporating zero trust principles. They aimed to address the insufficient effectiveness of traditional network security models due to the increased complexity and interconnection of modern hybrid cloud and on-premises information technology infrastructure. The framework detected anomalies, identified vulnerabilities, and prioritized risks with extremely high accuracy, allowing for rapid mitigation before potential threats escalate [5]. Vimal V et al. proposed a software-defined network architecture based on cognitive protocol networks. The architecture utilized random neural networks to extract information and make decentralized decisions while integrating into SerIoT technology to strengthen IoT encryption and access control, aiming to solve the security and energy efficiency issues of IoT caused by limited node energy. This network infrastructure was able to circumvent the system's unpredictable connections and node performance degradation, and generate an end-to-end anti-theft solution that meets predetermined circuit constraints based on actual statistical data [6]. Sheng C et al. proposed a self-growing attack traffic classification model based on a new density heuristic clustering method, and designed an effective Supervisory Control and Data Acquisition (SCADA) network

traffic representation method. This method aimed to address that existing methods are difficult to effectively classify attack traffic in SCADA networks due to the lack of attack samples and high real-time requirements. In key scenarios with only normal SCADA network traffic, the performance was better than the current advanced attack traffic classification methods [7].

Although existing research has made some progress, there is still room for improvement. At the data level, most methods still pay insufficient attention to the deep integration and real-time cleaning of multi-source heterogeneous security data, resulting in the phenomenon of data islands still existing, which in turn makes the situation understanding lack comprehensiveness. At the model level, existing prediction models often treat the network as a homogeneous whole and ignore its internal functional structure and topological correlation, resulting in the model lacking the ability to conduct fine-grained analysis of network micro-modes. These existing models also generally lack a mechanism that can simulate the cognitive process of security experts and effectively accumulate and reuse historical experience. As a result, their level of intelligence in forward-looking prediction and decision support is limited when dealing with complex and changing attack scenarios. To solve the above problems, this study proposes a security perception prediction model that integrates big data based on the Estimation Memory Control (EMC) framework and Structural Modal Modeling and Decomposition (SMMD), namely EMC-SMMD.

The reason why the EMC framework and SMMD are selected as the core method is that they have more theoretical and practical advantages in accordance with the nature of the problem than other cutting-edge technology paths, such as Graph Neural Network (GNN), attention-based time series model, or digital twin framework. Although GNN can describe the network topology, its processing efficiency for dynamic, real-time, multi-source stream data, and explicit modeling ability for the expert cognitive decision-making process are insufficient. The attention model is good at capturing temporal dependence, but it lacks explicit decoupling of network functional structure and systematic reuse mechanism of historical experience. The digital twin framework focuses on high-fidelity simulation and faces challenges in the lightweight of real-time perceptual prediction and the construction of closed-loop decision-making. In contrast, SMMD provides an intuitive and interpretable method to functionalize and decouple complex networks, which lays a structural foundation for fine-grained situation understanding. The EMC framework directly simulates the cognitive closed loop of human

experts "assessing the status quo – learning from experience – predictive control", providing a natural paradigm for historical experience knowledge and its application in forward-looking decision-making. The combination of the two achieves the organic unity of structural understanding and cognitive intelligence in theory, and balances the analysis granularity, real-time requirements, and decision-making interpretability in practice. It provides a solid methodological foundation for the perception prediction model constructed in this study.

Compared with the methods in the previous literature, the proposed model shows clear advantages. For example, compared with the RBF neural network model optimized by Chen Z et al. [2], the distributed flow processing framework and EMC-SMMD cognitive model adopted in this study can not only process static samples but also adapt to the data flow of continuous evolution. The reuse mechanism of historical experience is embedded in the model, which improves the prediction adaptability and decision intelligence in a dynamic environment. Compared with the unified conceptual security framework proposed by Oladosu SA et al. [3], this study provides a complete and implementable technical path and quantitative model from data fusion, structural analysis, to cognitive decision-making, rather than staying at the level of architecture design. Compared with the method that Sheng C et al. [5] focused on the classification of specific attack traffic of SCADA networks, this study has the ability to deconstruct and analyze the general situation of complex heterogeneous networks through SMMD and has a wider range of applications. More importantly, this model realizes the closed-loop of "perception memory prediction" through the EMC framework. It makes up for the deficiency of Vimal V et al. [4]'s decentralized decision-making method based on a stochastic neural network in global situation understanding and experience accumulation.

The core innovation element is to propose a new paradigm of "data structure cognition" three-tier collaborative NSS awareness. Specifically, at the data level, a distributed stream processing framework based on Kafka+spark is constructed to realize the real-time fusion and high-quality feature extraction of multi-source heterogeneous security data. At the structural level, SMMD is innovatively introduced into the field of network security. By decoupling the function and topology of the network, fine-grained situation analysis from macro to micro modes is realized. At the cognitive level, the EMC framework is introduced to simulate the assessment, experience reuse, and decision-making process of security experts, so that the model has the ability of cumulative learning and situation deduction similar to human

beings. The organic combination of these three levels together constitutes the innovation, different from the existing methods.

Based on the above, this study is not only an improvement of the existing technology path but also a forward-looking solution for future complex network threats.

The proposed model is mainly for large and medium-sized enterprise networks, cloud data centers, or network environments with clear functional partitions, and its design can adapt to the daily terabytes of multi-source secure data stream processing. The preconditions for model deployment include: (1) The multi-dimensional data such as network traffic, host logs, and security device alarms can be continuously collected by deploying agents or images; (2) It has a marked historical security event dataset for the initial training of the model and the construction of the pattern library; (3) It has moderate computing resources that can support distributed stream processing frameworks (such as Kafka+spark cluster) and model real-time reasoning. The advantage of this model is that it is the most significant scenario where resources are relatively sufficient, and the network structure can be divided reasonably. For environments with extremely heterogeneous, highly dynamic topologies, or where labeled data are extremely scarce (such as some large IoT networks), the direct applicability of this model may be limited, and targeted adaptation is required. Clarifying the scope and assumptions of these applications will help to define the practical feasibility of this method and the boundaries of its advantages.

2 Methods and Materials

2.1 Security Situation Data Stream Processing Framework Based on Big Data Platform

The accuracy and real-time performance of NSS perception fundamentally rely on the deep integration and real-time processing of massive, multi-source, and heterogeneous security data. Traditional data processing architecture often faces problems such as throughput bottlenecks, data islands, and processing delays when faced with multi-dimensional data flows such as high-speed network traffic, distributed system logs, security device alarms, and external threat intelligence. This makes it difficult to accurately characterize the macro and micro security situation. To overcome the above difficulties, this study designs a Data Stream Processing Framework (DSPF) based on distributed computing technology. This framework aims to realize

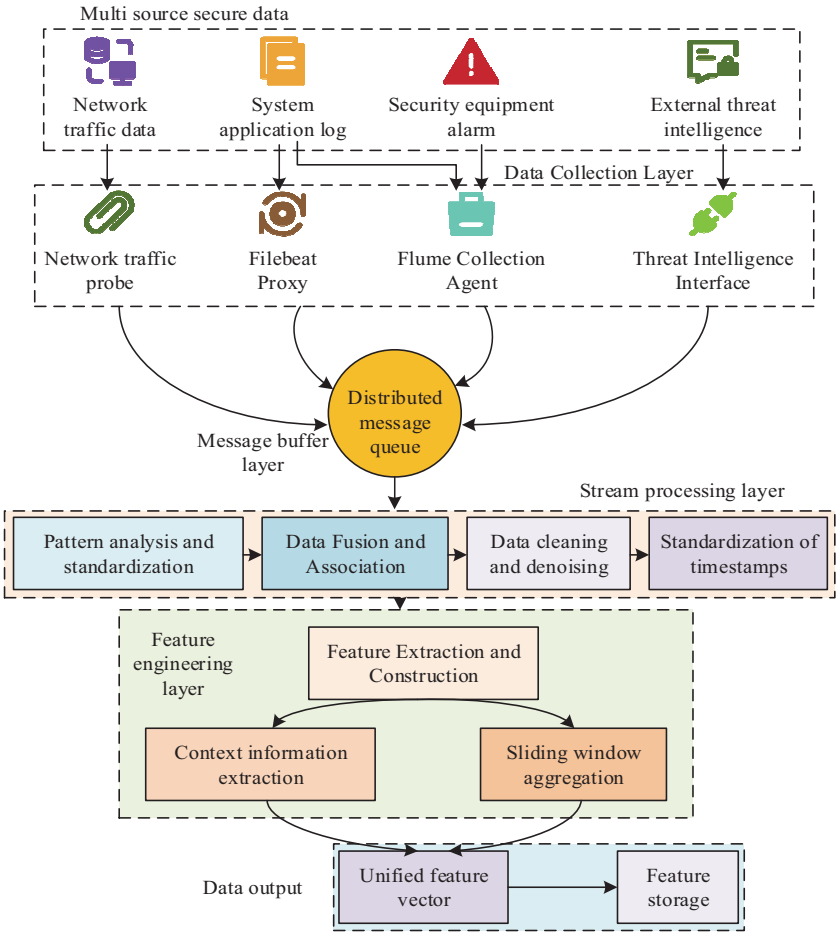


Figure 1 DSPF based on distributed computing technology.

the full-process, pipelined processing of security data from collection, integration, cleaning, to feature extraction. The architecture of DSPF based on distributed computing technology is shown in Figure 1.

In Figure 1, the operation of this framework starts with the real-time collection and access of multi-source security data. Data sources include raw traffic data in the network, operating system, and application logs, alarm information generated by security devices such as firewalls and intrusion detection systems, as well as external threat intelligence data obtained from open communities or commercial channels [8]. Data sources have a

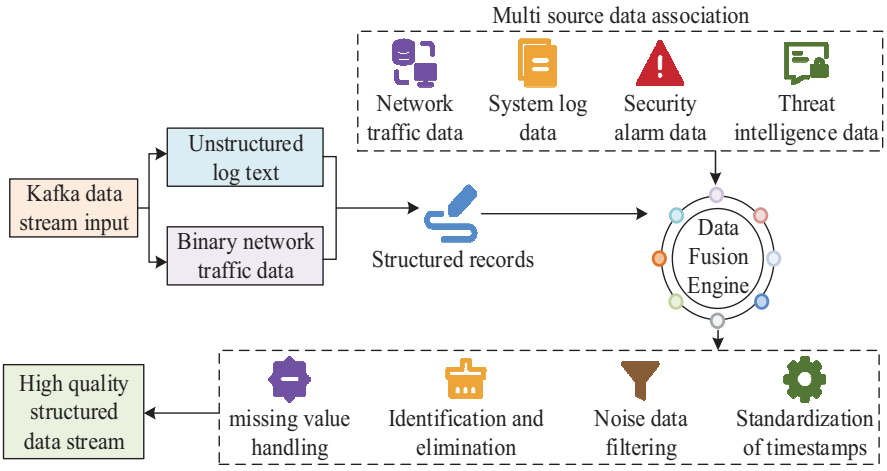


Figure 2 Streaming processing and data fusion stage.

certain degree of diversity, and data generation is continuous. Therefore, to avoid rate-mismatch problems due to the access of heterogeneous data, the research framework uses Apache Kafka (Kafka for short) as a unified high-throughput message queue for data buffering. Kafka is an open source distributed stream processing platform whose core function is high-throughput message publishing and subscription [9]. In this framework, Kafka is mainly responsible for accepting data streams pushed from various data collection agents and organizing these data into continuously updated data topics. Based on this design, the data production and consumption links can be effectively decoupled, so that the subsequent processing modules can stably pull data from Kafka for processing at their own speeds, and respond to the impact of data flood peaks. After completing the initial access of data, the streaming processing and data fusion stages follow, as shown in Figure 2.

In Figure 2, the research framework uses Apache Spark Structured Streaming (ASSS) as the core computing engine. The main reason is that ASSS can provide a high-level application programming interface that can perform complex operations similar to static data on infinitely growing data streams through micro-batch or continuous processing modes. After data are ingested from a specific Kafka topic to Spark, schema parsing and data normalization are first performed [10]. Unstructured log text uses regular expressions and natural language processing technology to extract key fields. Binary network traffic data can be converted into structured records through

deep packet inspection or flow statistical feature analysis. The next step is to perform data fusion operations, that is, to connect information from different data sources through preset association keys [11]. The main purpose of this process is to break down data silos and build a unified security event view, including network behavior, host activities, and security events, as shown in formula (1).

$$\mathcal{D}_{\text{merged}} = \sum_{k=1}^{N} \mathcal{D}_k \text{s.t. } \text{Key}(\mathcal{D}_i) \triangleright \triangleleft \text{Key}(\mathcal{D}_j), \quad \forall\, i, j \leq N \tag{1}$$

In formula (1), $\mathcal{D}_j$ is the parsed and standardized structured data of the k-th data source. Key is the default associated key. $\mathcal{D}_{\text{merged}}$ is data fusion. $\triangleright\triangleleft$ is the concatenation operator [12]. After completing the streaming processing and data fusion stages, to ensure data quality, the framework continues to perform data cleaning based on data fusion [13]. This stage will implement predefined rules, including processing missing values in the original data, identifying and eliminating duplicate records caused by network jitter or collector anomalies, and filtering noisy data that obviously does not conform to logic or known patterns based on the rule base and statistical outlier detection methods [14]. At the same time, the framework standardizes timestamps and uses them as the basis for event sorting. The structured data stream after cleaning and fusion will enter the feature engineering stage. The goal is to extract features from raw data that can effectively characterize the network security status, as shown in Figure 3.

In Figure 3, feature engineering includes two categories: statistical features and contextual semantic features. Statistical features mainly focus on the aggregation calculation of network traffic and host behavior within a sliding time window. For a given time window W_t, the traffic count and rate characteristics of a specific entity can be expressed as formula (2).

$$\begin{cases} \text{FlowCount}(e, W_t) = \sum\limits_{i \in W_t} \mathbb{I}(\text{entity}_i = e) \\ \text{PacketRate}(e, W_t) = \dfrac{\sum_{i \in W_t} \text{Packets}_i \cdot \mathbb{I}(\text{entity}_i = e)}{\Delta t} \end{cases} \tag{2}$$

In formula (2), e is the target entity. $\mathbb{I}$ is an indicator function, which takes a value of 1 when the entity field of data record i matches e, otherwise it takes a value of 0. $Packets_i$ is the number of packets recorded in the i flow [15]. Semantic features are dedicated to extracting deeper contextual information from raw data to capture semantic correlations between events,

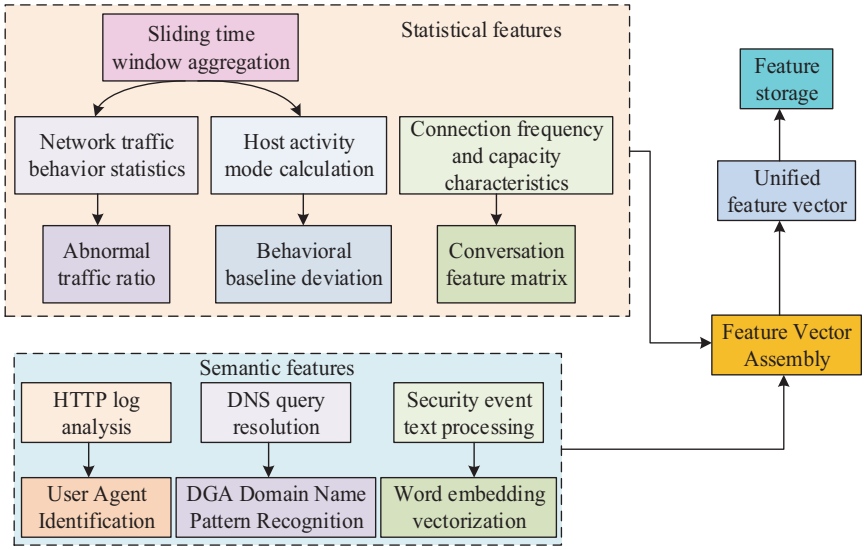


Figure 3 Feature engineering stage diagram.

including: extracting the User-Agent field from HTTP logs and analyzing whether it is a known malicious or crawler tool, identifying the pattern characteristics of dynamic domain name generation algorithm domain names from domain name query records, and using pre-trained word embedding models to vectorize security event description texts [16]. For the extraction of text semantic features, the security event description S can be mapped into a vector representation through the pre-trained word embedding model ϕ, as shown in formula (3).

$$\text{SemanticVector}(S) = \frac{1}{|S|} \sum_{w \in S} \phi(w) \tag{3}$$

In formula (3), w is a word describing text S. $\phi(w)$ is the word vector corresponding to word w, and finally the semantic feature vector of the event description is obtained through average pooling [17]. Finally, all generated statistical features and semantic features are assembled into a unified feature vector. This vector is continuously updated over time and constitutes a dynamic, quantitative description of network entities and their interrelationships. At this point, the feature data stream processed by the above framework can be output to a specific data storage for real-time estimation and prediction by subsequent models.

2.2 Security Perception Prediction Model Integrating EMC-SMMD

Although the above-mentioned DSPF can provide a high-quality data basis for the quantitative perception of NSS, further analysis of this data is required to achieve an in-depth understanding and forward-looking prediction of the security situation. This study constructs an EMC-SMMD model, aiming to simulate the cognitive decision-making process of human experts through the EMC framework, and uses SMMD to analyze the intrinsic structure of complex networks, thereby achieving dynamic estimation and accurate prediction of the security situation. The specific structure of the model is shown in Figure 4.

In Figure 4, the core of EMC-SMMD is to regard macroscopic NSS as the joint action of several interrelated structural modes. SMMD first divides the target network system into several relatively independent modes based on its functions and topological characteristics, including user access mode, core switching mode, data center service mode, and boundary protection mode [18]. Each modality is defined by its unique set of assets, traffic patterns, and security capabilities. By performing SMMD on the network, the model decomposes the complex network-wide security analysis problem into the analysis of a series of sub-modalities with clear structures and more consistent behaviors, laying the foundation for subsequent perception. After completing SMMD, the estimation module is responsible for quantitatively evaluating the safety status of each mode at time t. The specific process is shown in Figure 5.

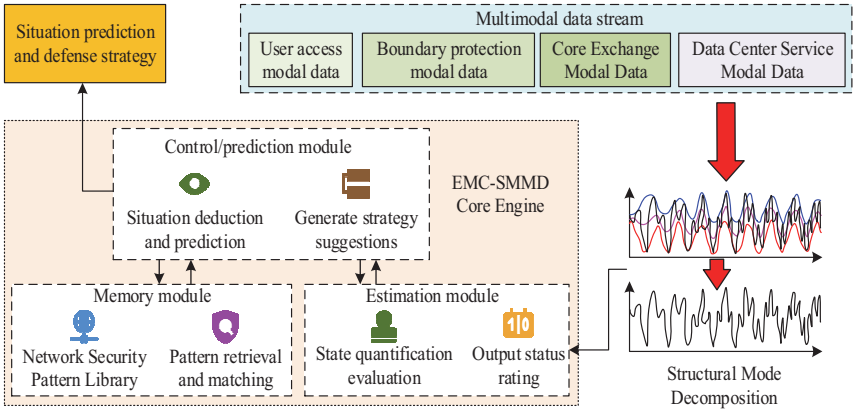


Figure 4 EMC-MMD model diagram.

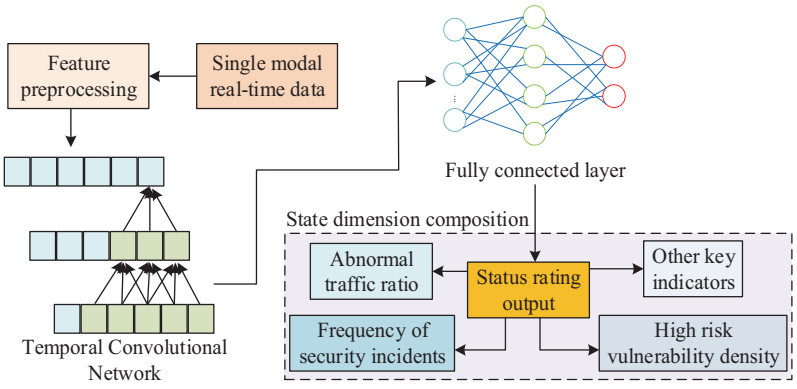


Figure 5 Modal assessment process based on safety status.

In Figure 5, this state is a multi-dimensional vector whose dimensions include key indicators such as abnormal traffic ratio within the mode, density of high-risk vulnerabilities, and frequency of security events. The estimation process relies on real-time feature data output from DSPF that has been divided by modality [19, 20]. Specifically, for each modality, the estimation module utilizes a lightweight machine learning model, namely a sequential convolutional network, to fuse multi-source features. After fusion processing, a comprehensive status score can be output. This state can be expressed as formula (4).

$$S_k(t) = f_{\text{estimate}}(\vec{X}_k(t); \Theta_e) \tag{4}$$

In formula (4), $S_k(t)$ is the safe state of mode M_k at time t. $\vec{X}_k(t)$ is the aggregated feature vector of M_k at time t. Θ_e is the parameter of the estimated model. f_{estimate} is the estimation function [21]. The memory module is mainly used to realize knowledge accumulation and evolution. Its core is a dynamically updated network security model library, as shown in Figure 6.

In Figure 6, the library stores historical and real-time learned attack scenario patterns, situation evolution sequences, and their corresponding effective response strategies. Each memory record can be represented as a triplet. When the estimation module outputs the current and recent multi-modal state sequences, the memory module retrieves the K most similar historical records from the pattern library by calculating vector similarity. These records can provide valuable prior knowledge for predicting future situations [22]. Finally, the control module is mainly used to comprehensively estimate the current status of the module, output similar historical patterns

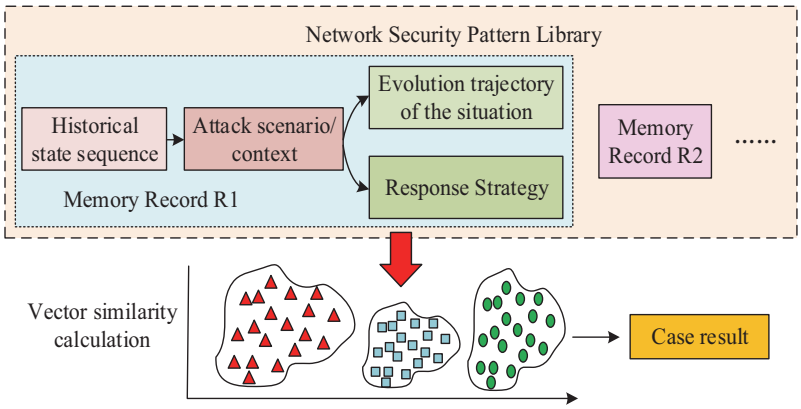


Figure 6 Network security model library.

provided by the memory module, and perform situation deduction and prediction. Specifically, the core task of this module is to generate a prediction of the comprehensive situation $Y(t+\Delta t)$ of the entire network after time Δt in the future. This prediction is achieved through a weighted fusion mechanism, as shown in formula (5).

$$Y(t+\Delta t) = \sum_{j=1}^{K} w_j \cdot \Delta S_j \tag{5}$$

In formula (5), w_j is the weight, and its value is determined by the matching degree between the current state sequence $S(t)$ and the historical sequence S_j. The higher the matching degree, the greater the weight. ΔS_j corresponds to the subsequent situation evolution stored in the historical record [23]. In addition, this module can also generate preliminary defense strategy suggestions based on the prediction results. Ultimately, EMC-SMMD transforms real-time data into status assessment through the closed-loop collaboration of estimation, memory, and control, realizes the integration of NSS perception, understanding, and prediction, and provides core decision support for building an active defense system.

3 Results

3.1 Model Performance Comparison Experiments and Analysis

To verify the validity, the EMC-SMMD model is tested. The experiment is conducted in a cluster environment consisting of 50 physical servers.

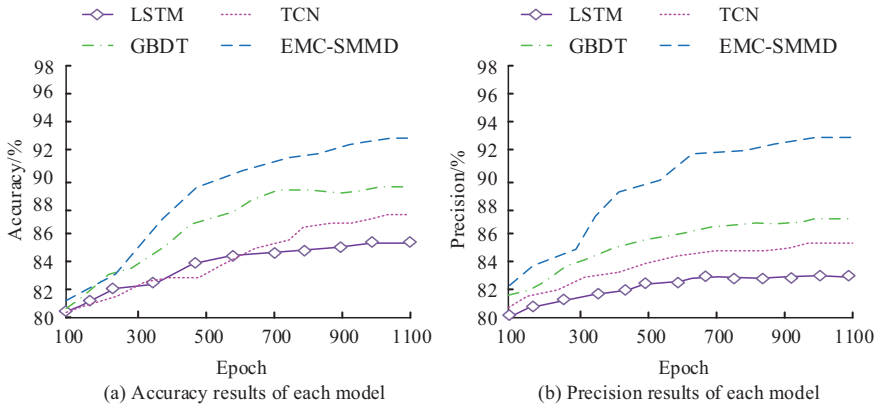


Figure 7 Comparison of macro security situation level prediction performance.

The cluster is equipped with Apache Kafka 3.5.0 and Apache Spark 3.4.0, which are used to deploy DSPF based on distributed computing. The experiment uses the CIC-IDS-2017 network intrusion detection dataset, which simulates the network traffic and system logs of normal and various typical attack behaviors in a real enterprise network environment. The experiment extracts 72 consecutive hours of data from it, totaling about 1.2TB of original data stream, and divides it into a training set, a verification set, and a test set in chronological order with a ratio of 7:2:1. The experiment adopts a comparison method and selects three representative NSS prediction models as comparison baselines, including: prediction model based on Long Short-Term Memory (LSTM), Temporal Convolutional Network (TCN), and Gradient Boosting Decision Tree (GBDT). The experiment first compares the classification performance of each model in the macro situation, as shown in Figure 7.

Figure 7(a) shows the performance of each model on the accuracy index. The accuracy of LSTM is 85.4%, TCN is 87.9%, and GBDT is 89.2%. In comparison, EMC-SMMD has the highest accuracy of 93.7%. Figure 7(b) shows the performance of each model on the precision index. The precision of LSTM is 82.1%, TCN is 85.6%, and GBDT is 87.8%. In comparison, EMC-SMMD has the highest precision of 92.5%. EMC-SMMD decouples complex networks through SMMD and combines it with the cognitive decision-making mechanism of the EMC framework to more accurately grasp the changing trend of the overall security status of the network. In addition, to further explore the model's ability to perceive specific attack types at the

Table 1 Comparison of F1-score for specific attack type identification

Model	DDoS Attack	Brute-Force	Web Attack	Port Scan	Botnet
LSTM	88.5	75.2	82.4	90.1	79.8
TCN	91.2	78.9	85.7	92.5	83.1
GBDT	92.8	81.4	87.3	93.9	85.6
EMC-SMMD	96.5	89.7	92.1	97.2	91.4

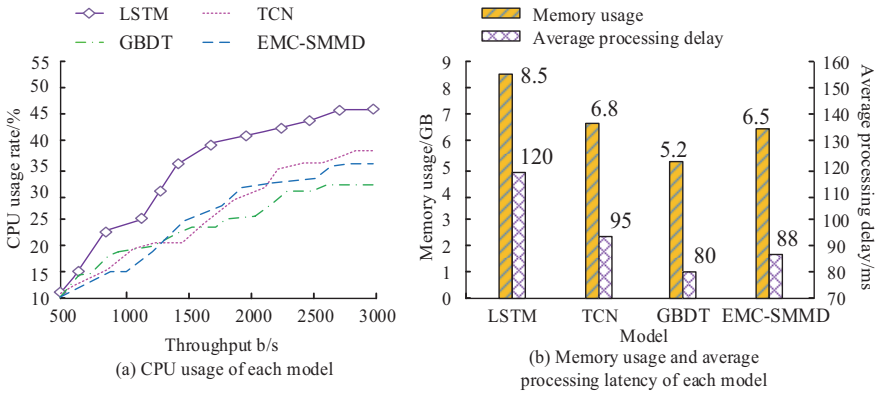


Figure 8 Comparison of model resource consumption and processing delay.

micro level, experiments are conducted to test the performance of each model in identifying five common attack types, as shown in Table 1.

In Table 1, compared with other models, EMC-SMMD achieves the highest F1-Score in five types of attack identification. Its F1-Score scores on DDoS attacks and port scanning are relatively high, at 96.5% and 97.2%. In terms of concealed brute force cracking and botnet identification, although EMC-SMMD's F1-Score is lower than the first two attack types, its F1-Score also reaches 89.7% and 91.4%. This shows that the research model has good advantages in fine-grained identification of various types of network attacks. Finally, the experiment evaluates the resource consumption and efficiency of each model in a real-time prediction environment, as shown in Figure 8.

Figure 8(a) shows the CPU usage of each model. The CPU occupancy rate of LSTM is 45.2%, TCN is 38.7%, GBDT is 32.1%, and EMC-SMMD is 35.5%. Figure 8(b) shows the memory usage and average processing delay of each model. The memory usage of LSTM, TCN, GBDT, and EMC-SMMD is 8.5GB, 6.8GB, 5.2GB, and 6.5GB, and the average processing delay is 120 ms, 95 ms, 80 ms, and 88 ms. Based on this result, it is found that despite the high complexity of EMC-SMMD, its resource consumption is controlled

within a reasonable range and can fully meet the real-time requirements for situational awareness in most network security scenarios.

3.2 Application Verification in Network Security Scenarios

To evaluate the actual support ability of the research model in the real network environment, this study uses the method of red and blue combat verification based on the network range. The core of this method is to integrate the EMC-SMMD into the workflow of the automated security operation center in a highly simulated medium-sized enterprise network shooting range environment with 2,000 terminals and 50 servers. During the seven-day test period, the system tests its full-chain capabilities under actual combat conditions by allowing security experts (blue team) to simulate and launch multiple rounds and multiple levels of real network attacks (such as network penetration, data leakage, advanced persistent threats, ransomware, etc.). Including: real-time perception and accurate identification of threats (corresponding to attack detection and classification) and situation understanding and risk prediction (corresponding to situation assessment and evolution). Ultimately, it is intended to drive automated responses or provide assisted decision support to analysts (corresponding to threat response time and attack containment effectiveness). Specifically, it is evaluated through three quantifiable dimensions: (1) Response timeliness, comparing the average response time of different threat levels; (2) Effectiveness of protection, statistics of successful containment rate of various attacks; (3) Analysis of the average event processing time and false alarm intervention rate. Through this practical, multi-dimensional quantitative evaluation framework, the actual support capabilities provided by the model in real complex network environments can be comprehensively and objectively measured.

The experiment adopts a comparative method, and the baseline model includes the traditional manual analysis process and the mainstream intelligent security operation platform. The experiment first tests the average response time of each method in response to different threat levels, as shown in Figure 9.

Figure 9(a) shows the performance based on the traditional manual analysis process. The response time of this method under high-risk, medium-risk, and low-risk threats is 45.2 min, 120.5 min, and 360 min, and the processing delay is long. Figure 9(b) shows the performance of the intelligent security operation platform. Regarding the three threats, the response times of the intelligent security operation platform are 12.5 min, 45.8 min, and 150.3 min,

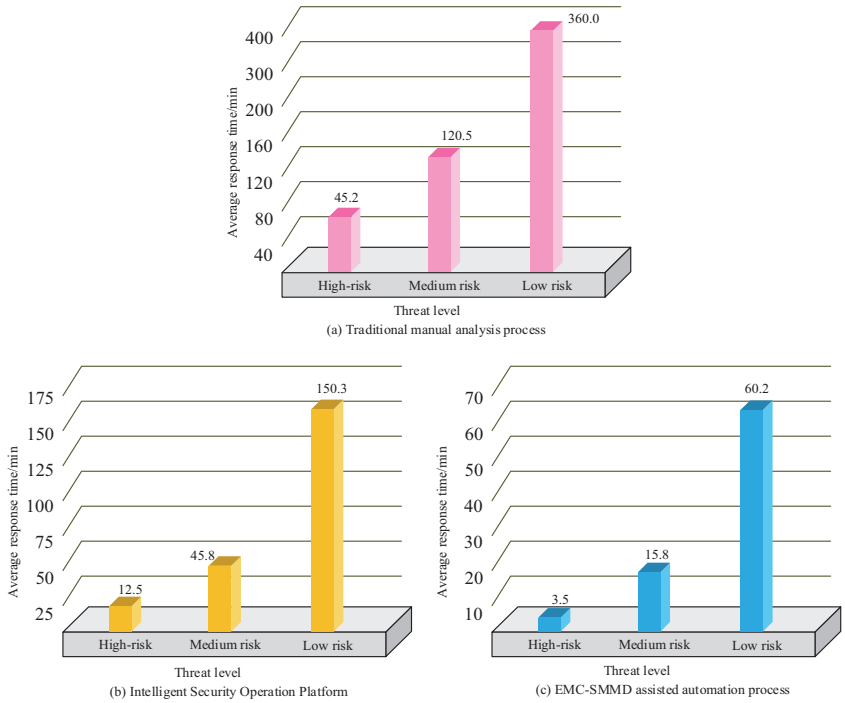


Figure 9 Comparison of average response times for different threat levels.

and the processing delay is relatively short. Figure 9(c) shows the performance of the EMC-SMMD-assisted automation process. Under this method, the processing delay is the shortest. The response time for high-risk threats is only 3.5 min, and the response time for medium-risk and low-risk threats is also controlled at 15.8 min and 60.2 min. To quantify the model's suppression effect on actual security risks, the experimental statistics are calculated on the proportion of various types of attacks launched by the blue team that are successfully contained during the test period, and compared with the data of historical benchmarks and intelligent security operation platforms that do not deploy the model, as shown in Table 2. The quantitative process follows the following definitions and calculation methods: first of all, "successful containment" refers to that at the critical stage of the attack chain (such as before the completion of initial intrusion, lateral movement, data theft or destruction), the automatic or semi-automatic blocking, isolation, authority revocation, and other disposal actions driven by the EMC-SMMD model are verified to be effective, making the attack unable to achieve its ultimate

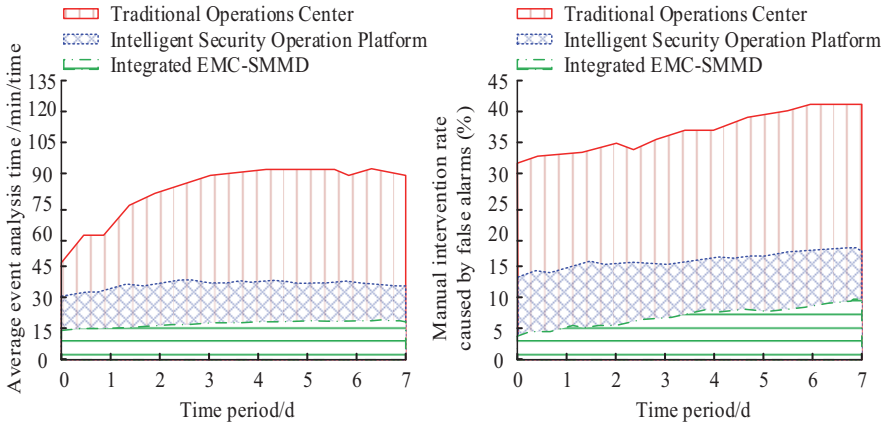
Table 2 Comparison of successful attack containment rates (%)

Attack Type	Traditional Manual Process	Intelligent Security Operations Platform	EMC-SMMD Model
Network Infiltration	65.8	85.5	92.5
Data Exfiltration	58.3	82.1	89.7
Advanced Persistent Threat (APT)	32.1	60.2	75.4
Ransomware	70.5	88.9	95.2

goal. Secondly, for each predefined attack type (such as network penetration, data leakage, etc.), the calculation formula of "attack success containment rate" is: (1 – the number of attacks successfully reaching the final goal/the total number of attacks of this type launched by the blue team) $\times$ 100%. This indicator directly reflects the actual ability to reduce business risks and prevent losses in real confrontation.

In Table 2, the containment rates of traditional manual processes for network penetration, data leakage, advanced persistent threats, and ransomware are 65.8%, 58.3%, 32.1%, and 70.5%. The performance of the intelligent security operation platform has improved, with containment rates reaching 85.5%, 82.1%, 60.2% and 88.9%. In comparison, EMC-SMMD has the best containment effect, with containment rates of 92.5%, 89.7%, 75.4% and 95.2%. This proves that the research model has excellent advantages in protecting against complex network attacks. Finally, the experiment evaluates the work efficiency of the security operation and maintenance team under each method, as shown in Figure 10.

Figure 10(a) shows the performance of each method on the average event analysis time indicator. It takes 90 min for a traditional operation center to process a single incident, but for the intelligent security operation platform, the time is relatively short, and it takes 35 min to process a single incident. In contrast, the platform integrated with the EMC-SMMD model has the lowest average event analysis time, only 15 min, which has a good analysis efficiency advantage. Figure 10(b) shows the performance of each method in terms of false alarm intervention rate. In terms of the rate of manual intervention caused by false alarms, the traditional operation center has the highest rate of false alarm intervention, reaching 40%. The false alarm intervention rate of the intelligent security operation platform is relatively low, at 18%. In comparison, the platform integrating the EMC-SMMD model has the lowest false positive intervention rate of 8%. This shows that the research model can greatly reduce the consumption of ineffective manpower.



(a) The average event analysis time of each method

(b) The manual intervention rate caused by false alarms under various methods

Figure 10 Comparison of security operation and maintenance manpower load.

4 Discussion and Interpretation

To improve the real-time performance and prediction accuracy of NSS perception, this study proposed an EMC-SMMD model. Experiments showed that the model had an accuracy of 93.7% and a precision of 92.5% in macro-situation level prediction. In the identification of micro-attacks, the F1-Score of the research model for DDoS attacks and port scanning reached 96.5% and 97.2%, and the F1-Score for the identification of brute force cracking and botnet attacks also reached 89.7% and 91.4%. In the real network range test, the model shortened the average response time for high-risk, medium-risk, and low-risk threats to 3.5 min, 15.8 min, and 60.2 min, and increased the attack containment rates of network penetration, data leakage, advanced persistent threats, and ransomware to 92.5%, 89.7%, 75.4%, and 95.2%. At the same time, it also reduced the average event analysis time to 15 min and controlled the false alarm intervention rate to 8%. Based on the above, the research model had good performance in situation awareness accuracy, response efficiency, and automation level.

In comparison with existing research, EMC-SMMD showed unique advantages in NSS perception. Alharbi F et al. [24] focused on specific scenarios of malware analysis and achieved quantitative analysis of honeypot data by building a visual dashboard. However, their method had limitations in real-time performance and multi-source data fusion. In contrast, the streaming processing architecture of this study, based on Kafka+spark could achieve

real-time fusion and feature extraction of multi-source security data, and had more advantages in the data processing dimension. Although the deep belief network model of Lv Z et al. [25] performed well in intrusion detection in cloud environments, it was mainly aimed at specific types of network attacks and relied on key hash self-synchronization technology. Its adaptability in complex network environments needed to be verified. This study used SMMD to decouple complex networks into multiple functional modes, and combined the cognitive decision-making mechanism of the EMC framework to achieve multi-level situational awareness from micro to macro, which has more advantages in comprehensive attack identification and environmental adaptability. In addition, the research model shortened the response time of high-risk threats to 3.5 min while maintaining high detection accuracy, which was significantly better than the comparison method, reflecting its practical value in actual operation and maintenance environments. Nonetheless, there was still room for improvement in this study. Due to the complex structure, its CPU usage was 35.5%, and its memory consumption was 6.5 GB. These two indicators were slightly higher than the GBDT model. Future work can further reduce resource overhead and improve deployment adaptability in resource-constrained environments through model lightweight design and feature selection optimization.

5 Conclusion

To improve the ability of NSS awareness and prediction, this study proposes a core assumption: by integrating the decision-making framework of distributed flow processing, network structure mode decomposition, and simulation expert cognition, a security awareness prediction model is constructed, which is superior to the existing methods in accuracy, speed, and intelligence. Around this assumption, this study introduces a new paradigm of "data structure cognition" collaboration. Its core innovation is to build a real-time data fusion framework based on Kafka+spark, apply SMMD to NSS decoupling, and integrate an EMC cognitive framework with the ability of empirical learning and situation deduction. The experimental study verified the effectiveness of the model. The main findings are summarized as follows: on the cic-ids-2017 dataset, the model achieved a 93.7% macro situation prediction accuracy and a maximum of 97.2% attack recognition F1-Score; In the real shooting range, the model shortened the response time of high-risk threats to 3.5 minutes, increased the containment rate of blackmail software and other attacks to 95.2%, reduced the average event analysis time

to 15 minutes, and controlled the false alarm intervention rate to 8%. These findings jointly show that the proposed model greatly improves the accuracy of NSS awareness, the timeliness of response, and the intelligence level of operation and maintenance, and provides effective technical support for the construction of an active defense system. Looking to the future, research work can be carried out in depth from the following aspects: (1) Lighter vehicle model deployment solutions can be explored to reduce resource consumption; (2) The model can be extended to more complex heterogeneous network environments, such as cloud native and IoT; (3) Reinforcement learning can be introduced into the EMC framework to further improve the model's adaptive decision-making capabilities under unknown threats.

Fundings

The research is supported by 2024 Guangdong Province University Characteristic Innovation Project "New Quality Intelligent Education – Multidimensional Teaching Evaluation System Driven by Intelligent Learning and Behavior Recognition" (Project Number: 2024KTSCX249).

References

[1] Xu H, Berres A, Yoginath S B, Sorensen H, Nugent P J, Severino J. Smart mobility in the cloud: Enabling real-time situational awareness and cyber-physical control through a digital twin for traffic. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(3): 3145–3156. DOI:10.1109/TITS.2022.3226746.

[2] Lill B, Sauerwein C, Mexis N, Langner K. A Comprehensive Review of Information Security Research regarding SMEs and Future Directions. Journal of Cyber Security and Mobility, 2025, 14(5): 1245–1288. DOI:10.13052/jcsm2245-1439.1459.

[3] Li Y F. Application Mode of Blockchain Technology in User Data Sovereignty and Privacy Protection. Journal of Cyber Security and Mobility, 2025, 14(5): 1199–1220. DOI:10.13052/jcsm2245-1439.1457.

[4] Chen Z. Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm. Journal of Computational and Cognitive Engineering, 2022, 1(3): 103–108. DOI:10.47852/bonviewJCCE149145205514.

[5] Oladosu S A, Ige A B, Ike C C, Adepoju P A. Next-generation network security: Conceptualizing a unified, AI-powered security architecture for cloud-native and on-premise environments. International Journal of Science and Technology Research Archive, 2022, 3(2): 270–280. DOI:10.53771/ijstra.2022.3.2.0143.

[6] Vimal V, Muruganantham R, Prabha R, Arularasan A N, Nandal P, Chanthirasekaran K, et al. Enhance Software-Defined Network Security with IoT for Strengthen the Encryption of Information Access Control. Computational Intelligence and Neuroscience, 2022, 2022(1): 4437507. DOI:10.1155/2022/4437507.

[7] Sheng C, Yao Y, Li W, Yang W, Liu Y. Unknown attack traffic classification in SCADA network using heuristic clustering technique. IEEE Transactions on Network and Service Management, 2023, 20(3): 2625–2638. DOI:10.1109/TNSM.2023.3238402.

[8] Nafees M N, Saxena N, Cardenas A, Grijalva S, Burnap P. Smart grid cyber-physical situational awareness of complex operational technology attacks: A review. ACM computing surveys, 2023, 55(10): 1–36. DOI:10 .1145/3565570.

[9] Mokayed, H., Quan, T. Z., Alkhaled, L., and Sivakumar, V. Real-time human detection and counting system using deep learning computer vision techniques. Artificial Intelligence and Applications. 2023, 1(4): 221–229. DOI:0000-0001-6158-3543.

[10] Wang X, Mei J, Cui S, Wang C X, Shen X S. Realizing 6G: The operational goals, enabling technologies of future networks, and value-oriented intelligent multi-dimensional multiple access. IEEE Network, 2023, 37(1): 10–17. DOI:10.1109/MNET.001.2200429.

[11] Li M, Naeem F, Kaddoum G, Hossain E. Metaverse communications, networking, security, and applications: Research issues, state-of-the-art, and future directions. IEEE Communications Surveys & Tutorials, 2023, 26(2): 1238–1278. DOI:10.1109/COMST.2023.3347172.

[12] Venkatesan K, Rahayu S B. Blockchain security enhancement: an approach towards hybrid consensus algorithms and machine learning techniques[J]. Scientific Reports, 2024, 14(1): 1149. DOI:s41598-024 -51578-7.

[13] Palbar Misas J D, Hopcraft R, Tam K, Jones K. Future of maritime autonomy: cybersecurity, trust and mariner's situational awareness. Journal of Marine Engineering & Technology, 2024, 23(3): 224–235. DOI:10.1080/20464177.2024.2330176.

[14] Lee C E, Baek J, Son J, Ha Y G. Deep AI military staff: Cooperative battlefield situation awareness for commander's decision making. The Journal of Supercomputing, 2023, 79(6): 6040–6069. DOI:10.1007/s11227-022-04882-w.

[15] Talukder M A, Islam M M, Uddin M A, Hasan K F, Sharmin S, Alyami S A, et al. Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction[J]. Journal of big data, 2024, 11(1): 33. DOI:10.1186/s40537-024-00886-w.

[16] Bringhenti D, Marchetto G, Sisto R, Valenza F. Automation for network security configuration: State of the art and research trends. ACM Computing Surveys, 2023, 56(3): 1–37. DOI:10.1145/3616401.

[17] Abou El Houda Z, Brik B, Senouci S M. A novel IoT-based explainable deep learning framework for intrusion detection systems[J]. IEEE Internet of Things Magazine, 2022, 5(2): 20–23. DOI:10.1109/IOTM.005.2200028.

[18] Aminu M, Akinsanya A, Dako D A, Dickson A. Enhancing cyber threat detection through real-time threat intelligence and adaptive defense mechanisms. International Journal of Computer Applications Technology and Research, 2024, 13(8): 11–27. DOI:10.7753/IJCATR1308.1002.

[19] Jiang J, Karran A J, Coursaris C K, Leger P M, Beringer J. A situation awareness perspective on human-AI interaction: Tensions and opportunities. International Journal of Human–Computer Interaction, 2023, 39(9): 1789–1806. DOI:10.1080/10447318.2022.2093863.

[20] Yang L, El Rajab M, Shami A, Muhaidat S. Enabling automl for zero-touch network security: Use-case driven analysis. IEEE Transactions on Network and Service Management, 2024, 21(3): 3555–3582. DOI:10.1109/TNSM.2024.3376631.

[21] Mohy-Eddine M, Guezzaz A, Benkirane S, Azrour M. An efficient network intrusion detection model for IoT security using K-NN classifier and feature selection. Multimedia Tools and Applications, 2023, 82(15): 23615–23633. DOI:10.1007/s11042-023-14795-2.

[22] Sarker I H, Khan A I, Abushark Y B, Alsolami F. Internet of things (iot) security intelligence: a comprehensive overview, machine learning solutions and research directions. Mobile Networks and Applications, 2023, 28(1): 296–312. DOI:10.1007/s11036-022-01937-3.

[23] Ofoegbu K D O, Osundare O S, Ike C S, Fakeyede O. Proactive cyber threat mitigation: Integrating data-driven insights with user-centric

security protocols. Computer Science & IT Research Journal, 2024, 5(8): 2083–2106. DOI:10.51594/csitrj.v5i8.1493.

[24] Alharbi F, Kashyap G S. Empowering Network Security through Advanced Analysis of Malware Samples: Leveraging System Metrics and Network Log Data for Informed Decision-Making. International Journal of Networked and Distributed Computing, 2024, 12(2): 250–264. DOI:10.1007/s44227-024-00032-1.

[25] Lv Z, Chen D, Cao B, Song H, Lv H. Secure deep learning in defense in deep-learning-as-a-service computing systems in digital twins. IEEE Transactions on Computers, 2023, 73(3): 656–668. DOI:10.1109/TC.2021.3077687.

Biographies

Changyao Yang, male, holds a Master's degree in Engineering from Chongqing University. Currently serving as an associate professor in the Department of Information Science at Zhanjiang Preschool Education College, with research interests in computer application technology, data analysis, etc. Published over ten articles at or above the provincial level, including two articles from Chinese core journals, edited two textbooks in related fields, and led and participated in multiple projects at or above the provincial and municipal levels.

Yuan Yan, female, received her Master of Engineering degree from Chongqing University. She is currently an Associate Professor at Zhanjiang Preschool Education College. Her research interests focus on digital media technology. She has published many academic papers, including 2 papers in Chinese core journals and 1 SCI-indexed paper, as well as several papers in provincial journals. She has presided over a number of provincial and municipal research projects. She is the corresponding author of this paper.